University of Reading

Representing the Arctic in Global Surface Temperature Time Series of Recent Climate Change

PhD in Atmosphere, Oceans and Climate

Department of Meteorology

Emma Dodd

May 2015

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

- Emma Dodd

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Chrisopher Merchant. I have greatly benefited from all the expertise and guidance he has provided throughout my time as a PhD student. Thanks are also due to my other supervisors, both past and present; Nick Rayner, Prof. Simon Tett and Dr Simone Morak. Furthermore I should thank my monitoring committee at the University of Reading and those on my confirmation panel at the University of Edinburgh for pointing out gaps in my knowledge.

I would like to express my gratitude to the many people whom I had the pleasure to meet during the course of my PhD, especially those at the University of Edinburgh, Met Office Hadley Centre, University of Reading, and the EarthTemp Network Workshops. Thank you for your advice, constructive comments, discussions and support. I would like to offer my special thanks to Dr. Colin Morice who provided guidance, insightful discussion, and data in the course of producing my first paper. I am also indebted to Dr. Josefino Comiso and Larry Stock who kindly shared their renormalised monthly AVHRR data.

Lastly some personal thanks. Thank you to my parents, who gave me the freedom to follow my own path, and my friends, for your company. Special thanks are due to my fellow PhD student Emma Knowland for her friendship, support and sympathy. I hope I was able to reciprocate in at least some small way. I would also like to thank all those who keep up the tradition of 11am coffee in ESSC, especially Drs Claire Bulgin, Claire MacIntosh, Kevin Pearson, Jonathan Mittaz and Owen Embury for their interesting conversation on many topics. And finally to my partner John, my deepest heartfelt appreciation for your unwavering support, encouragement and never failing belief in me.

Abstract

The Arctic is an important region in the study of climate change, but monitoring surface temperatures in this region is challenging, particularly in areas covered by sea ice. Here in situ, satellite and reanalysis data were utilised to investigate whether global warming over recent decades could be better estimated by changing the way the Arctic is treated in calculating global mean temperature. The degree of difference arising from using five different techniques, based on existing temperature anomaly dataset techniques, to estimate Arctic SAT anomalies over land and sea ice were investigated using reanalysis data as a testbed. Techniques which interpolated anomalies were found to result in smaller errors than non-interpolating techniques. Kriging techniques provided the smallest errors in anomaly estimates. Similar accuracies were found for anomalies estimated from in situ meteorological station SAT records using a kriging technique. Whether additional data sources, which are not currently utilised in temperature anomaly datasets, would improve estimates of Arctic surface air temperature anomalies was investigated within the reanalysis testbed and using in situ data. For the reanalysis study, the additional input anomalies were reanalysis data sampled at certain supplementary data source locations over Arctic land and sea ice areas. For the in situ data study, the additional input anomalies over sea ice were surface temperature anomalies derived from the Advanced Very High Resolution Radiometer satellite instruments. The use of additional data sources, particularly those located in the Arctic Ocean over sea ice or on islands in sparsely observed regions, can lead to substantial improvements in the accuracy of estimated anomalies. Decreases in Root Mean Square Error can be up to 0.2K for Arctic-average anomalies and more than 1K for spatially resolved anomalies. Further improvements in accuracy may be accomplished through the use of other data sources.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Recent Global Climate Change

Warming of the Earth's land surface was first noted by Callendar 1938, who used temperature observations from 200 meteorological stations to demonstrate that air temperatures over land had increased in the 50 years prior to that study (Hawkins and Jones, 2013). Subsequently, climate change over all surfaces of the Earth has become a major topic in contemporary scientific research. Many researchers and organisations are working to quantify post-industrialisation climate change both globally and regionally, assess how fast it is happening and what the consequences will be.

Currently it is estimated that global mean surface temperatures have risen by around 0.8K between 1850 and 2014 (Berkeley Earth, 2014; Hansen et al., 2010; Hartmann et al., 2013; Ishii et al., 2005; Jones et al., 2012; Kaplan et al., 1998; Kennedy et al., 2011b; Kent et al., 2013; Lawrimore et al., 2011; Morice et al., 2012; Muller et al., 2013; Rayner et al., 2003; Rohde et al., 2013b; Smith et al., 2008; Vose et al., 2012). This warming in global mean surface temperatures is illustrated in Figure 1.1, which shows the annual anomalies of global land-surface temperatures between 1880 to 2012 from three temperature anomaly datasets. The datasets differ slightly in their annual variations due to structural uncertainty (the uncertainty arising from the choice of methodology) but show a common trend of rising global average surface temperatures. This trend is also noted in regional surface temperature datasets (e.g. Menne et al., 2009; Vincent et al., 2012).

Consistent with the observed increases in surface temperatures other changes have been observed including: warming of the oceans and atmosphere, decreasing snow cover, decreases in Arctic sea ice thickness and extent, almost global reductions in glacier and ice sheet mass and extent, and sea level rise (Intergovernmental Panel on Climate Change, 2013). The Intergovernmental Panel on Climate Change (IPCC) has concluded in its Fifth Assessment Report (AR5) that "Warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia" (IPCC, 2013). This report also concluded that "It is extremely likely that human influence has been the dominant cause of the observed warming since the mid-

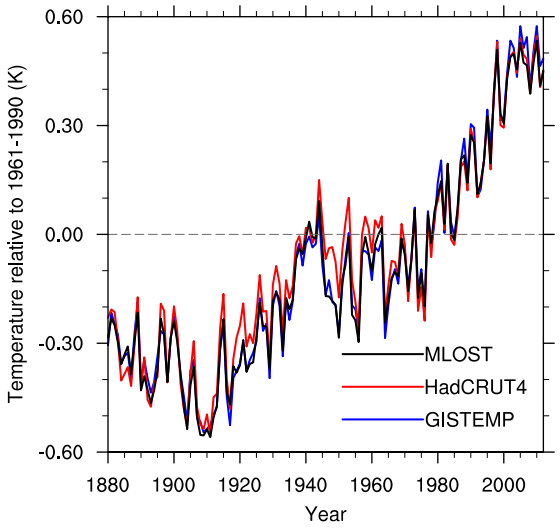

Figure 1.1: Annual global mean surface temperature anomaly (K) relative to 1961-1990 from three combined land-surface air and sea surface temperature datasets; the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) Merged Land-Ocean Surface Temperature Analysis (MLOST, Vose et al. 2012), the HadCRUT4 global temperature anomaly dataset (Morice et al., 2012) produced by the Met Office Hadley Centre and Climatic Research Unit (CRU) of the University of East Anglia, and the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP, Hansen et al. 2010).

20^{th} century" (IPCC, 2013) as a result of the increase in atmospheric concentrations of greenhouse gases since 1750.

1.2 Monitoring Recent Climate Change

Post-industrialisation changes in climate are often monitored using timeseries of global and regional mean surface temperatures, which are temperatures measured at, or near, the surface of the Earth. Timeseries of global and regional mean Surface Air Temperature (SAT) anomalies are one of the main metrics used to estimate recent climate change. SATs measured at meteorological stations are largely used to generate these timeseries over land and sea ice. For most of the ocean Sea Surface Temperatures (SSTs), collected in situ or derived from satellite instruments, are used. This section provides an overview of the types of surface temperatures most relevant to this thesis and how these temperatures are observed; some of the biases, quality control and data coverage issues associated with surface temperature records; and how surface temperatures are currently utilised to produce datasets and timeseries of global mean surface temperature change.

1.2.1 Surface Temperatures

There are various types of surface temperature that can be employed to monitor surface temperature change and therefore changes in climate. Surface temperatures are often

differentiated by the type of surface they are associated with (e.g. ocean, land, ice) and/or the height, or depth, at which they are measured as illustrated in Figure 1.2. These temperatures are measured from sensors on various different platforms. They may be measured in situ or derived from satellite sensor measurements. Surface temperatures, such as those in Figure 1.2, which could be used to monitor changes in climate in the Arctic (those most relevant for this thesis) are discussed here along with the platforms used to measure them.

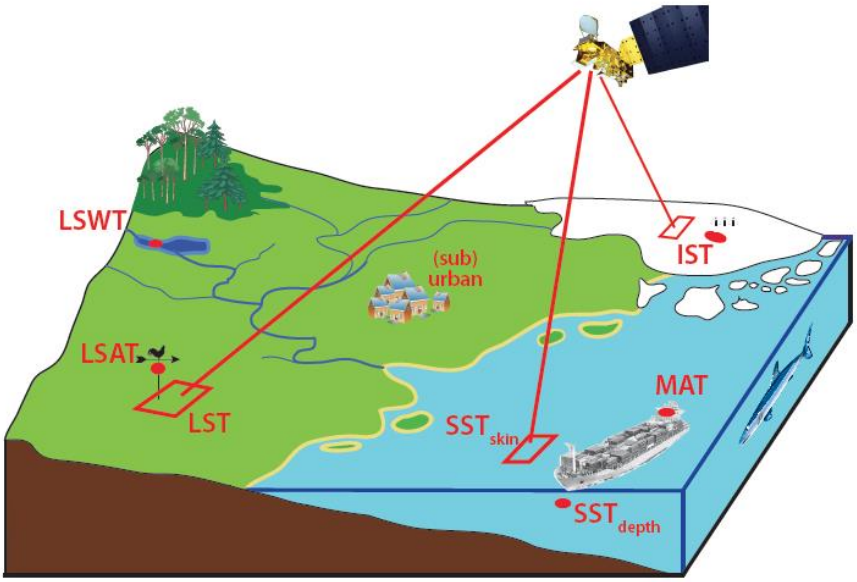

Figure 1.2: Different surface temperatures. Sea Surface Temperature (SST): at depth, measured in situ, or of the skin layer, measured by radiometers on ships or in space; Marine Air Temperature (MAT); Land Surface Temperature (LST); Surface Air Temperature (SAT); Lake Surface Water Temperature (LSWT); Ice Surface Temperature (IST). Fig. 1 from Merchant et al. 2013.

1.2.1.1 Surface Air Temperature

Of the temperatures illustrated in Figure 1.2, one of the most commonly used to measure global and regional mean surface temperature change is the SAT. SATs measured in situ are generally used to monitor global temperature change over land and sea ice areas. They provide the longest running instrumental records of temperature in the world; the Met Office Hadley Centre Central England Temperature (HadCET) dataset provides monthly temperatures back to 1659 (Manley, 1953). SAT is generally measured by shielded thermometers or temperature sensors at a height of between 1.25 and 2.0m above the surface of the Earth. These sensors may be placed in a Stevenson screen or mounted on other measuring platforms such as a meteorological tower. In situ SATs over areas of sea ice are measured from drifting buoys and drifting ice stations (e.g. Polashenski et al., 2011; Rigor et al., 2000; Uttal et al., 2002). As SATs are measured at around 1.5m height they are influenced by sensible and latent heat exchanges with the atmosphere and surface of the Earth, radiation emitted from the Earth's surface, solar radiation, and lateral advection.

Radiometric surface temperatures from satellite instruments can be related to SATs, but are not generally utilised at present for monitoring surface temperature change.

1.2.1.2 Sea Surface Temperature and Marine Air Temperature

Sea Surface Temperature (SST) and Marine Air Temperature (MAT) are the types of temperatures used to monitor global temperature change over the oceans. SST is a measurement of the water temperature close to the surface of the ocean, up to several metres below the surface (Kennedy et al., 2011c). MAT is a measurement of the air temperature above the surface of the ocean. Both SST and MAT are measured from in situ temperature sensors on platforms such as ships, oceanographic stations, moored buoys, drifting buoys and research vessels (Kent and Taylor, 2006; Kent et al., 2013; Kennedy et al., 2011b; Woodruff et al., 2011). In common with SATs, SSTs and MATs provide a long temperature record; for example, the Met Office Hadley Centre's sea surface temperature dataset version 3 (HadSST3) extends from 1850 to present (Kennedy et al., 2011b,c).

SST anomalies are more frequently employed to monitor global temperature change than MAT anomalies as SST measurements are thought to be more homogeneous (Rayner et al., 2003). Records of MAT change are almost exclusively produced using Night-time only Marine Air Temperature (NMAT) (Kent et al., 2013; Rayner et al., 2003). Daytime MATs are often affected by direct or indirect solar heating, although corrected daytime MAT records are available (Berry and Kent, 2009). MATs are also affected by biases resulting from increasing observing platform height as the average height of ship's decks have increased over time (Kent et al., 2013; Rayner et al., 2003). Additionally, SST observations are more plentiful (Rayner et al., 2003). As a result SSTs are generally used to monitor global temperature change over the oceans as a proxy for MAT. Air temperatures are significantly influenced by exchanges of heat and radiation from the surface of the Earth so SST anomalies can be expected to agree well with MAT anomalies, at least on seasonal and longer time-scales, as a result of the thermal lag of the oceans. SST anomalies have been compared to NMAT anomalies and this has been confirmed (Rayner et al., 2003; Kent et al., 2013). Nonetheless, MAT datasets are still produced and used to monitor post-industrialisation climate change (Berry and Kent, 2009; Hartmann et al., 2013; Kent et al., 2013).

SSTs can additionally be derived from infrared or microwave radiometers flown on satellites such as the AVHRR (Advanced Very High Resolution Radiometer, e.g. Casey et al. 2010), ATSR (Along Track Scanning Radiometer, e.g. Merchant et al. 2012) and AMSR-E (Advanced Microwave Scanning Radiometer for the Earth observing system, e.g. Wentz and Meissner 2000). Satellite instruments provide relatively detailed observations, with a greater spatial coverage, of the radiometric temperature of water (and ice, see Section 1.2.1.3) surfaces (Merchant et al., 2013). Satellite radiometers observe brightness temperatures in one or more wavelengths for the surface skin of the ocean

(within 1 mm of the ocean surface). The SST is derived from brightness temperatures using retrieval schemes employing either coefficients (e.g. Embury and Merchant, 2012; Wentz and Meissner, 2000, 2007) or optimal estimation (e.g. Merchant et al., 2008). SSTs derived from infrared instruments are more commonly used than those from microwave instruments as their spatial resolution is higher. However, infrared instruments cannot collect temperature information in the presence of cloud cover and their SST retrieval schemes also have to account for factors such as viewing angle and aerosols (Embury and Merchant, 2012; Merchant et al., 2008). Microwave derived brightness temperatures are negligibly affected by clouds, but the microwave SST retrieval schemes do have to account for factors including sea surface roughness, salinity and precipitation (Wentz et al., 2000; Wentz and Meissner, 2000, 2007). It should be noted that, when comparing satellite derived SST and in situ SSTs, a skin to bulk SST conversion will also need to be applied.

1.2.1.3 Ice Surface Temperature

Ice Surface Temperatures (IST) are a measurement of the temperature of the surface layer of ice or snow. ISTs are not often used to monitor surface temperature change, but could provide a useful temperature record in the sparsely observed polar regions. ISTs could be used as a proxy for SATs over ice and snow, similar to the way SST is used as a proxy for MATs. ISTs can be derived from in situ sensors on platforms as radiometers (Scambos et al., 2006) and Ice Mass Balance Buoys (IMBs, Polashenski et al. e.g. 2011; Richter-Menge et al. e.g. 2006). However, currently the majority of IST observations come from satellites - derived either from microwave or infrared radiometers.

As with SSTs, ISTs are derived from brightness temperatures retrieved from satellite radiometers (e.g. Comiso et al., 2003; Hall et al., 2004). Microwave radiometers are not affected by cloud cover or polar darkness and therefore are useful for monitoring snow and ice temperature change. However, microwaves penetrate into ice and snow beyond the surface, so temperature information derived from these sensors is not easily relatable to IST, and their relatively coarse spatial resolution limits their utility for studies of sea ice temperatures (Hall et al., 2004). Therefore, IST is generally estimated using infrared instruments such as AVHRR (Key and Haefliger, 1992; Lindsay and Rothrock, 1994) and the MODerate-resolution Imaging Spectroradiometer (MODIS, Hall et al. 2004).

1.2.1.4 Lower Tropospheric Temperature

Lower Tropospheric Temperature (TLT) is the temperature of the lower part of the troposphere; from the surface to around 3km altitude (Hartmann et al., 2013). In common with IST, upper air temperatures are not as commonly utilised as observations of SAT and SST for monitoring global temperature changes but could provide a useful temperature record in sparsely observed areas (Chapters 5). TLT datasets derived from in situ and satellite sources have been used to monitor upper air temperature change (e.g. Mears and Wentz, 2009), have been included in datasets of surface temperature change (Cowtan

and Way, 2014), and are included in the IPCC AR5 report (Hartmann et al., 2013).

TLT can be measured by in situ platforms (radiosondes or rawinsondes, Thorne et al. 2005) and estimated from satellite instruments such as the Microwave Sounding Unit (MSU, Christy et al. 2003; Mears and Wentz 2009). Satellite estimates of TLT are estimated from brightness temperatures up to 12.5km height, which are heavily weighted to be more representative of the temperature of the troposphere below 3km (Hartmann et al., 2013; Mears and Wentz, 2009). As TLT observations include the temperature of the atmospheric boundary layer, where temperatures are influenced by exchanges of heat and radiation from the surface of the Earth, the temperature trends and short term variations in upper air temperatures are highly correlated with those at the surface but are of a greater amplitude (Hartmann et al., 2013).

1.2.2 Biases, Quality Control and Incomplete Data Coverage

There are many types of surface temperature records which can be used to monitor recent climate change. However, measurements of surface temperatures are not consistent in space and time. They are subject to biases, quality issues, and may be spatially and temporally sparse.

SAT and SST data have been collected using various methods, instruments and platforms since records began in the 1800s (e.g. Folland and Parker, 1995; Kennedy et al., 2011b,c; Kent and Taylor, 2006; Kent and Kaplan, 2006; Trewin, 2010). Other changes have also occurred such as: urbanisation (Parker, 2010; Hansen et al., 2010; Menne et al., 2009; Trewin, 2010); increasing average ship speed (impacting engine room intake and hull contact sensor SST measurements, Kennedy et al. 2011c); land use and land cover change (Trewin, 2010); and the movement, addition or discontinuation of meteorological stations (Menne et al., 2009; Trewin, 2010). These changes result in time-varying biases which are not consistent temporally or spatially (Kent and Taylor, 2006; Trewin, 2010). Such changes may also not have been recorded adequately, leading to uncertainty in the measurements.

Additionally, temperature data coverage is not consistent across the globe. Figure 1.3 shows the location and length of record of all land-based meteorological stations in the International Surface Temperature Initiative (ISTI) Stage 3 Merged Recommended Monthly databank product (Thorne et al., 2011). The databank is a merged collation of all station records provided for the ISTI global land surface databank project. As illustrated in Figure 1.3, meteorological stations are not evenly distributed across land areas. Some areas, such as North America, are relatively densely populated with stations whilst in others, such as the Arctic, station coverage is sparse. Meteorological station coverage can also be temporally sparse. For example, Figure 1.3 shows that North America has many stations which report for 100 years or more whereas in Antarctica many stations report for 25 years or fewer. Figure 1.4 shows the total number of in situ observations of SST in 1850, 1960, 1980 and 2011 from HadSST3 (Kennedy et al., 2011b,c). Some ocean areas, such

as the North Atlantic are relatively well observed in most time periods whereas others, such as the Southern Ocean are poorly observed (Figure 1.4). The areas observed and numbers of observations also change temporally, and there are fewer observations earlier in the temperature record (Kennedy et al., 2011c; Rayner et al., 2006). SST observations derived from satellite measurements are more spatially complete than in situ SST observations. However, satellite derived SST observations are not temporally or spatially complete. Thermal infrared observations cannot be used when clouds are present and satellite SST records do not extend as far back in time as in situ measurements. Incomplete temporal and/or spatial coverage in both SAT and SST observations can lead to biases and uncertainties in temperature timeseries (Cowtan and Way, 2014; Rohde, 2013).

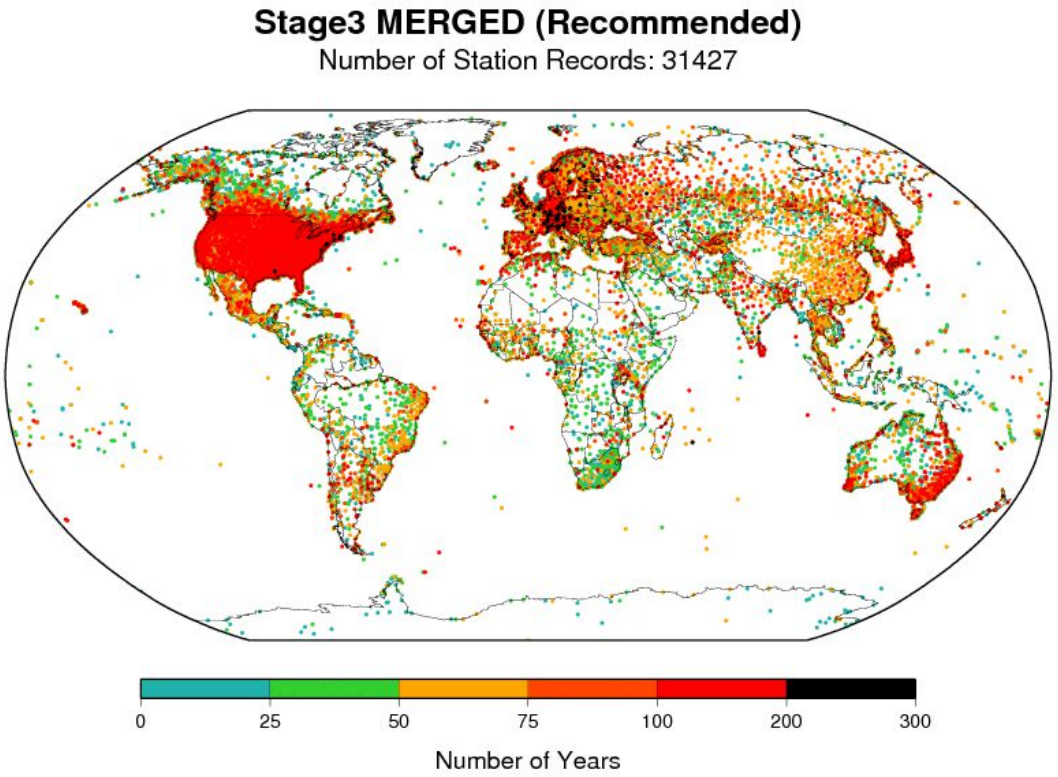

Figure 1.3: Map of the locations and length of temperature record of all stations in the ISTI Stage 3 Merged Recommended Monthly databank product (Thorne et al., 2011). Acquired from *ftp : //ftp.ncdc.noaa.gov/pub/data/globaldatabank/monthly/stage3/recommended/plots /merged_locations.gif*.

Lastly, temperature records are also affected by quality control issues and systematic uncertainties including, but not limited to: problems with instrumentation (Rigor et al., 2000); quality of station siting (Fall et al., 2011; Trewin, 2010); solar heating effects (Kent and Kaplan, 2006; Kent et al., 2013; Trewin, 2010); data imprecision (temperatures may be recorded at a lower precision than WMO standards recommend, Trewin 2010); and precision changes (Trewin, 2010). There also may be uncertainties resulting from random effects (Kent and Challenor, 2006).

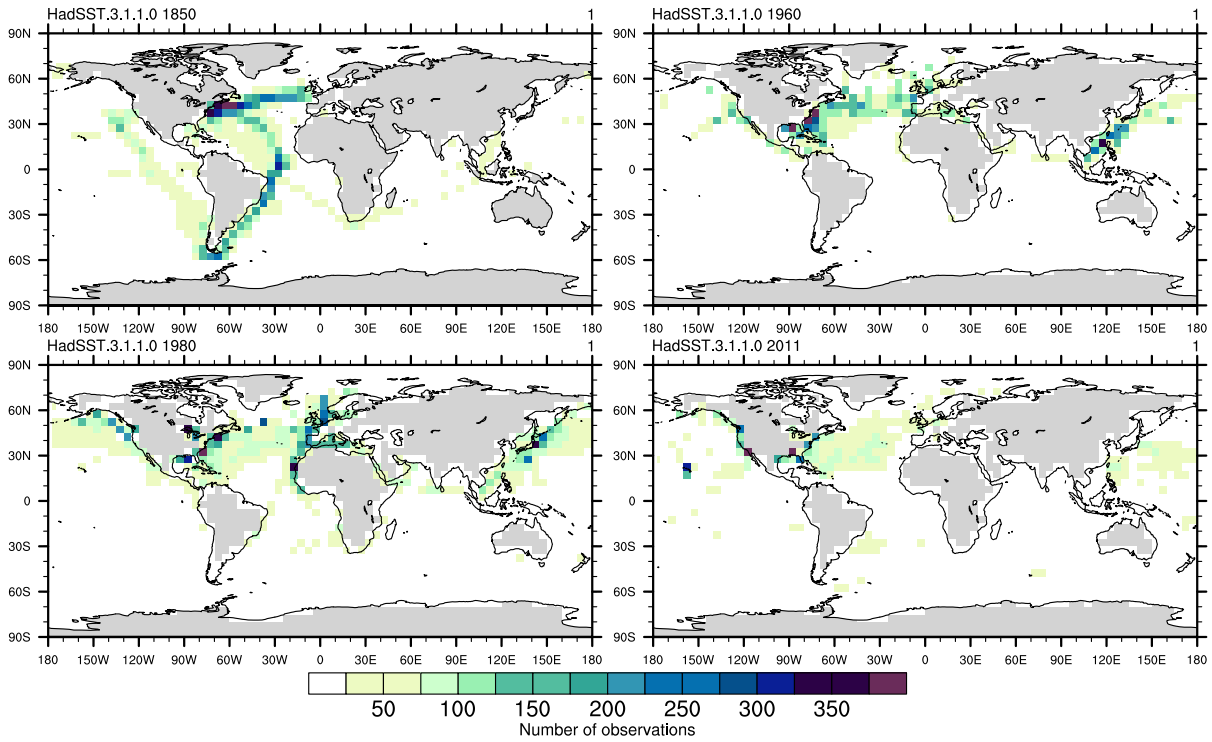

Figure 1.4: A map of grid cells where SST is recorded and the total number of observations in HadSST3 (Kennedy et al., 2011b,c) for 1850, 1960, 1980 and 2011.

1.2.3 Surface Temperature Timeseries of Recent Climate Change

Surface temperature measurements may be sparse, both temporally and spatially, as well as subject to various inhomogeneities, errors and uncertainties. To exclude these non-climatic variations in temperatures from estimates of actual climate change there are many different techniques, methods and adjustments that can be employed. As a result many different groups generate datasets of SAT anomalies using various techniques from a subset of the available surface temperature records (e.g. Berkeley Earth, 2014; Hansen et al., 2010; Morice et al., 2012; Vose et al., 2012).

Different SAT anomaly datasets have different quality control and bias adjustment methods. Some datasets largely utilise data which have been quality controlled and adjusted by other organisations. The National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) Surface Temperature Analysis, also known as GISTEMP, uses adjusted Global Historical Climatology Network - Monthly temperature dataset version 3 (GHCN-Mv3) over land, along with some additional data, and the Extended Reconstructed Sea Surface Temperature dataset version 3b (ERSSTv3b, Smith et al. 2008) over the ocean (GISS, 2014; Hansen et al., 2010). GISTEMP additionally employ their own urbanisation adjustment based on satellite data (Hansen et al., 2010). The CRUTEM4 land surface temperature anomaly dataset (produced by the Climatic Research Unit of the University of East Anglia and the Met Office Hadley Centre)

mainly utilises homogenised temperature records provided by national meteorological services (Jones et al., 2012). For remaining records, comparisons are made with neighbouring stations and adjustments are made accordingly (Jones et al., 2012). Some datasets use metadata to identify and adjust temperature records. For example, HadSST3 uses metadata to adjust record according to the measuring platform and methods used (Kennedy et al., 2011c). Finally, some datasets adjust temperature records without the use of metadata. The Berkeley Earth Surface Temperature dataset uses a method they call the 'scalpel' to identify large discontinuities in single station records (Rohde et al., 2013a, 2012). The records are then split at these points to create two separate record sections. Any adjustments to the records occur automatically as part of the statistical techniques they employ to produce their dataset (Berkeley Earth, 2014; Rohde et al., 2013a, 2012).

Different techniques are also employed to quantify surface temperature changes from records that are not globally or temporally complete. In situ temperature measurements can be used exclusively without spatial infilling of data. This technique is employed to create the HadCRUT4 global temperature anomaly dataset (Morice et al., 2012) which is produced by the Met Office Hadley Centre from a combination of CRUTEM4 over land and HadSST3 over ocean areas. Any grid boxes which do not have available in situ data are empty in these datasets for the time periods in which SSTs or SATs are unavailable (Morice et al., 2012). Some datasets interpolate and extrapolate surface temperatures. GISTEMP is a combination of linearly interpolated and extrapolated SAT data over land and sea ice, and SST data infilled using statistical methods over the ocean. SAT anomalies are interpolated between stations and extrapolated up to 1200km into land and sea ice regions with no measurements (Hansen et al., 2010; Lawrimore et al., 2011). ERSSTv3 is produced by combining the outputs of low- and high- frequency analyses, which derive decadal and seasonal to interannual SST variations respectively (Smith and Reynolds, 2003, 2004, 2005; Smith et al., 2008). The Berkeley Earth Surface Temperature dataset method uses temperature records from many pre-existing datasets and statistical techniques to produce a globally complete estimate of temperature anomalies over land and ocean areas (Berkeley Earth, 2014; Rohde et al., 2013a,b, 2012). It utilises all available data records over land; the other datasets discussed here require records for which a reference 'normal' can be produced to calculate temperature anomalies. A recent update to the dataset adds HadSST ocean data to the Berkeley land-area dataset and extrapolates SAT data from the land-area dataset over sea ice (Berkeley Earth, 2014). Finally, some datasets such as MLOST employ both interpolation and extrapolation with exclusion of data (Vose et al., 2012). MLOST utilises homogenised GHCN-Mv3 meteorological station data for SAT anomalies and ERSSTv3b for SST anomalies (Smith and Reynolds, 2005; Smith et al., 2008; Vose et al., 2012). The SAT data from GHCN-Mv3 are analysed using the same low- and high- frequency analysis employed for ERSSTv3b (Smith and Reynolds, 2005). Areas where data are determined to be insufficient are excluded from the low- and high- frequency analyses (Vose et al., 2012).

Finally it should be noted that there may also be differences in how each dataset produces temperature anomalies from temperature records. Many datasets use the Climate Anomaly Method (CAM) which uses a typically 30 year reference period to calculate temperature anomalies. Datasets may choose different climatology periods, however; HadCRUT4 uses 1961-90, Berkeley Earth and GISTEMP use 1951-1980 (Hansen et al., 2010; Morice et al., 2012; Rohde et al., 2013a). Other datasets and dataset versions have used anomaly methods such as the Reference Station Method (RSM) (Hansen and Lebedeff, 1987) and the First Difference Method (FDM) (Peterson et al., 1998).

Despite the structural uncertainty arising from the diversity of techniques used at each stage in producing estimates of recent surface temperature changes, the timeseries produced from temperature anomaly datasets largely result in similar post-industrialisation climate change trends as noted in Section 1.1 and illustrated in Figure 1.1. There is a relatively low impact of differences in methods on trends in global surface temperature timeseries. But, what are the impacts of choices of data and methodology on Arctic surface temperature anomalies? How should these datasets incorporate the appropriate contribution to global and regional-average temperature anomalies of the sea-ice regions of the Arctic Ocean?

1.3 The Importance of the Arctic

The Arctic is an important region in the study of climate change because of expected and observed changes in this region. Temperature changes are predicted, and observed, to be more rapid in the Arctic because of Arctic amplification. However, monitoring Arctic temperature change is difficult, particularly in areas covered by sea ice for all or part of the year. This is because in situ measurements of Arctic surface temperatures are sparse and the records are often short. In addition, utilising SAT records collected in this region can be challenging (Chapter 4). Here Arctic amplification is discussed, the various changes associated with climate change in the Arctic are described, and how Arctic surface temperature changes are currently monitored is outlined.

1.3.1 Arctic Amplification

Arctic amplification occurs as a result of various processes, feedback effects and characteristics of the Arctic region. The most notable of these processes is the albedo effect (Serreze and Barry, 2011). As temperatures increase and snow, ice and sea ice melt, darker surfaces (land, vegetation and ocean) are exposed which have a lower albedo and absorb more solar energy (illustrated in Figure 1.5). This lowering of surface albedo and corresponding increase in solar radiation absorption causes increased warming, which melts more snow and ice to expose darker surfaces, and so on in a positive feedback loop called the albedo effect. Albedo changes may also result from soot deposits on snow and other feedback effects may arise from causes such as changes in cloud cover and water

vapour content (Serreze and Barry, 2011).

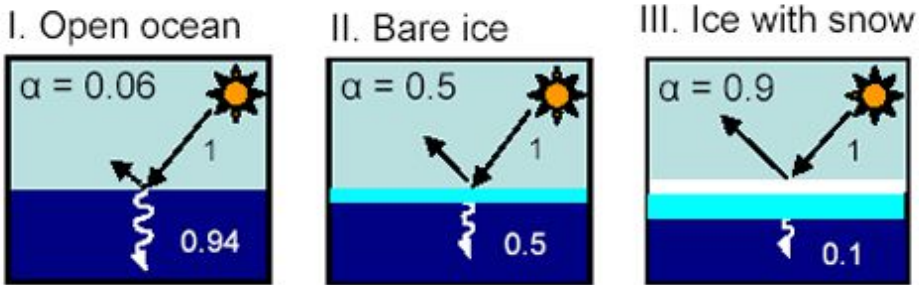

Figure 1.5: Image from the National Snow and Ice Date Center (NSIDC) showing the albedo (α) of open ocean, bare sea ice and sea ice with snow. Acquired from *http* : *//nsidc.org/cryosphere/seaice/processes/albedo.html* (NSIDC, 2013).

The following characteristics of the Arctic region can also lead to Arctic Amplification. In the Arctic a larger fraction of energy received at the surface goes into warming compared to areas such as the tropics where more energy goes into evaporation. Therefore a larger fraction of the extra energy received as a result of increasing concentrations of greenhouse gases will go into warming the region compared to other areas of the globe (ACIA, 2005). Also, the atmospheric layer that has to warm in the Arctic in order to warm the surface is much shallower than in the tropics (ACIA, 2005). In addition, because heat is transported to the Arctic by the atmosphere and oceans, alterations in their circulation patterns may increase (or decrease) warming in the Arctic region (ACIA, 2005; Anisimov et al., 2007). Furthermore, it is important to note that the processes and consequences of Arctic amplification will not be confined to the Arctic; there will be global implications. Loss of surface reflectivity in the Arctic will accelerate warming at the global scale, and changes in circulation patterns may increase (or decrease) warming globally (ACIA, 2005).

1.3.2 Arctic Surface Temperature Change

Currently it is estimated that Arctic mean surface temperatures over land have risen by around 3.0K between 1880 and 2012 (see Figure 3.2 of this thesis). This is more than double the warming seen for global mean surface temperatures (Figure 1.1) and is significant enough to affect global mean surface temperature estimates when large areas of the Arctic are excluded from temperature anomaly datasets (Cowtan and Way, 2014; Kennedy et al., 2011a). The area above the Arctic circle covers about 6% of the global area, so if we assume a warming of 3.0K for the Arctic region as a whole this would lead to a warming of global temperatures by 0.18K. This is 20% of the increase in global mean surface temperatures between 1850 and 2014.

1.3.3 Other Changes in the Arctic Associated with Climate Change

In addition to increasing Arctic surface temperatures over land, many other changes associated with climate change are already being recorded in this region. These changes include rising sea, soil, ice and permafrost temperatures (Comiso, 2001, 2006; Hinzman et al., 2005); warmer and shorter winters (Serreze et al., 2000); increasing length of melt season and area of melt (Comiso, 2006); increasing precipitation and changes in precipitation type (Serreze et al., 2000); decreasing snow cover (Callaghan et al., 2011; Serreze et al., 2000); thawing permafrost (Hinzman et al., 2005; Serreze et al., 2000); changing wind patterns (ACIA, 2005; Anisimov et al., 2007); changes in cloud cover and atmospheric pressure (Comiso, 2001; Przybylak et al., 2013; Serreze et al., 2000); increasing discharge from Arctic draining rivers (Hinzman et al., 2005); decreasing and thinning lake and river ice (Hinzman et al., 2005); more frequent wildfires (ACIA, 2005; Anisimov et al., 2007) and shifting vegetation (Hinzman et al., 2005; Serreze et al., 2000).

Sea ice cover has also seen major changes in recent years. The September sea ice extent (the annual minimum extent) has been decreasing by around 13% each decade since satellite records began in 1979 (Perovich et al., 2014). Figure 1.6 shows the Arctic sea ice extent as of September 17^{th} 2014, summertime daily ice extent data for 2010-2014, and the 1981-2010 average and standard deviations (NSIDC, 2014). The 2012 melt season broke the previous record for minimum sea ice extent, set in 2007, before the end of the melt season and as of September 5^{th} 2012 it was already below 4 million square kilometres in extent (NSIDC, 2012b; Parkinson and Comiso, 2013). The sea ice extent minimum for 2012 was reached on September 13^{th} at 3.40 million square kilometres (NSIDC, 2012a; Parkinson and Comiso, 2013). The 2013 and 2014 melt seasons have been nearer the 1981-2010 average as shown in Figure 1.6. Less ice is surviving the summer melt so multi-year ice is decreasing (Comiso, 2012; Kwok, 2007; Maslanik et al., 2007; Nghiem et al., 2007; Parkinson and Comiso, 2013; Polyakov et al., 2012). There is also evidence that sea ice in the Arctic is thinning and decreasing in volume (e.g. Kwok et al., 2009; Maslanik et al., 2007; Parkinson and Comiso, 2013). Some studies have indicated that the Arctic will have a sea ice free summer within 30 years (Overland and Wang, 2013; Wang and Overland, 2009). As the Arctic warms and sea ice decreases there is greater access for ships. Already ship traffic is increasing due to tourism and the development of natural resources such as gas, oil, and boreal forest resources (Brigham, 2011; Cressey, 2011; Kullerud, 2011). This leads to increased risk of pollution and oil spills as well as concerns about safety for people, the Arctic environment and the ships that are in these waters (Brigham, 2011). There are also political issues as countries try to lay claim to Arctic resources exposed by the retreating ice (Cressey, 2011). As sea ice extent decreases and the sea ice thins the difficulty of deploying and maintaining manned and unmanned measuring platforms will increase. These changes also shorten the length of time these sensors can operate on ice. This increases the already difficult task of monitoring temperature changes using in situ

data in sea ice regions.

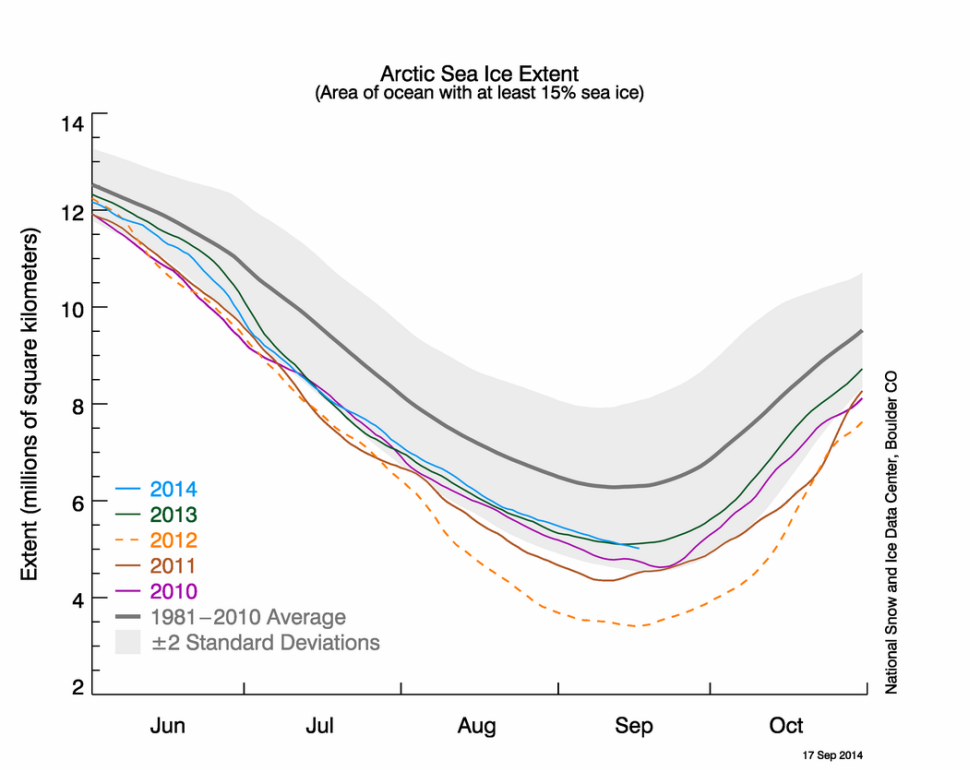

Figure 1.6: Arctic sea ice extent from NSIDC as of September 17^{th} 2014. Summertime daily sea ice extents are also shown for 2010-2014, along with the 1981-2010 average and standard deviations. Acquired from $http://nsidc.org/arcticseaicenews/2014/09/arctic-minimum-reached/$ (NSIDC, 2014).

Furthermore, changes in the Arctic do not just have an impact locally or regionally but also globally. In addition to the impacts of accelerated warming of the Arctic atmosphere both regionally and globally, greenhouse gas emissions may increase from the Arctic due to permafrost warming and the release of methane hydrates (ACIA, 2005; Anisimov et al., 2007). These emissions add to the greenhouse gas concentration in the Earth's atmosphere. Increases in glacial melt, ice sheet melt, ice cap melt and river runoff will contribute to sea level rise which will flood low-lying land. For example, the Greenland Ice Sheet is estimated to have contributed around 0.5mm per year to global sea level rise between 2002 and 2011 (Church et al., 2013; Hanna et al., 2013). This increase in melt will alter the freshwater budget of the Arctic which may affect ocean circulation (ACIA, 2005; Anisimov et al., 2007). There will also be other global impacts resulting from Arctic warming such as changes in animal diversity, ranges and distribution (ACIA, 2005). It is therefore important to monitor and quantify the changes that are happening in the Arctic. But, as the Arctic warms and sea ice extent and thickness decrease how should we monitor the temperature changes occurring in these dynamic and rapidly changing areas?

1.3.4 Monitoring Arctic Arctic Surface Temperature Change

As we have seen, monitoring climate change in the Arctic region is important because of expected and observed changes in this region. But, in situ measurements of Arctic surface temperatures, particularly in areas covered by sea ice for all or part of the year, are sparse and the records are often short. The sparseness of temperature data for the Arctic is noted as an issue by many researchers as it introduces uncertainty to the calculation of average temperature changes in this region (e.g. Cowtan and Way, 2014; Jones et al., 2012; Parker et al., 2009; Pielke et al., 2007). It is often difficult to assess how temperatures are changing in this region as well as where current climate trends fit in a long term perspective (Bromwich et al., 2007). Also, low density measurements make it hard to separate local changes from regional or continental ones (Bromwich et al., 2007). As described in Section 1.2.3 there are many different techniques and methods that are employed to quantify surface temperature changes from records that are not spatially or temporally complete. There are therefore many different techniques and methods used to quantify SAT changes over the Arctic from sparse in situ measurements.

The majority of currently available temperature anomaly datasets estimate Arctic temperature change using SAT and SST measurements. These surface temperature measurements are used exclusively and/or interpolated and extrapolated using various techniques. Datasets which use available in situ temperature measurements exclusively, for example many of the Met Office Hadley Centre temperature anomaly datasets such as HadCRUT4 (Morice et al., 2012) and HadSST3 (Kennedy et al., 2011b,c), do not spatially infill data. This means that large areas of the Arctic are unrepresented in these datasets even with recent updates to the datasets (Kennedy et al., 2011b; Morice et al., 2012). Temperature anomaly datasets which interpolate and extrapolate temperatures, such as the GISTEMP (Hansen et al., 2010; Lawrimore et al., 2011) and Berkeley Earth (Berkeley Earth, 2014; Muller et al., 2013; Rohde et al., 2013a, 2012) analyses, are spatially complete for the Arctic. Sea ice areas in both of these datasets are treated as if they were land areas with a geographical area which changes over time with sea ice extent. MLOST employs both spatial infilling and the exclusion of data. Temperature data are interpolated where possible and excluded where data are determined to be insufficient to interpolate temperature anomalies (less than 20% sampling), the target is polar land (above 75°N) with no data in the month of interest, and/or more than half the grid box is classified as sea ice (Vose et al., 2012). Therefore large areas of the Arctic are excluded (Vose et al., 2012). Temperatures over areas of Arctic sea ice are therefore either excluded or estimated by interpolating and extrapolating SATs over sea ice areas from land-based SAT sources. These datasets deal with the paucity of Arctic temperature data from traditional observations in various different ways. However, this sparseness of data also makes it difficult to investigate the impact of choices of data and methodology on Arctic surface temperature anomalies. How should Arctic sea ice areas be treated in global and regional averages of surface temperatures?

What are the uncertainties associated with using the various techniques and methods currently utilised? Can we quantify the uncertainties in extrapolation and interpolation, or not, of information over the Arctic? There are many other data sources and types of surface temperature which could be used to help quantify Arctic temperature changes. Would the inclusion of additional data sources over Arctic sea ice improve estimates of Arctic climate change? Also, which data sources would be most appropriate to monitor Arctic sea ice surface temperature change; in situ surface temperature records, satellite-based radiometric ice surface temperatures, other satellite observations, or other data sources? Could global warming over recent decades be better estimated by changing the way the Arctic is treated in calculating global mean temperature?

1.4 Aims of the Thesis

The purpose of this thesis is to investigate how the Arctic should be represented in global surface temperature time series of recent climate change. This work primarily focuses on how the sea ice regions of the Arctic should be treated in global and regional averages of surface temperatures, but also investigates estimates of surface temperature change over Arctic land and ice sheet areas. Specifically, the aims of this thesis are to:

1. Quantify the uncertainties in extrapolation and interpolation of information over the Arctic.

2. Explore whether additional data sources would improve estimates of surface temperature change over the Arctic.

3. Identify appropriate data sources which could be used to monitor Arctic surface temperature change over Arctic sea ice.

4. Provide some recommendations on how to monitor temperature changes in the Arctic, especially over sea ice areas.

In order to fulfil the aims this thesis is structured as follows. Chapter 2 considers available global and Arctic reanalyses and examines their representation of Arctic temperatures. This is in order to determine whether any of these reanalyses are suitable to use as a testbed for Chapter 3, "An Investigation into the Impact of using Various Techniques to Estimate Arctic Surface Air Temperature Anomalies", and Chapter 5, "Can the Inclusion of Additional Data Sources be used to Improve Estimates of Arctic Surface Air Temperature Anomalies?". Chapters 3 and 4 investigate the various techniques currently employed for estimating Arctic SATs and aim to quantify the uncertainties in extrapolation and interpolation, or not, of information over the Arctic. Chapter 4 additionally highlights some of the issues associated with utilising in situ data sources over Arctic sea ice areas. Chapter 5 explores whether additional data sources would improve estimates of surface temperature change over the Arctic. Potential additional data sources over

Arctic sea ice which could be employed for monitoring temperature changes in this region are also identified in Chapter 5. Finally, Chapter 6 investigates whether the inclusion of surface temperatures derived from infrared satellite sensors would improve estimates of SAT anomalies over Arctic sea ice. This chapter also explores the utility of interpolated Arctic SAT datasets as climatologies over Arctic sea ice areas. The results of this thesis are summarised in Chapter 7, conclusions are drawn from the results, and I provide some recommendations for monitoring Arctic temperature changes.

Chapter 2

Are ERA-Interim Data a Suitable Testbed for Studying How to Monitor Surface Air Temperatures in the Arctic?

2.1 Introduction

Meteorological reanalyses are datasets widely used in atmospheric science. They are produced using atmospheric forecast models run in hindcast mode to assimilate past climate and weather observations into a coherent dataset. They are data driven, dynamically constrained datasets that provide information on a wide variety of atmospheric and climatic variables. Reanalysis datasets have some advantages over more traditional, observation-only datasets for climate monitoring and research. They provide complete spatial and temporal coverage even in data sparse areas, can provide more frequent data estimates than from in situ observations, can generate information on unobserved climate variables, improve estimates of observed variables from data on related variables, and may be more consistent than traditional datasets depending on the variable (Bromwich et al., 2007; Dee et al., 2011a; Thorne and Vose, 2010).

Reanalysis data are very valuable for investigating temperatures in the Arctic, a data sparse area, due to the complete spatial and temporal coverage of reanalysis datasets. Jakobson et al. 2012 state that "reanalyses are arguably the best available source of integrated information on the four-dimensional structure of the atmosphere" for many variables. As a result, atmospheric reanalyses are often used in Arctic research to investigate climate variability, climate trends, large-scale circulations and teleconnections; to validate climate models; and to provide boundary conditions for ocean, sea ice, land-surface, and limited-area atmospheric models (Jakobson et al., 2012). However, it must be noted that reanalysis datasets contain only estimates of climatic variables. Reanalysis data do not exhibit the uncertainties and noise associated with actual measurements but

can contain biases and errors. Reanalysis is performed using the latest and most sophisticated atmospheric models, but even the best models have errors (Dee et al., 2011a). Reanalyses biases may be time varying, making these datasets unsuitable for long-term trend monitoring (depending on the time, region and variable of interest) (Thorne and Vose, 2010). Uncertainties in reanalysis data will also be higher in certain areas and for certain variables due to sparse observational data (Inoue et al., 2009). Meteorological reanalyses must therefore be used with care and with consideration of the limitations of these datasets.

The first two aims of this thesis are to quantify uncertainties in extrapolation and interpolation of information over the Arctic, and explore whether additional data sources would improve estimates of surface temperature change over Arctic sea ice. Due to the spatially and temporally sparse nature of in situ temperature measurements over large areas of the Arctic, and especially over sea ice regions, it would be difficult to investigate these topics using only in situ observations. As a result it was decided to examine these issues in the first instance using a reanalysis dataset as a testbed. The European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim) was identified as a potentially suitable testbed. But, how does this reanalysis perform in representing surface temperatures in the Arctic?

The objective of this chapter is, therefore, to examine how well ERA-Interim data represent surface temperatures in the Arctic. This is in order to determine whether ERA-Interim data are physically plausible, and therefore a suitable testbed for Chapters 3 and 5 of this thesis. The outline of this chapter is as follows. Firstly I give an overview of how surface temperatures are assimilated into, and estimated by, reanalyses in Section 2.1.1. This is to provide some insight into how surface temperatures are used in reanalyses and give some context to the comparisons detailed in the studies referenced here. Then I provide a review of the literature on reanalyses and their representation of Arctic temperatures is provided in Section 2.2. Section 2.3 describes an investigation I conducted into the representation of Arctic Surface Air Temperature (SAT) correlation length scales in ERA-Interim. Then in Section 2.4 I draw on the conclusions of Sections 2.2 and 2.3 to determine whether ERA-Interim data are a suitable testbed for studying SATs in the Arctic.

2.1.1 An Overview of Surface Temperature Assimilation and Estimation in Meteorological Reanalysis Datasets

Reanalysis of meteorological data was first found to be of use for atmospheric research after the First Global Atmospheric Research Program Global Experiment (FGGE) in 1979 (Dee et al., 2011b). Data from this experiment was reanalysed as part of work investigating how to make better use of observations to initialise numerical weather forecasts (Dee et al., 2011b). It was subsequently realised that the reanalysis datasets were themselves useful for atmospheric research (Dee et al., 2011b; Uppala et al., 2005). Since then many

atmospheric reanalysis datasets have been produced by many organisations on both global and regional scales. Global scale reanalysis datasets, which provide reanalysis data for the Arctic, include the ECMWF reanalyses ERA-15 (Gibson et al., 1999), ERA-40 (Uppala et al., 2005) and ERA-Interim (Dee et al., 2011b); the National Center for Environmental Prediction (NCEP) reanalyses R1 (Kalnay et al., 1996), R2 (Kanamitsu et al., 2002) and CFSR (Saha et al., 2006, 2010); the Japan Meteorological Agency (JMA) reanalyses JRA-25/JCDAS (Onogi et al., 2007; Ebita et al., 2011) and JRA-55 (Ebita et al., 2011); the NASA Modern Era-Retrospective Analysis for Research and Applications (MERRA) dataset (Rienecker et al., 2011); and the National Oceanic and Atmospheric Administration - Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES) 20^{th} Century Reanalysis (20CR, Compo et al. 2011). A regional reanalysis dataset has also been produced which covers the Arctic region: the Arctic System Reanalysis (ASR, Bromwich et al. 2010).

For reanalysis, forecast models are initially run with a set of boundary conditions. These boundary conditions typically include Sea Surface Temperatures (SSTs) sourced from globally complete products utilising in situ data, satellite data and interpolation techniques. Currently available reanalyses largely use the same types of surface temperature sourced from similar data products as boundary conditions for the forecast model. For example, ERA-Interim has used several different SST products produced from in situ and satellite records such as the Operational Sea Surface Temperature and Sea-Ice Analysis (OSTIA, Donlon et al. 2012) (Dee et al., 2011b), MERRA uses the Reynolds et al. 2002 SST product, and 20CR (Compo et al., 2011) uses the Met Office Hadley Centre's interpolated SST and sea ice concentration product (HadISST, Rayner et al. 2003). These initial runs produce forecast fields of atmospheric variables.

Available observations are then assimilated in order to correct the forecast fields produced by the forecast model (Jakobson et al., 2012; Kalnay et al., 1996). Data assimilation is the process by which "irregularly distributed atmospheric observations are combined with a general circulation model (GCM) to provide a comprehensive four-dimensional representation of the global atmosphere" (Bloom et al., 1996). Many currently available reanalyses use a variational analysis, either three-dimensional (3D-Var, e.g. MERRA, NCEP reanalyses) or four-dimensional (4D-Var e.g. ERA-Interim upper-air atmospheric state, JRA-55) to combined observations with the forecast model fields (Dee et al., 2011b; Ebita et al., 2011; Kalnay et al., 1996; Kanamitsu et al., 2002; Rienecker et al., 2011; Saha et al., 2006, 2010). In 4D-Var, forecast fields, produced using the forecast model, are combined with observations during a given timestep (often a 6 hour time period) at the time of each observation using a best linear unbiased estimate method (Lorenc and Rawlins, 2005; Bloom et al., 1996). In 3D-Var all observations are assumed to observe in the middle of each timestep (Lorenc and Rawlins, 2005; Bloom et al., 1996). The surface temperature observations used for data assimilation in reanalyses are generally in situ SST records (for example, from ships, fixed buoys, drifting buoys, ice-pack buoys, drifting

stations) and SAT records from land-based meteorological stations. These SST and SAT observations are very similar between reanalysis datasets and are often sourced from the same or similar existing databanks. For example, MERRA (Rienecker et al., 2011) and NCEP reanalyses both use SATs from surface land observations collected for NCEP, and SSTs collated in the Comprehensive Ocean-Atmosphere Data Set (COADS)or the newer International COADS (ICOADS).

Many reanalyses assimilate in situ surface temperature observations into forecast fields at surface or near-surface model levels. SAT is then estimated by interpolating between the analysed skin temperature and the lowest atmospheric model level temperature to 2m height (Lindsay et al., 2014). This occurs in the NCEP reanalyses (Kalnay et al., 1996; Kanamitsu et al., 2002; Saha et al., 2006, 2010), MERRA (Rienecker et al., 2011), JMA reanalyses (Onogi et al., 2007; Ebita et al., 2011), and ASR (Bromwich et al., 2014). Once data has been assimilated into the forecast fields, atmospheric and climatic variables are produced at various model levels. Several iterations of the forecast model may be used in the assimilation (Bromwich et al., 2014). The analyses produced are then used as the initial conditions for the next model run.

However, there are some exceptions. The ERA-Interim reanalysis assimilates surface temperature observations separately from the main atmospheric analysis. The SAT fields included in the ERA-Interim products are explicitly analysed using separate univariate optimal interpolation, rather than derived from temperatures at other model levels (Simmons et al., 2004). In situ SAT and SST observations from platforms such as land stations, buoys and ships are combined with background fields of SAT (forecast fields from the previous analysis time step of the forecast model interpolated between model levels) to create each surface analysis (Douville, 1998; ECMWF, 2007; Simmons et al., 2004). The difference between the observation and background value is calculated for each observation location. Analysis increments, which are added to the background field to create the surface analysis at the current timestep, for each model grid point are calculated from weighted background increments . Optimum weights for the background increments are calculated by solving matrix equations for each model grid point using estimated error covariances. The estimated error covariances between background values and the model grid point, and between background values at different locations, are then calculated using the distance between these locations to produce a surface analysis field. This method has similarities with the kriging methods employed in various chapters of this thesis and outlined in Chapter 3 and Appendix B.

Finally, 20CR does not assimilate surface observations of temperatures and therefore only produces forecast fields rather than assimilated fields of data (Jakobson et al., 2012). The 20CR dataset only assimilates surface pressure reports from the International Surface Pressure Databank (ISPD), with SST and sea ice distribution as boundary conditions, unlike other reanalyses which assimilate many in situ and satellite reports and many atmospheric variables (Compo et al., 2006, 2011). The SAT estimate from 20CR is therefore

not influenced by surface temperature observations.

2.2 Meteorological Reanalysis Datasets and their Representation of the Arctic

In this section I provide a review of the literature on each of the reanalyses mentioned above and their representation of Arctic temperatures and temperature trends. The literature review and the performance of the reanalyses in the Arctic is summarised in Table 2.2.7. Information about whether the comparison data used in each study are independent, or whether they are assimilated into each reanalysis, is also summarised in Table 2.2.7.

2.2.1 National Centers for Environmental Prediction Reanalyses

The first reanalysis produced by NCEP, in collaboration with the National Center for Atmospheric Research (NCAR), was the NCEP-NCAR reanalysis, or NCEP R1. This reanalysis covers the time period 1948 to present (Kalnay et al., 1996). Errors and biases were noted in NCEP R1 and a corrected version of R1, often referred to as the NCEP or NCEP/DOE R2 dataset (hereafter referred to as NCEP R2), was produced for the satellite era; 1979 to present (Kanamitsu et al., 2002; Saha et al., 2010). However, even with the correction of identified errors and updates to the system, only minor differences were found between NCEP R1 and R2 (Kanamitsu et al., 2002). These two reanalyses continue to be updated by NCEP. A newer, coupled global NCEP reanalysis, known as the NCEP Climate Forecast System Reanalysis (NCEP CFSR), is also available, which covers the same time period as R2 but with a higher temporal and spatial resolution (Saha et al., 2006, 2010).

2.2.1.1 NCEP R1

NCEP R1 SATs over sea ice were compared data from Arctic and Antarctic Research Institute (AARI) North Pole Drifting stations (NPDS) between 1954 and 2006 by Makshtas et al. 2007. Makshtas et al. 2007 found that SATs from NCEP R1 and North Pole Drifting stations largely showed good agreement; correlation was more than 0.8 in all seasons, except summer when the correlation value dropped to 0.55. The Mean Errors (MEs) were noticeably seasonal. Biases were as low as 0.2K (winter) and as large as -2.6K (autumn) (Makshtas et al., 2007). SATs from NCEP R1 have also been compared to independent in situ rawinsonde observations over the Arctic Ocean from the Coordinated Eastern Arctic Research Experiment (CEAREX, September 1988 to May 1989) by Bromwich and Wang 2005. Mean monthly SATs were underestimated for the first few months of CEAREX (September to November) by 2-3K, then for the remaining months of the study CEAREX

temperatures were overestimated by 1-2K with lower biases in warmer months (Bromwich and Wang, 2005). NCEP R1 was outperformed by two ERA reanalysis in terms of the Root Mean Square Error (RMSE); the NCEP RMSE was 2.8K while RMSEs were 2.4K and 1.7K for ERA-15 and ERA-40 respectively. The findings of the study by Bromwich and Wang 2005 correspond with those of Makshtas et al. 2007; autumn temperatures were underestimated by around 2-3K and winter/spring temperatures were overestimated by up to 2.5K on average.

Wang et al. 2011 compared SAT data from all NCEP reanalyses and ERA-40 to SATs from land-based meteorological stations in the Global Historical Climatology Network version 2 and the Climate Anomaly Monitoring System dataset (GHCNCAMS, Fan and Van den Dool 2008). R1 SATs were spatially correlated with those from GHCNCAMS over land and, excluding the Greenland ice sheet, the correlation was more than 0.7 for the majority of Arctic land areas and above 0.9 for north-west Russia (Wang et al., 2011). Spatial patterns of Arctic SAT were therefore fairly well estimated by NCEP R1, but it was the least representative of the reanalyses investigated (Wang et al., 2011). A similar study by Lindsay et al. 2014 compared SATs over Arctic land areas from various reanalyses, including NCEP R1, to monthly average land station data from the Climate Research Unit at the University of East Anglia and Met Office Hadley Centre land surface temperature anomaly version 3 dataset (CRUTEM3, Brohan et al. 2006). SATs from R1 had a monthly ME of between -2 and 1K when compared to Arctic land stations (Lindsay et al., 2014). Monthly SAT anomaly correlations were 0.75-0.91 (Lindsay et al., 2014). The biases in the study by Lindsay et al. 2014 had a similar seasonality, and were also similar in size, to those noted over Arctic sea ice areas by Makshtas et al. 2007. Biases were lowest in winter, increased in spring, decreased slightly in summer and increased again in autumn. SAT correlations in both studies were above 0.8 for most seasons except summer when correlations were lower (Lindsay et al., 2014; Makshtas et al., 2007).

Lower Tropospheric Temperature (TLT) trends in the Arctic derived from four reanalyses (ERA-40, ERA-Interim, JCDAS and NCEP R1) were compared to Remote Sensing Systems (RSS) Microwave Sounding Unit (MSU) TLT by Screen and Simmonds 2011. TLT trends are known to be highly correlated with surface temperatures, which are the focus of this review (Hartmann et al., 2013). In order to compare the reanalysis data, which is produced for many pressure levels in the atmosphere, with MSU data the reanalysis temperature fields were vertically weighted using MSU weighting functions, i.e., heavily weighted to be representative of the tropospheric temperature below 3km. R1 produced realistic trends in Arctic temperatures between 1979 to 2009 in the lower troposphere when compared to MSU data (Screen and Simmonds, 2011). Biases were up to $\pm$0.3K and the annual temperature trend in R1 was within the 95% confidence interval of the MSU data (Screen and Simmonds, 2011). SAT trends in the Arctic from NCEP R1 were also compared to the International Arctic Buoy Programme/Polar Exchange at the Sea Surface (IABP/POLES) dataset (Rigor et al., 2000) by Alexeev et al. 2012 between

1979 and 2008. It should be noted that monthly mean SAT data from NCEP R1 were included in the IABP/POLES dataset to supplement in situ observations in areas where observations were sparse, such as in the North Atlantic. Spatially the SAT trends across the Arctic were consistent in most seasons (Alexeev et al., 2012). There were some quite large differences between R1 and IABP/POLES data, e.g. for the Beaufort Sea in Winter where R1 overestimated the rate of warming and errors could be up to 10K in some years before about 1995 (Alexeev et al., 2012). However, SATs from IABP buoy data prior to 1992 may not be reliable due to the design of the buoy in the early years of this program (Rigor et al., 2000). In addition, there are other known biases in the IABP/POLES dataset, such as due to snow cover on buoys, and work is still ongoing to identify all biases in these data (EOL, 2014; Rigor, 2014). Therefore I believe the differences between NCEP R1 and IABP/POLES noted by Alexeev et al. 2012 may not be a true reflection of the performance of R1 for Arctic SAT trends.

2.2.1.2 NCEP R2

SATs from NCEP R2 (and ERA-40) were investigated by Liu et al. 2008 and compared to IABP/POLES dataset over the Arctic Ocean between 1979 and 1999. R2 showed good agreement with the IABP/POLES dataset. There was a slight warm bias of 0.7K for annual mean SATs over the Arctic Ocean (half of the warm bias in ERA-40) and SATs were within 2K of IABP/POLES annual SATs spatially (Liu et al., 2008). Seasonal variations were also reproduced well and the seasonal biases were smaller than those for ERA-40. R2 had a warm bias in seasonal mean SAT for most seasons of between 0.54 and 1.83K and a slight cold bias of -0.24K in summer (Liu et al., 2008). However, as mentioned before, NCEP R1, which is quite similar to NCEP R2, is included in the IABP/POLES dataset where data is sparse and therefore the comparison data are not completely independent. NCEP R2 and several other reanalyses were compared to vertical air temperature profiles collected by tethersondes at ice station Tara between 2006 and 2007 by Jakobson et al. 2012, and their relative performance was compared over the central Arctic ocean. The surface warm bias noted by Liu et al. 2008 for the central Arctic Ocean in NCEP R2 was also noted in the study by Jakobson et al. 2012. NCEP had a warm bias of 0.2K on average and a RMSE of 1.8K for near surface temperatures at 10m height (Jakobson et al., 2012). NCEP R2 produced the second smallest RMSEs compared to the other reanalysis datasets investigated in that study for temperatures close to the surface (Jakobson et al., 2012). However, significant errors in SAT were also observed for some NCEP R2 profiles; for example spring SAT errors could be around 15K.

NCEP R2 data were compared to land area SATs from GHCNCAMS (Wang et al., 2011). R2 showed the same spatial correlation patterns with GHCNCAMS as NCEP R1 (>0.7 generally, above 0.9 for north-west Russia), and therefore also produced realistic reconstructions of Arctic SAT patterns (Wang et al., 2011). The study of Lindsay et al. 2014 compared R2 data to SATs over Arctic land areas from CRUTEM3 stations. R2

had very similar biases in SAT for spring and summer compared to R1 (Lindsay et al., 2014). In autumn the cold bias reduced from -2 to -1K in September and October, and November had a warm bias of 0.5K in R2. This means that R2 was more representative of autumn SATs on average than R1. However, in winter the warm bias increased from 0.5K to 2K which made R2 less representative on average in this season. Monthly SAT anomaly correlations were higher, or the same as R1, in all months. Therefore R2 was slightly more representative than R1 but it was not the most representative reanalysis for Arctic land areas in that study (Lindsay et al., 2014).

In addition to investigations of temperature patterns and observations, the Arctic warming trend contained in NCEP R2 has been investigated (Chung et al., 2013). SAT anomalies from several reanalyses, including NCEP R2, were compared to NASA Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP) dataset (with a 250km smoothing) between 1979 and 2011 (Chung et al., 2013). All reanalysis warming trends were largely within 0.5K of each other and GISTEMP (Chung et al., 2013). The results were very close for all investigated reanalyses in that study and no one reanalysis was notably better than the others. The spread in the warming trends in the reanalyses was larger in the winter than in the summer (Chung et al., 2013). Liu et al. 2008 compared R2 SAT trends to those in the IABP/POLES dataset over Arctic ocean areas. NCEP R2 overestimated SAT trends compared to IABP/POLES; the annual mean was overestimated by 0.15K and seasonal trends, whether positive or negative, were overestimated by 0.01 (winter and spring) to 0.29 (autumn) (Liu et al., 2008). IABP/POLES data for the Arctic in general were also compared to NCEP R2 data by Alexeev et al. 2012. As with R1, many of the warming patterns were similar across the Arctic compared to IABP/POLES but R2 had more areas of negative temperature changes and there were also some areas of disagreement (Alexeev et al., 2012). R2 showed winter SAT changes over the Beaufort Sea which were much closer to IABP/POLES than R1 SAT trends (within 5K) (Alexeev et al., 2012).

2.2.1.3 NCEP CFSR

NCEP CFSR was compared to atmospheric profiles collected at drifting ice station Tara by Jakobson et al. 2012. The results for CFSR were very similar to those for NCEP R2; CFSR had a slight warm bias of around 0.3K on average for near surface temperatures, which was a slightly larger bias than for R1, but a RMSE of 1.6K, which showed slightly better agreement between CFSR and the in situ data (Jakobson et al., 2012). NCEP CFSR performed better for air temperatures below 200m and the profiles did not contain the significant errors in SAT noted for some profiles from R2 (Jakobson et al., 2012). SATs in the Arctic from NCEP CFSR were also compared to SATs from the GHCNCAMS dataset (Wang et al., 2011). CFSR was the most representative of the NCEP analyses for spatial patterns of global SAT in general, as well as performing better over the Arctic (Wang et al., 2011). Much of the Arctic had a SAT correlation with GHCNCAMS of

0.9 or more in CFSR (Wang et al., 2011). The study by Lindsay et al. 2014 also noted that CFSR outperformed R1 and R2 compared to land-based meteorological stations in terms of correlation (0.85-0.95 compared to 0.75-0.92 for R1/R2). Furthermore, CFSR had smaller biases compared to R1 and R2 in between April and October (up to -1K), and was one of the most representative reanalyses in the study of (Lindsay et al., 2014). Temperature trends in CFSR were investigated by Chung et al. 2013 and all reanalyses were within about 0.5K of each other and GISTEMP, as mentioned previously (Chung et al., 2013). CFSR therefore produced realistic Arctic SAT warming trends (Chung et al., 2013).

2.2.2 Japan Meteorological Agency Reanalyses

The JMA's numerical assimilation and forecast system has been used to create several global atmospheric reanalyses. JRA-25 was the first JMA reanalysis and is available for the time period 1979-2004 (Onogi et al., 2007). Since 2004 JRA-25 has been continued under a new name, JCDAS (JMA Climate Data Assimilation System reanalysis), and continues to be updated (Ebita et al., 2011). Production of a new 55 year reanalysis spanning 1958 to present, JRA-55, started in 2010 (Ebita et al., 2011). Data for JRA-55 were made available in January 2014 (JRA Project, 2014) and a comprehensive report on this new reanalysis has been produced (Kobayashi et al., 2015).

2.2.2.1 JRA-25/JCDAS

The JCDAS reanalysis was validated by Jakobson et al. 2012 for the central Arctic Ocean. JCDAS had the largest bias for average near surface temperatures (-3K), the largest RMSE for near surface temperatures (4.5K), and the largest RMSE values in that study for the whole temperature profile (Jakobson et al., 2012). Significant errors in SAT were also observed for some JCDAS profiles (Jakobson et al., 2012). However, when compared to monthly average land station SAT data by Lindsay et al. 2014, JCDAS was found to perform quite well and was one of the four most representative reanalyses in that study (Lindsay et al., 2014). JCDAS had smaller SAT biases than NCEP R1, R2, CFSR (in some months), and the 20CR reanalysis (Lindsay et al., 2014). It also had very good SAT correlation with CRUTEM3 station observations in most months; correlations were between 0.88 and 0.93 and were more consistent seasonally than for the NCEP reanalyses and 20CR (Lindsay et al., 2014). Therefore, JCDAS seems to perform well over land areas in the Arctic for SATs but has relatively large errors for temperatures over sea ice areas. Temperature trends in the Arctic as reconstructed by JCDAS were compared to MSU TLTs by Screen and Simmonds 2011. All reanalyses produced realistic trends in Arctic temperatures in the lower troposphere when compared to MSU data, apart from ERA-40 (Screen and Simmonds, 2011). JCDAS produced biases in annual TLTs which were similar to NCEP R1 and ERA-Interim (up to $\pm$0.3K) and the annual temperature trend was again within the 95% confidence interval of the MSU data (Screen and Simmonds,

2011). However, ERA-Interim outperformed all other reanalyses, including JCDAS, in terms of realistic trends in that study.

2.2.2.2 JRA-55

Currently, one study has investigated the performance of the JRA-55 reanalysis for SATs in the Arctic. Simmons and Poli 2014 explored the representation of near-surface and TLT change over the Arctic in several datasets for the past 35 years. For SAT and lower tropospheric anomaly timeseries above 70°N, JRA-55 was within 0.5K of ERA-Interim, MERRA, CRUTEM4 and the Cowtan and Way temperature anomaly dataset (Cowtan and Way, 2014). JRA-55 was also similar to ERA-Interim (but with a smaller warm bias) and the Cowtan and Way dataset for spatial anomalies (Simmons and Poli, 2014). Therefore, JRA-55 shows promise for investigating Arctic surface temperatures in the future. However, JRA-55 was released after work had started on Chapter 3 and was therefore not utilised in this thesis. In addition, I think further studies would be beneficial in terms of affirming how well this new reanalysis represents Arctic temperatures and temperature change.

2.2.3 The NASA Modern Era-Retrospective Analysis

MERRA is NASA's reanalysis project for the satellite era (1979 to present) (Rienecker et al., 2011). MERRA is created using version 5.2.0 of the Goddard Earth Observing System (GEOS) atmospheric model and data assimilation system (Rienecker et al., 2011). A new reanalysis from 1980 to present, MERRA2, is in production (Reanalyses.org, 2014). MERRA2 will use a new version of GEOS-5 and will assimilate additional observations in order to extend NASA reanalyses into the future (Reanalyses.org, 2014).

MERRA air temperature data were compared to observations at ice station Tara by Jakobson et al. 2012. The RMSE for MERRA air temperatures was found to be as good as the NCEP reanalyses overall for the temperature profile (Jakobson et al., 2012). However, the RMSE near the surface was above 3K (only the RMSE for JDCAS was larger) and MERRA was found to have a significant warm bias of about 2K for air temperature below 200m (Jakobson et al., 2012). When compared to SATs from Arctic land-based meteorological stations, MERRA provided good estimates of SATs and SAT anomalies and was one of the most representative reanalyses according to Lindsay et al. 2014. The monthly ME in MERRA compared to CRUTEM3 stations was ±0.5K; lower errors than all other reanalyses investigated in that study, except ERA-Interim which was comparable (Lindsay et al., 2014). Monthly correlations were also consistently good with values of 0.89 to 0.93 (Lindsay et al., 2014). Only ERA-Interim and NCEP CFSR (in some months) had higher correlation values (Lindsay et al., 2014). MERRA was also found to perform well for SAT and lower tropospheric anomalies by Simmons and Poli 2014 and this reanalysis was close (within 0.5K) to the timeseries and anomalies from ERA-Interim, JRA-55, CRUTEM4 and the Cowtan and Way dataset. However, MERRA

showed weaker surface warming in recent years compared to other datasets investigated in that study (Simmons and Poli, 2014). MERRA Arctic warming trends were investigated were found to be realistic (Chung et al., 2013).

2.2.4 The NOAA-CIRES 20th Century Reanalysis

The 20^{th} Century Reanalysis project, or 20CR, is a comprehensive global atmospheric circulation dataset covering the time period 1871 to present (Compo et al., 2011). The 20CR dataset uses SSTs and sea ice concentration as boundary conditions and only assimilates surface pressure reports, unlike other reanalyses which assimilate many in situ and satellite derived observations of atmospheric variables (Compo et al., 2006, 2011).

20CR was another of the reanalyses investigated for Arctic SATs over land by Lindsay et al. 2014. 20CR was the least representative reanalysis in this study. A large winter bias of up to 6K was present in 20CR while summer biases were up to 1K (Lindsay et al., 2014). The correlation with SAT anomalies was up to 0.87 in some months, but could be as low as 0.63 (May) whereas all other reanalyses had correlation values above 0.7 in all months (Lindsay et al., 2014). Lindsay et al. 2014 attributed this to a bad sea ice specification. This indicates that 20CR may also have problems for SATs over Arctic Ocean and sea ice areas, but this has not yet been confirmed in the literature. But, 20CR has been shown to provide realistic estimates of global SAT trends. 20CR provided an independent confirmation, without the use of in situ temperature observations, of the global warming observed over land (Compo et al., 2013). Comparison with annual SAT trends over land from 8 different datasets showed temporal correlations of 0.84 to 0.92, and correlations of 0.74-0.81 when a 7 year running mean was used (Compo et al., 2013). Therefore, 20CR represents global SAT trends well but was found to be less representative of Arctic SATs over land than many currently available reanalyses.

2.2.5 The Arctic System Reanalysis

ASR is a high resolution regional reanalysis for the Arctic (Bromwich et al., 2010). This Arctic specific analysis is created using a Polar optimised Weather Research and Forecasting (Polar WRF) model developed and tested for Arctic ice sheets, ocean/sea ice and land areas (Bromwich et al., 2009, 2010; Hines and Bromwich, 2008; Hines et al., 2011; Wilson et al., 2011).

SATs have been investigated for model simulations of the Polar WRF model version 2.2. SATs were found to show fairly good agreement with in situ observations from the Surface Heat Budget of the Arctic (SHEBA) drifting ice station for January, June and August 1998 (Bromwich et al., 2009). Polar WRF SATs were compared to SHEBA 2.5m temperatures (temperatures interpolated to 2.5m from tower observations). The correlation values were 0.81, 0.48 and 0.61 for January, June and August respectively (Bromwich et al., 2009). Biases were between -1.8 and 0.2K and RMSE values were 0.9-

4K (Bromwich et al., 2009). SATs from the Polar WRF model version 3.1 simulations were compared with in situ observations from the National Climatic Data Center and ERA-Interim data (Wilson et al., 2011), and in situ observations from the north slope of Alaska as well as other Arctic land areas (Hines et al., 2011). Polar WRF v3.1 was in general agreement for annual SATs in terms of spatial patterns over land areas, with correlations of 0.76-0.81 for annual SATs, but underestimated annual SATs by up to 1.6K with a RMSE of 4.4 (Wilson et al., 2011). For monthly SATs, both studies noticed a warm bias in summer but disagreed about the presence of a warm bias in winter, which was noted by Hines et al. 2011 but Wilson et al. 2011 observed a cold bias in most seasons including winter and for most of the area studied (attributed to the model version used for this reanalysis). Both studies do agree, however, about the size of monthly correlation values (0.6 to 0.9) and RMSE values (in the range 2-6.7K) (Hines et al., 2011; Wilson et al., 2011). ME values showed much larger warm biases, especially for certain stations in the study by Hines et al. 2011. Therefore, the investigations completed so far suggest that ASR may provide realistic estimates of Arctic temperatures, given the performance of its model simulations. However, more research is needed to confirm this and some large errors were noted in the study by Hines et al. 2011.

2.2.6 European Centre for Medium-Range Weather Forecasts Reanalyses

ECMWF first produced a reanalysis as part of the FGGE in 1979 (Dee et al., 2011b). After the value of reanalysis was recognised, ECMWF continued to produce reanalysis datasets. ERA-15, a 15 year reanalysis for 1979 to 1993, was the first ECMWF reanalysis (Gibson et al., 1999). The ERA-40 reanalysis project, 1957 to 2002, followed ERA-15 and benefited from an enhanced observing system as well as updates and changes which were informed by the ERA-15 project (Uppala et al., 2005). ERA-Interim is the current ECMWF reanalysis dataset. It was designed as an interim reanalysis of the period 1989-present in preparation for an extended reanalysis due to replace ERA-40 (Dee et al., 2011b). ERA-Interim has since been extended to 1979. Two new ECMWF reanalyses, ERA-20C and a high resolution, near real-time satellite era reanalysis, are currently in production as part of the ERA-CLIM2 project (ERA-CLIM, 2014; Reanalyses.org, 2014).

2.2.6.1 ERA-15

ERA-15 was one of three reanalyses compared to in situ rawinsonde datasets over the Arctic Ocean by Bromwich and Wang 2005. ERA-15 was found to overestimate mean monthly surface temperatures for the first few months of CEAREX (September to November 1988) by the same amount as NCEP R1, about 2-3K (Bromwich and Wang, 2005). After this temperatures were overestimated by both ERA-15 and NCEP R1 by 1-2K (Bromwich and Wang, 2005). ERA-15 outperformed NCEP R1 overall as RMSEs were

lower (2.4K for compared to 2.8K), but was in turn outperformed by ERA-40 which had a RMSE of 1.7K (Bromwich and Wang, 2005).

2.2.6.2 ERA-40

SATs from ERA-40 were investigated in the Arctic by comparing them to IABP/POLES observations from 1979-1999 by Liu et al. 2008. The reanalysis generally showed good agreement with the in situ observations of SATs and seasonal variations were reproduced well, as was also found for NCEP R2 in that study (Liu et al., 2008). However, ERA-40 had a consistent warm bias in all seasons (0.48 to 2.05K), as well as a warm bias in the annual mean (1.48K) which was twice the size of the the annual mean warm bias in NCEP R2 (Liu et al., 2008). The warm biases were seen for almost every area of the Arctic and interannual variability was underestimated (Liu et al., 2008). But, as IABP/POLES includes NCEP R1 data in some areas, which is very similar to NCEP R2, this may impact conclusions about the relative performance of ERA-40. As mentioned previously, ERA-40 was also compared to rawinsonde data by Bromwich and Wang 2005. It was found that all three reanalyses had very similar ME values, but ERA-40 outperformed both the ERA-15 and NCEP R1 for RMSE values. Again a warm bias in ERA-40 was noticed for many months and the biases were of similar sizes in those two studies (Bromwich and Wang, 2005; Liu et al., 2008).

ERA-40 was compared to SATs from GHCNCAMS globally and was found to have the highest correlation values for SATs over land (Wang et al., 2011). ERA-40 had correlations of more than 0.9 for the majority of Arctic land areas, excluding the Greenland ice sheet, with only small areas of 0.7-0.9 correlation values (Wang et al., 2011). ERA-40 data have also been compared to radiosonde profiles located above 70°N by Dee and Uppala 2009. Near surface temperature from ERA-40 had a very small warm bias of 0.2K (Dee and Uppala, 2009). However, it was outperformed slightly in this regard by ERA-Interim (Dee and Uppala, 2009).

ERA-40 has been investigated for surface temperature trends. Compared to SAT seasonal and annual timeseries in the IABP/POLES dataset, ERA-40 overestimated annual warming by 0.07K, underestimated spring warming by the same amount, and overestimated summer/autumn warming and winter cooling trends by up to 0.15K (Liu et al., 2008). It therefore outperformed NCEP R2 in terms of annual and seasonal temperature trends compared to IABP/POLES. However, issues were noted for Arctic temperature trends in ERA-40 by Screen and Simmonds 2011. The performance of ERA-40 in the Arctic was assessed using various reference datasets from satellites, radiosondes and other reanalyses (Screen and Simmonds, 2011). A discontinuity occurring in 1997 in ERA-40 was shown to lead to exaggerated warming of the Arctic troposphere (Screen and Simmonds, 2011). ERA-40 had a warm bias of up to 0.6K relative to MSU TLT after 1997 and as a result was outside the 95% confidence interval for decadal trends (Screen and Simmonds, 2011). ERA-40 can also be more than 1K warmer for decadal trends than

other reanalyses after 1997 (Screen and Simmonds, 2011). The discontinuity was also noted when atmospheric temperatures at various levels above the surface were compared with radiosonde data (Screen and Simmonds, 2011). These issues were not observed in ERA-Interim (Screen and Simmonds, 2011). This means that ERA-40 is not a good choice of reanalysis for examining Arctic temperature trends in recent decades (Screen and Simmonds, 2011).

2.2.6.3 ERA-Interim

ERA-Interim was found to outperform other reanalyses (ERA-40, JCDAS, NCEP CFSR, NCEP R2 and MERRA) in biases and RMSE for SATs (Lindsay et al., 2014). Biases in ERA-Interim were comparable to MERRA (±0.5K) but correlation values were higher (0.92-0.97) than noted for MERRA and all other reanalyses investigated in that study in all months (Lindsay et al., 2014). Lindsay et al. 2014 suggested that this is because ERA-Interim uses the observed SAT and background temperatures from the previous analysis time step in an optimal interpolation scheme rather than estimating the SAT by interpolating between the skin and lowest modelled temperatures. ERA-Interim has also been found to perform well in depicting tropospheric (<3K bias) and near surface (1000hPa) temperatures (0.1K warm bias) over the Arctic (Dee and Uppala, 2009), fairly well for temperatures from the surface (2K warm bias) up to 2km (<3K warm bias, Lüpkes et al. 2010 and fairly well (2K warm bias, 2.4K RMSE) for near surface temperatures at 10m height (Jakobson et al., 2012). However, ERA-Interim does suffer from a consistent warm bias in the lower 400m layer of the atmosphere of up to 2K (Jakobson et al., 2012). This bias was also noted by Lüpkes et al. 2010 and Simmons and Poli 2014, and is a long standing problem for the ECMWF reanalyses (Beesley et al., 2000; Curry et al., 2002; Jakobson et al., 2012; Lüpkes et al., 2010; Screen and Simmonds, 2011; Tastula and Vihma, 2011; Vihma et al., 2002).

ERA-Interim Arctic climate change trends were investigated along with other reanalysis datasets and compared to both the Cowtan and Way and GISTEMP datasets. All investigated reanalyses, including ERA-Interim, showed similar warming trends in the Arctic when compared to each other and to SAT and TLT datasets (Chung et al., 2013; Simmons and Poli, 2014). All reanalyses were largely within 0.5K of each other and other investigated datasets (Chung et al., 2013; Simmons and Poli, 2014). ERA-Interim was also found to reconstruct temperature trends in the Arctic well compared to other reanalyses and satellite datasets for ME values (Screen and Simmonds, 2011). ERA-Interim was found to produce the most realistic trends in Arctic temperatures in the mid- and lower troposphere compared to MSU data (Screen and Simmonds, 2011).

2.2.7 Discussion

For surface temperatures in the Arctic, ERA-Interim performed very well compared to in situ observations and other reanalyses. Despite a consistent warm bias, ERA-Interim

produced generally realistic estimates of surface temperatures and surface temperature temperature trends over Arctic land and sea ice areas. In Table 2.2.7 it can be seen that ERA-Interim was ranked first in all studies investigating Arctic SATs at around 2m height. Biases and RMSE values were relatively low and correlation values were higher compared to other reanalyses. ERA-Interim outperformed, or was comparable to, other reanalyses in terms of SAT patterns, anomalies, and SAT trends in the studies identified. ERA-Interim also performs quite well for SATs at other heights and TLT and is often ranked first.

However, in the study by Jakobson et al. 2012, NCEP CFSR was ranked first when reanalyses were compared to SATs at 10m height from tethersondes. ERA-Interim was also outperformed by NCEP R2. This suggests that ERA-Interim may not be the most realistic reanalysis for air temperatures above 2m. It must also be noted that, although ERA-Interim provides realistic estimates of Arctic temperatures and temperature trends for areas of the Arctic studied thus far it is not necessarily a good representation of Arctic SATs in other uninvestigated areas. Temperatures at other model levels are used as background fields for the separate surface analysis in ERA-Interim. If these background fields are not realistic then ERA-Interim temperature fields may not provide a good estimate of SATs far from observations. This may be a problem for the Arctic where temperature observations are spatially and temporally sparse; the assumed horizontal correlation coefficients of background error covariances in ERA-Interim are around 0.5 at 350km and reach 0 at 1000km (Appendix B, Simmons et al. 2004). However, ERA-Interim and its precursor reanalysis ERA-40, do provide realistic estimates of Arctic surface temperatures when compared to independent data sources and ERA-Interim surface temperatures above 2m did not contain errors much larger than those for SATs at 2m height. This suggests that ERA-Interim will provide realistic Arctic SATs even in unobserved areas.

There is the potential for newer reanalyses, such JRA-55, to provide better estimates of Arctic SATs compared to ERA-Interim. JRA-55 was found to be comparable to ERA-Interim in the studies identified (Table 2.2.7). However, as these reanalyses have only become available in the last few years, sufficient research is not yet available to draw firm conclusions about the relative performance of these new reanalyses compared to ERA-Interim. In addition, JRA-55 was released after work started on Chapter 3 of this thesis, so although it appears to perform well in the Arctic it was not included. Therefore, the literature discussed here suggests that ERA-Interim was the most suitable reanalysis to use as a testbed for investigating Arctic SATs of the reanalyses available at the time of the study detailed in Chapter 3.

Table 2.1: A summary of the performance of currently available reanalyses for Arctic temperatures, including information on whether the comparison data for each study is independent or assimilated into the reanalysis. Acronyms used in this table are: Decadal (D), Annual (A), Seasonal (SNL), Monthly (M), Spatial (SPA) and Timeseries (TS).

Variable	Time Period	Comparison Data	Indepe-ndent?	Rank	Reanalysis	Bias/ME (K)	RMSE (K)	Correlation	Reference	Comments
Arctic Ocean and Sea Ice										
SAT (SNL)	1957-2006	AARI NPDS	N		NCEP R1	-2.6 - 0.2		0.55-0.92	(Makshtas et al., 2007)	
SAT (M)	1988-1989 (sep-nov, dec-may)	CEAREX rawinsonde	Y	1	ERA-40	-2 - -3, 1 - 2	1.7		(Bromwich and Wang, 2005)	
				2	ERA-15	-2 - -3, 1 - 2	2.4			
				3	NCEP R1	-2 - -3, 1-2	2.8			
SAT	1979-1999	IABP/POLES	Y	1	NCEP R2	0.7 (A), -0.24 - 1.83 (SNL)			(Liu et al., 2008)	IABP/POLES: includes NCEP R1 data, known biases.
				2	ERA-40	1.48 (A), 0.48 - 2.05 (SNL)				
SAT	1998 (Jan, Jun, Aug)	SHEBA	N		Polar WRF v2.2 (ASR)	-1.8, 0.2, 0.1		0.81, 0.48, 0.61	(Bromwich et al., 2009)	
SAT (D)	1981-2013	Ships/Buoys, Ice stations	N		ERA-Interim	-0.5-1.4, 3.1-4.5			(Simmons and Poli, 2014)	
SAT (10m)	2006-2007	Ice Station Tara tethersondes	Y	1	NCEP CFSR	0.3	1.6		(Jakobson et al., 2012)	NCEP R2, JDCAS: very large errors (15K) in certain profiles.
				2	NCEP R2	0.2	1.8			
				3	ERA-Interim	2	2.4			
				4	MERRA	2	3.3			
				5	JDCAS	-3	4.5			
SAT (30m), TLT	1996, 2001, 2007	Polarstern, rawinsondes	N		ERA-Interim	$\leq$2 (30m), <3 (to 2km)			(Lüpkes et al., 2010)	

SAT (TS)	1979-1999	IABP/POLES	Y	1	ERA-40	0.07 (A), 0.07 - 0.15 (SNL)			(Liu et al., 2008)	IABP/POLES: includes NCEP R1 data, known biases.
				2	NCEP R2	0.15 (A), 0.01 - 0.29 (SNL)				
Arctic Land										
SAT (M)	1981-2010	CRUTEM3	Y	1	ERA-Interim	±0.5		0.92-0.97	(Lindsay et al., 2014)	20CR: bad sea ice specification (Lindsay et al., 2014)?
				2	MERRA	±0.5		0.89-0.93		
				3	JCDAS	-0.5 - 2		0.88-0.93		
				4	NCEP CFSR	-1 - 2		0.85-0.95		
				5	NCEP R2	±2		0.75-0.92		
				6	NCEP R1	-2 - 1		0.75-0.91		
				7	20CR	-0.5 - 6		0.63-0.87		
SAT	2006-2007	ERA-Interim, NCDC Observations	Y, N		Polar WRF v3.1 (ASR)	A: -1.3, -1.6 (inland). M: -2-0.1	A: 4.4, 4.5 (inland). M: 3-6	A: 0.76, 0.81 (inland). M: 0.6-0.9	(Wilson et al., 2011)	
SAT (M)	2006-2007	NSA and other in situ	N		Polar WRF v3.1 (ASR)	-3.3-6.4 (winter)	2-6.7 (winter)	0.67-0.96 (M)	(Hines et al., 2011)	
SAT	1981-2013	Land stations	Y		Background ERA-Interim forecast fields	-0.1-2.9 (D), ±2 (A, anomalies)			(Simmons and Poli, 2014)	
SAT (SPA)	1979-2008	GHCNCAMS	Y	1	ERA-40			>0.9, some areas >0.7	(Wang et al., 2011)	
				2	NCEP CFSR			>0.9, some areas 0.5-0.7		
				3	NCEP R2			>0.7, some areas 0.5-0.9		
				3	NCEP R1			>0.7, some areas 0.5-0.9		

TLT (TS, A)	1979-2009	RSS MSU	Y	1	ERA-Interim	±0.3			(Screen and Simmonds, 2011)	
				1	NCEP R1	±0.3				
				1	JCDAS	±0.3				
				2	ERA-40	±0.3, ≤ 0.6 after 1997				
Arctic in General										
SAT (SPA)	1981-2013	Cowtan and Way, other reanalyses	Y		ERA-Interim, MERRA, JRA-55	±0.5			(Simmons and Poli, 2014)	
SAT (TS)	1979-2008	IABP/POLES	Y	1	NCEP R2	≤5			(Alexeev et al., 2012)	IABP/POLES: includes NCEP
				2	NCEP R1	≤10				
SAT (TS, A)	1979-2010	GISTEMP (250km)	Y		NCEP R2, CFSR, MERRA	±0.5			(Chung et al., 2013)	R1 data, known biases.
SAT, TLT (TS)	1981-2013	Cowtan and Way, HadCRUT4, other Reanalyses	Y	1	ERA-Interim	±0.5 (A), ±1 (M)			(Simmons and Poli, 2014)	
		ERA-Interim, MERRA		1	JRA-55	±0.5 (A)				
		ERA-Interim, JRA-55		2	MERRA	±0.5 (A)				
SAT	1989-2007	Radiosondes	N	1	ERA-Interim	0.1 (1000hPa), <3 (TLT)			(Dee and Uppala, 2009)	
				2	ERA-40	0.2 (1000hPa), ±4 (TLT)				

2.3 An Investigation of the Representation of Arctic Surface Temperature Anomaly Correlation Length Scales in the ERA-Interim Reanalysis

2.3.1 Introduction

Chapter 3 of this thesis investigates the impact of using several different estimation techniques, which are related to current temperature anomaly datasets, to estimate Arctic SAT anomalies over land and sea ice areas. Both 'interpolating techniques', which produce complete fields of Arctic SAT anomalies, and 'non-interpolating' techniques, which do not produce complete SAT anomaly fields, are investigated. Estimating temperature anomalies from sparse in situ data using interpolating techniques allows us to produce spatially complete estimates of climate change in the Arctic, which are not available solely from in situ data. This means that the whole of the Arctic can be represented in global averages of temperature change, rather than being represented by sparse in situ data which only cover a small subset of the region. However, in order to interpolate and extrapolate Arctic SAT anomalies using these techniques, some assumptions about the correlation of SAT variations with distance are needed.

An observation of a geophysical variable at a certain location will be more highly correlated with nearby observations of the same variable than observations further away; the correlation decreases with distance. Correlation functions describe this correlation with distance. Correlation length scales are defined as the distance at which mean correlations fall below 1/e (Rigor et al., 2000; Belousov et al., 1971). Some interpolation techniques assume that SAT anomalies can be interpolated up to a certain distance from a station and use a correlation length scale or correlation function, generally determined from available in situ data, to determine this distance. For example, the linear interpolation technique (also known as kernel smoothing using a conical filter) employed by the GISTEMP dataset assumes that monthly SAT anomalies can be interpolated and extrapolated up to 1200km distance from a station as this was the distance at which, on average, the correlation coefficient fell below 0.5 in middle and high latitudes (Hansen et al., 2010; Hansen and Lebedeff, 1987). Other datasets use the correlation function to determine the weighting for an interpolation method. The Berkeley Earth Surface Temperature dataset utilises a Simple Kriging method in which a model correlation function is used to weight observations of SAT anomalies by distance (Berkeley Earth, 2014; Rohde et al., 2013a, 2012). The correlation function is assumed to be a good approximation for the covariance function as long as the variance changes slowly with time (Rohde et al., 2012, 2013a).

So, when investigating the impact of using various techniques to estimate Arctic SAT anomalies using reanalysis data as a testbed, we need to check that the correlation length scales of the reanalysis data are close to the reality. If the correlation length scale in the testbed data is not close to the reality then this will result in conclusions which do not

reflect the actual performance of the investigated estimation techniques in practice. For instance, if the reanalysis data had a correlation length scale which was similar to the correlation length scale assumed for meteorological station data in a given interpolating technique, when in reality the correlation length scale was quite different, then the interpolating technique would appear to provide a very good reconstruction of anomalies when used in the testbed whereas for interpolating real data it would perform poorly. Consequently, it is important to investigate correlation with distance in ERA-Interim for the Arctic and establish how close it is to the reality.

Here I compare monthly and annual correlation functions for Arctic SATs as represented in ERA-Interim to reference correlation functions in order to determine how close the correlation length scales in ERA-Interim are to Arctic correlation length scales in reality. Section 2.3.2 describes the data and techniques used in this study. Sections 2.3.3 and 2.3.4 evaluate the representation of Arctic annual and monthly correlation functions respectively in ERA-Interim. Finally, Section 2.3.5 discusses the results and provides a summary and conclusions for this correlation length scale study.

2.3.2 Data and Techniques

SAT anomaly fields were created from ERA-Interim data so that each grid cell had a timeseries of monthly and annual SAT anomalies between 1979 and 2011. These SAT anomaly fields were sub-sampled to produce 'station location' timeseries: ERA-Interim grid cell SAT anomaly timeseries sampled at Arctic meteorological station locations. The Arctic was defined in this study as the area above 65°N, which approximately matches the area northward of the Arctic circle while matching cleanly the grid cell edges of ERA-Interim. The station location timeseries were cross correlated for each grid cell to produce monthly and annual cross correlation fields. Monthly and annual correlation functions were calculated from the cross correlation fields for each station location. These station location correlation functions were compared to reference correlation functions, along with their implicit correlation length scales, and their representativeness analysed.

2.3.2.1 Reference Correlation Length Scales

2.3.2.1.1 Annual The correlation of meteorological station temperature records with distance was investigated as part of the development of the GISTEMP (Hansen and Lebedeff, 1987) and Berkeley Earth (Rohde et al., 2012) temperature anomaly datasets. Both studies (Hansen and Lebedeff, 1987; Rohde et al., 2012) investigated the annual correlation function at various latitude bands across the earth. The correlation function for Arctic latitudes in both studies (>64°N for Hansen and Lebedeff 1987 and the latitude band centred at 68°N for Rohde et al. 2012) had a correlation value of around 0.98 at 0km station separation, decreasing to 0.5 at 1200km, and reaching a correlation of 0 at about 3000km. I fitted a curve to these parameters (Figure 2.1) and this constituted the reference annual correlation function to which annual station location correlation

functions were compared. From this curve I determined the reference annual correlation length scale (the distance at which mean correlations fall below 1/e): 1570km.

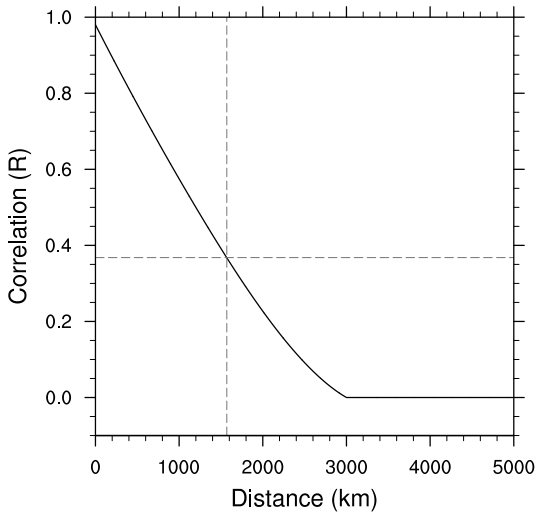

Figure 2.1: The annual correlation function in high latitudes produced using the parameters identified by Hansen and Lebedeff 1987 and Rohde et al. 2012. This was the reference correlation function for annual Arctic SAT anomalies. The reference annual correlation length scale was 1570km.

2.3.2.1.2 Seasonal The IABP/POLES SAT dataset is an optimally interpolated dataset of SATs in the Arctic (Rigor et al., 2000). The correlation with distance was investigated between 12 hourly meteorological station, IABP buoy and NPDS observations over different Arctic surfaces (coastal land, interior land and ocean) by Rigor et al. 2000 prior to production of the dataset. Monthly correlation length scales were calculated for correlations between different types of observations; for example ocean platforms with ocean platforms, ocean platforms with coastal platforms, interior platforms with coastal platforms, and so on. The correlation length scales for all observation types were found to be almost constant for boreal spring, autumn and winter but was shorter in boreal summer (Rigor et al., 2000). This seasonal pattern in the correlation length scales for monthly Arctic SATs has also been noted elsewhere (e.g. Przybylak, 1997, 2003a).

As the correlation length scales were fairly constant on seasonal scales in that study, I compared the monthly station location correlation length scales to reference seasonal correlation length scales in this study. In addition, in this study station location timeseries over land (coastal and interior stations) were correlated with grid cell timeseries over ocean, coast and interior Arctic areas only. I therefore calculated the average of the monthly correlation length scales in each season from Rigor et al. 2000 for all observation types relevant to this study (all but ocean with ocean). The average correlation length scale values were 1020km (spring, autumn, winter) and 710km (summer) to the nearest 10km. The study of Rigor et al. 2000 does not describe the expected average correlation

value at 0km, unlike the studies of Hansen and Lebedeff 1987 and Rohde et al. 2012. Therefore, in order to produce a reference seasonal correlation function, I assumed a correlation value of 0.98 at 0km here. This is based on the results of the studies by Hansen and Lebedeff 1987 and Rohde et al. 2012; studies which, in common with the study of Rigor et al. 2000, employed in situ data. Curves were fitted to the parameters noted above and these are shown in Figure 2.2. These curves constituted the reference seasonal correlation functions to which monthly station location correlation functions were compared.

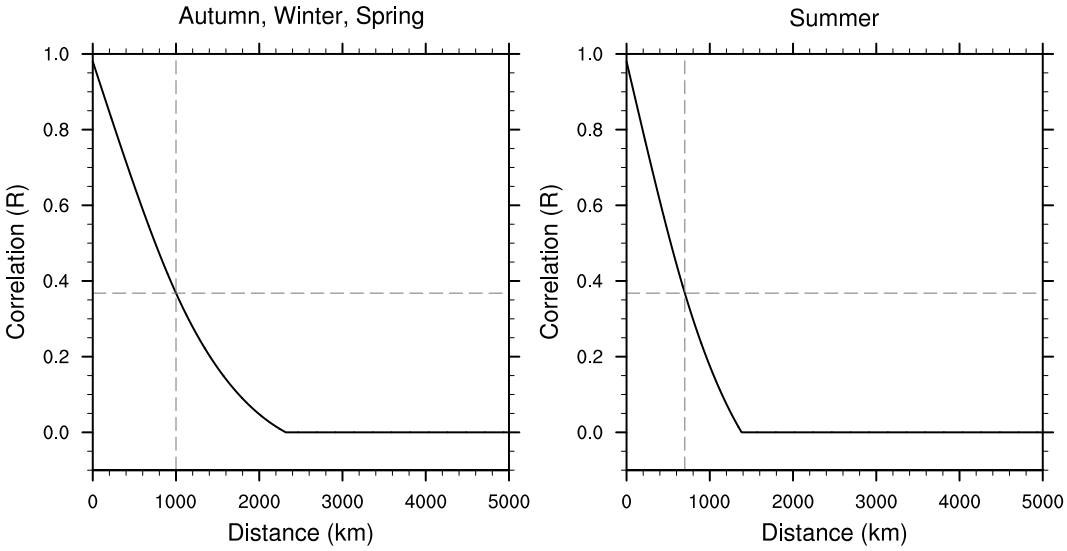

Figure 2.2: The seasonal correlation function in high latitudes based on the correlation functions identified by Rigor et al. 2000. These were the reference correlation functions for monthly Arctic SAT anomalies. The reference monthly correlation length scales were 1020km (spring, autumn, winter) and 710km (summer).

2.3.2.2 Station Location Correlation Length Scales

Station location correlation functions were produced from ERA-Interim grid cell monthly and annual SAT anomaly timeseries, sampled at Arctic meteorological station locations. Monthly SAT anomalies were produced for each ERA-Interim grid cell from 6 hourly resolution ERA-Interim 2m air temperature data between 1979 and 2011 relative to a 10 year climatological period (1990-1999). A 10 year climatological period was used instead of the conventional 30 year period in order to simulate the use of a conventional climatological period in a longer dataset using ERA-Interim data which only covers 33 years. Annual anomalies for each grid cell were created from monthly anomalies for each calendar year. The anomalies were subsequently detrended so that the correlation value did not include the correlation arising from longer term trends in the temperature anomalies. This produced detrended ERA-Interim SAT anomaly fields.

The locations in latitude and longitude of all meteorological stations in the GHCN-M version 3.1.0 (GHCN-Mv3.1.0) were obtained from NOAA NCDC in February 2012. I

chose station locations from GHCN-Mv3.1.0 for this study as many temperature anomaly datasets use the GHCN-M temperature records as the basis for their temperature anomalies. The meteorological station locations used in this investigation were not the same as those used for Chapter 3. Chapter 3 utilised the locations of all meteorological stations in the CRUTEM.4.1.1.0 land surface temperature anomaly dataset (produced by the Climatic Research Unit of the University of East Anglia and the Met Office Hadley Centre, Jones et al. 2012) databank, which was released after work on this investigation had begun. The CRUTEM.4.1.1.0 databank contains a higher number of Arctic meteorological stations than GHCN-Mv3.1.0 (Jones et al., 2012), so I decided that the former would be a more suitable input for the study in Chapter 3. These datasets have the majority of stations in common above 65°N (illustrated by comparing Figure 2.3 and Figure 3.3 of this thesis) so I do not expect the difference between the station locations in this study and those used in Chapter 3 to have a large impact on the results of this investigation, nor on the final conclusion of this chapter. All stations in GHCN-Mv3.1.0 located in the Arctic (above 65°N) were identified, regardless of whether they reported temperatures between 1979 and 2011. Some of the land stations were located within the same ERA-Interim grid cell, and therefore produced identical correlation fields. These stations were identified and duplicates were ignored. The locations of all identified, non-duplicate stations are shown in Figure 2.3.

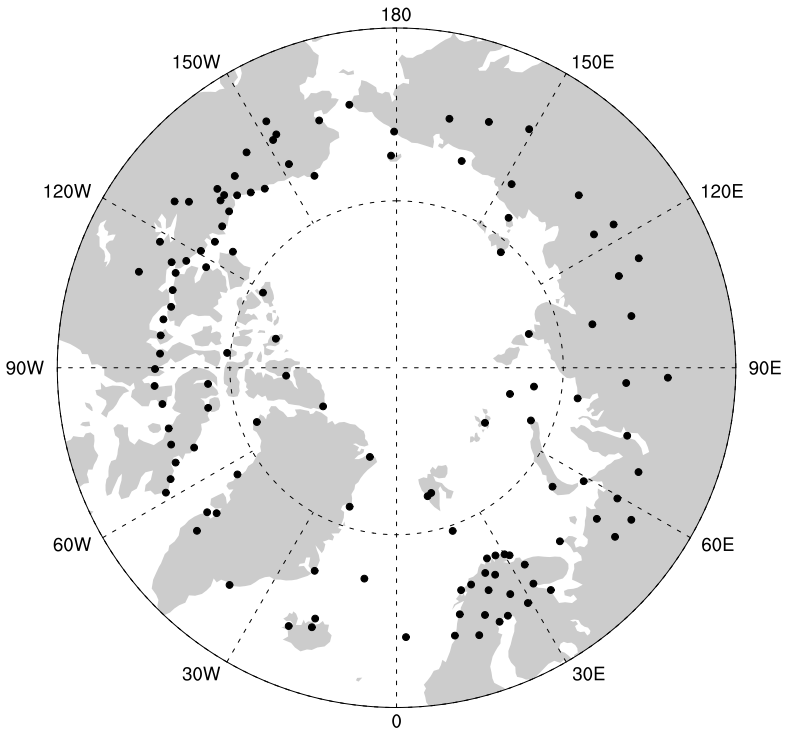

Figure 2.3: The locations of all Arctic (above 65°N) meteorological stations in the GHCN-Mv3.1.0 dataset.

In the studies by Hansen and Lebedeff 1987; Rigor et al. 2000; Rohde et al. 2012,

correlation with distance was determined from available in situ data by correlating the temperature records of many land-based meteorological stations. The station location correlation functions in ERA-Interim were determined using a slightly different method. Here I cross correlated 'station location' timeseries, ERA-Interim grid cell anomaly timeseries sampled at the identified Arctic meteorological station locations, with the anomaly timeseries at all other ERA-Interim grid cells to produce monthly and annual cross correlation fields. This was done in order to explore spatially complete Arctic station location correlation fields. However, this means that there may be differences between the reference and station location correlation functions and implicit correlation length scales which result from the different sample of timeseries correlated rather than differences in the correlation length scale.

The monthly and annual correlation values for each ERA-Interim grid cell were plotted against distance for each station location. A model function was fitted to these monthly and annual correlation values with distance to calculate monthly and annual correlation functions for each station location. The form of the correlation values with distance were examined visually and the majority were very similar to a "spherical" model function (described by Eq. (2.1); d is the distance in km and $C(d)$ is the model correlation at distance d). This corresponds with the study by Rohde et al. 2012. The parameters a, d_{max} and μ were determined by fitting the function to the station location correlation values. The MPFIT IDL least-squares curve fitting routines by (Markwardt, 2009) were used to determine the most appropriate values of a, d_{max} and μ for each model correlation function. The function was constrained to 1 at 0km. Without a constraint the MPFIT IDL routines would determine parameters for a model function with a correlation above 1 at 0km. I chose a value of 1, rather than 0.98 as in the reference, because station location timeseries were correlated with themselves at 0km in this study as a result of using gridded data instead of meteorological station data (used to produce the reference correlation length scales). Therefore, it should be noted here that the station location timeseries are expected to have slightly higher correlation values at each distance compared to the reference correlation function. If the correlation with distance decreases at the same rate in both the station location and reference correlation functions, we can expect slightly higher correlation values at each distance for ERA-Interim derived correlation functions. Furthermore, ERA-Interim data do not contain the inhomogeneities, uncertainties, noise and local effects associated with meteorological station records and can therefore be expected to have longer correlation length scales. The spherical model function was found to be a good model fit for the majority of station locations. A visual examination of the model functions and the correlation values with distance for each station location (monthly and annual) found that they were generally very similar. Additionally, a chi-squared goodness of fit test gave a result of no significant difference (at the 95% confidence level) between the model function and the distribution of correlation values with distance for more than 80% of station locations.

$$C(d) = \begin{cases} a(1 - \frac{d}{d_{max}})^2(1 + \frac{d}{2d_{max}}) + \mu & \text{for all } d < d_{max} \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

The station location correlation functions and maps of station location correlation fields produced using the methods detailed above were examined and compared to the reference correlation functions.

2.3.3 The Correlation Length Scale of Arctic Surface Air Temperature Anomalies at Annual Timescales

At annual timescales, some station location timeseries' correlation fields and correlation functions corresponded with the reference annual correlation functions. Figure 2.4 shows the cross correlation fields for three example station locations. Figure 2.5 shows the scatter plots of correlation values with distance and the station location correlation functions for the same three stations. These stations had correlation values of $1/e$ or more for all, or almost all, of the area within 1570km of the station location, while the correlation functions for the three stations had correlation values of around 0.6 at 1570km. This was slightly higher than the correlation value of the reference correlation function at this distance, as expected for ERA-Interim derived correlation functions in this study (Section 2.3.2.2). But, there were also some station location timeseries with annual spatial correlations and correlation functions which did not fit as well as with the reference, nor with the anticipated behaviour of ERA-Interim derived correlation functions relative to the reference. Figures 2.6 and 2.7 shows the cross correlation fields and correlation functions for a different set of three station locations. These stations had correlation values of $1/e$ or more for relatively small areas within 1570km of the station locations, or had very directional correlation fields. The correlation length scales for the three stations were also shorter than the reference; correlation values were 0 to 0.1 at 1570km. Were the station locations with correlation functions similar to, or slightly higher than, the reference in the minority or do they make up a significant proportion of Arctic station location annual correlation functions?

Binned averages of all station location timeseries' correlation values with distance were calculated on monthly and annual timescales and a spherical model function was fitted to these averages (as in Section 2.3.2.2). On average, the station location correlation length scales were very similar to the reference, but with slightly higher correlation values at each distance as anticipated (Figure 2.8). Both curves had correlations of around $1/e$ at 1570km distance and the average station location correlation length scale was slightly longer than the reference (1700km compared to 1570km). The cross correlation at 1570km for each station location was graphed and compared to the reference correlation value at this distance (Figure 2.9). 45.3% of the analysed station location timeseries had annual correlation values of $1/e$ or more at 1570km distance. Of the remaining stations, most had

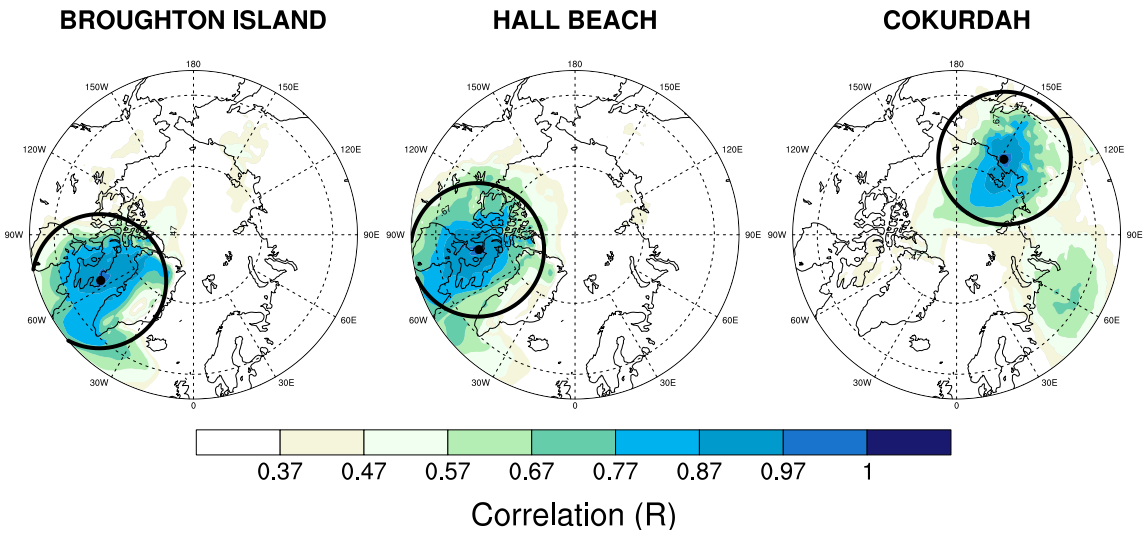

Figure 2.4: The cross correlation of grid-cell annual average anomalies from the grid cell nearest three station locations (Rowley Island, Hall Beach and Corkudah) with the annual average anomalies for all other ERA-Interim grid cells. Black dots show station locations and black circles show the area which is 1570km distance from each station location.

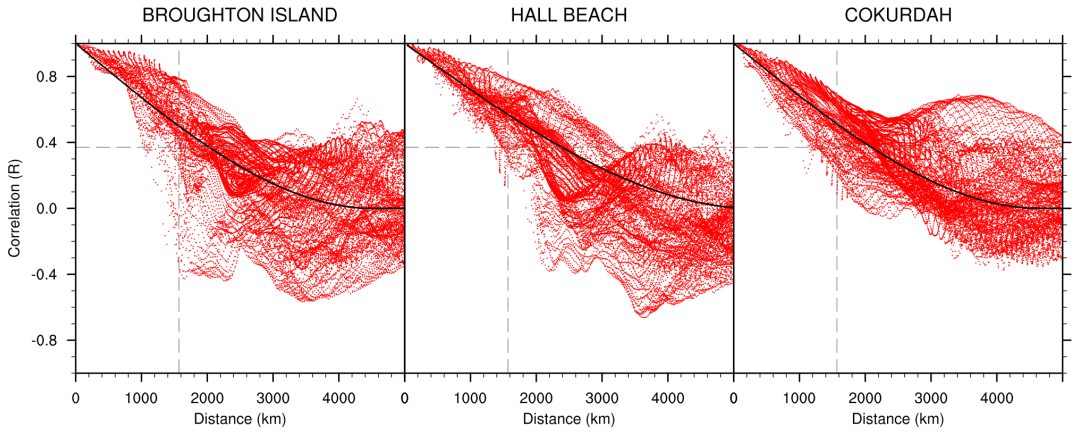

Figure 2.5: Scatter plot of the correlation between grid cell annual average anomalies from the grid cell nearest three Arctic station locations (Rowley Island, Hall Beach and Corkudah) with the annual average anomalies for all other ERA-Interim grid cells. The black line gives the polynomial fit to the data and the dashed lines show the reference correlation length scale.

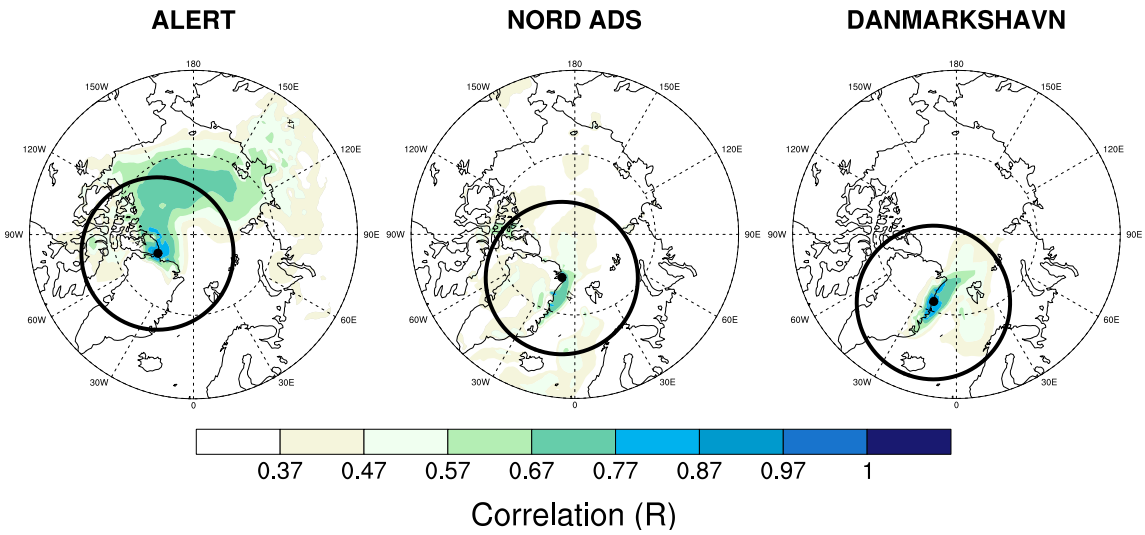

Figure 2.6: The cross correlation of grid-cell annual average anomalies from the grid cell nearest three station locations (Alert, Nord Ads and Danmarkshavn) with the annual average anomalies for all other ERA-Interim grid cells. Black dots show station locations and black circles show the area which is 1570km distance from each station location.

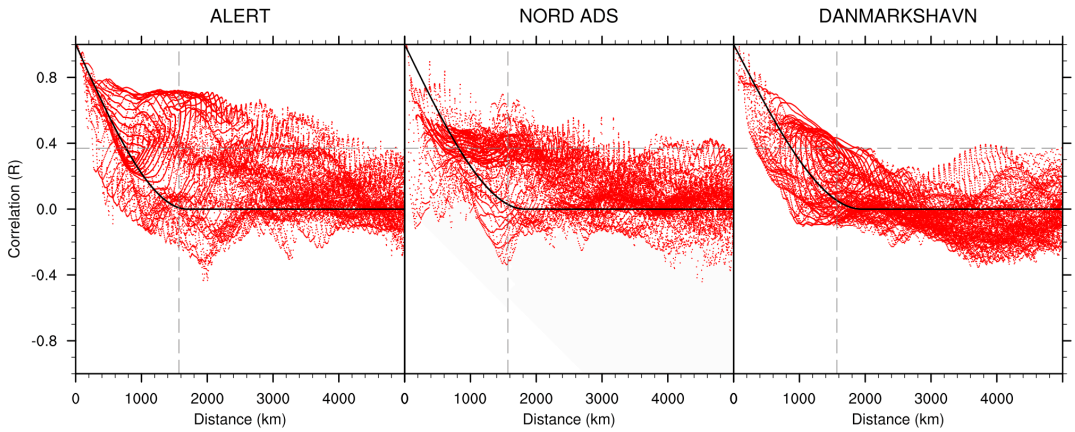

Figure 2.7: Scatter plot of the correlation between grid cell annual average anomalies from the grid cell nearest three Arctic station locations (Alert, Nord Ads and Danmarkshavn) with the annual average anomalies for all other ERA-Interim grid cells. The black line gives the polynomial fit to the data and the dashed lines show the reference correlation length scale.

correlation values which were fairly close to the reference; 72.2% of all stations investigated had correlation values of ≥0.27 or more at 1570km and 87.4% had values of 0.22 or more. The majority of investigated station location CLS had correlation values at 1570km that were more than or equal to the reference value at this distance, or were within 0.15 of this correlation value.

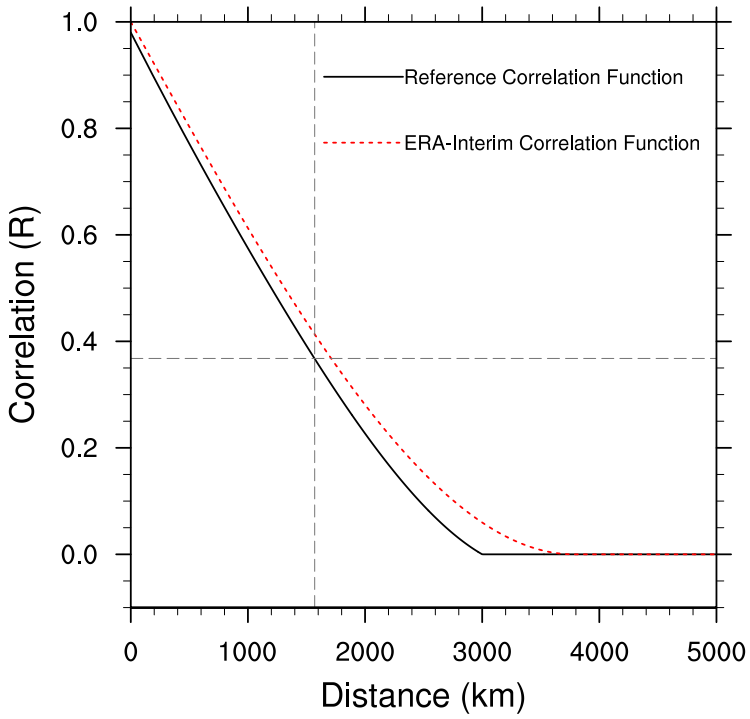

Figure 2.8: Annual reference and ERA-Interim correlation functions. The dashed lines show the reference correlation length scale.

So, on average the correlation function for annual Arctic SAT anomalies as represented in ERA-Interim was very similar to the expected. Just under half of the station locations fitted the reference correlation length scales, or had the higher correlation values at each distance predicted for ERA-Interim derived correlation length scales. Most of the remaining stations had correlation values close to the reference value at this distance.

2.3.4 The Correlation Length Scale of Arctic Surface Air Temperature Anomalies at Monthly Timescales

As noted for the reference seasonal correlation functions, in many of the station location timeseries the boreal summer months (June, July and August) had a shorter correlation length scale than in other seasons. For example, Figure 2.10 shows the monthly cross correlation fields for one example station location (Tuktoyaktuk, N.W.T) with black dots to indicate the station locations and black circles to show the area which is 1020km (spring, autumn, winter) or 710km (summer) distance from each station location. For most months, this station had correlation fields with values of at least $1/e$ for much of the

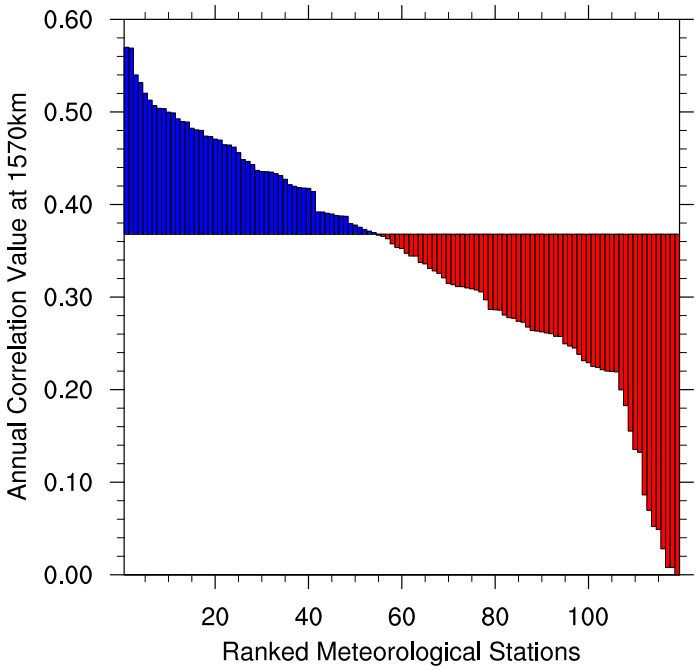

Figure 2.9: Graph of the correlation value of grid-cell annual average anomalies at 1570km (the reference correlation length scale) for all station locations above 65°N with a reference line representing the reference correlation value associated with the correlation length scale.

area within 1020km of the station (Figure 2.10). However, in boreal summer (June, July, August) the area for which correlation values were $\geq 1/e$ was smaller. This pattern of smaller correlation length scales in the summer months was also observed in Figure 2.11 for monthly scatter plots and correlation functions for the same example station location.

Binned averages of all station location timeseries correlation values with distance were calculated for monthly SAT anomalies in the Arctic and a polynomial was fitted for each month. On average the station location correlation values were around 0.6 at 1020km and 0.5-0.6 at 710km (summer), whereas the reference correlation values were $1/e$ at these distances. This was a larger difference between the curves than was noted for the similar annual correlation function in Figure 2.9; 0.2-0.3 difference in monthly average correlation values at the distances of interest rather than 0.1 for annual timescales. I repeated the comparison using monthly, rather than seasonal, reference correlation functions (Figure 2.13). For many months the monthly average station location correlation functions were slightly closer to monthly rather than seasonal reference correlation functions. However, the differences between the reference and station location correlation functions were still larger than those noted for the annual correlation functions. The monthly correlation value for each station location at 1020km and 710km (summer), and the average monthly correlation value across all station location correlation functions at these distances, were produced and are plotted in Figure 2.14. The average monthly correlation value in each month was above the reference value of $1/e$. The correlation values at the appropriate distances were 0.5-0.6 in all months, which fits with the results shown in Figure 2.12. For winter and spring months, all station location correlation functions were at or above

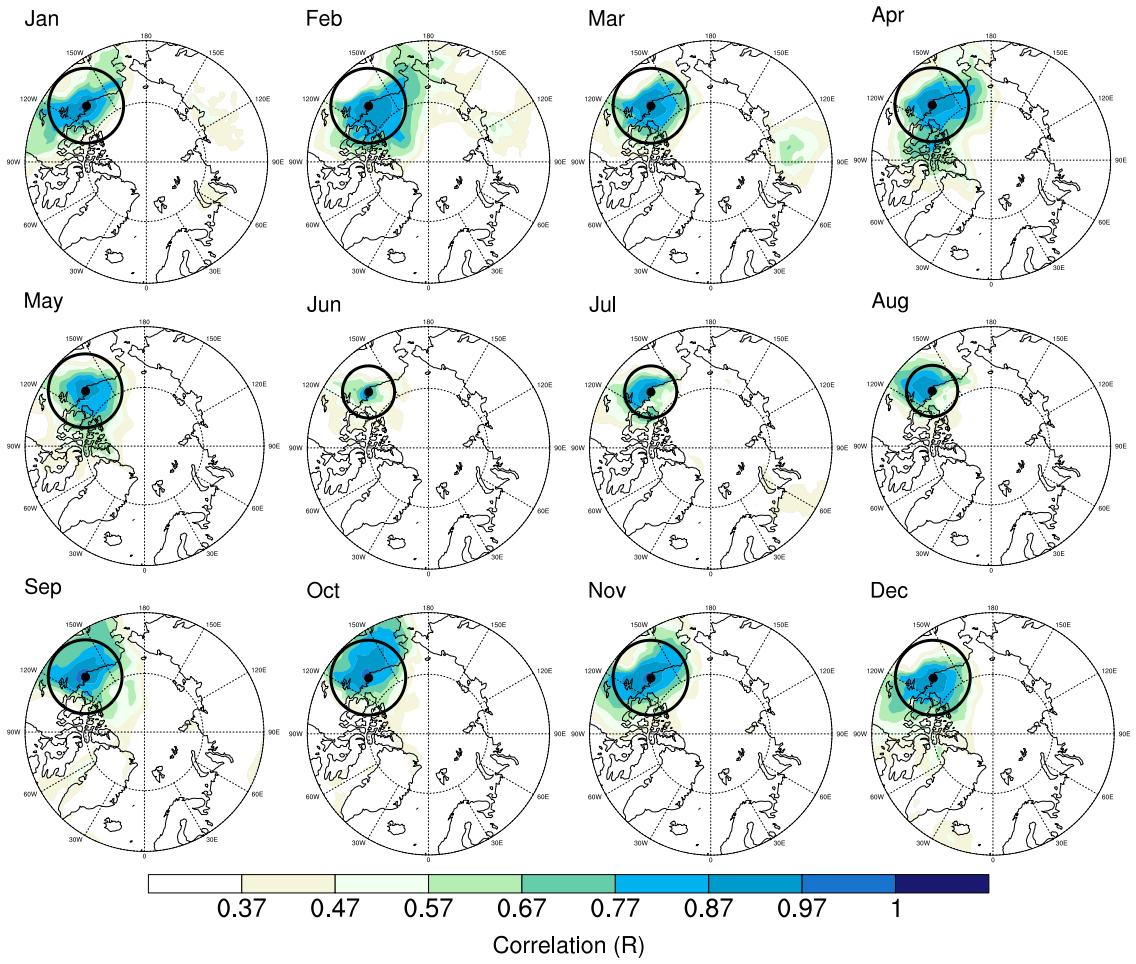

Figure 2.10: Panel plot of the cross correlation of grid-cell monthly average anomalies from the grid cell nearest the Tuktoyaktuk, N.W.T station location with the monthly average anomalies for all other ERA-Interim grid cells. Black dots show the station location and the black circles show the area which is 1020km (spring, autumn, winter) or 710km (summer) distance from the station location.

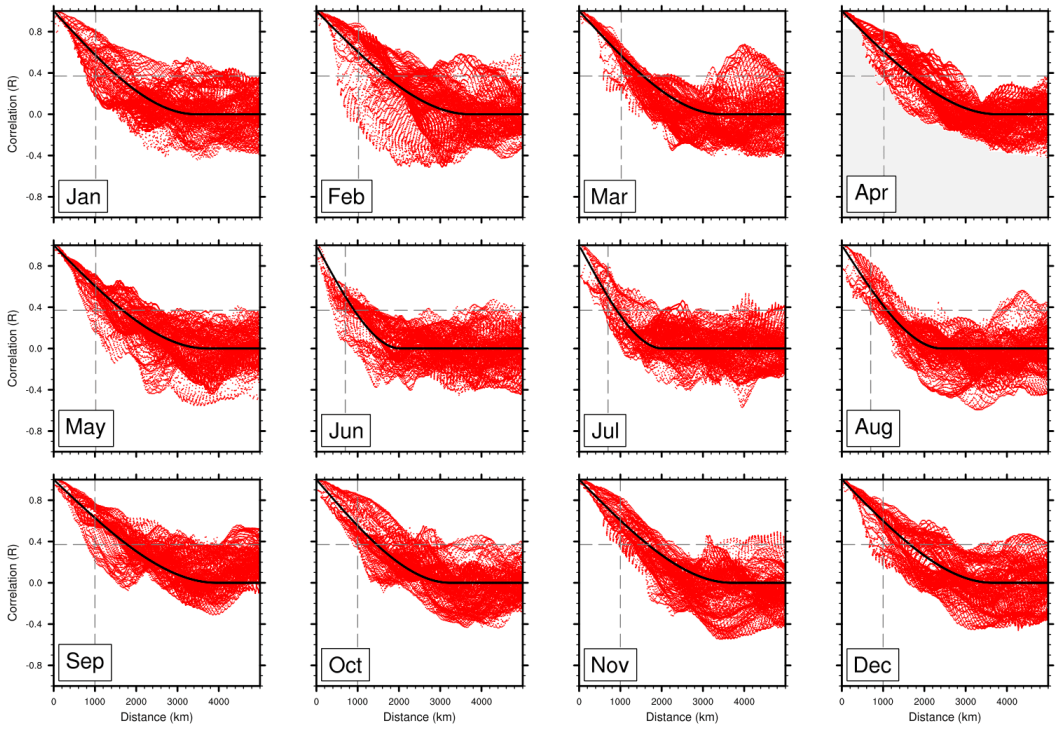

Figure 2.11: Scatter plot of the correlation between grid cell monthly average anomalies from the grid cell nearest the Tuktoyaktuk, N.W.T station locations with the monthly average anomalies for all other ERA-Interim grid cells. The black line gives the polynomial fit to the data and the dashed lines show the reference correlation length scale.

the reference value. Many had a correlation value of 0.6. For the remaining months the majority (at least 84%) of station location correlation values were $\geq 1/e$.

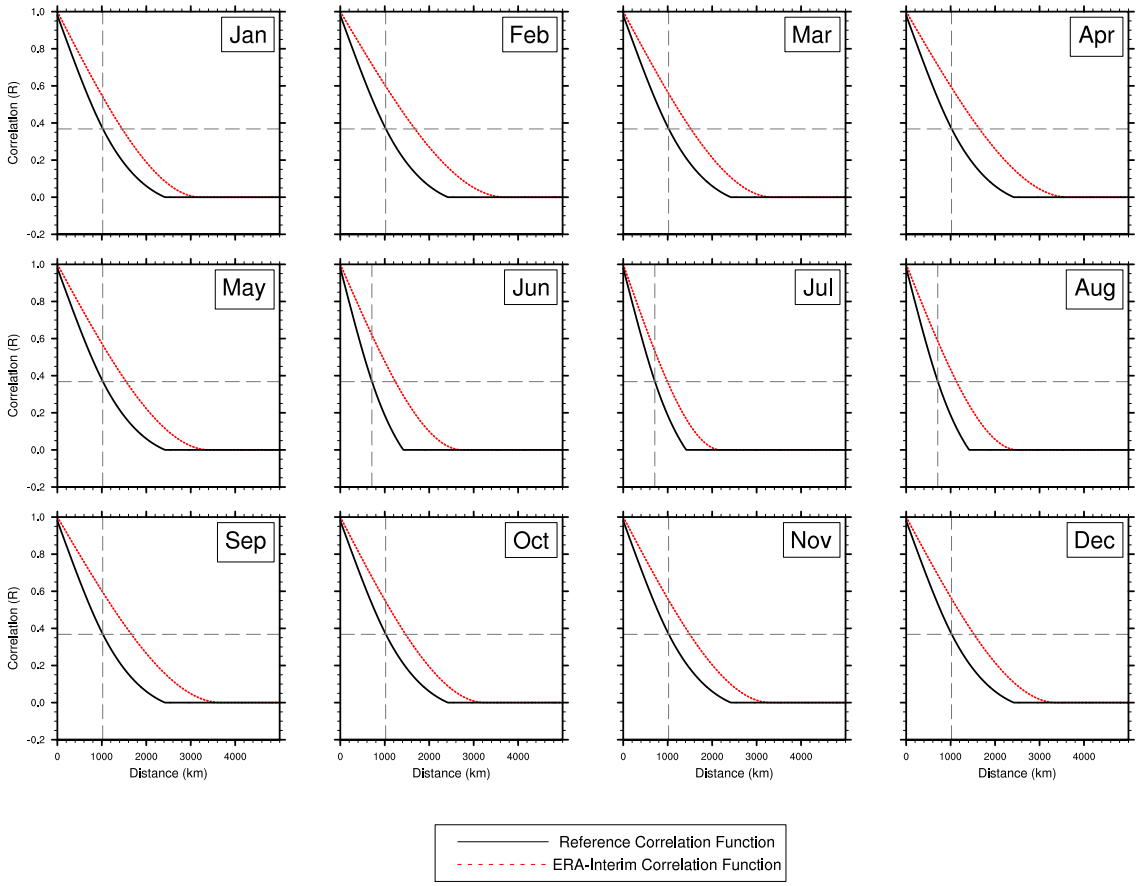

Figure 2.12: Seasonal reference and monthly ERA-Interim correlation functions. The dashed lines show the reference correlation length scale.

So, the correlation length scales of monthly Arctic SAT anomalies as represented in ERA-Interim was longer than the reference values but longer than anticipated given the results for annual correlation length scales. It could be that ERA-Interim data were less representative of monthly correlation length scales than annual correlation length scales. However, the annual correlation functions were quite similar to the reference and, as the monthly ERA-Interim correlation functions were consistently the same or higher on average and for the majority of stations, it seems unlikely that the more similar annual correlation function was due to cancellation of errors. This suggests that the reference seasonal correlation functions might not be representative of the areas over which the station location timeseries were being correlated. In this study it was assumed that an even mix of correlation length scales of all observation types from the study of Rigor et al. 2000, excluding ocean with ocean stations, would provide an appropriate reference correlation length scales for this study on seasonal (and monthly) timescales. However, as the station locations used in this study were all land based, it is likely that the station location timeseries were correlated with a higher percentage of land-based grid cells (coastal or interior) the shorter the distance from the station. Therefore assuming

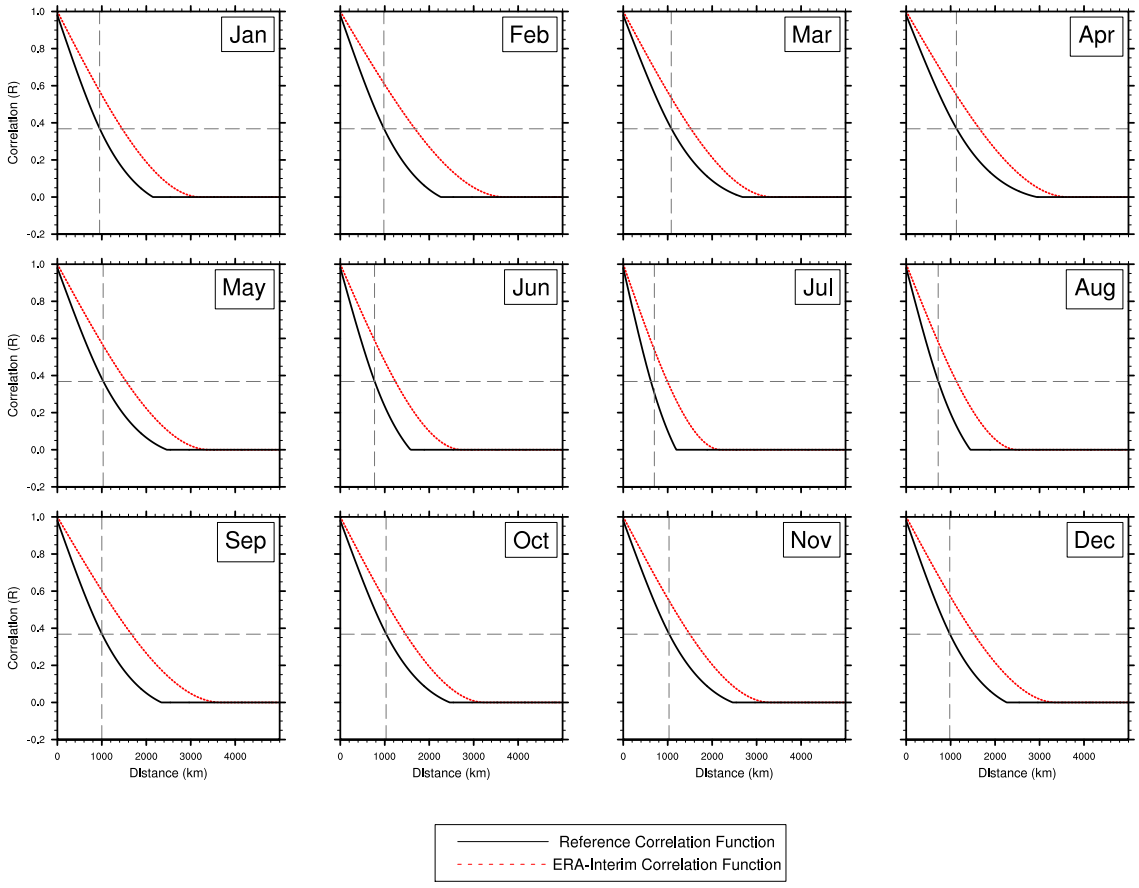

Figure 2.13: Monthly reference and ERA-Interim correlation functions. The dashed lines show the reference correlation length scale.

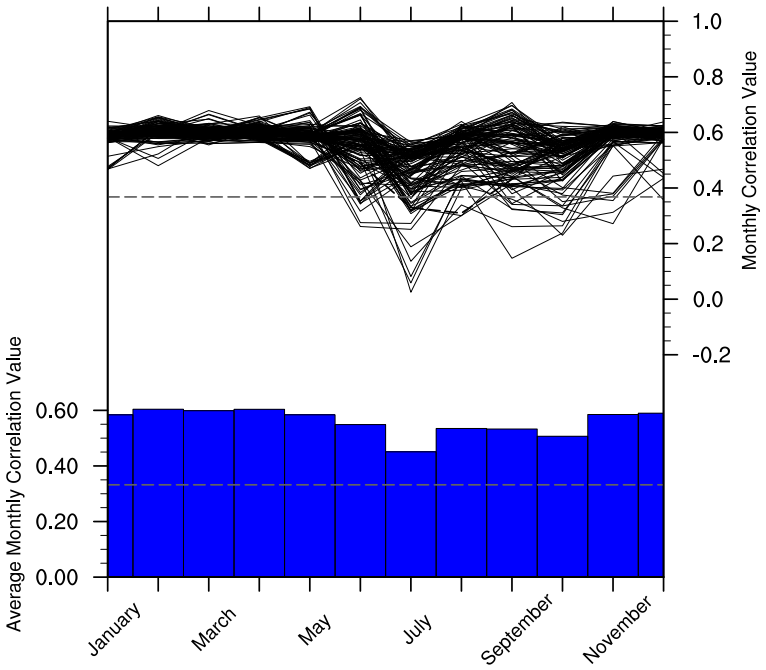

Figure 2.14: Graph of the monthly cross correlation value at 1020km and 710km (summer) for each station location above 65°N and the average monthly cross correlation value at these distances across all station locations above 65°N. Dashed grey lines show the reference correlation value at the distances of interest.

an even mix of interior, coastal and ocean grid cell correlations for the relatively short distances of interest (710 and 1020km) may not be appropriate.

In conclusion, there was a seasonal pattern in the correlation length scale on average, and for the majority of, the station location timeseries analysed. On average the correlation length scales as represented in ERA-Interim on monthly timescales were fairly similar to, but longer than, the reference seasonal correlation length scales. The monthly ERA-Interim correlation functions were often slightly closer to monthly reference correlation functions when investigated if used, but the differences between the reference and ERA-Interim correlation functions were still larger than noted for the annual correlation functions. This may be due to the assumptions made in identifying reference seasonal correlation length scales.

2.3.5 Discussion

For Arctic SAT anomalies as represented in ERA-Interim, the annual and monthly correlation functions were generally similar to the reference annual and seasonal correlation functions, and their implicit correlation length scales. For annual Arctic SAT anomalies the station location correlation length scales were, on average and for nearly half of the station locations analysed, very similar to the reference CLS but with slightly higher correlation values at all distances until 0 correlation was reached. The slightly higher correlation values at each distance in ERA-Interim were expected; station location timeseries were correlated with themselves at 0km in this study and ERA-Interim data are more consistent than meteorological station records. Therefore ERA-Interim data can be expected to have longer correlation length scales for Arctic SAT anomalies.

In common with the seasonal reference correlation functions there was a seasonal pattern in the ERA-Interim correlation length scales on average and for the majority of the station location timeseries analysed. Summer correlation length scales were shorter than those in other seasons. As noted for annual correlation functions the monthly correlation functions derived from ERA-Interim Arctic SAT anomalies were fairly similar to, but with higher correlation values at each distance than, the seasonal reference correlation functions. The higher correlation values were expected, but there was a larger difference between the correlation functions than noted for annual anomalies. Correlation values in monthly ERA-Interim correlation functions had correlation values 0.2-0.3 higher on average than those of the seasonal reference correlation functions at the distances of interest rather than the 0.1 noted for annual timescales. The monthly ERA-Interim correlation functions were often slightly closer to monthly reference correlation functions if used, but the correlation values still had greater differences at the distances of interest compared to the annual correlation functions. This may be due to the choice of the reference correlation functions to which the ERA-Interim correlation functions was compared to; the reference correlation function might not be representative of the areas over which the station location timeseries were being correlated (Section 2.3.4).

Some of the differences noted between the ERA-Interim and reference correlation functions could be related to differences in the sample of timeseries used to create the functions. The reference correlation functions, here assumed to be the 'true value', were calculated by correlating meteorological station temperature timeseries only, whereas the ERA-Interim correlation functions in this study were produced by correlating all grid cell timeseries with each station location timeseries. This creates uncertainty in the comparisons. Nonetheless the correlation with distance in ERA-Interim Arctic SAT anomalies is similar to the reality, which in this study are reference correlation functions produced using parameters identified in studies by Hansen and Lebedeff 1987, Rohde et al. 2012 and Rigor et al. 2000 using in situ observations of Arctic SATs. This shows that ERA-Interim is physically plausible in terms of correlation length scales in Arctic SATs.

2.4 Summary

Despite a consistent warm bias of around 2K compared to temperature observations near the surface, ERA-Interim produces generally realistic estimates of Arctic temperatures and temperature trends. ERA-Interim was ranked first in all identified studies which investigated Arctic SATs at 2m height and outperformed, or was comparable to, other reanalyses in terms of SAT patterns, anomalies and SAT trends. Simmons and Poli 2014 found that ERA-Interim SATs in the Arctic were within 0.5K of SATs from the Cowtan and Way dataset (Cowtan and Way, 2014), HadCRUT4 (timeseries only), and other reanalyses spatially and for SAT timeseries. ERA-Interim SATs over Arctic sea ice had ME values of -0.5 to 1.4K compared to ships and buoys, although the ME compared to ice stations could be up to 4.5K (Simmons and Poli, 2014). ERA-Interim SATs over Arctic land areas were found by Lindsay et al. 2014 to be the most representative, out of the seven reanalyses investigated in that study, of monthly SATs from CRUTEM3 meteorological stations. ERA-Interim data were within 0.5K of CRUTEM3 monthly SATs were very well correlated with this in situ data; monthly correlation values were 0.92-0.97 (Lindsay et al., 2014). ERA-Interim also performs well in terms of the representation of annual and monthly Arctic SAT anomaly correlation length scales. The reanalysis produced correlation functions with correlation values within 0.1-0.3 of the reference at each distance on average and for the majority, or near majority, of station locations investigated. Based on the literature discussed and the Arctic SAT correlation length scales investigation described in this chapter, ERA-Interim was found to be a physically plausible reanalysis and consequently determined to be a suitable reanalysis to use as a testbed for studying SAT anomalies in the Arctic.

Chapter 3

An Investigation into the Impact of using Various Techniques to Estimate Arctic Surface Air Temperature Anomalies

This study was accepted for publication as a paper in the Journal of Climate on the 23^{rd} September 2014 (Dodd et al. 2015; ©American Meteorological Society, used with permission). This chapter contains the final accepted version of the paper with some modifications. Slight changes have been made so that this chapter conforms to the formatting of this thesis. Changes have also been made to the introduction and discussion in order to avoid unnecessary repetition between chapters. An additional section (Section 3.4.2.1) has also been added. This section was not included in the final draft of the paper as there is some uncertainty surrounding the results, as well as concerns about the potential for the results being taken out of context despite appropriate caveats. A more comprehensive investigation and a more robust method would allow more confidence in these results. Nonetheless, they are included here as they do raise an interesting point about the need for caution when using estimation techniques in extremely data-sparse regions, such as the Arctic.

3.1 Introduction

As noted in Chapter 1, the Arctic is an important region in the study of climate change. However, monitoring Arctic temperature change is challenging, particularly in areas covered by sea ice for all or part of the year. Timeseries of global and regional mean Surface Air Temperature (SAT) anomalies are one of the main metrics used to estimate recent climate change. But, in situ measurements of Arctic SATs are sparse, especially early in the temperature record (Figure 3.1), and the records are often short. This introduces uncertainty to the calculation of average temperature changes in this region.

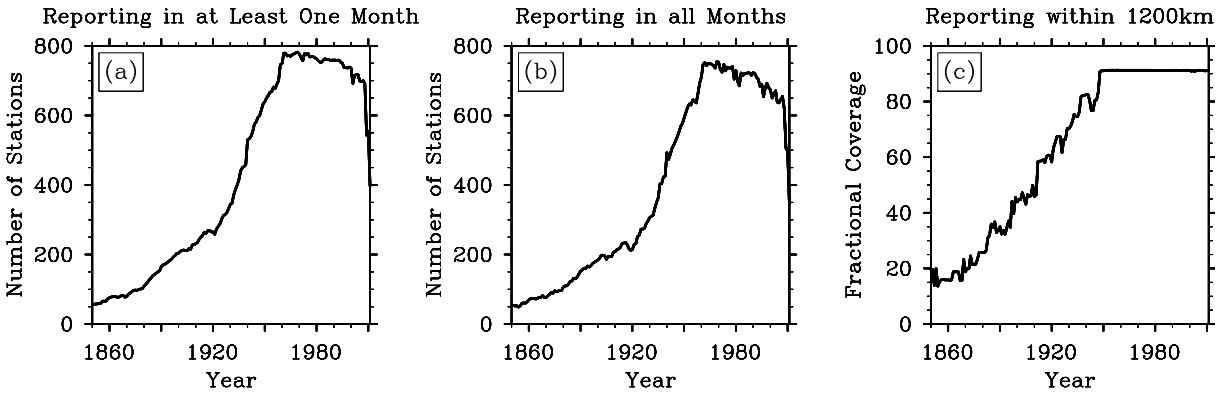

Figure 3.1: The number of stations (a) reporting at least one temperature in each year, (b) reporting temperatures in all months of each year, and (c) the percentage of grid cells with at least one station reporting within 1200km ('Fractional Coverage').

Many types of surface temperature, measured from sensors on various different platforms, could be used to monitor SAT changes in the Arctic (Chapter 1). For example, satellite instruments can provide relatively consistent, continuous and detailed observations of the radiometric temperature of ice and water surfaces in the Arctic (Merchant et al., 2013). There are also platforms which monitor SATs in Arctic sea ice areas such as Arctic and Antarctic Research Institute North Pole Drifting Stations. However, satellite derived temperatures and in situ temperature observations over Arctic sea ice are not generally utilised at present for producing datasets of SAT anomalies. Instead, many different groups generate datasets of SAT anomalies using various techniques and methods to quantify SAT changes over the Arctic from sparse in situ measurements as described in Chapter 1.3.4. These result in timeseries of temperature changes which generally produce similar trends in post-industrialisation climate change for the Arctic region (see Figure 3.2).

How effectively do these techniques reconstruct Arctic SATs and Arctic SAT change? Previous studies have compared different techniques used to reconstruct surface temperatures at different temporal and spatial scales. Kriging techniques were often found to produce good estimates of surface temperatures. For example, Cowtan and Way (2014) and Rohde (2013) found that kriging techniques outperformed both the Met Office Hadley Centre and Climatic Research Unit of the University of East Anglia (CRU) global temperature anomaly dataset version 4 (HadCRUT4) method of estimation and Linear Interpolation in the style of the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) Surface Temperature Analysis, or GISTEMP, for global monthly mean surface temperature anomalies. Hofstra et al. (2008) found that kriging in combination with splines was the more effective technique for interpolating temperature data at daily resolution for European land areas. Other techniques have also been used to interpolate surface temperatures including Angular Distance Weighting (e.g. New et al., 2000), regression based methods (e.g. Vicente-Serrano et al., 2003), splines (e.g. Price et al., 2000) and optimal interpolation (e.g. Kaplan et al., 1998). However,

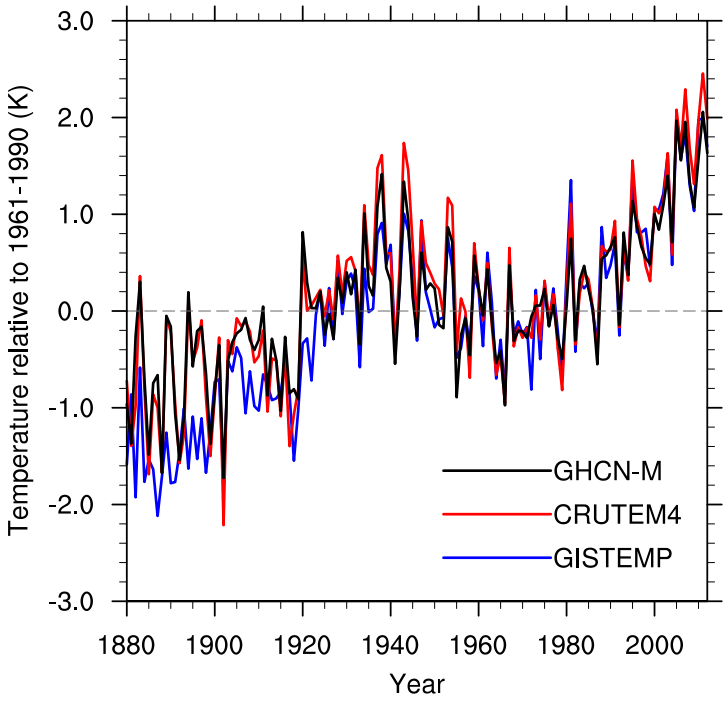

Figure 3.2: The annual Arctic SAT anomaly (K) over land relative to 1961-1990 from several temperature anomaly datasets; Global Historical Climatology Network-Monthly temperature dataset (GHCN-M, Lawrimore et al. 2011), National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP, Hansen et al. 2010), and the Climatic Research Unit of the University of East Anglia and Met Office Hadley Centre land surface temperature anomaly dataset version 4 (CRUTEM4, Jones et al. 2012). The timeseries are produced from the dataset grids using grid boxes north of 65°N over land and converted to be relative to 1961-1990. The dataset versions used for this figure are GHCN-M.3.2.2.20140729, GISTEMP (downloaded on the 29th July 2014) and CRUTEM4.2.0.0.

despite the range of studies comparing surface temperature estimation techniques at different temporal and spatial scales, as well as in various different areas, no previous study has specifically looked at estimating surface temperatures in the Arctic.

The purpose of this work was therefore to investigate the impact of using several different estimation techniques to estimate Arctic SAT anomalies over land and sea ice areas. Five estimation techniques are investigated in this study, which are based on the techniques used for existing temperature anomaly datasets, such as CRUTEM4, GISTEMP and Berkeley Earth, that assimilate only in situ data sources. I explored the degree of difference arising from using these different estimation techniques to estimate Arctic SAT anomalies using European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA) project interim reanalysis, or ERA-Interim, data as a testbed. In Chapter 2 of this thesis I found that ERA-Interim is consistent with independent observations of Arctic SATs and provided realistic estimates of Arctic temperatures, temperature trends and correlation length scales. ERA-Interim outperformed, or was comparable to, all reanalyses available at the time this chapter was produced. I therefore identified ERA-Interim data as a suitable testbed for this study. Two investigations were undertaken. Firstly, the performance of the estimation techniques for the time period for which ERA-Interim is available (1979-2011) was investigated, hereafter referred to as the 'recent decades' experiment. Secondly the interaction of changing station coverage with estimation techniques was investigated using historical meteorological station coverages between 1850 and 2011, hereafter referred to as the 'historical coverage' experiment. The outline of this chapter is as follows. Section 3.2 describes the data and techniques used in this study. Section 3.3 evaluates estimation technique performance in recent decades. Section 3.4 analyses the effect of the historical coverage of meteorological stations. The final section discusses the results and a summary is provided.

3.2 Data and Techniques

The objective was to compare the accuracy of Arctic SAT anomalies produced using five estimation techniques: Linear Interpolation (LI), Global Ordinary Kriging (GOK), Global Simple Kriging (GSK), a restricted finite volume interpolation technique (the 'Binning' technique), and Not Interpolating (NI). A description of these estimation techniques is given in Section 3.2.3.

The estimation techniques were applied to 'input anomalies', which are monthly SAT anomalies from ERA-Interim sampled at Arctic meteorological station locations. In this study the Arctic was defined as north of 65°N, which approximately matches the area northward of the Arctic circle while matching cleanly the grid cell edges of many gridded datasets of relevance. Two investigations were undertaken in this study: one focusing on estimation technique performance in recent decades, and one looking at the effect of using historical coverage. The same input anomalies were used as an input for both

investigations with some slight modifications specific to each investigation (Section 3.2.2).

The estimation techniques yield 'estimated anomalies'; estimates of the ERA-Interim SAT anomaly at several temporal resolutions and, excluding the Binning technique, for both investigations. Two types of anomaly were investigated: 'Arctic-average' anomalies, which are area-weighted averages of SAT anomalies across the Arctic region, and 'spatially resolved' anomalies, which are complete fields of Arctic SAT anomalies. Techniques which produced spatially resolved anomalies as well as Arctic-average anomalies (LI, GOK and GSK) are described collectively as 'interpolating' techniques. The Binning and NI techniques were used to produce Arctic-average SAT anomalies only and are hereafter collectively referred to as 'non-interpolating' techniques. The target areas for the interpolation were land areas and areas of sea ice with a sea ice concentration of more than 15%, these being the conventional areas not addressed by SST anomalies when creating surface temperature datasets. The 'truth' for the target area against which the estimated anomalies were compared was the SAT anomaly from ERA-Interim, described hereafter as 'reference' anomalies.

3.2.1 Reference Anomalies

To compare the accuracy of Arctic SAT anomalies produced using different estimation techniques, a reference dataset was produced; the 'truth' to which the anomalies produced by each estimation technique will be compared. The reference for this study was the SAT anomaly from ERA-Interim, produced using the method described in this section.

Monthly SAT anomalies were produced for each ERA-Interim grid cell from 6 hourly resolution ERA-Interim 2m air temperature data between 1979 and 2011 relative to a 10 year climatological period (1990-1999). A 10 year climatological period was used instead of the conventional 30 year period in order to simulate the use of a conventional climatological period in a longer dataset using ERA-Interim data which only cover 33 years. This was necessary if the performance of the estimation techniques outside of the climatological period was to be investigated. To validate the reconstruction of spatially complete fields of estimated anomalies at various temporal scales, the spatially resolved monthly anomalies were used to create annual anomalies for each calendar year and seasonal anomalies for boreal winter (December, January and February), spring (March, April, May), summer (June, July and August) and autumn (September, October and November). To validate the reconstruction of estimated Arctic-average anomalies, area-weighted averages of the monthly anomalies for the Arctic region were produced for each month using a cosine of latitude weighting. Annual and seasonal Arctic-average anomalies were produced from the monthly Arctic-average anomalies. These monthly, seasonal and annual Arctic anomalies for each ERA-Interim grid cell and for the Arctic area as a whole constitute our reference anomalies.

3.2.2 Input Anomalies

'Input anomalies' is the term used in this study for ERA-Interim data treated as if they were meteorological station data. Using ERA-Interim data as input anomalies instead of actual station measurements enables us to isolate and evaluate the limitations of the estimation techniques. The input anomalies were ERA-Interim grid cell anomalies, sampled at Arctic meteorological station locations.

The locations in latitude and longitude of all meteorological stations in the CRU and Met Office Hadley Centre land surface temperature anomaly dataset version 4 (CRUTEM4) databank were obtained from the list of CRUTEM4 meteorological stations in the International Surface Temperature Initiative (ISTI) Stage 2 dataset (Thorne et al., 2011). The dataset version used in this study was CRUTEM.4.1.1.0. Some of these stations are not included in the gridded CRUTEM4 temperature anomaly dataset as they are missing required information. Nonetheless, we identified all land stations situated north of 53 °N which provided information on Arctic temperatures necessary for the interpolating techniques. Some of the land stations in the CRUTEM4 databank were located within the same ERA-Interim grid cell. These duplicates would have caused a problem for certain methods investigated in this study such as kriging. Therefore, we identified all stations that were duplicates at the ERA-Interim resolution and merged the records. Where only one station reported an anomaly this comprised the merged record for that month and year. Otherwise a simple average of the duplicate station anomalies was used. This creates a single station record comprising the reporting record of all stations located in the same grid cell. The location associated with the merged stations is the latitude and longitude of the first station listed. For visualisation purposes, the locations of the identified, non-duplicated stations are shown in Figure 3.3. Monthly anomalies from the ERA-Interim grid cell containing each identified meteorological station were extracted from the reference dataset. These station anomaly timeseries were masked and used to create input anomaly datasets for estimating anomalies in recent decades and using historical coverages.

3.2.2.1 Recent Decades

Not every station identified from the CRUTEM4 databank for use in creating the input datasets reported a temperature in every year or month between 1979 and 2011. This is due to: stations being moved, added and discontinued; problems with instrumentation; and reports which are not communicated in real time. To account for this, information on which stations report in each year and month between 1979 and 2011 was extracted from the CRUTEM4 databank and the input dataset was masked accordingly. There was no required minimum number of reports. The locations of all stations that contributed to the input dataset for investigating the performance of the various techniques in recent decades (stations listed as reporting between 1979 and 2011) are shown in Figure 3.3.

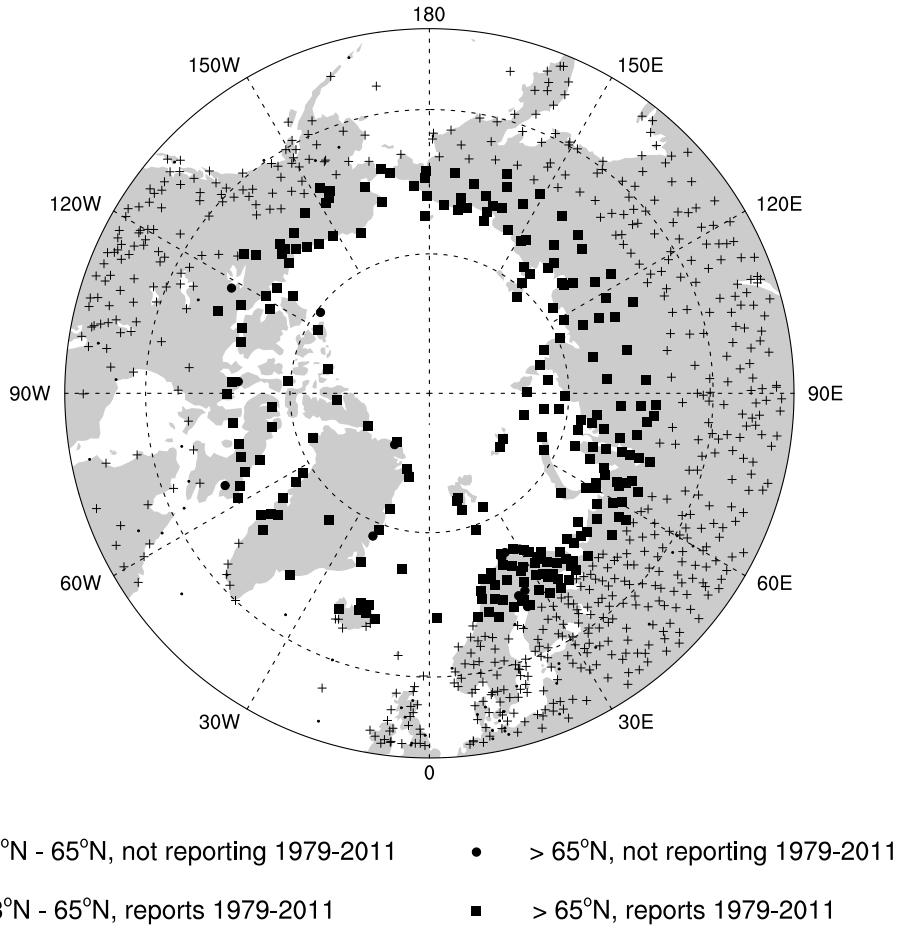

Figure 3.3: The locations of all meteorological stations in the CRUTEM4 databank. Different markers are used to show the locations of the meteorological stations depending on whether they are above 53°N or above 65°N and whether or not they report a single temperature between 1979 and 2011.

3.2.2.2 Historical Coverage

As ERA-Interim data are only available back to 1979 it was not possible to investigate the effect of using historical coverage to estimate Arctic SAT anomalies before 1979 using ERA-Interim data contemporary with the meteorological station coverage. Instead we apply each year's historical station coverage (1850 to 2011) to the estimation of monthly Arctic anomalies for all years of the ERA-Interim period (1979-2011).

Each ERA-Interim year was used to produce an ensemble member. Each ensemble member is a set of repeated instances of the input anomalies for the same 12-month year between 1979 and 2011 (33 ensemble members). As the input anomalies for each repeated year were masked according to the historical station coverage for 1850 to 2011 (162 years), each ensemble member is 162 times 12 months long, with the anomalies from the same 12 months repeated throughout. Information on which input stations reported in each year and month between 1850 and 2011 was extracted from the CRUTEM4 databank. The locations of all stations that contributed to the input dataset for investigating the effect of using historical coverage are shown in Figure 3.3. The monthly input anomalies for each repeated year of an ensemble member were masked according to whether each station reported in each month of each year of historical coverage. For example, the input anomalies for January to December 1979 were first masked according to the station coverage in January to December 1850, then according to the station coverage in January to December 1851, in 1852, and so on. To aid understanding, a diagram of this process is given in Figure 3.4. For each year's historical coverage this gives an ensemble of 33 results (ensemble members) whose statistical properties can be investigated.

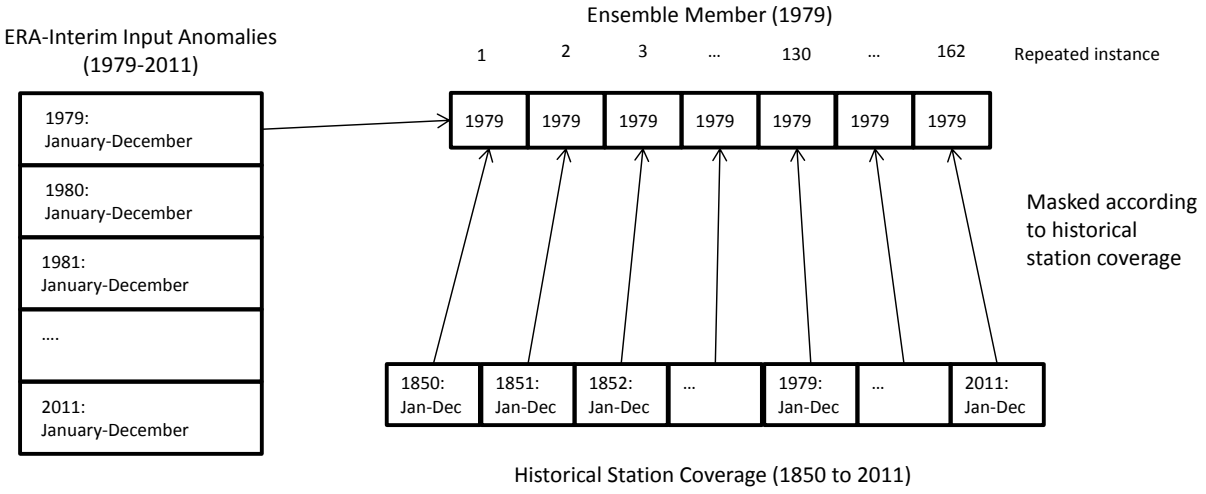

Figure 3.4: A diagram illustrating the creation of an ensemble dataset of input anomalies using ERA-Interim data masked according to station coverage between 1850 and 2011. Each ensemble member comprises a set of repeated instances of one 12-month year from the period 1979-2011, masked according to the station coverage between 1850 and 2011.

3.2.3 Estimation Techniques

3.2.3.1 Linear Interpolation

LI is used to combine input data and estimate SAT anomalies over land and sea ice at unsampled points in the GISTEMP dataset (Hansen et al., 2010). The anomaly at each GISTEMP grid cell is a weighted average of the input anomalies within a certain radius. The LI technique used in this study and described below was based on the GISTEMP technique.

The anomaly estimated at each grid cell is a weighted average of the input anomalies within a 1200km radius of that grid cell. The weight w decreases linearly from unity for input anomalies located at the grid cell being analysed (distance, d, between the input anomaly and the grid point being analysed is 0km) to 0 for input anomalies located more than or equal to the Radius (1200km) away (Eq. (3.1)). This weighting method and the 1200km radius is the same as used for GISTEMP.

$$w = 1 - \frac{d}{\text{Radius}} \quad \text{when } d \leq \text{Radius} \tag{3.1}$$

3.2.3.2 Kriging

Kriging refers to a set of geostatistical methods of interpolation which interpolate variables using a given covariance or semivariance structure (Cressie, 1990). The variance structure is often determined from available observations of the variable. Two forms of kriging were investigated in this study: Global Ordinary Kriging and Global Simple Kriging. The differences between the two kriging methods applied in this study are described in subsections 3.2.3.2.1 and 3.2.3.2.2. A full description of the equations and variables used for both kriging methods is available in Appendix B. A summary of the approach detailed in Appendix B is as follows:

- A semivariance function was determined from the input anomalies. This function prescribed the weighting for this interpolation; stations up to 3585.9km away can contribute to the kriged anomaly and the station weighting halves at about 1200km.
- The semivariance function was used to construct a semivariance matrix between the input anomalies as well as a vector of semivariances between the input anomalies and the grid cell to be analysed.
- Covariances were calculated from the semivariance matrix and semivariance vector.
- A vector of optimal weights was produced by solving the system of linear equations.
- An estimate of the anomaly at each ERA-Interim grid cell was given by the dot product of the vector of optimal weights and the input anomalies.

3.2.3.2.1 Global Ordinary Kriging The first kriging method investigated is Global Ordinary Kriging (GOK). 'Ordinary' kriging assumes stationarity of the mean of the variable, where the mean is unknown a priori. The unknown mean is determined during interpolation by constraining the optimal weights so that they sum to 1. 'Global' refers to the fact that one single covariance function is used for all grid cells. In this study the same covariance function is also used for all months and years to be interpolated.

3.2.3.2.2 Global Simple Kriging GSK is similar to GOK, but with the assumption that the mean of the variable is known rather than unknown. This means that the weights are not required to sum to 1 and they are calculated without this constraint. The mean of the variable is instead added to the dot product of the vector of optimal weights and the input observations to produce an estimate of the variable.

The Berkeley dataset is produced using a method which includes a Simple Kriging technique. In the Berkeley method the temperature anomaly field over land is conceptualised as the sum of the global mean land temperature over time, the climate at each location, and a 'weather' field for each location over time (Rohde et al., 2012, 2013a). Simple Kriging is applied to the weather field using a global correlation function instead of a covariance function (Rohde et al., 2012, 2013a). The correlation function is assumed to be a good approximation for the covariance function as long as the variance changes slowly with time (Rohde et al., 2012, 2013a). This study also uses a method of Simple Kriging with a global variance function. However, GSK in this study interpolates the temperature anomalies rather than a weather field and uses the same global covariance function as GOK rather than a correlation function. A mean of 0 was assumed for GSK as the anomaly observations were expected to have this mean value.

3.2.3.3 Non-Interpolating Techniques

Non-interpolating techniques were the final techniques applied in this study. In this study 'non-interpolation' techniques are those which do not produce spatially complete fields of data. The Met Office Hadley Centre and Climatic Research Unit of the University of East Anglia datasets such as CRUTEM4, HadCRUT4, and the Met Office Hadley Centre SST dataset version 3 (HadSST3) use a non-interpolating method. In this study two non-interpolating methods were explored. Comparing the results from these two methods allows us to explore the impact of spreading information through the use of 5° grid boxes.

3.2.3.3.1 The Binning Technique The first method, the 'Binning' method, is similar to the technique used by the CRUTEM4 dataset. The input anomalies were gridded to the 5° by 5° grid used for CRUTEM4, instead of the ERA-Interim grid used by all other estimation techniques in this study, so that each grid box anomaly is a simple average of all available station anomaly values within that grid box (Jones et al., 2012).

3.2.3.3.2 Not Interpolating In the second method, the Not Interpolating (NI) method, input anomalies were treated as estimated anomalies and analysed without modification.

3.2.4 Estimated Anomalies

The estimation techniques used in this study yield estimates of ERA-Interim monthly Arctic SAT anomalies when applied to the input anomalies. All estimation techniques used in this study produced estimated monthly anomalies gridded to the ERA-Interim grid, except for the Binning technique (Section 3.2.3.3). These monthly anomalies were used to produce estimates of spatially resolved Arctic anomalies and Arctic-average anomalies at monthly, seasonal and annual timescales for both recent decades and historical coverages.

The estimated monthly anomalies from all interpolating techniques were masked using information on monthly average sea ice concentration from ERA-Interim. Only anomalies from Arctic land areas and areas of sea ice with a sea ice concentration of more than 15% were retained as the target areas for this study were areas of land and sea ice. Estimated monthly, seasonal and annual Arctic-average anomalies were produced by area-weighting the estimated anomalies at these timescales using a cosine of latitude weighting. These estimated monthly, seasonal and annual spatially resolved Arctic anomalies and Arctic-average anomalies were compared to the reference anomalies.

3.2.5 Comparison of Estimated Anomalies to Reference Anomalies

To investigate the performance of our chosen estimation techniques for both recent decades and historical coverages the estimated anomalies were compared to the reference anomalies. Errors were calculated for estimates of both spatially resolved and Arctic-average anomalies by subtracting the relevant reference anomalies from the estimated anomalies. These errors were assessed at monthly, seasonal and annual timescales. Two error metrics were calculated: the Root Mean Square Error (RMSE) and Compound Relative Error (CRE). RMSE is a metric which measures the absolute error. CRE measures the relative error i.e. the error variance as a fraction of the expected variance. CRE was calculated using Eq. (3.2). The estimated anomalies are represented by e, r designates the reference anomalies and $\bar{r}$ is the mean of the reference anomalies. CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

$$\mathrm{CRE} = \frac{\sum_{i=1}^{n}(e_i - r_i)^2}{\sum_{i=1}^{n}(r_i - \bar{r})^2} \tag{3.2}$$

Some additional metrics which measure the absolute error were also calculated, but were found to be extremely similar to the RMSE and are therefore not reported here. This similarity between metrics shows that there are few outliers in the errors and that the errors are of a similar magnitude.

3.3 The Performance of Estimation Techniques in Recent Decades

The performance of all estimation techniques, both interpolating and non-interpolating, in recent decades was investigated for the reconstruction of Arctic-average anomalies. Only interpolating techniques were investigated in terms of spatially resolved anomalies in recent decades. Their performance was analysed by comparing the estimated anomalies to the reference anomalies.

3.3.1 Arctic-Average anomalies

To investigate the performance of estimation techniques for estimating Arctic-average anomalies in recent decades, timeseries of estimated and reference anomalies, as well as the errors in the estimated anomalies, were produced and examined.

The timeseries for annual anomalies are shown in Figure 3.5. NI was the least accurate technique for annual Arctic-average anomalies. NI produced estimated annual anomalies with errors of up to nearly 1K, whereas the errors for all other estimation techniques were below 0.60K. Interpolating techniques were more accurate than non-interpolating techniques; their errors were below 0.30K. The errors produced by the different interpolating techniques were very similar to each other. Figure 3.6 shows the estimation errors for monthly Arctic-average anomalies. One representative month is shown for each season. The relative performance of the techniques for monthly anomalies was the same as for annual anomalies. Interpolating techniques were more accurate than non-interpolating techniques, with errors generally below 1K, and the errors produced by the different interpolating techniques were very similar to each other. NI produced monthly anomalies with the largest errors; up to nearly 4K in some months.

The errors in estimated monthly Arctic-average anomalies were found to have a seasonal variation; errors were largest in winter and smallest in the summer (Figure 3.6). This seasonality arises from seasonality in Arctic temperatures. Figure 3.7 is a box-whisker plot of monthly area-weighted Arctic SATs over land and sea ice areas from ERA-Interim. As shown in Figure 3.7, Arctic SATs are smaller in magnitude and have less variability in the summer compared to winter months. In summer the SAT over sea ice varies around the freezing point, as a result of latent heat effects from melting sea ice (Przybylak, 2003a). SATs over sea ice are therefore smaller in magnitude and variability in the summer than in the winter, when their variability is more dependent on the atmospheric circulation than on insolation and latent heat effects. SATs over land areas also show less variability and are smaller in magnitude in the summer so the same seasonal pattern is observed over land (Przybylak, 2003a). This seasonality in Arctic SATs leads to seasonality in the anomalies produced from these temperatures, and therefore also in the size of the errors in the estimated anomalies.

The performance of estimation techniques in recent decades was further investigated

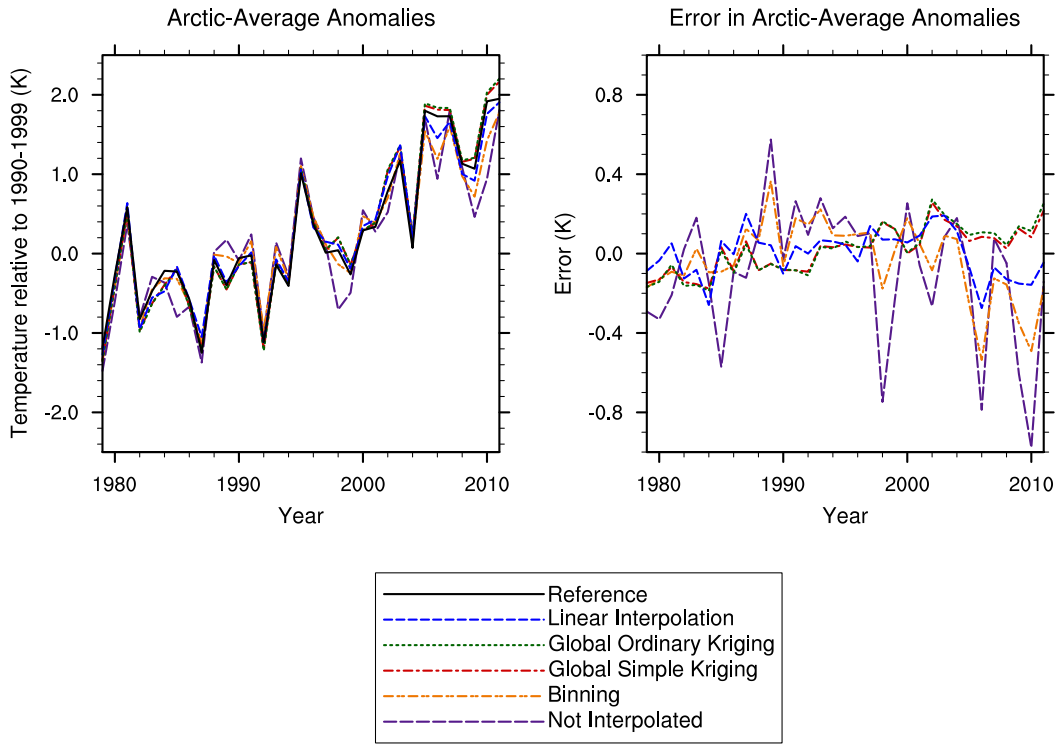

Figure 3.5: Left: time series of annual Arctic-average anomalies between 1979 and 2011 produced using each estimation technique investigated in this study and from reference anomalies. Right: the errors in estimated anomalies relative to the reference anomalies.

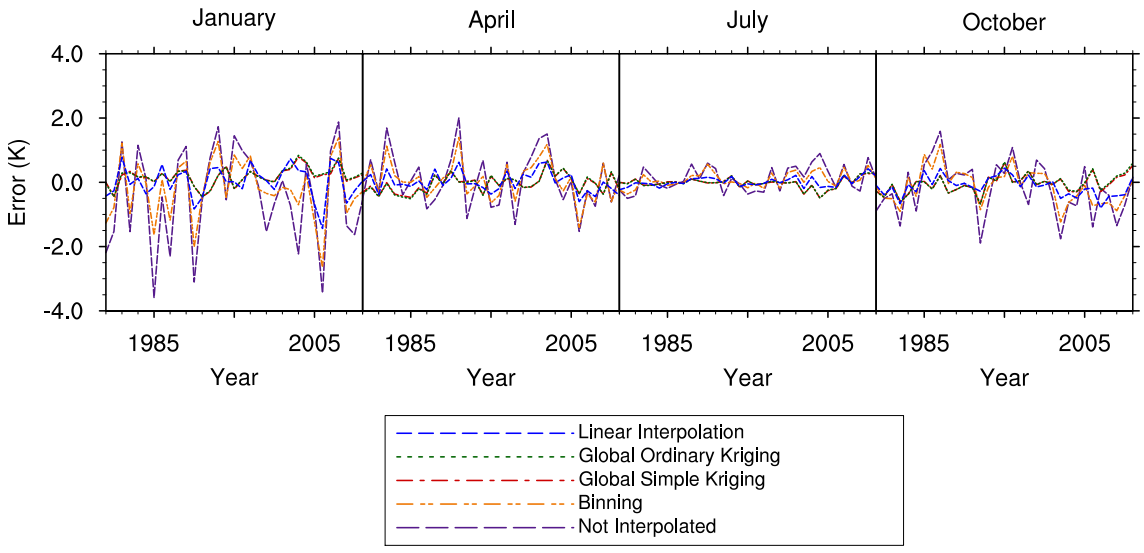

Figure 3.6: Timeseries of the errors in estimated monthly Arctic-average anomalies relative to the reference anomalies between 1979 and 2011 for each estimation technique investigated in this study. One representative month for each season is shown.

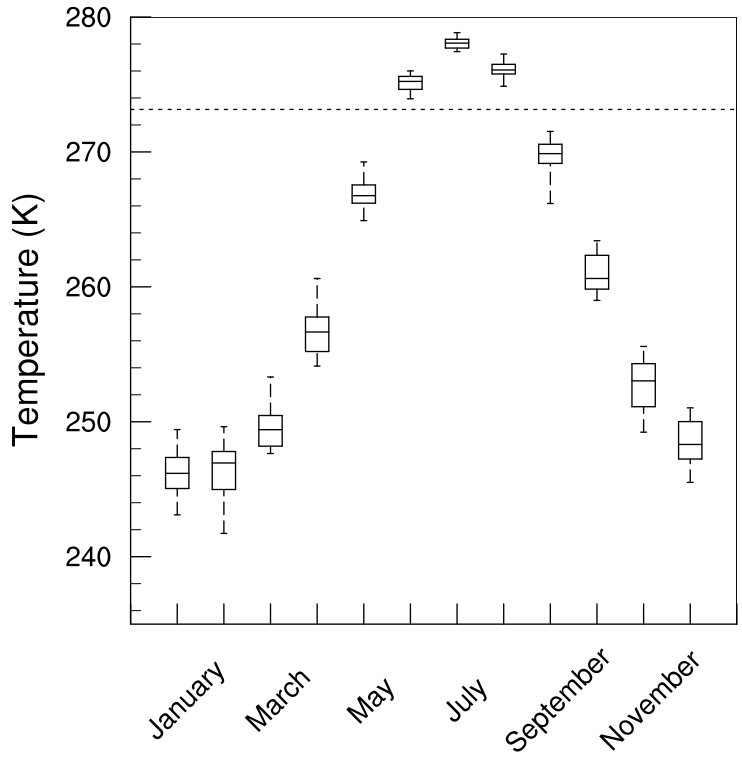

Figure 3.7: A box-whisker plot of the range, median and lower and upper quartiles of monthly area-weighted Arctic Surface Air Temperature averaged over land and sea ice from ERA-Interim between 1979 and 2011. A reference line is included at 273.13K.

by producing a Taylor diagram from the errors in estimated annual and monthly Arctic-average anomalies (Figure 3.8). Taylor diagrams are a way of graphically summarising how well estimated variables match a reference dataset. The Taylor diagram confirms the results mentioned above for estimated monthly Arctic-average anomalies. In addition it shows that kriging techniques were slightly more accurate for the majority of the Taylor diagram metrics compared to LI. However, as noted previously, the differences between the interpolating techniques were small. For example, the cross correlation for annual anomalies produced by both kriging techniques compared to the reference was 0.996 while for LI the cross correlation was 0.991. Kriging techniques were therefore slightly more accurate than LI in general for Arctic-average anomalies and GSK was the most accurate technique. However, none of the interpolating techniques were notably more accurate than the others. This agrees with the findings of similar studies (Cowtan and Way, 2014; Rohde, 2013).

The sizes of the errors produced by the interpolating techniques were fairly consistent in recent decades. However, after 2005 the errors for LI increased slightly while the errors for kriging techniques remained relatively constant as observed in Figure 3.5. One possible explanation for this decrease in technique performance is the impact of the changes that have been seen in the seasonal cycle of Arctic sea ice cover, sea ice extent and heat fluxes since 2005 as a result of rapid ice loss events (Stroeve et al., 2012). However, as both of

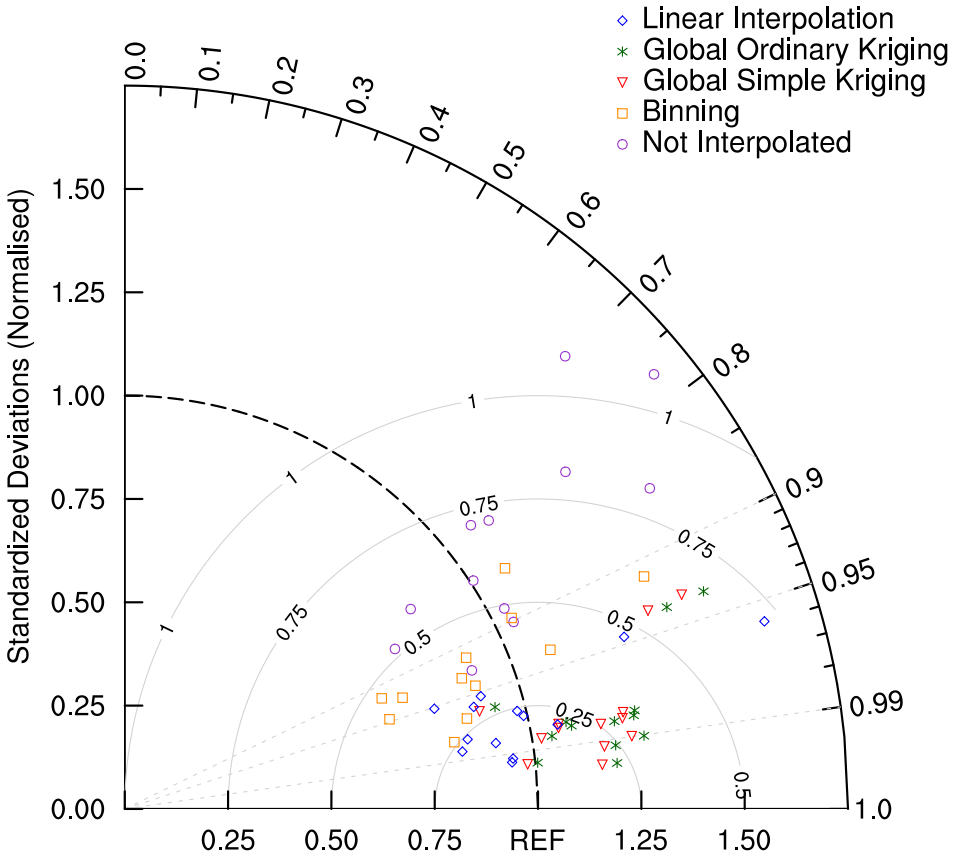

Figure 3.8: A Taylor Diagram comparing estimated Arctic-average monthly and annual anomalies produced by each estimation technique investigated in this study to the reference anomalies. Each symbol plotted represents a month of the year or the annual value. Cross correlation is shown by the angle with respect to the x-axis. The standard deviations (normalised with respect to the reference standard deviation) can be read from the y-axis. The Root Mean Square Error (K) of the estimated anomalies is proportional to the distance to the point on the x-axis identified as REF (shown by the concentric circles marked 0.25 to 1). The values for July estimated by Not Interpolating and Not Interpolating and Regridding are off the scale of this diagram.

the interpolation techniques investigated here use fixed variance functions for all grid cells and for all timesteps, the changes in Arctic sea ice areas due to rapid ice loss events would not explain the relative change in performance of LI compared to the kriging techniques. Therefore it is more likely that the reduction in accuracy of LI after 2005 is due to the impact on this technique of a decrease in the number of input station records as illustrated in Figure 3.1. LI may be more sensitive to reductions in temperature record coverage than kriging techniques. This is explored more fully in Section 3.4.2.

To summarise, the interpolating techniques used in this study provided a more accurate estimate of Arctic-average anomalies than non-interpolating techniques. Kriging techniques were found to provide slightly more accurate estimates than LI and GSK was the most accurate. However, the choice of technique did not make a meaningful difference to the accuracy of the results, especially for annual anomalies.

3.3.2 Spatially Resolved Anomalies

To investigate the performance of interpolating techniques for reconstructing spatially resolved anomalies in recent decades the RMSE and CRE for each grid cell in the Arctic was mapped and examined. Figure 3.9 shows the mapped RMSE and CRE for estimated annual anomalies. All interpolating techniques investigated produced estimates of annual anomalies with RMSEs below 2K for more than 99% of grid cells. For monthly anomalies (not shown), RMSEs were below 2K for 47-99% of grid cells depending on the month of the year; on average 83% of grid cells had RMSEs below 2K. Therefore all interpolating techniques investigated produced estimates of monthly and annual anomalies that were, for the majority of grid cells, within 2K of the reference. The area-weighted average of the RMSE and CRE across the Arctic was calculated for estimated annual and monthly anomalies using a cosine of latitude weighting. Figure 3.10 shows the monthly and annual area-weighted averages for RMSE and CRE. All interpolating techniques produced estimated anomalies that were, on average, within 2K of the ERA-Interim reference as shown in Figure 3.10. The area-weighted averages of the metrics were found to contain a seasonal variation; RMSEs were larger in winter than for other seasons and smallest in summer. This seasonality was also observed for errors in Arctic-average anomalies in recent decades and is explained in Section 3.3.1. CRE, which measures relative error, however, had a seasonal pattern opposite to that of the RMSE; while the absolute errors are smallest in the summer, the relative error is very small for most months and largest in the summer. Consequently, despite the low absolute error for estimated anomalies in summer, there is lower confidence in these anomalies compared to other months as the error is large compared to the size of the anomaly being estimated. This implies that monthly or seasonally varying covariances may be beneficial for interpolating Arctic anomalies.

On average LI was found to produce the poorest results for all metrics for both annual and monthly anomalies. The larger errors associated with LI can be seen in Figure 3.10. The annual and monthly anomaly error metrics for LI were up to 0.15K and 0.77K larger

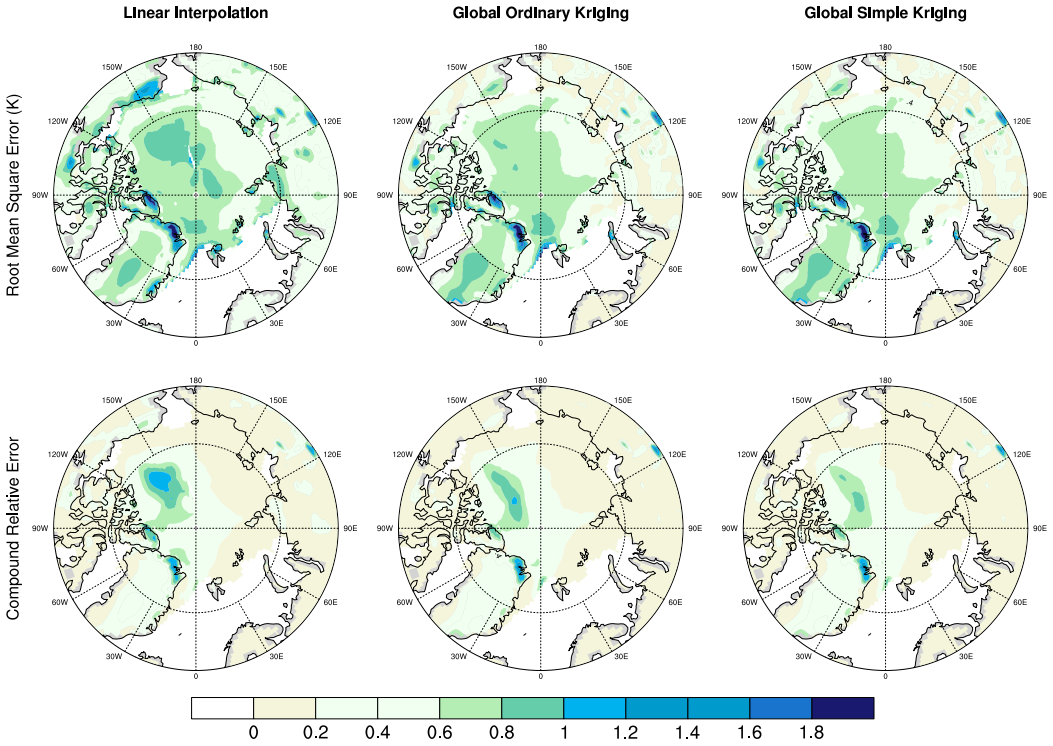

Figure 3.9: The Root Mean Squared Error (K) and Compound Relative Error between spatially resolved annual Arctic anomalies estimated using the investigated interpolating techniques and reference anomalies in recent decades (1979 - 2011). CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

respectively than for kriging techniques. The larger errors in anomalies estimated using LI can also be seen for the majority of grid cells in the maps of RMSE and CRE in Figure 3.9. Kriging techniques therefore produced estimated anomalies closer to the reference than LI.

The kriging methods investigated were equally good at estimating anomalies on average. The difference between the monthly and annual error metrics for the two kriging methods was, for their area-weighted averages, less than 0.10K. In addition, at the grid cell scale, GSK errors were, for the majority of grid cells in the Arctic, within 0.10K of GOK errors (Figure 3.9). However, when the maps of RMSE and CRE in Figure 3.9 were compared for the kriging methods over sea ice areas only, GSK was found to provide a slightly more accurate estimate of anomalies over sea ice areas than GOK. This is due to the choice of a representative mean for GSK which influenced the anomalies produced in the Arctic for regions, such as areas of sea ice, where no observations of SAT were available.

When estimating spatially resolved Arctic anomalies the choice of interpolation method was found to influence the accuracy of the estimated results, unlike for Arctic-average anomalies where the choice of interpolation method did not make a large difference to the results. This suggests that the errors in different sub-regions of the Arctic cancel each other out for Arctic-average anomalies estimated using interpolation techniques.

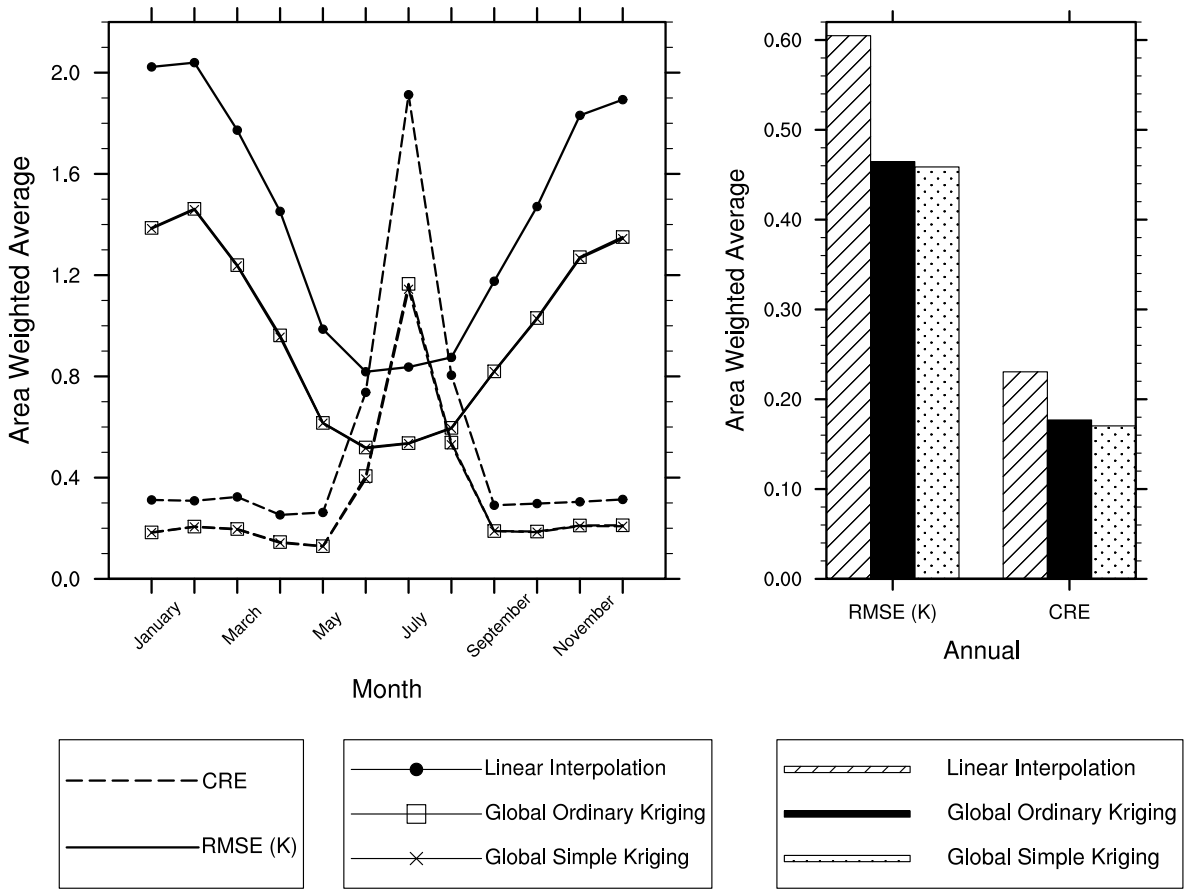

Figure 3.10: The area-weighted average of the Root Mean Square Error (K) and Compound Relative Error between estimated spatially resolved Arctic anomalies (estimated using the investigated interpolating techniques) and reference anomalies in recent decades (1979-2011). Left: monthly anomalies. Right: annual anomalies. CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

Comparing the errors spatially avoids this cancellation of errors and therefore emphasises differences in the performance of the interpolation techniques.

3.4 The Effect of Historical Meteorological Station Coverage on SAT Indices

The second objective of this study was to investigate the interaction of historical station coverage with several techniques used for estimating anomalies. The impact of changing station coverage was simulated by creating an ensemble dataset of input anomalies using ERA-Interim data masked according to station coverage between 1850 and 2011. Each ensemble member comprises a set of repeated instances of one 12-month year from the period 1979-2011, masked according to the station coverage between 1850 and 2011. In other words, each ensemble member is 162 times 12 months long, with the anomalies from the same 12 months repeated throughout, masked according to the station coverage in each month of successive years from 1850 to 2011. The performance of the estimation techniques and the effect of historical coverage was analysed by comparing the estimated anomalies for each ensemble member to the corresponding reference anomalies. This gives an indication of the error statistics of each estimation technique based on the simulation.

For recent decades the performance of all estimation techniques was investigated for the reconstruction of Arctic-average anomalies. The performance of interpolating techniques only was investigated for anomaly pattern reconstructions. The results in Section 3.3 showed that NI was the least representative technique compared to the reference anomalies for estimating Arctic-average anomalies. As a result, in this section only the performance of the other four techniques is described.

3.4.1 Relative Performance of Estimation Techniques

3.4.1.1 Arctic-Average anomalies

The interaction of changing historical station coverage with the chosen estimation techniques was investigated for Arctic-average anomalies by comparing the estimated annual and seasonal Arctic-average anomalies for each ensemble member to the corresponding reference anomalies. This produced ensemble datasets of errors where each ensemble member had an error value for each year or season of historical coverage. An example of an ensemble dataset of errors is shown graphically in Figure 3.11. The RMSE and CRE across ensemble members for each year of historical coverage are shown in Figure 3.12.

When using historical coverages, conclusions about the relative performance of the techniques for Arctic-average anomalies were very similar to those for recent decades. Interpolating methods generally provided a more accurate estimate of Arctic-average anomalies than non-interpolating methods. For historical coverages before 1930 the RMSE and CRE were generally smallest for kriging methods; kriging error metrics were smaller than

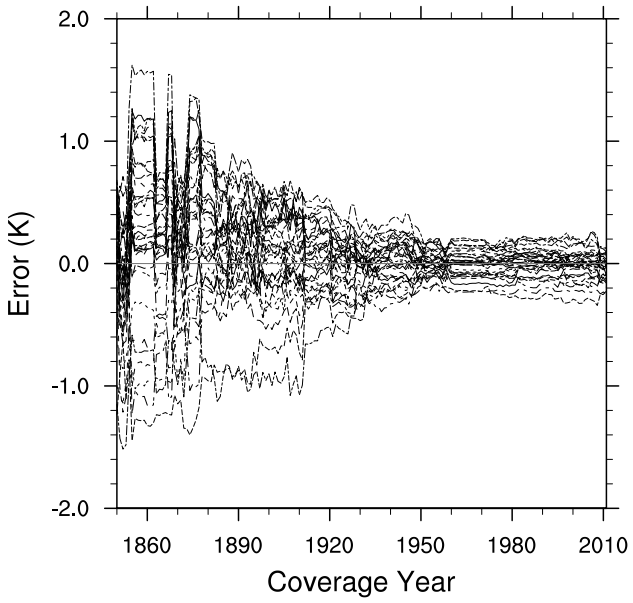

Figure 3.11: The error in annual Arctic-average anomalies estimated by Linearly Interpolating each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) masked using historical station coverages (1850-2011). Similar graphs for all estimation techniques and seasons are provided in Appendix C.

those for LI for between 62% and 92% of coverage years before 1930. The errors for kriging methods also changed less during this time period and exhibited less interannual variability. GSK was comparable to GOK and often produced anomaly estimates which were more representative (Figure 3.12). However, after 1930 (and particularly between 1930 and 1950), LI was more likely than kriging methods to produce more accurate estimates of anomalies; seasonal and annual metrics were smaller for 89% of coverage years on average. Coincident with this was a rapid increase in station coverage between 1930 and 1950 as shown in Figure 3.1. The LI technique used in this study was therefore more sensitive to reductions in station coverage than the kriging techniques. This fits with the results detailed in Section 3.3.1. Nevertheless, the errors for the techniques are, for the majority of years, within 0.20K of each other and LI is not substantially better than the kriging techniques in this time period, especially after 1950.

So, for Arctic-average anomalies estimated using historical coverages the results are similar to those for recent decades. Errors are generally smallest for kriging methods and GSK often produced the smallest errors of the two kriging methods.

3.4.1.2 Spatially Resolved Anomalies

The interaction of historical coverage with interpolating techniques was also investigated for spatially resolved anomalies. A field of RMSE was calculated from spatially resolved errors for each year of historical coverage across all ensemble members. The area-weighted average RMSE was then calculated for each year of historical coverage and the results are shown in Figure 3.13.

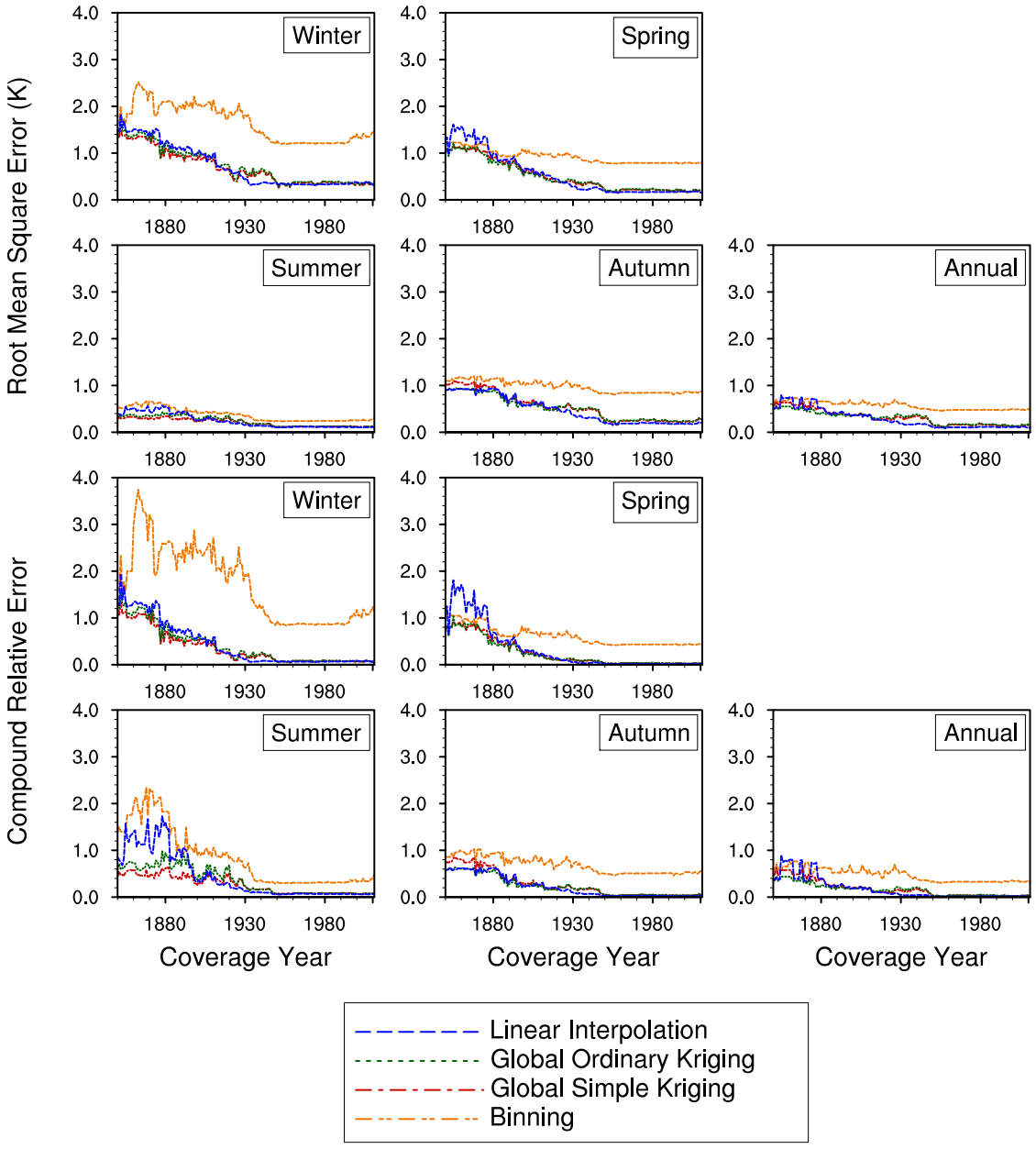

Figure 3.12: The Root Mean Square Error (K) and Compound Relative Error across ensemble members (each year of ERA-Interim anomalies 1979-2011) in each historical coverage year for estimated seasonal and annual Arctic-average anomalies. CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

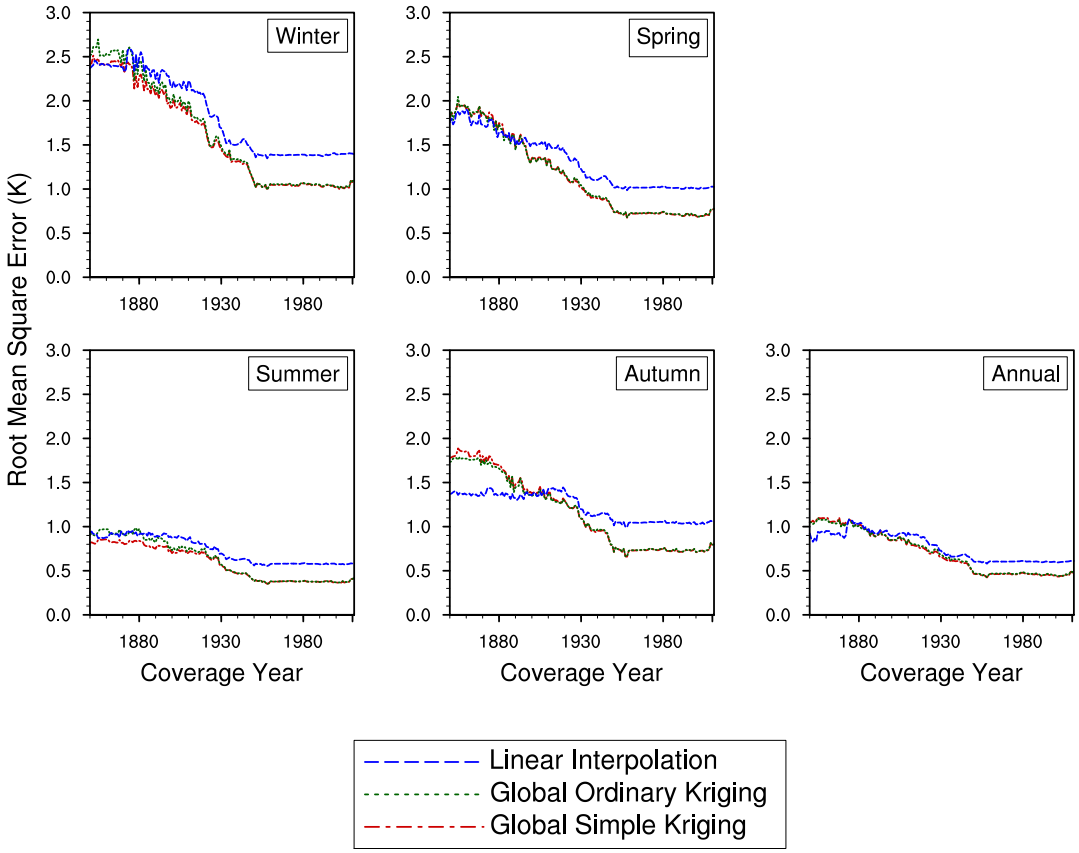

Figure 3.13: The area weighted Root Mean Square Error (K) for spatially resolved Arctic anomalies across ensemble members (each year of ERA-Interim anomalies 1979-2011) produced by the investigated interpolating techniques.

For spatially resolved anomalies estimated using historical coverages the results are again very similar to those for recent decades. For most months and coverage years kriging techniques were more likely to produce spatially resolved anomalies with greater accuracy than LI. Errors produced by kriging techniques were smaller for 62-91% of coverage years after 1890, on average by 0.2-0.4K. For coverages before 1890, LI was slightly more likely to produce estimated anomalies with greater accuracy than the kriging techniques, except in summer. However, for most months the errors are only about 0.20K smaller and LI is therefore not notably better than kriging techniques in this time period, except for autumn anomalies. Neither kriging technique was substantially better in terms of estimating spatially resolved anomalies on average. However, when maps of the RMSE were examined the performance of the kriging techniques was not constant in time. For coverages prior to 1910, GOK often produced slightly more accurate estimates of anomalies for sea ice areas than GSK as illustrated in Figure 3.14. After this GSK was more representative of SATs over Arctic sea ice.

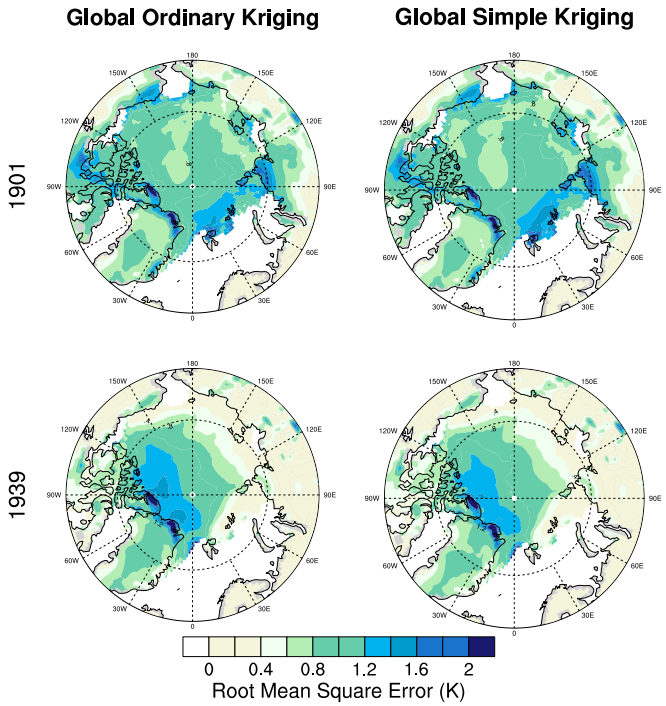

Figure 3.14: The mapped Root Mean Square Error (K) across an ensemble of 33 different years of ERA-Interim data (1979-2011) for spatially resolved annual Arctic anomalies produced using Global Ordinary Kriging and Global Simple Kriging for two example historical coverage years (1901, top; 1939, bottom).

In conclusion, kriging techniques produced the smallest errors, in general, for spatially resolved seasonal and annual anomalies when using historical station coverages between 1850 and 2011. Neither kriging technique was substantially better in terms of estimating spatially resolved anomalies on average.

3.4.2 The Interaction of Historical Coverage with Estimation Techniques

All techniques, regardless of their relative performance, were found to have larger values of RMSE and CRE for earlier historical coverages. For Arctic-average anomalies (Figure 3.12) the largest error metric values occurred before 1890. After this they decreased until about 1950 when the smallest values were reached and the values remained relatively constant. Using pre-1950 historical coverages caused larger errors and a greater uncertainty, as measured by the spread of the errors across the ensemble members; errors can be up to 2K larger for earlier historical coverages. This effect was observed for all four of the estimation techniques investigated for Arctic-average anomalies. In addition, the errors were more variable prior to 1950 which shows that the historical coverage impacted both the magnitude and interannual variability of the errors. For spatially resolved anomalies (Figure 3.13) the temporal evolution of error metrics was very similar to the temporal evolution seen in Arctic-average anomaly estimations. The largest metric values occurred in earlier historical coverage years and decreased until about 1950 before reaching their smallest values. Errors in anomaly patterns can be up to 2.5K larger on average for coverages before 1950 as well as showing more interannual variability.

The interaction of historical coverages prior to 1950 with interpolating techniques given example (1979-2011) anomaly fields resulted in larger errors and a greater uncertainty in the estimated anomalies. This was observed for both Arctic-average and spatially resolved anomalies and for all techniques. Therefore, there were no substantial changes in the relative performance of the estimation techniques when historical coverages were used but the general performance of the techniques did change. This shows that reductions in station coverage have an impact on all estimation techniques investigated in this study. As the number of stations observing decreases the observing network is increasingly likely to miss the key features of weather patterns. This will introduce larger errors and uncertainties into anomalies estimated from sparse data coverage.

3.4.2.1 Artificial Trends

The interaction of historical coverages prior to 1950 with interpolating techniques given example (1979-2011) anomaly fields may introduce artificial trends between the start and end of each ensemble member (which are repeated occurrences of the same year, so we would not expect any significant trend in their reconstruction).

In order to investigate this further the Mean Error (ME) for each ensemble member was calculated for estimated seasonal and annual Arctic-average anomalies in each year of historical coverage. The Percentage of ensemble members with Positive Errors (hereafter referred to as the PPE) was also calculated for each year of historical coverage. This was done for each of the estimation techniques investigated in this section and the results are shown in Figure 3.15. A negative ME shows that there is a general underestimation

of the anomaly values in this ensemble simulation and vice versa, because 1979-2011 has generally positive anomalies. The PPE indicates whether the majority of ensemble members show the same sign in errors as the ME (the average of these errors). If the PPE does not match the sign of the ME then this is an indication that the distribution of errors may be skewed.

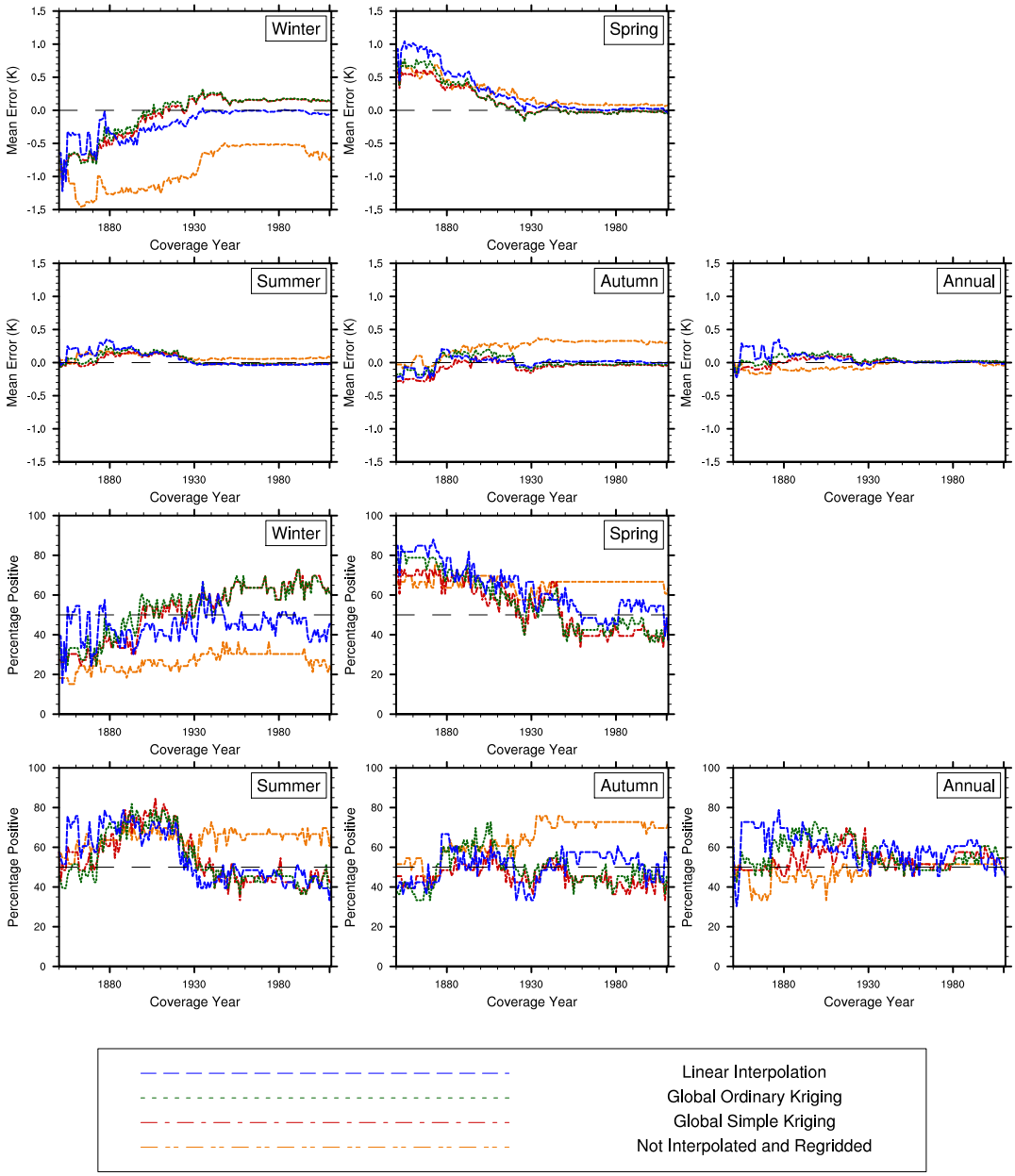

Figure 3.15: The Mean Error (K) for ensemble members (each year of ERA-Interim anomalies 1979-2011) and the percentage of ensemble members with positive errors in each historical coverage year for estimated seasonal and annual Arctic-average anomalies. Anomalies were estimated using all estimation techniques investigated in this study excluding Not Interpolating.

Estimates of summer, autumn and annual anomalies have ME values that are very small. The temporal pattern in ME corresponds with the PPE for all techniques for all three seasons. These seasons not only have small biases, shown by the small ME

values, but also small absolute errors as shown in Figure 3.12. This means that no significant artificial trends will be observed for summer, autumn or annual anomalies over the course of each ensemble member in this simulation. In winter the estimated Arctic-average anomaly MEs are negative for all techniques based on coverages prior to about 1930. For coverages after 1930 the kriging techniques slightly overestimate anomalies on average while the ME for LI is very small. For spring anomalies all techniques overestimate anomalies on average for coverage until 1930. After 1930 the ME is either slightly positive or near 0K. For both winter and spring the PPE corresponds with the ME for all techniques, except GOK anomalies in spring where larger negative errors were influencing the ME. This means that not only are there biases for anomalies in both seasons but that the MEs are also large compared to all other seasons and annual anomalies. These larger errors were also observed in the RMSE for estimated Arctic-average anomalies (Figure 3.12). Therefore, not only are biases present in the estimates of anomalies for winter and spring, but also that the errors are large enough that these biases may produce artificial trends when estimating Arctic climate change in these seasons using the estimation techniques investigated here.

The potential artificial trends in this simulation for winter, spring and annual Arctic-average anomalies were investigated using anomalies produced by GSK (the most representative estimation technique in this study). Although the biases in annual anomalies were not found to be large they were investigated as annual anomalies are an important metric of climate change. The ME for coverages between 1850 and 1950 was plotted and a trendline was fitted through it. The trend in the errors for each ensemble member was also investigated and histograms of this information were plotted. These are shown in Figure 3.16. The ME shows that there is the potential for winter and spring anomalies to artificially increase and decrease by 1.19K and 0.72K respectively, on average, over the course of the ensemble due to the effect of changing station coverage. The distribution of errors in each ensemble member shows that the trends could be -1.66 to 5.05K and -3.77 to 1.71K for winter and spring anomalies respectively. To put these values into perspective, the average artificial trends in summer and winter Arctic-average anomalies estimated using coverages between 1850 and 1950 in this simulation are of the same order of magnitude as the trends observed in real-world temperature anomaly datasets for annual Arctic-average anomalies during this time; CRUTEM4 shows an increase of 1.75K in Arctic temperatures between 1850 and 1950. For annual anomalies the ME shows that there is the potential for an artificial increase of 0.15K (error trends are between -1.37 and 2.28K) on average over 100 years due to the effect of changing station coverage in this simulation. This is about 10% of the trend seen in annual Arctic-average SAT anomalies from existing temperature anomaly datasets between 1850 and 1950.

However, as explained in Section 3.4 the results described in this section only indicate the error statistics of each estimation technique based on the simulation, not real world errors in estimation. The extrapolation of the conclusions of this simulation to

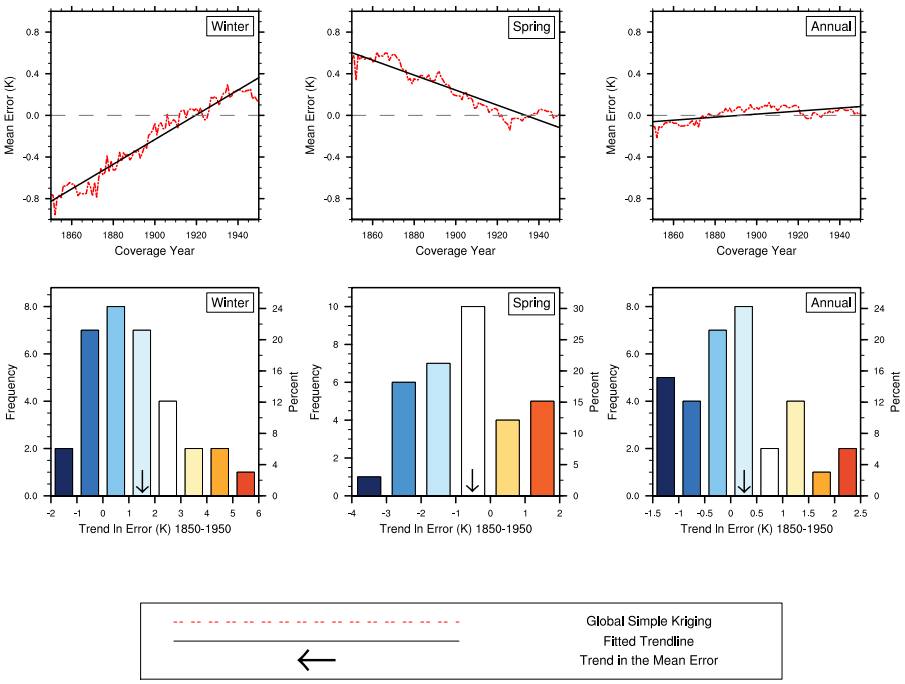

Figure 3.16: The trend in (top) Mean Error (K) across the ensemble and (bottom) errors in each ensemble member. Each ensemble member comprises repeated instances of Winter, Spring and Annual Arctic-average anomalies for an individual year (1979-2011, see text for further details) estimated using Global Simple Kriging.

real world errors in Arctic climate change estimates would depend on the assumption that 1979-2011 temperature anomaly patterns, a period during which changes to Arctic temperatures and sea ice variables are known to have been rapid, are representative of 1850-1950 temperature anomaly patterns. The results presented here should therefore not be interpreted as concluding that the same artificial trends are present for the Arctic in existing temperature anomaly datasets as a result of the use of pre-1950 historical coverage.

3.5 Discussion

In this study the choice of technique used to estimate Arctic SAT anomalies over land and sea ice areas was found to have an effect on the accuracy of the estimated anomalies produced. For Arctic-average anomalies, in both recent decades and for historical coverages, it was found that interpolating methods were most representative of the ERA-Interim reference. This is a result of the sparseness of temperature data in the Arctic, particularly over sea ice. Sparse sampling of a region means that large areas are unrepresented when non-interpolating techniques are used to estimate anomalies. This introduces uncertainty to the calculation of global and regional average temperature changes, especially if the unrepresented areas are likely to be warming at a faster (or slower) rate than sampled regions, such as in the Arctic. Therefore, as long as an interpolation technique provides a reasonable estimate of the anomalies in unsampled areas of the Arctic it will provide a

more representative estimate of the Arctic-average anomaly. The aforementioned sparse sampling of Arctic temperatures also explains the more accurate results from the Binning technique compared to the NI technique. In gridding the anomalies to a larger 5° grid there is some spatial infilling of anomalies, leading to a slightly more representative estimate of Arctic-average anomalies.

Of the interpolating techniques investigated, kriging techniques provided the smallest errors in estimates of Arctic anomalies overall. GSK was often the most accurate kriging method. This was observed for Arctic-average anomalies in both recent decades and for historical station coverages. Nonetheless, the choice of interpolating technique did not make a substantial difference to the accuracy of the results, especially for annual anomalies. Kriging techniques also produced the most representative estimates of spatially resolved anomalies. For both investigations the kriging techniques produced anomalies that were, on average, equally good estimates. GSK was generally more representative over sea ice regions. However, SAT anomalies over sea ice were slightly better estimated by GOK prior to about 1910 in the historical coverage investigation. Yet GOK is not notably better and, as the impact of changing station coverage on estimation techniques was investigated using an ensemble dataset of example (1979-2011) anomaly fields, the results should be regarded with caution.

LI was the least representative of the interpolation techniques, both for Arctic-average anomalies and spatially resolved anomalies. This is a result of both the sparseness of temperature data in the Arctic as well as the characteristics of the LI technique. Firstly, the radius used for LI, which is based on the GISTEMP technique, is smaller than the radius for the kriging techniques, which was identified from the variance of the input data. Some areas of the Arctic Ocean are at the extreme limits of the extrapolation radius and, as a result, the interpolation is influenced heavily by a very small number of stations distant from the location being estimated which increases the error in the estimated anomalies. Secondly, in sparsely observed regions where the distance to the nearest station is large, kriging techniques relax towards a prior value, either zero or the mean of the variable. This effectively bounds the maximum weight of an individual station and leads to a smoother interpolation in regions which are poorly observed. LI, however, does not relax towards a prior value and this increases the influence of a small number of stations on the interpolation. Therefore both the sparseness of temperature data and the features of the LI technique cause the LI technique to be less representative than kriging techniques for Arctic SAT anomalies.

Of the kriging techniques investigated in this study, results from GSK were the most representative of the reference anomaly fields. The difference between the kriging techniques is in the prior value they relax to. GOK relaxes to the best linear unbiased estimator of the mean of the variable whereas for GSK the mean is defined a priori. Therefore GOK should be a more robust choice of technique as the results are not dependent on the choice of a representative mean. There are a few possible explanations for the superior

performance of GSK compared to GOK. If the central Arctic is sufficiently isolated from the weather in the rest of the Arctic, which informs the estimated mean for GOK, then the a priori mean chosen for GSK may be a better fit than the calculated GOK mean. Related to this, the input anomalies include stations down to 53°N. This includes regions that are showing a cooling signal during a similar time period to the Arctic warming (e.g. Cohen et al., 2012). This could also affect the estimated GOK mean and lead to the GSK mean being more representative. The better accuracy of GSK over GOK could also be a feature of the relatively short time period and climatology period used in this study. A mean of 0 might be a good approximation for the 33 years of data and 10 year climatology used in this study but may be less representative of a longer time period, in which case GOK or the use of a slowly changing climatology as the prior might provide better anomaly estimates. GSK provides a better estimate of Arctic anomalies in this study but GOK may be a better choice of kriging technique in general as it is not dependent on the choice of a representative, and, in this study, constant, mean.

NI was the least representative technique compared to the reference for Arctic-average anomalies in both investigations. However, anomaly estimates from non-interpolating techniques are still helpful; they serve as useful checks for confirming whether estimates from interpolating techniques are reasonable. Also, using different techniques to estimate anomalies allows us to look at structural uncertainty in estimates of climate change. Furthermore, it must be noted that, while kriging techniques were found to provide more accurate estimates of Arctic anomalies in general, the choice of an estimation technique to apply to Arctic SAT anomalies will depend on the time period investigated, whether anomaly patterns or Arctic-average anomalies are being studied and which area of the Arctic is to be researched. For example, if a study's objective were to estimate spatially resolved Arctic SAT anomalies over sea ice in recent decades the results of this study suggest that GSK should be applied. If land area anomalies were to be estimated instead then the results of this study suggest that either kriging technique would be equally accurate. Also the parameters, or variograms in the case of kriging, chosen for each technique will have an impact on its estimation; as part of this study LI with a larger radius of 3585.9km (informed by the kriging semivariogram) was investigated and found to produce anomaly estimates with much larger errors.

All techniques, regardless of their relative performance, were found to have larger values of RMSE and CRE for earlier historical station coverages. The interaction of historical coverage with all estimation techniques leads to larger errors and a greater uncertainty in the anomalies produced. This was observed for both Arctic-average and spatially resolved anomalies. Sparser observing networks of temperature data are more likely to miss key features of weather patterns and will therefore introduce larger errors and uncertainties into anomalies estimated from sparse data coverage. These results were produced using an ensemble dataset which simulates the interaction of meteorological station coverage with estimation techniques. This means that these results only indicate the error

statistics of each estimation technique based on the simulation, not real world errors in estimation. The extrapolation of the conclusions of this simulation to real world errors in Arctic climate change estimates would depend on the assumption that 1979-2011 temperature anomaly patterns, a period during which changes to Arctic temperatures and sea ice variables are known to have been rapid, are representative of 1850-1979 temperature anomaly patterns. In addition, interannual meteorological variability is not included in each ensemble, which is the anomalies from the same 12 months repeated throughout. Interannual meteorological variability may mask the coverage bias impact. Using a longer time period of temperature anomaly patterns, especially ones outside the rapid changes of the past few decades, with more ensembles would provide a more robust investigation of this subject. Nonetheless, the results do tell us to be cautious when using such estimation techniques in extremely data-sparse regions, such as the Arctic. They also support efforts to increase data sharing and data rescue, for example ISTI, the international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative and the Canadian historical data typing project, to increase the coverage of temperature records, particularly in the Arctic (Allan et al., 2011; Thorne et al., 2011; Slonosky, 2014).

In addition to acknowledging the limitations of the above results for historical coverages, it must also be noted that, as a result of using reanalysis data, the uncertainties and noise associated with actual data are not present in this study's test data. Reanalysis data can, however, contain biases, uncertainties and errors and some of these issues may be present in the ERA-Interim data used in this study (Dee et al., 2011a; Inoue et al., 2009; Jakobson et al., 2012; Liu et al., 2008; Thorne and Vose, 2010). Furthermore, although ERA-Interim provides realistic estimates of Arctic temperatures and temperature trends for areas of the Arctic studied thus far (Chung et al., 2013; Dee and Uppala, 2009; Jakobson et al., 2012; Lindsay et al., 2014; Lüpkes et al., 2010; Screen and Simmonds, 2011), it is not necessarily a good representation of Arctic SATs in other, uninvestigated, areas.

3.6 Summary

In this study it was found that the technique chosen to estimate Arctic SAT anomalies over land and sea ice had an impact on the accuracy of the estimated anomalies produced in our ERA-Interim testbed. This was observed for both recent decades and when using historical station coverages. Interpolation techniques produced the most accurate estimates of anomalies compared to the ERA-Interim reference data. Kriging techniques provided the smallest errors in estimates of Arctic anomalies and Simple Kriging was often the best kriging method in this study, especially over sea ice. Non-interpolating techniques provided the least representative anomaly estimates. However, estimates of anomalies from these techniques are still beneficial as they are useful checks for confirming whether estimates from interpolating techniques are reasonable. The interaction of

meteorological station coverage between 1850 and 2011 with estimation techniques was simulated using an ensemble dataset comprising repeated individual years (1979-2011). All techniques, regardless of their relative performance, were found to have larger values of RMSE and CRE for earlier historical coverages. Reduced station coverage introduced larger errors and uncertainties into anomalies estimated from this sparser data coverage. This supports calls for increased data sharing and rescue, especially in sparsely observed regions such as the Arctic.

Chapter 4

Can we Assess the Impact of Estimating Arctic Surface Air Temperature Anomalies with Global Simple Kriging using In Situ Data?

4.1 Introduction

In Chapter 3 I investigated the impact of using several different estimation techniques, based on those used in current temperature anomaly datasets, to estimate Arctic Surface Air Temperature (SAT) anomalies over land and sea ice areas using European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA) project interim reanalysis, or ERA-Interim, data as a testbed. In this 'reanalysis study' the choice of technique had an affect on the accuracy of the estimated anomalies, relative to the reanalysis reference.

Reanalysis data are an estimate of climatic variables, and as such will not contain the same errors from noise and sampling as in situ data sources. But, they may still contain biases, uncertainties and errors from sources such as input data biases, atmospheric forecast model errors, and sparse sampling of input variables. This could mean that the conclusions reached in the reanalysis study about the accuracy of estimated anomalies may not be similar to the accuracy of these techniques in the real world. The gap between the reanalysis and the reality in terms of the results presented in Chapter 3 should be assessed.

Therefore the objective of this chapter was to repeat the recent decades investigation of Chapter 3 using in situ data and compare the results of this 'in situ data study' to the results of the reanalysis study. This is in order to investigate whether we can validate the results of the reanalysis study using in situ data. I applied Global Simple Kriging (GSK, the most representative technique in the reanalysis study) to monthly SAT anomalies from Arctic meteorological stations. I then assessed the accuracy of these in situ data

study estimated anomalies using independent, in situ data sources. Two investigations were undertaken: one evaluating the estimated anomalies over Arctic land areas; and one validating these anomalies over Arctic sea ice. The outline of this chapter is as follows. Section 4.2 describes the data and techniques used in this study. Sections 4.3 and 4.4 evaluate the accuracy of the estimated anomalies over Arctic land and sea ice areas respectively and compare the results to those of the reanalysis study. The final two sections (Sections 4.5 and 4.6) discuss the results and provide a summary and conclusions.

4.2 Data and Techniques

GSK was applied to 'input anomalies', which in this study are monthly SAT anomalies from Arctic meteorological stations for the period 1950-2013. An overview of this estimation technique is provided in Section 3.2.3 of this thesis and more details are provided in Appendix B. Model function fitted parameter values (0 for s, 7.1 for y and 3683.0 for ϵ) were calculated from the Arctic meteorological station SAT records used as input anomalies. GSK yielded estimates of spatially resolved Arctic SAT anomalies, which are complete fields of SAT anomalies gridded to an Equal-Area Scalable Earth grid (EASE-Grid, Section 4.2.3.3). The target areas for the interpolation are the same as for the reanalysis study: land areas and areas of sea ice with a sea ice concentration of more than 15%. The accuracy of the estimated anomalies was assessed using independent in situ data sources. The validation data over land were monthly SAT anomalies from independent Arctic meteorological stations. The validation data over sea ice were data from 'drifting platforms'; North Pole Drifting Stations (NPDS) and U.S Army Cold Regions Research and Engineering Laboratory (CRREL) Ice Mass Balance Buoys (IMBs). The estimated anomalies and validation data were compared; difference metrics and double difference metrics were calculated from validation data over land and sea ice respectively. These difference, and double difference, metrics were compared to reanalysis study error metrics, which were produced relative to a reanalysis 'truth'.

4.2.1 Input Anomalies

The input anomalies for this in situ data study were monthly SAT anomalies from Arctic meteorological stations contained in the CRUTEM4 land surface temperature anomaly dataset databank (produced by the Climatic Research Unit of the University of East Anglia and the Met Office Hadley Centre, Jones et al. 2012). I chose this dataset for consistency with the reanalysis study. The station records that constitute the CRUTEM4 databank were downloaded from the Met Office Hadley Centre Observations Dataset website ($http://www.metoffice.gov.uk/hadobs/crutem4/data/download.html$). The dataset version was CRUTEM4.2.0.0. Stations situated north of 53 °N (which provide temperature information required for interpolating techniques) were identified. The locations of the identified stations are shown in Figure 4.1. All stations were found to have

temperature records between 1950 and 2013, which was the time period of interest for this study. This is not the same time period as investigated in the reanalysis study, to which the results of the in situ data study will be compared to. However, a longer time period of interest was required here so that sufficient validation data was available over sea ice. Monthly SAT anomalies 1950-2013 were produced for all input stations using station normals provided in the CRUTEM4 databank. The station normals in CRUTEM4 are produced for the period 1961-1990, which is the climatological period in this study. These station anomalies comprise the input anomaly dataset for this study. Some of the identified input stations were located at the same latitude and longitude. As duplicate locations would have caused a problem for GSK, these 'duplicate' stations were identified and the station anomaly records merged. As in Chapter 3 a single station record is produced by averaging the duplicate station records; when one station reports an anomaly this comprises the merged record, otherwise the merged record is a simple average of the duplicate station anomalies.

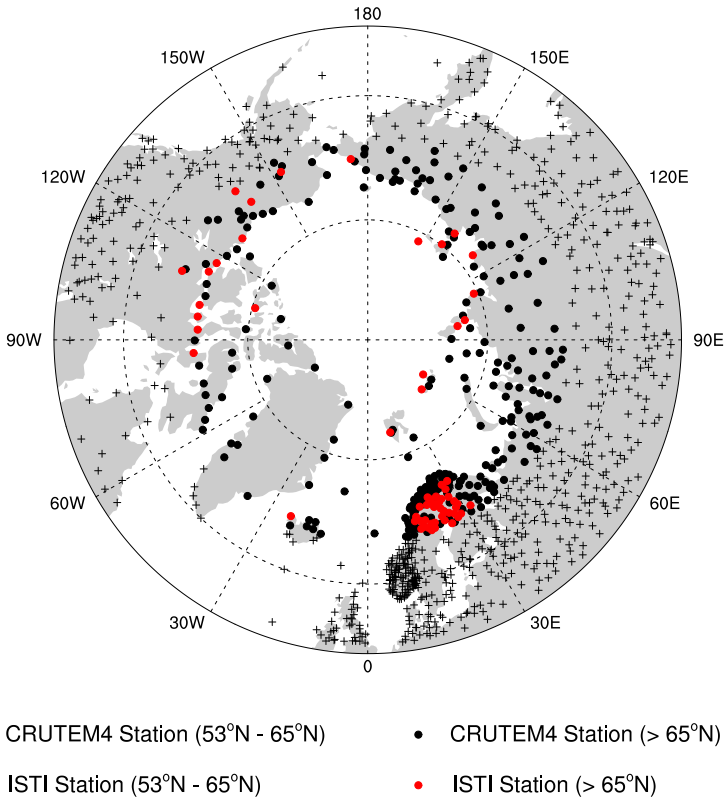

Figure 4.1: The locations of all meteorological stations in the CRUTEM4 databank (black), and the 63 independent land stations from ISTI (red) that constitute the land validation data in this study. Different marker styles are used to show the locations of the meteorological stations depending on whether they are above 53°N or above 65°N. All stations had records in the time period of interest (1950-2013).

4.2.2 Validation over Land

To evaluate this study's estimated anomalies over land, independent validation data were required. The independent validation data over land were anomalies from Arctic meteorological stations which were not used as input anomalies. Metadata and SAT records for Arctic meteorological stations over land were extracted from the International Surface Temperature Initiative (ISTI) merged Stage 3 version 1.0.0 recommended monthly databank product (Thorne et al., 2011). This is a single consolidated databank of in situ surface temperature observations from many different sources. I compared the metadata for all Arctic stations in the ISTI databank to the metadata for the input anomaly stations. I identified independent stations, which were not included in the CRUTEM4 databank, with data records in the time period of interest (1950-2013) and at least 14 years of data in the climatological period (1961-1990). This threshold for the production of station normals is the same as used for CRUTEM4 (Section 4.2.1).

In order to isolate errors resulting from the estimation of anomalies using GSK, the station records used to produce the validation data were required to be reliable and of good quality. Of the 88 independent stations suitable for this study, 28 stations were classified in the metadata as having raw, unprocessed data; 47 stations had been quality controlled by the originator; 10 had been homogenised by the originator and 3 were unclassified. Additional homogenised versions of 13 station records were identified from the Global Historical Climatology Network-Monthly temperature dataset version 2 (GHCN-Mv2) data in the ISTI Stage 2 dataset (GHCN-Mv2.1.1.20120508). The ISTI Stage 2 dataset contains data from various sources converted to a common data format, but not consolidated. I selected all stations with homogenised records as validation stations. I then examined the SAT records and metadata for the remaining, unhomogenised stations. Any station which had a quality controlled record, no large gaps in the data record and no obvious inhomogeneities was also selected as a validation station. Using these criteria, I identified 63 validation stations. Their locations are given in Figure 4.1. Anomalies were produced from the validation station SAT records with respect to 1961-1990 and these constitute the validation data over land.

4.2.3 Validation over Sea Ice

The accuracy of in situ data study estimated anomalies over sea ice areas was assessed using temperatures from 'drifting platforms' as validation data. Drifting platforms is the term I use here to describe platforms housing temperature sensors at around 1.5m height which are deployed on Arctic sea ice (SAT is generally measured at between 1.25 and 2.0m above the surface of the Earth).

I employed two validation data sources here; North Pole Drifting Stations (NPDS) and the U.S Army Cold Regions Research and Engineering Laboratory (CRREL) network of Ice Mass Balance Buoys (IMBs). I identified NPDS data as suitable validation data as

these manned, drifting meteorological stations are considered to be a reliable source of SAT observations for the Arctic Ocean (Rigor et al., 2000). However, because they are manned platforms, NPDS validation data are sparse. There was also a gap in NPDS deployment from 1991 to 2003, between the dissolution of the Soviet Union and the reinstatement of the programme by Russia. Therefore I decided to perform a second sea ice validation using temperature records from CRREL IMBs. CRREL IMBs have been frequently deployed on Arctic sea ice since 1997 and provide radiation shielded temperature observations at 1m height (Elder, 2013). Radiation shielding prevents the temperature observations being influenced by direct or reflected solar radiation and precipitation, which can be an issue for in situ temperature sensors (e.g. Hall et al., 2015). However, it must be noted that much of the CRREL IMB data is preliminary, and therefore may not be quality controlled. In addition, these are automatic sensors which are generally left unattended without routine maintenance, and therefore may be subject to various issues which affect accuracy such as rime ice. Furthermore, I found that there can be large differences between NPDS and IMB data, which may indicate that accuracy issues are indeed present in the IMB records. As a result I decided to produce two independent validations over sea ice, one using NPDS and one using CRREL IMBs. If both validations show similar results this will give us more confidence in the conclusions.

Due to the relatively short record of each drifting platform and their movement on sea ice across the Arctic Ocean it is impossible to produce anomalies from traditional normals (generally used as a reference for temperature anomalies) for these data. Datasets which could provide a gridded climatology for the Arctic were considered but all such datasets were either land only, ocean only, or interpolated. Rather than using interpolated data, which would contain uncertainties resulting from the use of an interpolation technique, I decided to investigate the use of another statistic to determine the differences between the estimated anomalies and the validation temperature data. I estimated the accuracy of the estimated anomalies using a 'double difference' statistic which is described in Section 4.2.3.4.

4.2.3.1 North Pole Drifting Station Data

NPDS 6 hourly SAT data were downloaded from the Arctic and Antarctic Research Institute (AARI) website ($http://www.aari.nw.ru/default\_en.asp$). I identified all drifting station observations between 1950 and 2013. The tracks of each relevant NPDS are shown in Figure 4.2. The NPDS SAT records were gridded to an EASE-Grid (described in Section 4.2.3.3) and calendar month average SATs were calculated where the minimum monthly sampling was fulfilled (see Section 4.2.3.5). I identified all grid cells with an average temperature for at least two years in the same month and grid cell. These grid cells were used to estimate the accuracy of the estimated anomalies over sea ice, compared to NPDS records, using the double difference statistic (Section 4.2.3.4).

4.2.3.2 Ice Mass Balance Buoy Data

Air temperature records from IMBs, which are generally recorded every couple of hours, were downloaded from the CRREL website (*http* : *//imb.erdc.dren.mil/buoysum.htm*). I identified all IMB records which had both meteorological and location data available, reported for more than a month on Arctic sea ice (excluding land fast ice), and did not have any noted problems with the air temperature sensor. There were 59 suitable IMB records, providing data between 2003 and 2013. The tracks for each relevant IMB are shown in Figure 4.2. Some IMBs did not have concurrent air temperature and location observations; air temperatures were often recorded on the hour every couple of hours and the position of the buoy was recorded irregularly with a sub-hourly frequency. In order to utilise these air temperatures and increase the amount of validation data available, I associated each air temperature observation with the latitude and longitude recorded nearest the time of the air temperature observation. This introduced some uncertainty to the location of the IMB temperature observations. But, sea ice drift is about 2% of wind speed, wind speeds in the central Arctic rarely exceed 25m/s (Przybylak, 2003b), and the locations of the buoys are recorded at sub-hourly frequency so the location errors are extremely likely to be less than 1km. This is small compared to the nominal grid cell size of the EASE-Grid used here. As for NPDS records, IMB air temperature observations were gridded to an EASE-Grid, monthly average SATs were calculated and these data were used to estimate the accuracy of the in situ data study estimated anomalies over sea ice using the double difference statistic (where possible).

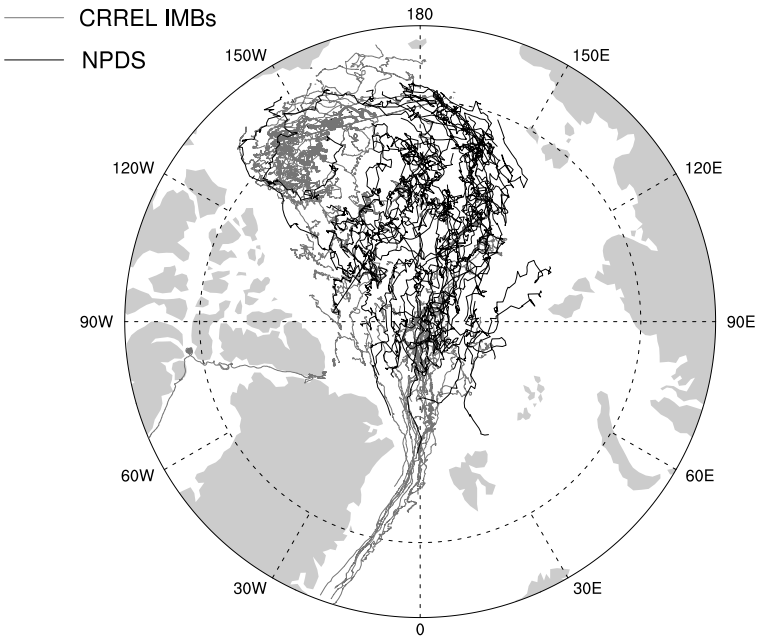

Figure 4.2: The tracks of each relevant North Pole Drifting Station (1950-2013) and CRREL IMB (2003-2013).

4.2.3.3 The Equal-Area Scalable Earth Grid

In situ measurements of Arctic temperatures are sparse, particularly in sea ice areas, and the records are often short. This leads to poor grid cell coverage, a problem which is exacerbated by the use of equal-angle grids. This is an issue for the drifting platforms used as validation data over sea ice in this chapter. Therefore, in order to have sufficient grid cell coverage over sea ice, it was decided to use an equal area gridding system.

The Equal-Area Scalable Earth Grid (EASE-Grid) system is an equal area gridding system developed at the National Snow and Ice Data Center (NSIDC). It includes three equal-area map projections (Northern Hemisphere, Southern Hemisphere, and Global Cylindrical) which can be used, in combination with defined variables such as nominal cell size and map origin, to define numerous equal area grids. A predefined EASE-Grid, which is a polar subset of the Northern Hemisphere Azimuthal Equal-Area Projection and has a 100km nominal grid cell size, was chosen for this study. A predefined grid was utilised here, rather than defining a new grid, as this reduced computational time. A 100km grid cell size provides a sufficient number of grid cells with more than 1 year of drifting platform data in each month, while being smaller than the expected correlation length scale. There is also a precedent for using this type of grid; a Northern Hemisphere EASE 100km grid was used for the IABP/POLES SAT dataset which also utilised drifting platform temperature records (Rigor et al., 2000). An EASE-Grid version 2.0 is available, which is defined with a different projection ellipsoid and has some advantages over the original EASE-grid: it allows exact, rather than nominal, grid cell sizes and does not have undefined corner cells in the azimuthal grids. However, it was decided to use an original EASE-Grid due to the fact that a predefined Northern Hemisphere polar subset is available in the original gridding system. I do not expect this to influence the results.

4.2.3.4 The Double Difference Statistic

In order to estimate the accuracy of estimated anomalies over sea ice, a 'double difference' statistic was utilised. Temperature anomalies are often produced using the Climate Anomaly Method (CAM). This method requires traditional station normals or a gridded climatology. However, there are other techniques which can be used to create anomalies from station temperatures: the Reference Station Method (RSM, Hansen and Lebedeff 1987) and the First Difference Method (FDM, Peterson et al. 1998). RSM requires at least one long record of data to refer other, shorter, station records to, and is therefore not suitable for calculating anomalies from the validation data over sea ice. FDM requires only monthly first differences: two successive years with a temperature value in the same calendar month for each station or grid cell analysed. The monthly temperature for a grid cell or station in one year is subtracted from the temperature for the same month and grid cell in the subsequent year. A reference period is not required and grid cells can be used instead of static meteorological stations. The concepts of the FDM were incorporated into the double difference statistic, which was calculated using Equation

(4.1). $s(x,t)$ is the monthly average SAT from a drifting platform at grid cell x in year t. $a(x,t)$ is the monthly average estimated SAT anomaly at grid cell x in year t. These two variables were compared with the monthly average temperature and monthly average anomaly respectively for the same grid cell and calendar month in another year $t+i$. The difference between the estimated anomalies in years t and $t+i$ was then subtracted from the difference between the drifting platform data in the same two years to create the double difference statistic dd (Eq. (4.1).

$$dd = [a_{(x,t+i)} - a_{(x,t)}] - [s_{(x,t+i)} - s_{(x,t)}] \tag{4.1}$$

It should be noted here that I anticipate a factor difference between the 'single differences' of the land validation and reanalysis study, and the double differences used for the sea ice validation. The difference from the validation data will contain uncertainties arising from both sides of the dd statistic; there will be additional errors from comparing estimated anomalies in different years as well as from comparing validation temperatures in the same way.

4.2.3.5 Minimum Sampling for a Monthly Average Surface Air Temperature

The drifting platforms used as validation data sources over Arctic sea ice in this study are not static. This means that a whole month of observations is unlikely to be associated with a single grid cell. Sparse temporal sampling of monthly averages may lead to unrepresentative values being used as validation data and affect the results of this study. Therefore the impact of sparse sampling on monthly averages produced from the validation data over sea ice was investigated and an appropriate minimum sampling for monthly average SATs was determined for this study. Ideally I would have identified the minimum sampling, and quantified the uncertainties, from partial records by subsampling full month records. However, the sample of full month records was very small so a different method, described below, was employed.

dd (the double difference statistic) was calculated between the estimated anomalies and the validation data over sea ice. dd metrics (double difference metrics, Section 4.2.4) were plotted against the average, minimum and maximum number of days observed by the validation data in a calendar month. The average, minimum and maximum number of days observed are derived from the number of days contributing to each of the two monthly average temperature measurements used to calculate dd. The results were very similar for all of the calculated dd metrics so the results for Root Mean Square dd (RMSdd) are shown as an example. The RMSdd for validation data for both NPDS and CRREL IMBs are shown in Figure 4.3. The NPDS and IMB metrics have larger values when the average, minimum and maximum number of days observed are lower (Figure 4.3). The RMSdd can be above 9.0K when the monthly average temperatures were calculated using less than 5 days of data, whereas the RMSdd is less than 5.0K for monthly average temperatures calculated using more than 30 daily temperatures.

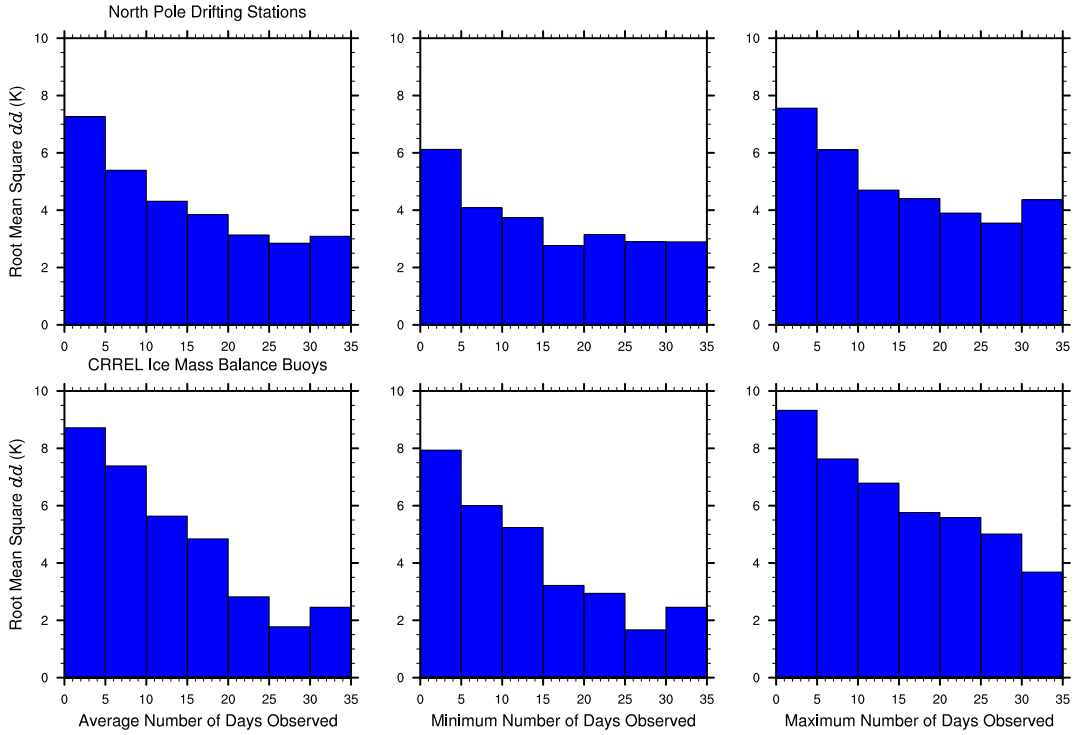

Figure 4.3: The Root Mean Square *dd* (K) between drifting platform temperatures and estimated anomalies plotted against the average, minimum and maximum number of days observed. NPDS temperatures are compared between 1950 and 2013 and CRREL IMB temperatures are compared between 2003 and 2013.

NPDS RMS*dd* were found to be almost constant when the average, minimum and maximum number of days observed is around, or above, 20. Figure 4.3 suggests a threshold of 15 days is sensible for NPDS as above this the RMS*dd* reaches its lowest values and is relatively consistent. For IMBs the RMS*dd* does not reach a plateau after a certain number of daily temperature observations, but the values are more constant after about 15 days are observed. So, NPDS and IMB RMS*dd* values suggest a sample of at least 15 days of data is adequate for calculating monthly average temperatures in this study. However, if we look at the scatter plots of the monthly absolute errors for the CRREL IMBs in Figure 4.4, using this minimum sample for monthly averages leads to extremely small samples of this validation data in most months and an inability to calculate *dd* metrics in March. This was also found for NPDS temperatures (Figure 4.8). Therefore it was decided to use a 10 day minimum sample for monthly averages for both the NPDS and IMB validation data. This was a compromise between the increase in apparent anomaly error arising from undersampling of the validation data and having enough data available to calculate *dd* metrics for the estimated anomalies.

4.2.4 Comparison of Estimated Anomalies to Validation Data

The estimation technique used in this study yielded estimates monthly Arctic SAT anomalies when applied to the input anomalies over Arctic land and sea ice areas. These in situ data estimated anomalies were validated using independent data sources over Arctic land and sea ice areas on a monthly timescale. The results of the validations were compared to those from Chapter 3. For Arctic land areas, monthly differences were calculated by comparing the estimated anomalies in the EASE-Grid cell nearest the location of each validation data station. These differences were used to calculate difference metrics: the Root Mean Square Difference (RMSD), Standard Deviation of Difference (SDD) and Compound Relative Difference (CRD). For estimated anomalies over sea ice areas, monthly *dd*s were calculated for each grid cell in which there are at least two years of monthly average SAT observations in the same month and grid cell from the validation data. This was done for both validation data sources over sea ice. *dd* metrics were calculated: the RMS*dd*, Standard Deviation of *dd* (SD*dd*) and Compound Relative *dd* (CR*dd*). The metrics noted above were also used in the reanalysis study, calculated using errors rather than the differences used here (see Section 3.2.5 of this thesis for more information). These error metrics are: the Root Mean Square Error (RMSE), Standard Deviation of Error (SDE) and Compound Relative Error (CRE). Some additional difference, *dd* and error metrics were calculated, but were not reported due to their similarity to other metrics as in Chapter 3.

The difference and *dd* metrics were compared to corresponding error metrics from the reanalysis study. The error metrics were created from reanalysis study estimated anomalies (produced using GSK) by gridding the estimated anomalies and the reanalysis reference to the EASE-grid used in this study and comparing them. The error metrics were

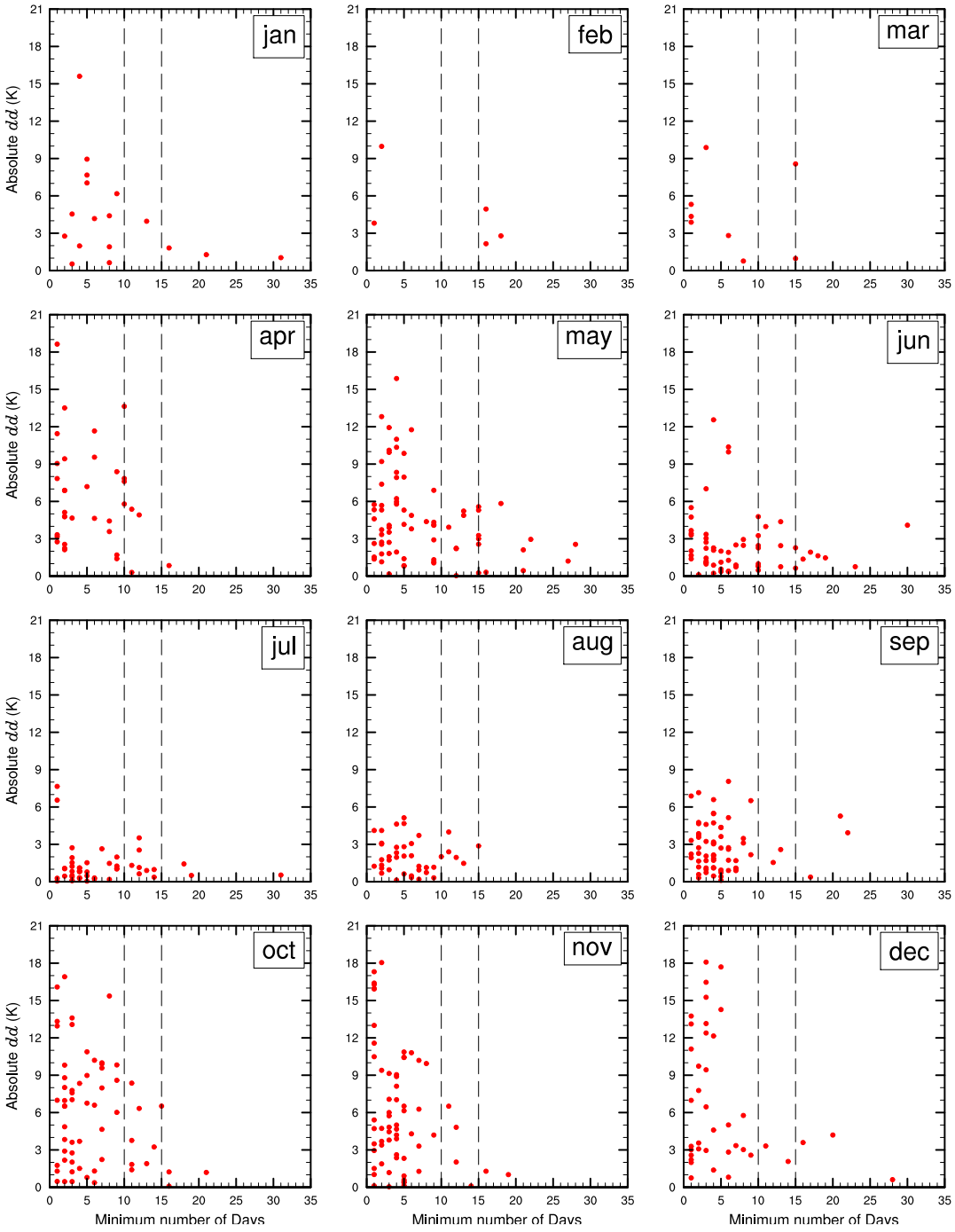

Figure 4.4: A scatter plot of the Absolute *dd* (K) between CRREL IMB temperatures and anomalies estimated using GSK between 2003 and 2013 against the minimum number of days observed. The minimum number of days observed is calculated from the number of observations which contribute to the two monthly average temperature measurements from CRREL IMBs used to create the *dd*. Two reference lines show 10 and 15 days observed.

produced using all grid cells in land, or sea ice, areas unless otherwise stated. Spatially resolved and spatial-average metrics were produced for both the in situ data and reanalysis studies and were compared over land and sea ice areas. Area-weighting of the averages was not required here due to the use of an equal-area grid. Metrics from this chapter and those from Chapter 3 were compared and the results are given in Section 4.3 and 4.4 for Arctic land and sea ice areas respectively.

4.3 The Validation of Estimated Anomalies over Arctic Land Areas

Difference metrics, produced by validating the estimated anomalies over Arctic land areas, were compared to corresponding error metrics from the reanalysis study detailed in Chapter 3. The average monthly RMSE and CRE from the reanalysis study, and RMSD and CRD from the land validation investigation, are shown in Figure 4.5. 95% confidence interval estimates were calculated for the RMSD and RMSE using the SDD and SDE respectively. The sample size of the observations is large enough ($n \geq 30$) to calculate confidence intervals using the standard deviation of the sample according to the Central Limit Theorem.

The monthly metrics from the two studies have the same seasonal pattern; the RMSE/RMSD is largest in the boreal winter months (December, January and February) and smallest in boreal summer, while the CRE/CRD is low for most months and slightly larger in the summer (Figure 4.5). This seasonal pattern is due to seasonality in Arctic temperatures and was explained in Section 3.3.1. There are some slight differences between the two studies; for example the CRD shows an unexpected increase for March and October. These differences may be due to incomplete spatial coverage of validation data as shown in Figure 4.1. Many of the independent land stations are in Scandinavia, Arctic Canada and on islands and coastal regions of Russia. This also means that there is a smaller sample of comparisons for the in situ data study compared to the reanalysis study, which is spatially complete over Arctic land areas. There may be inhomogeneities and noise present in the validation data over land, although this is unlikely as the validation data have been quality controlled, checked and some have been homogenised. These differences may also be due to the slightly different time periods of interest of the two studies. Finally, there may be local effects from comparing gridded estimated anomalies with independent station records which may not be representative of the grid cell they are located in. Nonetheless, the general pattern for the metrics from the in situ data study is very similar to that of the reanalysis study metrics. The sizes of the metrics are also very similar; the metric values from both studies are within 0.25(K, CRE/CRD are unitless metrics) of each other and RMSE values are often within the estimated confidence intervals of RMSD values (Figure 4.5). This shows the effectiveness of the ERA-Interim reanalysis as a testbed over Arctic land areas.

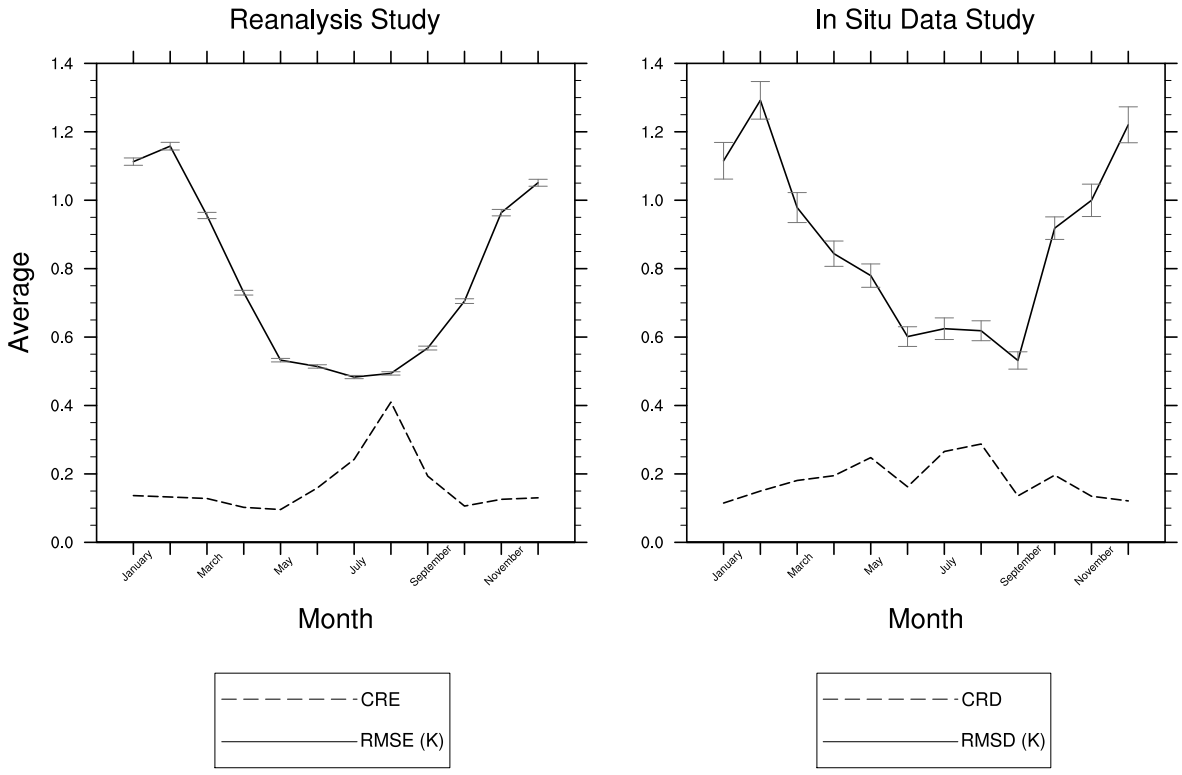

Figure 4.5: Monthly average of the Root Mean Square Error (K) and Compound Relative Error for the reanalysis study (left) and the the Root Mean Square Difference (K) and Compound Relative Difference for the land validation investigation (right). 95% confidence intervals are given in grey. CRD is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

The metrics in each month from both studies were also analysed spatially. The difference metric values for each land station were mapped and compared to spatially resolved error metrics derived from the reanalysis study. The mapped RMSE and RMSD is shown for several example months in Figure 4.6. The difference metrics from the in situ data study are a little larger in most months, but the values are generally within 1.0(K, CRD is a unitless metric) of the error metrics. However, in some cases they may be 2.0(K) larger than the reanalysis study metrics. There is no obvious spatial pattern to the metric values for the sample of validation stations used.

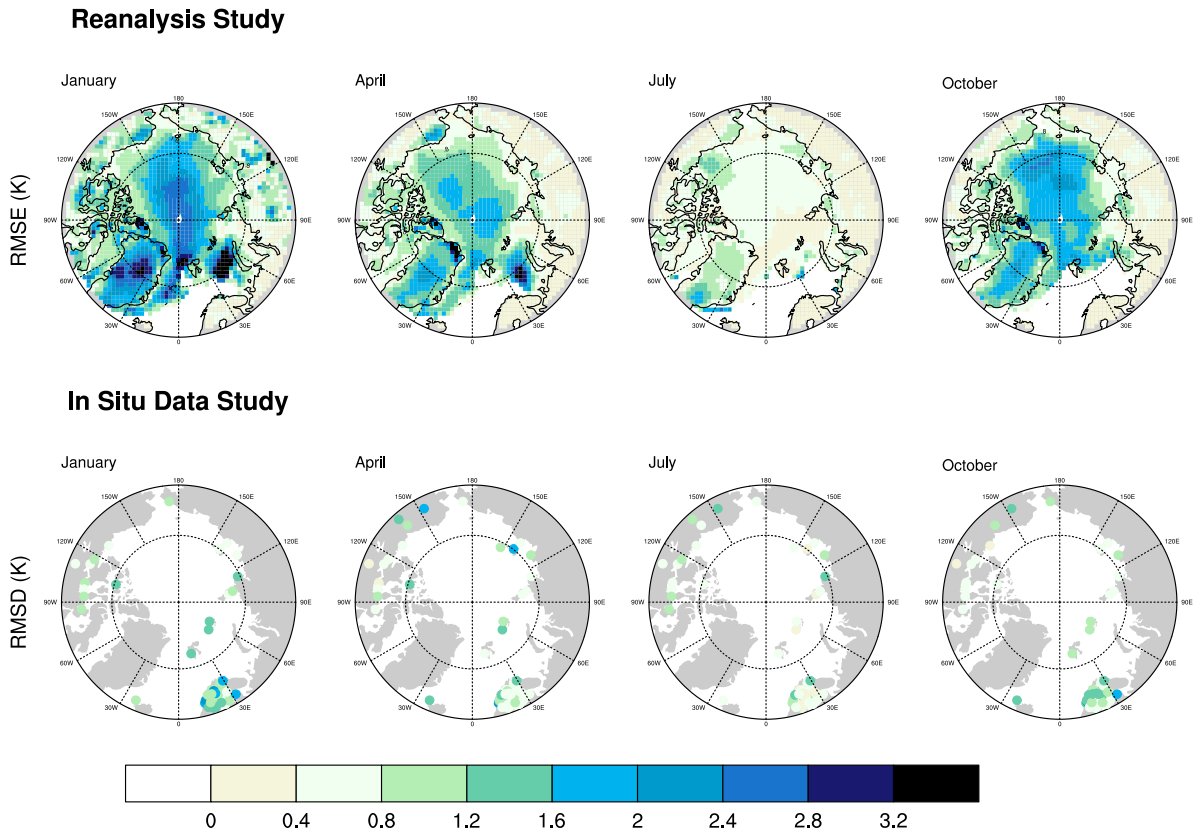

Figure 4.6: The spatially resolved Root Mean Square Error (K) from the reanalysis study and the Root Mean Square Difference (K) for each independent land station for several example months.

4.4 The Validation of Estimated Anomalies over Arctic Sea ice Areas

Double difference (*dd*) metrics, produced by validating the estimated anomalies over Arctic sea ice areas, were compared to corresponding error metrics from the reanalysis study (Chapter 3). The results from validating the estimated anomalies using two independent validation data sources, NPDS and CRREL IMBs, are given separately.

4.4.1 North Pole Drifting Stations

The monthly average RMS*dd* over sea ice from NPDS validation data was plotted and compared with corresponding error metric from the reanalysis study (Figure 4.7). The reanalysis study metrics over sea ice were calculated across all EASE-Grid cells over sea ice; across only grid cells and months present in the validation data (using random years between 1979 and 2011); and from simulated double differences for the same grid cells, months and years where possible (validation data years before 1979 or after 2011 were replaced randomly with years within this time period). For the reanalysis study metrics calculated across all EASE-Grid cells the sample size was large ($n \geq 30$) so the 95% confidence interval estimates for the metrics were calculated using the SDE. For the in situ data study, and the reanalysis metrics sampled to the same grid cells and month, the sample size was relatively small ($n < 30$) so Student's t-distribution was employed to estimate the confidence intervals from the SD*dd* or SDE as appropriate.

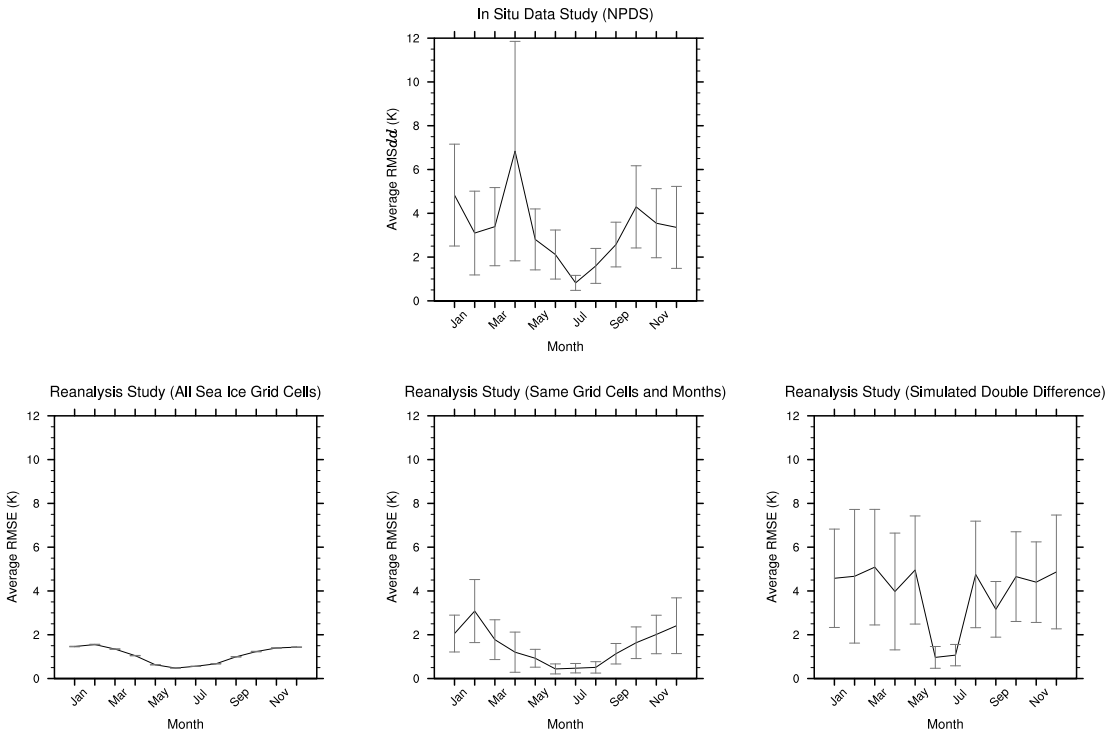

Figure 4.7: Monthly average of the Root Mean Square *dd* (K) for the NPDS sea ice validation data (top), and the Root Mean Square Error (K) for the reanalysis study (bottom) calculated across all EASE-Grid cells over sea ice (left), across only grid cells and months present in the validation data (middle, using metrics for random years between 1979 and 2011), and from simulated double differences for the same grid cells, months and years where possible (right, validation data years before 1979 or after 2011 were replaced randomly with years within this time period). 95% confidence intervals are given in grey.

Again both studies show a broadly similar seasonal pattern in the monthly metrics (Figure 4.7). The NPDS sea ice validation metric values in April are larger than would be expected given the general seasonal pattern in the RMSE. I investigated this further by plotting monthly absolute *dd* against the minimum number of days observed (Figure 4.8).

It was found that there are few *dd* values in April with more than 10 days of data contributing to them. This is because fewer NPDS report for the whole of April than in any other month (Figure 4.9) and even fewer fulfil the sampling requirements (Figure 4.8). Many NPDS were deployed, retired or abandoned during April, especially those deployed by the Soviet Union, leading to a small sample of data in this month. This reduces data availability in this sparsely sampled area resulting in an extremely small sample of *dd*s. In addition, within this small sample of April *dd* values there are some absolute *dd* values which are larger than would be expected given the seasonal pattern (Figure 4.8). These probably result from the interpolation missing key features of the weather field due to the sparse observation network, and therefore not being as representative of the anomalies in a given grid cell, year and month for one or more months used to calculate the *dd*. This is backed up by the absence of a peak in the April RMSE for the reanalysis study plots in Figure 4.7. If the larger error was a consistent feature of a grid cell and/or month, arising from the sampling of validation data, then larger RMS*dd* values for April would also be noted for the reanalysis study metrics.

The monthly RMS*dd* values from the in situ data study are around twice as large as RMSEs from the reanalysis study produced using all grid cells over sea ice (Figure 4.7). The difference between the metrics from the two studies are around 2.0K in general and can be up to 5.8K. A factor difference between single and double differences was expected (Section 4.2.3.4), and in addition there is a smaller sample of data used for the RMS*dd* values (due to the relatively small sample of NPDS validation data). Both have an impact as illustrated in the three reanalysis plots in Figure 4.7. When the reanalysis metrics are produced using simulated *dd*s for the same grid cells, months and years (where possible) the difference between the metrics reduces to around 1.0K in general. The metrics are within 2.0K of each other for 9 of 12 months, and are within each others confidence intervals. The differences can, however, still be up to 3K for some months. March and August RMSEs were found to be slightly larger for the reanalysis study, when simulated *dd*s are used, than would be expected given the general seasonal pattern; they are larger than the April and September values respectively. This could be an artefact of the decision to simulate the *dd* for validation data years before 1979 or after 2011 (outside the time period of the reanalysis data) using random years that occur during this time period, which was a period of large sea ice changes in the Arctic.

So, in conclusion, the seasonal patterns in the reanalysis and in situ data study metrics are very similar. The metric values are larger for the in situ data study, however this seems to be explained by the small sample of validation data and the use of the *dd* statistic instead of directly comparing anomalies. In addition, the fact that the validation data over sea ice typically sample less than the whole of a month may lead to an increase in *dd* of around 1.0K on average (shown in Figure 4.3, Section 4.2.3.5). Finally, it must be noted that the confidence intervals for the in situ study metrics are large and the sample size is small so caution is needed in drawing conclusions from these data. However the seasonal

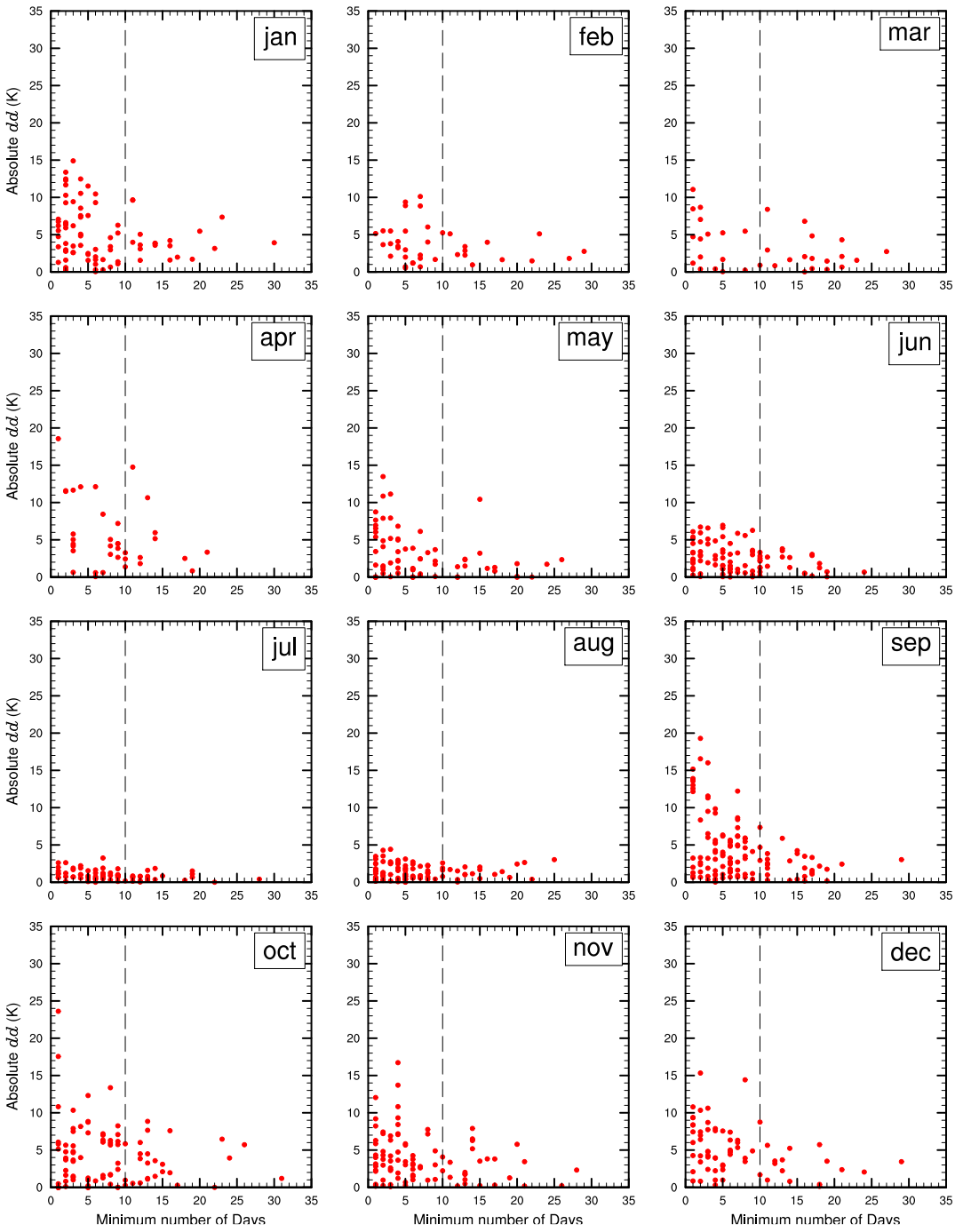

Figure 4.8: The monthly Absolute *dd* (K) between NPDS temperatures and estimated anomalies estimated using GSK between 1950 and 2013 plotted against the minimum number of days observed. The minimum number of days observed was calculated from the number of daily temperatures which contribute to the two monthly average temperature measurements from NPDS used to calculate the *dd* statistic. A line shows the minimum required number of days observed in this study.

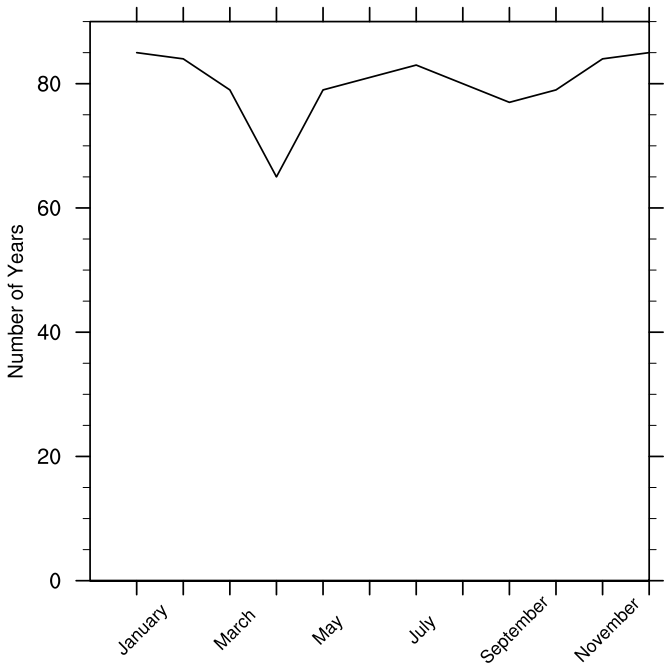

Figure 4.9: The number of years between 1950 and 2013 where NPDS report for the whole of each month.

pattern and the sizes of the metrics are very similar, and the confidence intervals overlap, when the reanalysis study metrics are created using the *dd* statistic and re-sampled.

4.4.2 Ice Mass Balance Buoys

The monthly average RMS*dd* from the in situ data study was plotted and compared with that from the reanalysis study (Figure 4.10), in the same way as for the NPDS validation (Section 4.4.1). A very similar result to the NPDS investigation was found here; there is a similar seasonal pattern in the metrics for the two studies over sea ice, but also differences, most notably in March (Figure 4.10). This is for the reasons noted for April metrics in Section 4.4.1. This result also illustrates the influence of both small sampling and a larger than anticipated *dd*. There are other months with very few IMB SAT observations and consequently very large confidence intervals, for example February and April, but the *dd* metrics are not as large as in March. In March one particular *dd* is larger than would be expected given the general seasonal pattern, whereas in February and April the *dd*s fit the seasonal pattern. The monthly RMS*dd* values from the in situ data study are again larger than reanalysis study RMSEs, as anticipated if the sampling and effect of using a *dd* statistic are not accounted for. When the reanalysis metrics are produced using simulated *dd*s the difference between the metrics is around 1.5K in general. The metrics are often within 2.0K (for 9 of 12 months), but differences can be more than 4K for some months. The conclusions for the validation of the estimated anomalies over Arctic sea ice areas using CRREL IMBs are the same as those resulting from the use of NPDS validation

data.

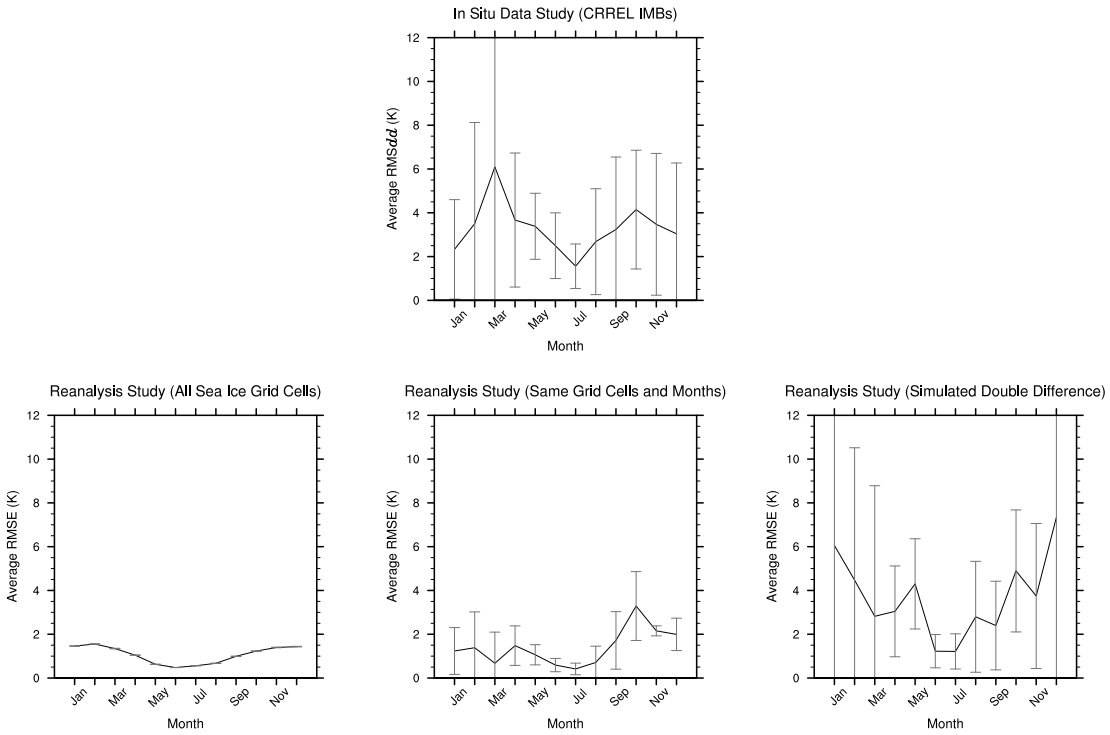

Figure 4.10: Monthly average of the Root Mean Square *dd* (K) for the CRREL sea ice validation data (top) and Root Mean Square Error (K) for the reanalysis study (bottom); (left) sampled to the same grid cells and months as the validation data using anomalies for random years between 2003 and 2013, (right) simulated *dd* for the same grid cells, months and years where possible. 95% confidence intervals are given in grey.

4.5 Discussion

In this study it was found that the accuracy of monthly SAT anomalies in the Arctic, estimated from in situ meteorological station SAT records using Global Simple Kriging, was similar to the accuracy of the anomalies simulated using the same technique within a reanalysis testbed. The difference and double difference (*dd*) metrics have the same seasonal pattern as the equivalent metrics from the reanalysis study over land and sea ice respectively. Some differences were noticed in the sea ice validation (Section 4.4), but this was attributed to small samples of validation data as well as some *dd*s with large errors. The size of the metrics was also found to be similar. For the land validation the monthly average metrics and spatially resolved metrics from both studies were within 0.25(K) and 1.0(K) of each other respectively. In addition, RMSE values from the reanalysis study were often within the estimated confidence intervals of RMSD values from the in situ data study. This is despite using a fairly small sample of validation data (63 meteorological stations) which have incomplete spatial coverage (most of the independent land stations are located in Scandinavia). This shows that the ERA-Interim reanalysis is an effective testbed over Arctic land areas. For the validation over sea ice the differences between the

metrics of the two studies were larger than those noted for the land investigation; around 2 or 3K in general. Much of this difference seems to be explained by the small sample size of the reference data, and the use of the *dd* statistic (which was expected). Differences between the two studies reduce when the reanalysis data are subsampled and the double differences are simulated. When using NPDS as validation data the difference between the monthly average metrics from the two studies was around 1.0K if the reanalysis metrics are created using simulated *dd*. For the IMB validation they were within 1.5K of each other in general under the same conditions. The accuracy of the in situ data estimated anomalies are remarkably similar to those from within the reanalysis testbed over sea ice areas when the sample size and the use of the *dd* are accounted for. Within the sea ice validation data there are uncertainties resulting from drifting platforms not sampling a whole month in a given grid cell. Even with the required minimum sampling for a monthly average, metric values may be 1.0-2.0K larger when less than a whole month of observations are available. Uncertainties may also result from local effects, noise, and inhomogeneities as in the land validation. In addition, much of the CRREL IMB data is preliminary and may not be quality controlled. These platforms were also not designed for producing meteorological records. The confidence intervals for drifting platform metrics are large and the sample size is very small so caution is needed in drawing conclusions from this data. However, the two independent datasets produced very similar results despite the small samples of these datasets.

The issues arising from using drifting platforms in this study highlight the difficulty of investigating SATs over Arctic sea ice. SAT data records over sea ice are temporally and spatially sparse. Here this necessitated the use of an EASE-Grid with an 100km grid cell size. In addition the drifting nature of these platforms increases the uncertainties associated with producing monthly average temperatures from their SAT observations. There are also uncertainties and issues related to the data rather than the spatial coverage of data. Sometimes, in situ data are not available. The possibility of including IABP data was also considered as this programme has maintained a network of drifting buoys in the Arctic Ocean since 1979. However, raw data from these buoys is no longer made available on the IABP website, so IABP data could not be used as a validation dataset in this study. Another issue experienced with drifting platform data in this study was differences in data format and metadata. For CRREL IMBs, 10 different data file formats were used for the data between 2003 and 2013. 22 of the IMBs had air temperatures and location data which were not recorded at the same time. The air temperatures were often recorded on the hour every couple of hours and the position of the buoy was recorded irregularly with a sub-hourly frequency. In addition, some IMBs had essential data files missing, metadata and data were not always consistent within data files, and some data quality information may be missing as the IMB data are often preliminary. Furthermore, automatic sensors, such as IMBs, which are installed in the Arctic and left unattended without routine maintenance may be impacted by issues such as rime ice and snow cover

which will impact their accuracy (Hall et al., 2015).

Finally, additional errors were added to the results in this in situ data study due to use of the double difference statistic as noted previously. This statistic used as a result of the impossibility of creating traditional normals for drifting platform data. It was also decided not to use another dataset as a climatology as only datasets which could be identified that provide Arctic SATs for the normal period (1961-1990) were either land only, ocean only, or interpolated. Interpolated data, for example the IABP derived datasets (Rigor et al., 2000; EOL, 2014), would contain uncertainties resulting from the use of an interpolation technique. It was therefore decided to try and use another statistic to determine the differences between the estimated anomalies and the validation temperature data. But, using the double difference may have increased the error in the results by around 2.0K which may be larger than the error that would be introduced by using an interpolated climatology.

4.6 Summary

The in situ study metrics over land are very close to the reanalysis study metrics. Over sea ice the metrics have a similar seasonal pattern and size (once the sample size and double difference statistic are accounted for). However, the confidence intervals are large and the sample size is very small so caution is needed in drawing conclusions from this data. This highlights the difficulty of investigating SATs over sea ice where data records are sparse. In addition, there are many other complications that might arise when attempting to use SAT data over sea ice areas including different types of monitoring platform, a lack of uninterpolated data, and many different data and metadata formats.

Chapter 5

Can the Inclusion of Additional Data Sources be used to Improve Estimates of Arctic Surface Air Temperature Anomalies?

5.1 Introduction

As noted in Chapter 1, temperature anomaly datasets generally use SATs measured at meteorological stations to generate the datasets over land and sea ice. However, many of these datasets use only a subset of the available meteorological station records; for example 88 Arctic meteorological stations independent from the 956 stations present in the CRUTEM.4.2.0.0 dataset (utilised in Chapter 3, the 'reanalysis study') were identified in the ISTI merged Stage 3 version 1.0.0 recommended monthly databank product in Chapter 4. Furthermore, at present none of the major temperature anomaly datasets use SAT data recorded over Arctic sea ice. Even those data sources considered to be reliable such as Arctic and Antarctic Research Institute (AARI) North Pole Drifting Station (NPDS) data are not utilised (Rigor et al., 2000). Utilising these supplementary data sources, especially those reporting over sea ice areas, might provide better estimates of SATs in this data sparse region. If it can be shown that including extra data in the Arctic leads to an improvement in the accuracy of estimated SAT anomalies in this region, this would indicate a need to further investigate potential sources of data over Arctic sea ice and how to utilise these data sources for improved monitoring of Arctic climate change.

Therefore, the objective of this study was to investigate how sensitive various estimation techniques are to the inclusion of additional sources of data in the Arctic and whether their inclusion has an impact on the accuracy of the estimated anomalies. My original concept for this study was to investigate the objective using in situ data. However, I decided that the uncertainties and issues associated with in situ data noted in Chapter 4 would make it difficult to isolate the impacts of additional data on the accuracy of

the techniques. Therefore the study detailed in this chapter repeats the recent decades investigation of the reanalysis study with additional ERA-Interim data sampled at supplementary data source locations and compares the results of this 'additional data study' to those of the reanalysis study. I undertook two investigations; one using the locations of independent land-based meteorological stations, hereafter referred to as the 'additional land data' investigation, and another using NPDS data locations, hereafter referred to as the 'additional sea ice data' investigation. These are data sources likely to provide climate quality records of SATs. The outline of this chapter is as follows. Section 5.2 describes the data and techniques used in this study. Section 5.3 evaluates the impact of including additional land data on the accuracy of the estimation techniques. Section 5.4 evaluates the impact of including additional data over sea ice areas. The final sections discuss the results and provide a summary and conclusions.

5.2 Data and Techniques

This additional data study repeats the recent decades investigation of the reanalysis study in Chapter 3 with additional ERA-Interim data sampled at supplementary data source locations and analyses the results. Five estimation techniques were applied to 'input anomalies' combined with 'additional input anomalies'. The input anomalies were monthly SAT anomalies from ERA-Interim sampled at Arctic (north of 65°N) meteorological station locations. The additional input anomalies are monthly SAT anomalies from ERA-Interim sampled at supplementary data source locations over Arctic land, or sea ice, areas. The supplementary data source locations were the locations of independent meteorological stations over land, and NPDS observation locations over sea ice.

The estimation techniques yield 'estimated anomalies'; estimates of the ERA-Interim SAT anomaly at several temporal resolutions for both investigations. 'Arctic-average' anomalies are area-weighted averages of SAT anomalies across the Arctic region, and 'spatially resolved' anomalies are complete fields of Arctic SAT anomalies. As in Chapter 3, techniques which produced spatially resolved anomalies as well as Arctic-average anomalies are described collectively as 'interpolating' techniques, while those producing only Arctic-average SAT anomalies were referred to as 'non-interpolating' techniques. The target areas for the interpolation were land areas and areas of sea ice with a sea ice concentration of more than 15%.

The estimated anomalies were compared to monthly SAT anomalies produced for each ERA-Interim grid cell. These ERA-Interim anomalies are the 'truth' for the target area and are the same 'reference' anomalies used for Chapter 3 (Section 3.2.1). Errors and error metrics were produced for both the additional land data and additional sea ice data investigations as described in Section 3.2.5. These errors and error metrics were then compared to the reanalysis study errors and error metrics.

5.2.1 Reference Anomalies

To compare the accuracy of Arctic SAT anomalies produced using different estimation techniques, a reference dataset was produced. This is the 'truth' to which the anomalies produced by each estimation technique will be compared. The reference anomalies for this study were monthly SAT anomalies produced for each ERA-Interim grid cell between 1979 and 2011 relative to a 10 year climatological period (1990-1999). This is the same reference dataset as used for the reanalysis study, and as such is described in further detail in Section 3.2.1 of this thesis.

5.2.2 Input Anomalies

In this additional data study the input anomalies used for the reanalysis study in Chapter 3, which were ERA-Interim grid cell anomalies sampled at the location of all Arctic meteorological stations in the CRUTEM4 databank (Section 3.2.2), were combined with additional input anomalies produced using the locations of supplemental data sources. Using ERA-Interim data as input anomalies instead of actual station measurements enables us to isolate the limitations of the estimation techniques from errors arising from using in situ measurements. These additional input anomalies over land and sea ice areas were produced using the methods described below.

5.2.2.1 Additional Input Anomalies over Land

In Chapter 4 independent Arctic meteorological stations over land, which were not included in the CRUTEM.4.2.0.0 databank, were identified from the the ISTI merged Stage 3 version 1.0.0 recommended monthly databank product. 63 validation stations were identified in that study with data records in the time period of interest (1950-2013), and at least 14 years of data in the climatology period (1961-1990), and temperature records which were either homogenised or quality controlled with no large gaps in the data record and no obvious inhomogeneities (Section 4.2.2).

The locations in latitude and longitude of the identified independent meteorological stations were obtained from the ISTI Stage 3 dataset (Thorne et al., 2011). I combined these additional station locations over land with the CRUTEM.4.1.1.0 databank station locations used to produce the input anomalies for Chapter 3. Monthly anomalies from the ERA-Interim grid cell nearest the combined station locations were extracted from the reference anomaly dataset. These station location anomaly timeseries were then masked according to whether the stations reported a temperature in each month or year between 1979 and 2011. Some of the stations were located within the same ERA-Interim grid cell. As in Chapter 3 all stations that were duplicates at the ERA-Interim resolution were identified and the station location anomaly timeseries were merged. These timeseries constitute the input anomalies for this additional land data study. 56 additional station location timeseries were added to the 852 utilised in Chapter 3 (Figure 5.1).

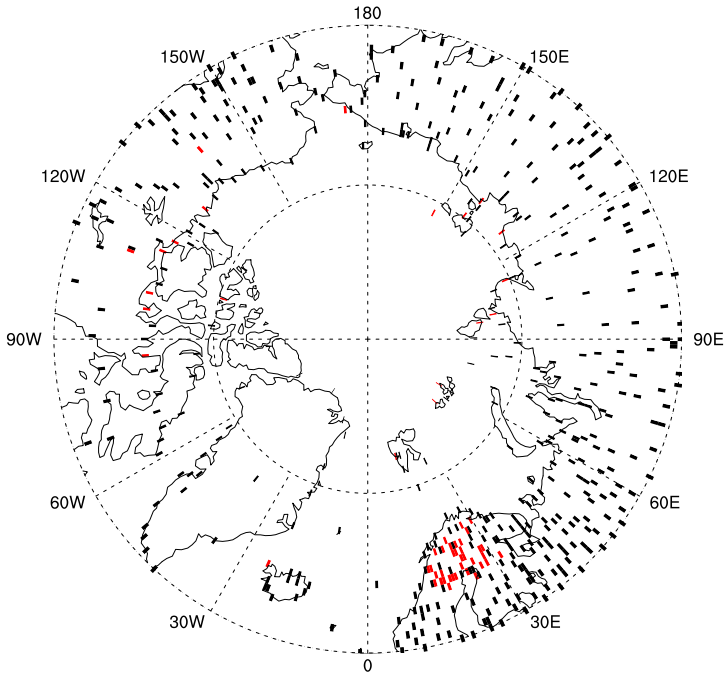

Figure 5.1: The locations of input anomalies for Chapter 3 (black) and additional input anomalies (red) over land between 1979 and 2011.

5.2.2.2 Additional Input Anomalies over Sea Ice

NPDS 6 hourly SAT data between 1979 and 2011 were downloaded from the AARI website. I identified the ERA-Interim grid cell nearest each NPDS temperature observation and extracted monthly anomalies for these identified grid cells from the reference anomaly dataset. The anomaly timeseries were then masked according to whether the NPDS reported a 6 hourly temperature in each month or year between 1979 and 2011, as done for the additional input anomalies over land. These timeseries constitute the input anomalies for this additional sea ice data study.

It should be noted that this is slightly unrealistic in terms of the coverage of monthly temperatures from NPDS; these drifting platforms are unlikely to sample a whole month of temperatures in a single grid cell even if a coarser, equal area grid is used (such as the 100km EASE grid utilised in Chapter 4) rather than the ERA-Interim grid utilised here. However, if the same 10 day minimum sample for monthly averages used in Chapter 4 was required here, only 207 ERA-Interim grid cells would contain a monthly average between 1979 and 2011. Therefore, I assumed a monthly average observation in the additional input anomaly timeseries over sea ice if there was at least 1 NPDS 6 hourly temperature record in a grid cell, month and year. This is a compromise between the need to use data and techniques that are as close as possible to the previous chapter and having enough extra input data to investigate the impact of adding additional data. The locations of these additional input anomalies over sea ice are shown in Figure 5.2.

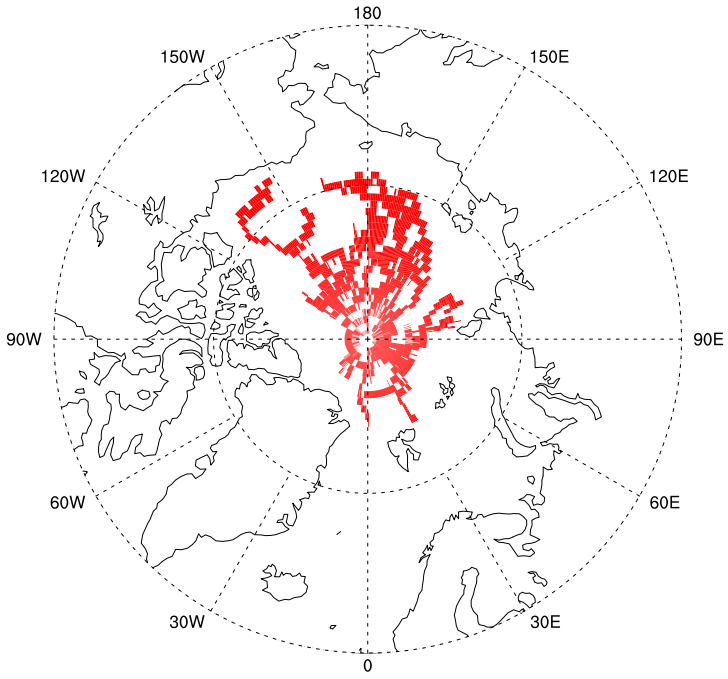

Figure 5.2: The locations of additional input anomalies over sea ice between 1979 and 2011.

5.2.3 Estimation Techniques

The estimation techniques applied to the input anomalies for the two additional data investigations in this study were the same techniques investigated in Chapter 3 of this thesis. The five estimation techniques were Linear Interpolation (LI), Global Ordinary Kriging (GOK), Global Simple Kriging (GSK), a restricted finite volume interpolation technique (the 'Binning' technique), and Not Interpolating (NI). A description of the estimation techniques is provided in Section 3.2.3 of this thesis, with additional information about the kriging techniques in Appendix B. The model function fitted parameter values employed in Chapter 3 (see Appendix B) were used here.

5.2.4 Comparison of Estimated Anomalies to Reference Anomalies

The estimation techniques used in this study yield estimates of ERA-Interim monthly Arctic SAT anomalies when applied to the input anomalies for the additional land and additional sea ice data investigations in this study. All estimation techniques used here produced estimated monthly anomalies gridded to the ERA-Interim grid, except for the Binning technique. The monthly anomalies were used to produce estimates of spatially resolved Arctic anomalies and Arctic-average anomalies over land and sea ice areas (sea ice concentration of more than 15%) at monthly, seasonal and annual timescales for recent decades in the same way as for the reanalysis study.

The anomalies estimated in the additional data study were compared to the reference anomalies and errors were calculated for estimates of both spatially resolved and Arctic-average anomalies at all timescales. As in Chapter 3, several error metrics were calculated which assess the absolute and relative error in the estimated anomalies. The Root Mean Square Error (RMSE) and Compound Relative Error (CRE) were reported here; the other absolute error metrics were again found to be extremely similar to the RMSE and as a result are not reported here.

5.3 The Impact of Including Additional Land Data on the Performance of Estimation Techniques

5.3.1 Arctic-Average Anomalies

Timeseries of reference and estimated Arctic-average anomalies produced in this additional land data investigation, as well as the errors in the estimated anomalies, at both annual and monthly timescales were plotted and compared to the results from the reanalysis study for the Arctic in recent decades. It was found that for most techniques it was difficult to visibly distinguish any differences in the timeseries and estimation errors from the two studies. A Taylor diagram, which graphically summarises how well estimated variables match a reference dataset, was produced from the estimated anomaly errors of both the additional land investigation and reanalysis study (Figure 5.3). All interpolating techniques and the Binning technique show slight decreases in RMSE and increases in cross correlation for annual estimated Arctic-average anomalies, and for most ($\geq$ 8) calendar months, when supplemental input anomalies over land are used. However, the changes in accuracy were very small; a maximum of 0.01K for RMSE and 0.004 for cross correlation. The accuracy of the anomalies estimated using the NI technique decreased for all months and for annual anomalies. The increases in RMSE were larger than the changes seen for the other techniques; the increases were between 0.05 and 0.4K, which is not trivial. The cross correlation decreased by up to 0.2. The increases in error for NI are due to certain characteristics of this technique. When Arctic-average anomalies are estimated using NI, which does not spatially infill, areas which are well observed have more influence on the Arctic-average anomalies. This means that if additional anomaly timeseries are utilised in already well observed regions, the influence of these regions will increase and lead to a decrease the accuracy of the estimated anomalies. This is assuming that the anomalies in the well observed region are not representative of the Arctic-average anomaly. For other techniques the additional anomaly timeseries contribute to grid cells which already contain data (Binning), or are interpolated (Kriging techniques and LI), which means that the anomalies of one region have less influence on the estimated Arctic-average anomalies compared to those produced using NI.

In conclusion, all interpolating techniques and the Binning technique show slight de-

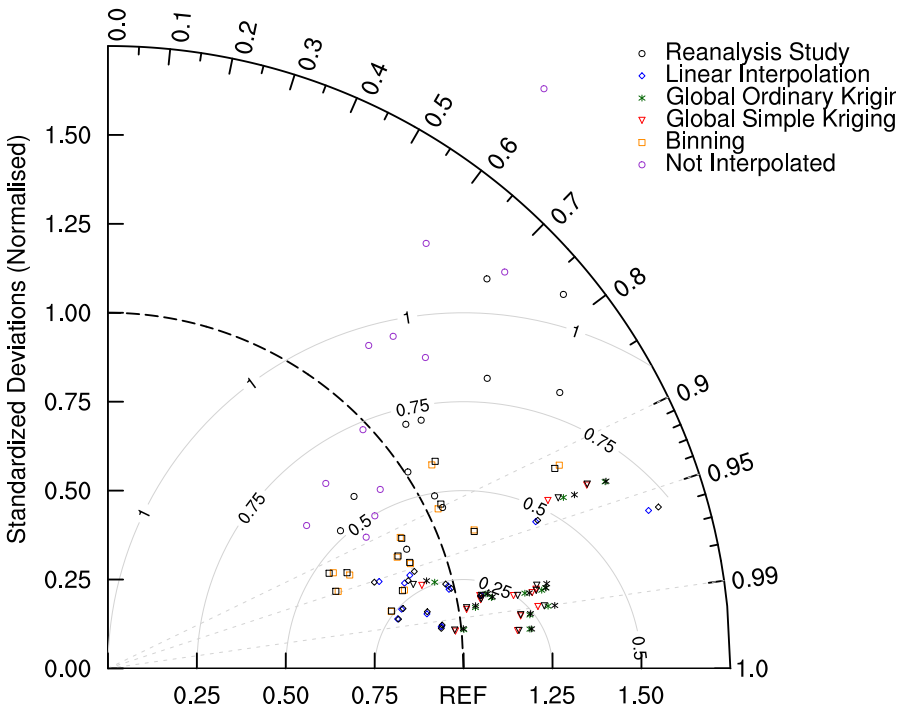

Figure 5.3: A Taylor diagram comparing the monthly and annual Arctic-average anomaly timeseries produced in the additional land investigation (colour) and from the reanalysis study (black). Each symbol plotted represents a month of the year or the annual value. Cross correlation is shown by the angle with respect to the x-axis. The standard deviations (normalised with respect to the reference standard deviation) can be read from the y-axis. The Root Mean Square Error (K) of the estimated anomalies is proportional to the distance to the point on the x-axis identified as REF (shown by the concentric circles marked 0.25 to 1).

creases in RMSE for annual Arctic-average anomalies, as well as for most months, when supplemental station location anomalies were added over land areas. They also show small increases in the cross correlation values. The accuracy of estimated Arctic-average anomalies produced using the NI technique decreased for all months and for annual anomalies in this additional land data investigation compared to the reanalysis study. However, the differences for most techniques are small and not notable.

5.3.2 Spatially Resolved Anomalies

The RMSE and CRE for spatially resolved annual anomalies produced in this additional land data investigation were mapped and compared to the corresponding results from the reanalysis study. The change in the annual error metrics between the two studies is given in Figure 5.4. The annual anomalies estimated in the additional land data investigation, for all estimation techniques, were slightly more accurate compared to the reference anomalies than those from the reanalysis study. The RMSE and CRE for these estimated anomalies decreased by up to around 0.2(K) for areas of Scandinavia, and a decrease in RMSE of 0.2K was observed for sea ice areas near Franz Josef land. The LI technique error metrics additionally decreased by up to about 0.1(K) in the central Arctic Ocean.

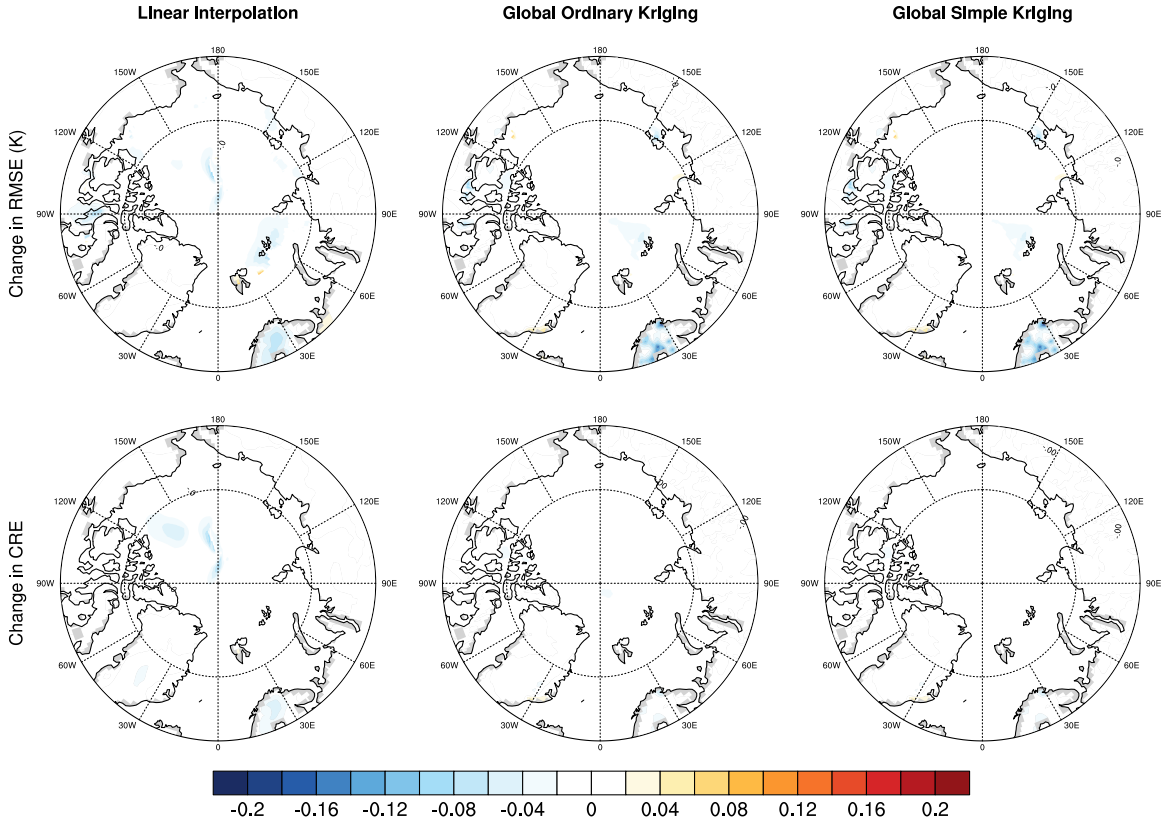

Figure 5.4: The change in Root Mean Square Error (K) and Compound Relative Error for spatially resolved annual SAT anomalies in the Arctic between the additional land data investigation and the reanalysis study.

The changes in RMSE and CRE between estimated monthly anomalies from the additional land data investigation and the reanalysis study are given in Figures 5.5 (GOK) and 5.6 (LI) for several months of interest. The changes for GSK were not plotted here as the results for both kriging techniques were nearly identical. There is a general increase in accuracy for all interpolating techniques; the GOK and LI error metrics decreased in size by around 0.2(K) in general. The GOK error metrics decreased by up to 0.6K for RMSE and more than 1.0 for CRE (Figure 5.5) while the LI error metrics decreased by up to 0.4K and more than 1.0 for RMSE and CRE respectively (Figure 5.6). Both metrics have improvements in accuracy extending from extra stations in the Franz Josef Land archipelago into the central Arctic Ocean. The decrease in CRE also extends into the Laptev and East Siberian Seas in July and can be relatively large (more than 1.0). RMSE also shows large decreases in Scandinavia (where many of the additional land stations are located) for both techniques. This increase in accuracy over Scandinavia is more extensive for LI compared to GOK. This is due to differences in the size of the initial reanalysis study metrics; for Scandinavia the kriging metrics were already very close to 0 and therefore cannot show as much improvement as seen for LI. In addition to the areas already mentioned, GOK (Figures 5.5) also showed small areas of localised decreases in metric values in Canada for both metrics. These results show that it is not only the amount of additional stations added which have an influence, as in Scandinavia, but also additional stations in a beneficial location, such as those extra stations in the Franz Josef Land archipelago. Two stations provided data in this archipelago in the input anomalies for the reanalysis study, but these stations were situated on the other side of the archipelago from the Arctic Ocean. Therefore, adding these new stations which are nearer the sparsely observed Arctic Ocean has lead to a reduction in errors in this area.

However, some decreases in accuracy also occur. Increases in error metrics were noted for RMSE in March and for CRE in July and August (Figures 5.5 and 5.6). For GOK these were small changes and often local in extent. The decreases in accuracy for LI were larger (up to 0.6K for RMSE in sea ice edge areas in the Atlantic) and slightly more extensive geographically than observed for GOK. The increase in error results from differences in the way LI and kriging techniques treat spatially clustered stations. Kriging accounts for clustering of stations and assigns these stations lower weights than those in sparsely observed areas. For LI the extra station location anomalies in Scandinavia increase the influence of these anomalies in nearby areas which these anomalies may not be representative of, such as Arctic sea ice areas.

For spatially resolved estimated anomalies at annual and monthly timescales, adding supplemental station location anomaly timeseries over land generally increased the accuracy of the anomalies relative to the reanalysis study. The RMSE and CRE values generally decrease by about 0.1-0.2(K). Larger improvements were seen in some areas of the Arctic, such as northern Scandinavia and near Franz Josef Land. In these areas RMSEs could decrease by up to 0.6K which is not trivial. Decreases in accuracy were

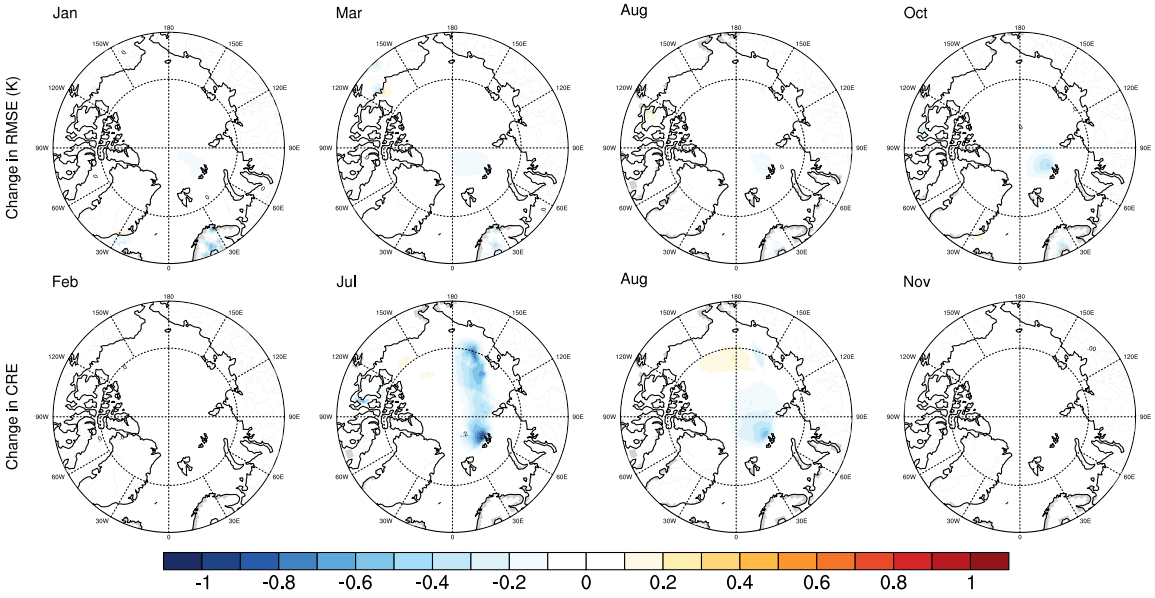

Figure 5.5: The change in the Root Mean Square Error (K) and Compound Relative Error between the reanalysis study and the additional land data investigation for the Global Ordinary Kriging technique. Maps are plotted for several months of interest for each error metric.

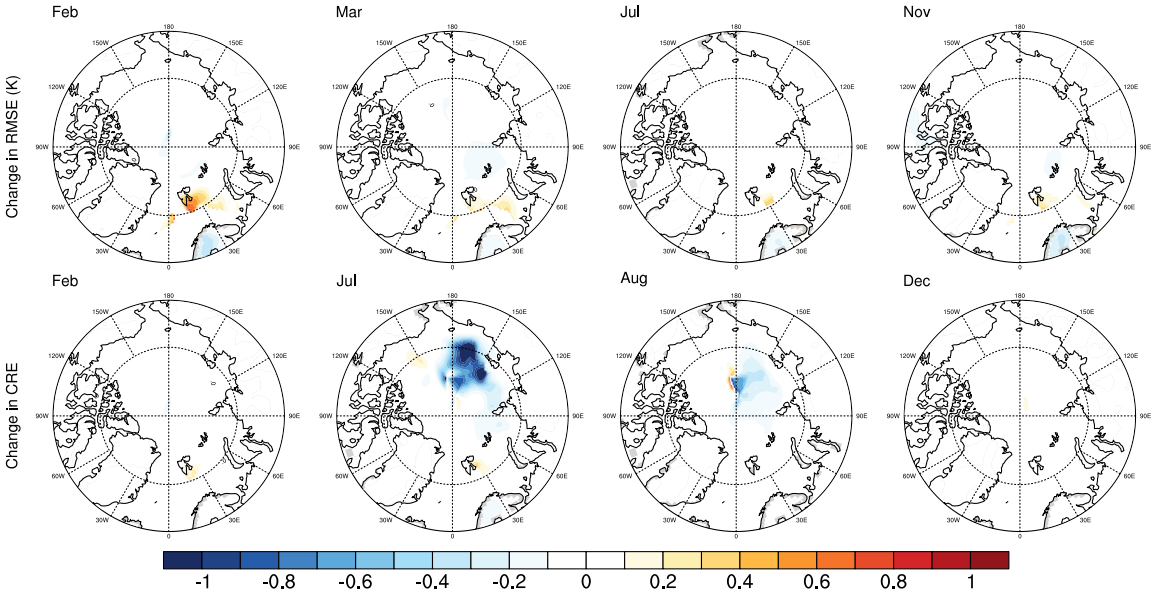

Figure 5.6: The change in the Root Mean Square Error (K) and Compound Relative Error between the reanalysis study and the additional land data investigation for the Linear Interpolation technique. Maps are plotted for several months of interest for each error metric.

noted for anomalies estimated using LI in the present of additional data over land. This shows that caution may be needed if additional land stations are to be added to datasets using the LI technique. Moreover, there may be issues associated with large changes in the distribution of the observing network for this technique.

5.4 The Impact of Including Additional Sea Ice Data on the Performance of Estimation Techniques

5.4.1 Arctic-Average Anomalies

Timeseries of reference and estimated Arctic-average anomalies, as well as the errors in the estimated anomalies, from this additional sea ice data investigation were compared with those from the reanalysis study. This was done for annual and monthly Arctic-average anomalies. Similar to the additional land data investigation, for most techniques it was difficult to distinguish any differences in the errors so a Taylor diagram was produced (Figure 5.7). All interpolating techniques and the Binning technique showed slight decreases in the RMSE and slight increases for the cross correlation for annual anomalies and for most, if not all, months. These improvements in accuracy were also observed for the NI technique, in contrast to the results for additional data over land (Section 5.3.1). The changes in accuracy observed for this additional sea ice investigation were fairly small, but were slightly larger than those generally noted in Section 5.3.1. Decreases in RMSE were up to 0.1K. Increases in cross correlation were slightly smaller than for Arctic-average anomalies produced with additional land data; 0.03 compared to 0.2.

In conclusion, all techniques show slight decreases in RMSE for annual Arctic-average anomalies, as well as for most or all months, when supplemental anomalies were added over sea ice areas. They also show very small increases in the cross correlation values. Unlike for the additional land data investigation, the improvement is universal across all techniques. In addition, the improvements in RMSE values are larger than noted when supplemental station location anomalies are added over land. The decrease in RMSE is still fairly small, up to 0.1K, but it is not entirely trivial.

5.4.2 Spatially Resolved Anomalies

The change in RMSE and CRE for annual, spatially resolved anomalies produced by the additional sea ice data investigation and the reanalysis study were mapped and examined. The change in the annual anomaly error metrics is shown in Figure 5.8. All techniques show both increases and decreases in the error metrics. There are decreases in the RMSE of up to 0.3K and decreases in the CRE of more than 0.5 over Arctic sea ice areas. The increases in RMSE and CRE are around 0.2-0.3(K). These decreases in accuracy generally occur in western Arctic sea ice areas, on the North American side of the Arctic, near the Beaufort Sea for both Kriging techniques and off the north coast of Greenland for LI. The

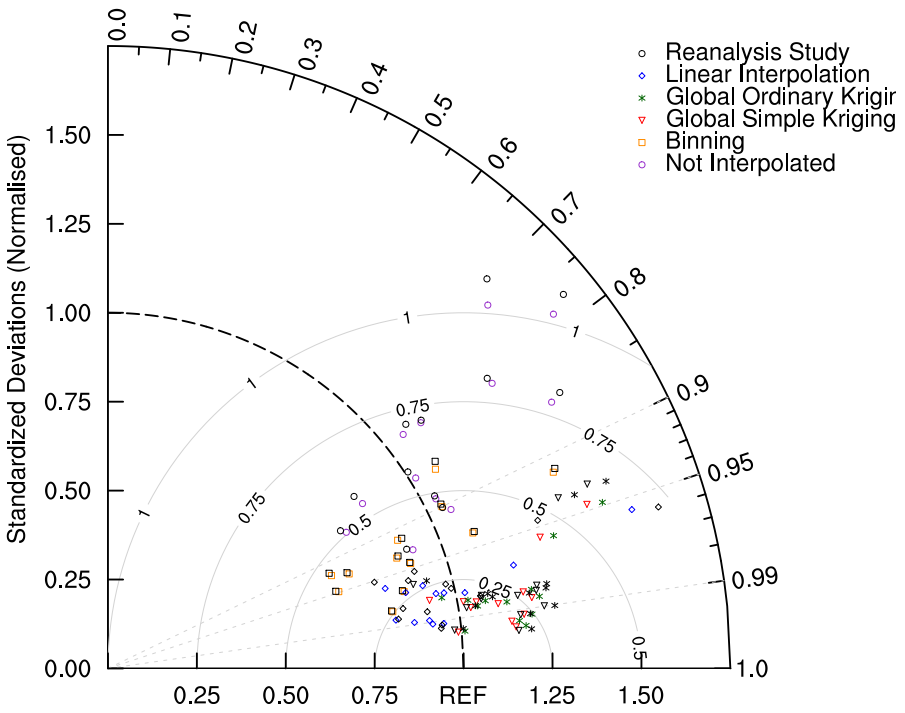

Figure 5.7: A Taylor diagram comparing the monthly and annual Arctic-average anomaly timeseries from this additional sea ice investigation (colour) and from the reanalysis study (black). Each symbol plotted represents a month of the year or the annual value. Cross correlation is shown by the angle with respect to the x-axis. The standard deviations (normalised with respect to the reference standard deviation) can be read from the y-axis. The Root Mean Square Error (K) of the estimated anomalies is proportional to the distance to the point on the x-axis identified as REF (shown by the concentric circles marked 0.25 to 1).

geographical extent of the area where accuracy has decreased, and the size of the increase in the error metrics is larger for LI than for the kriging techniques. To investigate this further, maps of the error and the location of NPDS were examined for annual anomalies. The decreases in accuracy in all techniques result from a combination of the climate in this region of the Arctic and the locations of the supplementary anomalies added over sea ice in this investigation. The area of decreased accuracy is in the Canadian climatic region, a more continental climate than the Interior Arctic where much of the NPDS data is reported (Przybylak, 2003c,a). Reference anomalies located in the Beaufort and Lincoln Seas are likely better represented by anomaly timeseries located over land in the Canadian climatic region than by NPDS observation anomaly timeseries located in the interior Arctic. This effect is worse for LI than for Kriging; LI is vulnerable to clustering effects, as noted in Section 5.3.2.

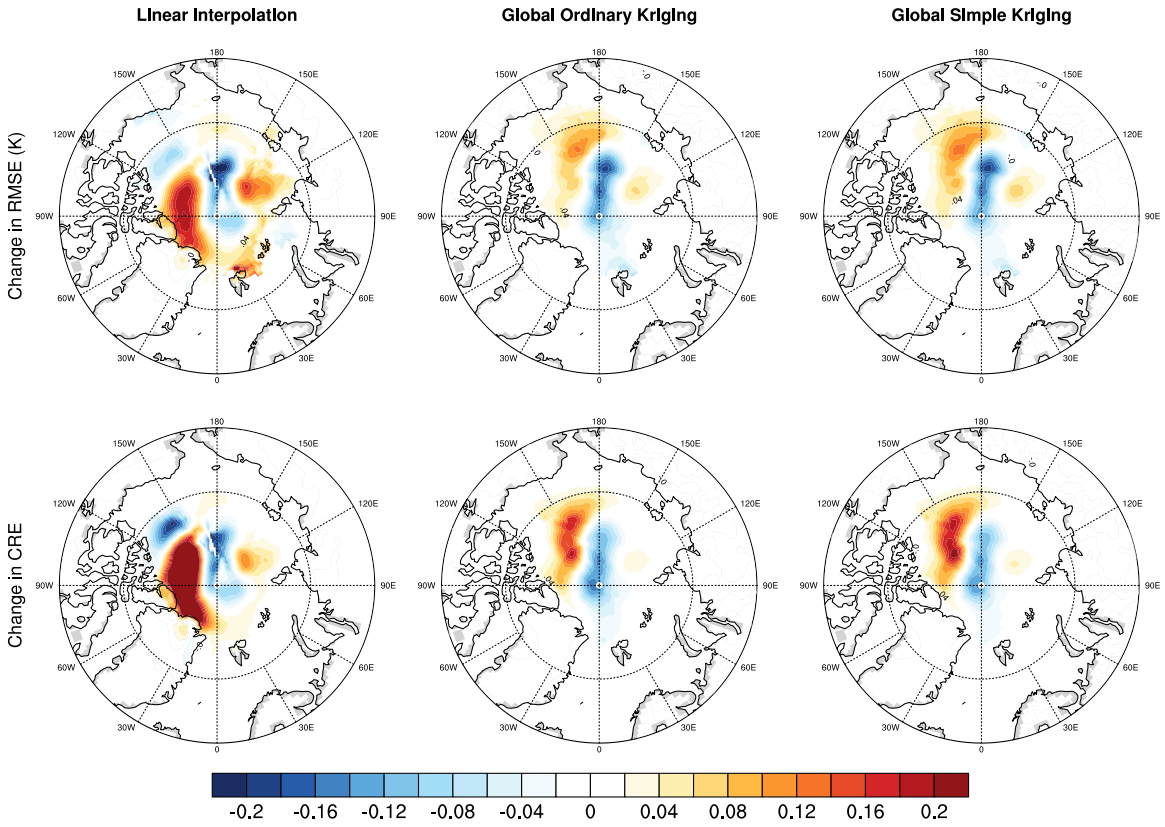

Figure 5.8: The change in Root Mean Square Error (K) and Compound Relative Error for spatially resolved annual SAT anomalies in the Arctic between the additional sea ice data investigation and the reanalysis study.

The changes in RMSE and CRE between estimated monthly anomalies are given in Figures 5.9 and 5.10 for GOK and LI respectively for several months of interest. GSK was again not plotted. As with the additional land data study, most of the areas of change between the metrics of the two studies show an increase in accuracy for interpolation techniques. The decrease in RMSE and CRE values was around 0.4(K) in general, with a decrease of more than 1.0(K) for the metrics in some areas. This is a notable increase

in accuracy. The largest decrease in error metric values was seen for the central Arctic Ocean where the NPDS report temperatures, while modest improvements in accuracy also occur in other areas of the Arctic Ocean. This suggests that adding NPDS data can lead to a considerable increase in accuracy for estimated spatially resolved monthly anomalies. The RMSE for the estimated anomalies reduces from 2.0-3.0K in the reanalysis study to around 1K for the additional sea ice investigation in this study. There are, however, some areas where the metric values increased. For GOK these increases were in the Beaufort Sea area and generally small; less than 0.2K for RMSE and up to 0.5 for CRE. For the LI technique, error metric values increased for the Arctic Ocean near to Greenland and in sea ice edge areas. These increases in RMSE and CRE values can be more than 1K and up to 0.6 respectively. As mentioned previously, this is probably due to a lack of NPDS data in this area and the slightly more continental climate in this region, as well as the sensitivity of LI to station clustering.

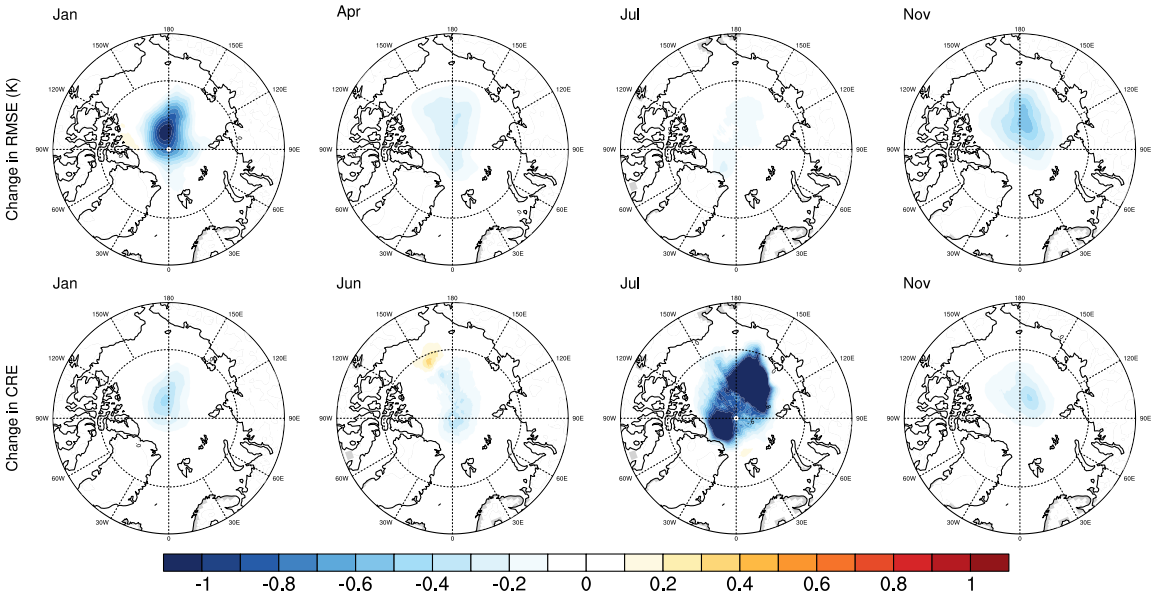

Figure 5.9: The change in the Root Mean Square Error (K) and Compound Relative Error between the reanalysis study and the additional sea ice data investigation for the Global Ordinary Kriging technique. Maps are plotted for several months of interest for each error metric.

Finally, the error metrics associated with each interpolation technique from both the additional sea ice investigation and the reanalysis study were area-weighted and compared (Figure 5.11). This was also done for the additional land data investigation but there were no noticeable differences in the metrics. For kriging techniques, represented in Figure 5.11 by GOK, there is a slight decrease in the RMSE and CRE for most months. This improvement in accuracy is, however, very small. The error metric values decrease by less than 0.05(K), except for CRE in August and July where the improvement can be up to 0.3. There is a very slight increase in the GOK RMSE for December when additional observation location anomaly timeseries are added over sea ice. However this is around 0.02K, and therefore negligible. The difference in the annual area-weighted averages of

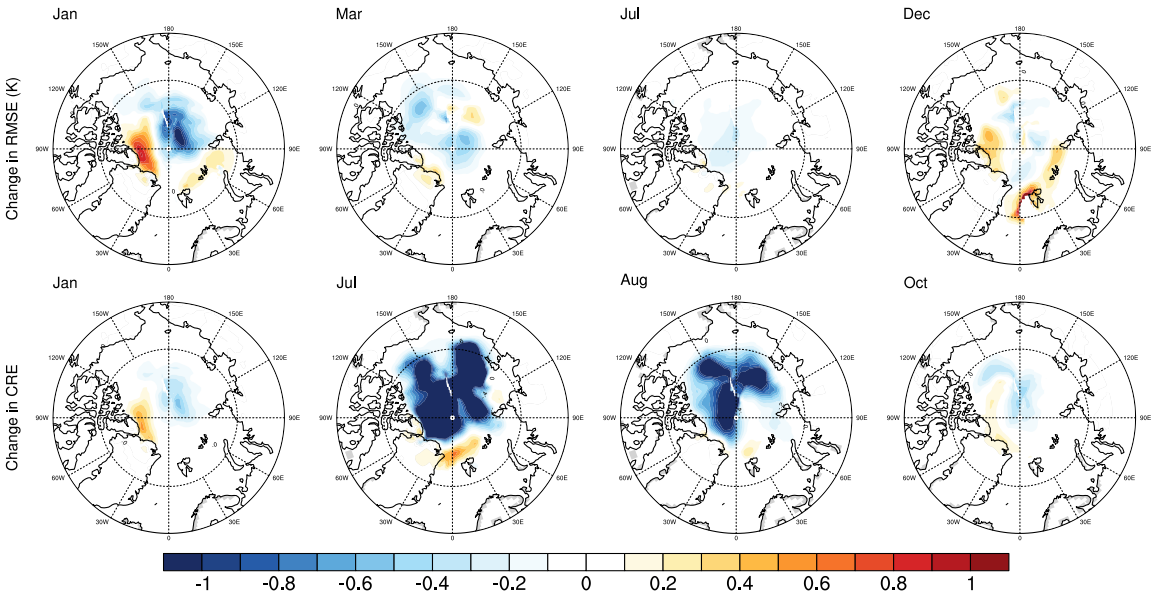

Figure 5.10: The change in the Root Mean Square Error (K) and Compound Relative Error between the reanalysis study and the additional sea ice data investigation for the Linear Interpolation technique. Maps are plotted for several months of interest for each error metric.

the error metrics is essentially 0. The results were very similar for LI.

In conclusion, the use of NPDS observation locations in this additional sea ice data investigation generally leads to an increase in the accuracy of estimated anomalies. This increase in accuracy is particularly notable for monthly anomalies, where there may be up to a 1K improvement in the RMSE in areas where station location anomalies were added over sea ice. This is a substantial decrease in the error considering that the NPDS observation locations are only added to 23 of the 33 years of this study's time period of interest. However there are increases as well as decreases in error metric values for anomalies produced using the LI technique when NPDS observation locations are employed. These increases in error can be large; more than 1K for the RMSE for the Arctic Ocean near to Greenland and for sea ice edge areas. It is likely that these errors are again due to the influence of station anomalies in areas they are not representative of. This is caused by the weighting system used in LI, which also makes this technique vulnerable to station location effects such as clustering.

5.5 Discussion

The addition of supplementary station location anomaly timeseries over land and sea ice to the input anomalies from the reanalysis study had an impact on the accuracy of anomalies estimated using various techniques. Generally there was an improvement in accuracy when additional data locations were added. In some areas of the Arctic there were substantial or non-trivial improvements in the anomalies. There were also, less

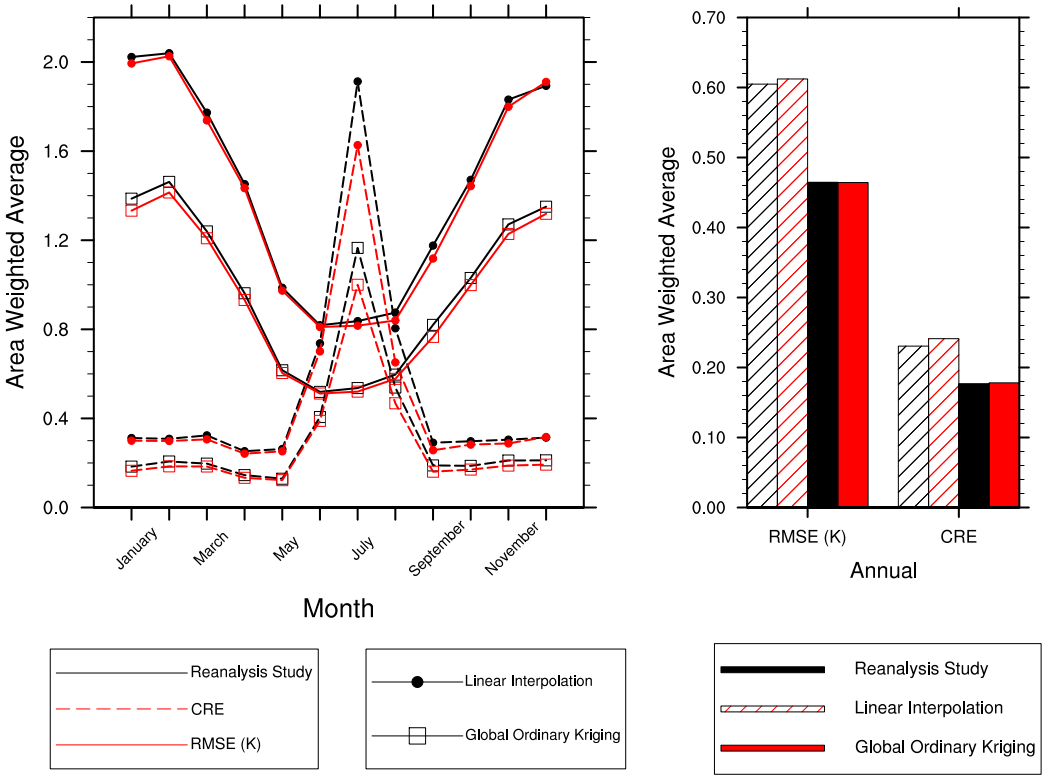

Figure 5.11: Area weighted average of the spatially resolved RMSE and CRE from the reanalysis study and the additional sea ice data investigation.

frequently, decreases in anomaly accuracy.

The improvements resulting from adding supplemental anomaly timeseries over land were smaller and less consistent than noted for the additional sea ice data investigation. This was due to the locations and coverage of the independent ISTI meteorological stations used as supplemental station locations over land. Many of these stations are clustered in Scandinavia and other areas where input anomalies were already present. The use of input anomalies over land which were clustered in certain areas was a limitation in this chapter. For several of the estimation techniques, the changes to the input anomalies for the additional land data investigation had little or no effect on the accuracy of the estimated anomalies produced. However, this spatial clustering of input and additional input anomalies was an issue for some techniques.

For Arctic-average anomalies estimated using NI, areas which are well observed have more influence on the Arctic-average anomalies. Additional anomalies which are clustered in an already well observed area will increase the influence of these regions and decrease the accuracy of the estimated anomalies if the additional anomalies are not representative of the Arctic-average anomaly. In contrast, for the Binning technique these additional anomaly timeseries contribute to grid cells which already contain data. The Arctic-average anomaly is not excessively influenced by a greater coverage of anomaly timeseries in one area. The addition of anomaly timeseries over land led to an increase in accuracy for spatially resolved anomalies estimated by kriging techniques, but a decrease in accuracy for anomalies estimated using LI in some sea ice edge areas. This is due to the fact that the linear weighting used for LI does not account for clustering of stations. The addition of anomaly timeseries over land in an already well observed area near to a sparsely observed area (e.g. the Atlantic sector of the Arctic) causes an increase in the influence of those clustered data within the sparsely observed region, whose climate can be distinct, and hence an increase in the size of the error metrics. This means that caution is needed if additional SAT records are used in conjunction with LI to estimate Arctic SAT anomalies. Moreover, this indicates potential issues may arise for anomalies estimated using LI if there are large changes in the distribution of the observing network.

It is not just clustered station locations that can have an impact on the estimation of anomalies using the five estimation techniques employed in this study. Two additional station location anomaly timeseries located in Franz Josef Land reduced the error for a relatively large area of sea ice coverage for spatially resolved estimated anomalies produced using all interpolation techniques. These station locations are in the particularly sparsely observed Atlantic region of the Arctic and are situated on the Arctic Ocean side of this archipelago. The location and number of station locations added, and their location relative to that of existing input anomalies, are both factors which contribute to the impact of additional data on various estimation techniques, as well as features of the techniques themselves.

For the investigation using additional data over sea ice, the additional input anomalies

were added in areas of the Arctic which were not observed by the reanalysis study input anomalies. As a result the improvements in accuracy were generally larger as well as more consistent spatially and across techniques. The estimated Arctic-average anomalies produced using supplemental observation locations over sea ice and all investigated estimation techniques had RMSEs up to 0.1K smaller. This was noted for annual anomalies as well as most, or all, months. For spatially resolved anomalies, the use of NPDS stations led to a decrease in the errors for estimated anomalies. This is most noticeable for monthly anomalies, where there may be up to a 1K improvement in the RMSE for interpolating techniques in areas where NPDS report. This is a substantial decrease in the error, not just considering the size of the improvement but also because the NPDS stations only provide a small amount of extra data in each month and year for 23 of the 33 years of this study's time period. In addition there are fewer increases in the metrics compared to the additional land data investigation.

However, LI of the additional input anomalies over sea ice still leads to increases in error for some regions. These increases can be large; more than 1K for the RMSE or the Arctic Ocean near to Greenland and for sea ice edge areas. This is probably due to a lack of NPDS data in this area and the slightly more continental climate in this region, as well as the weighting used for LI in this study. NPDS do not report in all areas of the Arctic Ocean. These are manned stations which cannot operate where sea ice thickness and/or sea ice concentrations are small. Some areas of Arctic sea ice, such as in the Beaufort Sea, have a more continental climate than the Interior Arctic where much of the NPDS data is reported. Therefore the addition of many station location anomalies in the interior Arctic will decrease the accuracy in estimated anomalies for areas better represented by anomalies from land based stations when the LI technique is used.

So, the addition of extra data over land and sea ice often led to an improvement in the performance of the investigated estimation techniques. This is especially true for additional data sources located over sea ice or on islands in the Arctic Ocean where data are particularly sparse. The results of Chapter 4 suggest that a similar improvement in accuracy would be seen for actual in situ data as well as the reanalysis data used here. Further improvements in accuracy may be accomplished by the addition of other data sources such as automated drifting platforms; for example the U.S Army Cold Regions Research and Engineering Laboratory network of ice mass balance buoys which provide near-surface air temperatures for areas of sea ice which the NPDS, as manned stations, cannot reach. In Appendix D I identify and summarise potentially suitable sources of further SAT data over sea ice areas in the Arctic from in situ observations, satellite sensors and reanalysis datasets. These include in situ data sources from manned platforms (such as ships and manned drifting ice stations) and unmanned platforms (buoys, automatic radiometeorological stations, radiosondes/rawinsondes and ice mass balance buoys); satellite thermal infrared (TIR) and microwave (MW) instruments measuring Ice Surface Temperatures (ISTs), near ISTs and lower tropospheric temperatures; and reanalysis datasets which

provide estimated, globally complete fields of SATs. The focus for satellite sensors mentioned in Appendix D is on those which are used for surface or atmospheric temperature anomaly datasets, or those which can provide a long and/or reliable record of temperature observations. This Appendix identifies many sources of surface temperatures over Arctic sea ice areas which could be employed to improve estimates of SAT anomalies. More data sources may become available as part of efforts to increase data sharing and data rescue (e.g. Allan et al., 2011; Thorne et al., 2011; Slonosky, 2014).

5.6 Summary

The addition of extra data over land and sea ice to the input anomalies from the reanalysis study detailed in Chapter 3 had an impact on the accuracy of anomalies estimated using various techniques. Generally there was a small improvement in the error metrics when additional data was added. However, in some area of the Arctic there were also substantial, or non-trivial improvements, in the anomalies as well as, less frequently, decreases in anomaly accuracy. The improvements resulting from adding extra data over sea ice were larger and more consistent than for adding extra data over land. This is due to the locations of the additional land stations which are often clustered in already well-observed areas whereas no data over sea ice was included in the reanalysis study input anomalies. Therefore, the addition of data in an area which is not previously represented in the input anomalies led to more accurate estimates of anomalies in general.

However, the performance of the techniques in response to the additional data was not consistent. Kriging techniques showed consistent improvements in the accuracy of their estimated anomalies in the presence of extra data while LI error metrics both increased and decreased in accuracy. This is because kriging techniques account for clustering of stations whereas LI can be quite vulnerable to the negative impacts of this. NI also had issues with the addition of extra land data due to the clustering of stations in some areas. This meant that some areas had a larger influence on the Arctic-average anomalies. If these areas are not representative of the Arctic-average anomaly then the increase in their influence will lead to larger errors. In contrast, the Binning technique accounts for clustering of stations in a small area as a result of regridding the anomalies.

Chapter 6

Can the Inclusion of Surface Temperatures Derived from Satellite Sensors be used to Improve Estimates of Arctic Surface Air Temperature Anomalies?

6.1 Introduction

In Chapter 5 I found that the inclusion of additional data sources improved the estimation of Arctic SAT anomalies when reanalysis data were used as a testbed. Generally there was an increase in accuracy when additional data were added, with larger and more consistent improvements when supplementary input anomalies were added over sea ice or on islands in the Arctic Ocean. The inclusion of additional data impacted the performance of the estimation techniques to different extents; some techniques were negatively impacted by station clustering and/or station weighting issues when additional data were added, whereas others were more robust. The results of Chapter 4 suggest that a similar improvement in accuracy would be seen when the techniques were applied to in situ data assuming the additional data were of sufficient quality. I identified potential additional sources of data (from in situ data sources, satellite instruments and reanalyses) in Appendix D.

The addition of valuable land-based meteorological stations in the Arctic, such as those identified on the Franz Josef Archipelago (Chapter 5), to the calculation of global and regional mean SAT anomalies would be straightforward. SATs from meteorological stations are one of the main data sources for current temperature anomaly datasets. For data sources over Arctic sea ice, which are not currently employed in the most prominent timeseries of temperature change, utilising the data is a more complex issue. In situ temperature records from drifting platforms are spatially and temporally sparse, and

traditional normals cannot be produced from these drifting observations (Chapter 4). In addition, issues were noted with the double difference statistic utilised in Chapter 4 to explore the difference between estimated anomalies and validation data from drifting platforms.

Surface temperatures derived from satellite sensors are available over Arctic sea ice for the last few decades. Ice Surface Temperatures (ISTs) are not generally used to monitor surface temperature change, but could provide a useful temperature record for the Arctic sea ice regions. Thermal infrared (TIR) derived surface temperatures observe the temperature of the surface skin, and are therefore not the same as SAT (measured at around 1.5m height). But TIR derived ISTs could be used as a proxy for SATs over ice and snow, similar to the use of SST as a proxy for MAT (Section 1.2.1.2), as changes in these two types of temperatures are often well correlated. Indeed, TIR surface temperature records show promise for estimating Arctic sea ice temperature change (Comiso, 2003, 2006; Comiso and Hall, 2014; Hansen et al., 2010) and there are many TIR satellite instrument derived surface temperature datasets and data products already available (e.g. ESA, 2014; Heidinger et al., 2014; Merchant et al., 2012). However it must be noted that TIR sensors cannot observe surface temperatures in the presence of cloud.

Therefore, this chapter investigates whether the addition of ISTs derived from TIR sensors on satellites could be useful for improving temperature anomaly dataset accuracy in the Arctic. In this 'additional satellite data' study I focus on the accuracy of estimates of monthly SAT anomalies over Arctic sea ice. This chapter also investigates whether datasets such as the International Arctic Buoy Programme/Polar Exchange at the Sea Surface (IABP/POLES) SAT and ARCSS Surface Air Temperature (ARCSS-SAT) analyses could be used as a climatology over Arctic sea ice areas. This would allow surface temperature anomalies to be calculated from in situ observations in these areas using the commonly employed Climate Anomaly Method (CAM). I applied Global Simple Kriging (GSK) to Arctic meteorological station records, supplemented with a sample of surface temperature data over sea ice areas derived from the Advanced Very High Resolution Radiometer (AVHRR). I assessed the accuracy of estimated anomalies from this 'AVHRR interpolation' using independent validation anomalies over Arctic sea ice. The validation anomalies are Surface Air Temperature (SAT) anomalies produced from in situ SATs and the IABP/POLES dataset as a climatology. The results were then compared to those produced from validating a 'reference interpolation'; GSK was applied to monthly SAT anomalies from Arctic meteorological stations only. The reference interpolation is similar to the in situ data study interpolation (Chapter 4). The input data are the same but a different climatological period is used. The outline of this chapter is as follows. Section 6.2 describes the data and techniques used in this study. Section 6.3 then provides the results. The final section discusses the results and provides a summary and conclusions.

6.2 Data and Techniques

Global Simple Kriging (GSK) was applied to 'input anomalies'; which are in situ monthly SAT anomalies from the CRUTEM4 databank between 1981 and 2013, either used exclusively (the 'reference interpolation') or with surface temperature timeseries derived from AVHRR for the same time period (the 'AVHRR interpolation'). An overview of GSK is provided in Chapter 3 and more details are provided in Appendix B. The same model function fitted parameter values used for GSK in Chapter 4 were employed here. GSK yields 'estimated anomalies'; estimates of spatially resolved Arctic SAT anomalies, which are complete fields of Arctic SAT anomalies gridded to an 100km Equal-Area Scalable Earth (EASE) grid (see Chapter 4). Using this grid allows the calculation of monthly average validation data with sufficient observations in each month and grid cell while being smaller than the expected correlation length of anomalies. The estimated anomalies from both interpolations were compared to validation anomalies over sea ice areas, produced from various in situ SAT data sources identified in Appendix D and the IABP/POLES dataset. Error metrics were produced for both interpolations and the accuracy of the estimated anomalies compared.

6.2.1 Input Anomalies

6.2.1.1 Meteorological Station Surface Air Temperature Anomalies

Input anomalies over land were produced from station records contained in the CRUTEM.4.2.0.0 land surface temperature anomaly dataset (Jones et al., 2012), downloaded from the Met Office Hadley Centre Observations Dataset website (*http* : *//www.metoffice.gov.uk/hadobs/crutem4/data/download.html*). The input anomalies are similar to those for Chapter 4; the data source is the same but a different climatological period is used.

The CRUTEM4 station records contain the monthly average temperatures for each station in the databank, as well as station normals 1961-1990 (if available) for each calendar month. Some of the datasets used in this study (AVHRR data and IABP/POLES) do not contain data back to 1961. I therefore calculated meteorological station anomalies from the station records with respect to station normals for the time period common to all data used to produce a climatology or normal in this study: 1981-2004. This ensures that differences between the anomalies are not a result of differing climatological periods. Station normals for 1981-2004 were created for each meteorological station where at least 14 years of data were available in this time period. This follows the method used to produce station normals 1961-1990 in the CRUTEM4 dataset (Jones et al., 2012). Figure 6.1 shows the stations for which valid normals could be produced between 1981 and 2004. Of the 956 stations in the CRUTEM4 databank which are located above 53°N (which provided information on Arctic temperatures necessary for the interpolating techniques), 856 had valid normals for 12 months while 19 had valid normals for only certain months

of the year. 81 had no valid monthly normals between 1981-2004. However, many of the stations with fewer than 12 months of station normals are located in areas which are relatively well-observed, such as Scandinavia, or are located near to stations which have full normals, for example in Greenland where the station with no normals is within 300km of three stations with normals for 12 calendar months. Therefore, I do not expect a notable loss of accuracy in the interpolation resulting from changes to the climatology period.

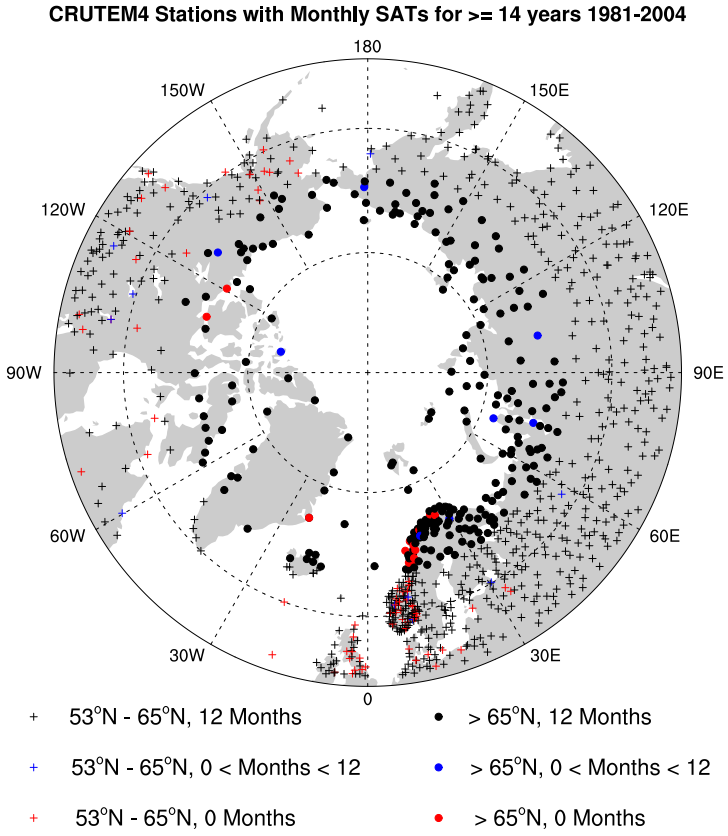

Figure 6.1: The locations of all meteorological stations in the CRUTEM4 databank. Different markers are used to show the locations of the meteorological stations depending on whether they are above 53°N or above 65°N and whether station normals are available between 1981 and 2004.

Monthly station SAT anomalies between 1981 and 2013 (the time period for which AVHRR data were available) were produced, where possible, with respect to 1981-2004. In the CRUTEM4 databank the same location in latitude and longitude can be associated with more than one station record. These duplicates would have caused a problem for the kriging method utilised in this study. Therefore all stations that are located at the same latitude and longitude were identified and the station anomaly records merged as in previous chapters of this thesis. These resulting meteorological station anomalies constitute the input anomalies over land for this additional satellite data study.

6.2.1.2 AVHRR Surface Temperature Anomalies

Input anomalies over sea ice, utilised in the AVHRR interpolation, were produced from a monthly surface temperature dataset derived from AVHRR data. The AVHRR satellite instruments are across-track scanning radiometers measuring visible and infrared radiation. They have been deployed on various satellites since 1978. Surface temperatures have been derived from the thermal infrared channels of AVHRR sensors using various different algorithms (e.g. Comiso and Hall, 2014; Comiso, 2003; Key and Haefliger, 1992). AVHRR surface temperatures are generally produced back to 1981, when AVHRR/2 was launched, due to the complexities and issues in producing surface temperatures from only one thermal infrared channel.

6.2.1.2.1 The Re-normalised AVHRR Monthly Surface Temperature Dataset of Comiso and Hall 2014 An AVHRR monthly surface temperature dataset (Comiso and Hall, 2014) was kindly provided for this study by Dr J. Comiso. This AVHRR dataset was produced using surface temperatures derived from AVHRR/2 and AVHRR/3 radiometric temperatures.

Cloud masking was performed using a combination of a thresholding technique and a daily differencing technique (Comiso, 1994). The thresholding technique utilises the difference in radiance between the channels centered at 3.7μm and 10.8μm, and the channels centered at 10.8μm and 12.0μm (Comiso, 2001). The daily differencing method masks clouds using the difference in daily orbital data from the 10.8μm channel with a given threshold; clouds are identified based on the changes in observed radiances caused by their movement (Comiso, 2000). This cloud masking method, which has an emphasis on the use of the daily differencing method, was found by Comiso 2000 to be the most effective method for cloud masking in a polar environment. Different thresholds were used to mask open ocean, sea ice, and ice sheet surfaces (Comiso, 2000). Day and night data were analysed separately and additional masking was applied over open ocean areas (Comiso, 2000). Atmospheric attenuation was corrected for using the split-window algorithm over open ocean, and a slightly modified version over areas of ice (Comiso, 2000). The AVHRR radiances were then used to derive surface temperatures using the same emissivities for ice and water as used by Comiso 1994. Monthly average surface temperatures were then produced.

The AVHRR instruments were, however, originally intended to be for meteorological purposes rather than climate monitoring (Embury, 2015), in common with meteorological stations. Orbital drift in the satellites carrying the AVHRR sensors, and discontinuities arising from the deployment of new satellites, affect long records of surface temperatures derived from these sensors (e.g. Jin and Treadon, 2003). In addition, the selection of a satellite orbit which avoids direct sunshine on the instruments (Jin and Treadon, 2003; Price, 1991), but is subject to orbital drift, may not continue to avoid direct radiation in the event of orbit drift. Furthermore, as the AVHRR instruments were not meant for

long-term monitoring they may not be calibrated adequately for climate data purposes, sensor degradation may occur, and calibration may differ between individual instruments. Therefore AVHRR surface temperatures are affected by calibration and degradation problems.

To account for changes in calibration as well as degradation by Comiso and Hall 2014 used "established techniques" (Comiso, 2001). In addition, the monthly AVHRR surface temperatures were re-normalised so that they are more consistent with in situ measurements (Comiso, 2001, 2014b). The re-normalisation was completed using the following procedure. In situ data were acquired from the World Meteorological Organization (WMO) for meteorological stations above 60°N (Comiso, 2014a). The station records were quality controlled so that data which were too noisy or different from neighbouring stations were removed (Comiso, 2014a). A regression was calculated between the AVHRR and in situ data and calibration coefficients were derived in order to adjust the AVHRR surface temperatures (Comiso, 2001, 2014a). Adjustments were made on a year to year basis using individual WMO station records (Comiso, 2014a). A normalisation based on monthly averages of all station records was produced as a final refinement (Comiso, 2014a). These re-normalised monthly average AVHRR surface temperatures were found to be consistent with meteorological stations and other in situ data sources (Comiso, 2014b). However, it must be noted that these AVHRR data have some serious cloud masking deficiencies resulting from instrument limitations which the re-normalisation technique employed by Comiso and Hall 2014 will not correct. Furthermore there may also be algorithm limitations.

6.2.1.2.2 A Comparison of Re-normalised AVHRR Surface Temperatures and Temperature Anomalies to Other Data Sources Re-normalised AVHRR monthly surface temperatures (Comiso and Hall, 2014), along with SATs from the IABP/POLES dataset (used as a climatology in this study) and European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA) project interim reanalysis, or ERA-Interim, data in the Arctic were mapped and are shown in Figure 6.2. In common with the study of Comiso and Hall 2014, I found that the spatial distribution of surface temperatures in the AVHRR dataset was consistent with expectations and similar to the spatial distribution of SATs in the IABP/POLES and ERA-Interim datasets. However, as shown in Figure 6.3, the spatially resolved Mean Difference between the datasets can be large.

I compared the monthly surface temperatures from AVHRR to in situ SAT data over sea ice, which constitute the validation data for this study. The AVHRR surface temperatures were generally within 3K of the validation data as shown in Figure 6.4. There is, however, often a warm bias in AVHRR surface temperatures, which can be up to 2K on average for temperatures around 240K, as well as a cold bias of approximately 1K for temperatures around 270K. The RMSE for the whole time period was 2.6K. In boreal summer (June, July, August) the RMSEs were 1.5K, while for all other months the

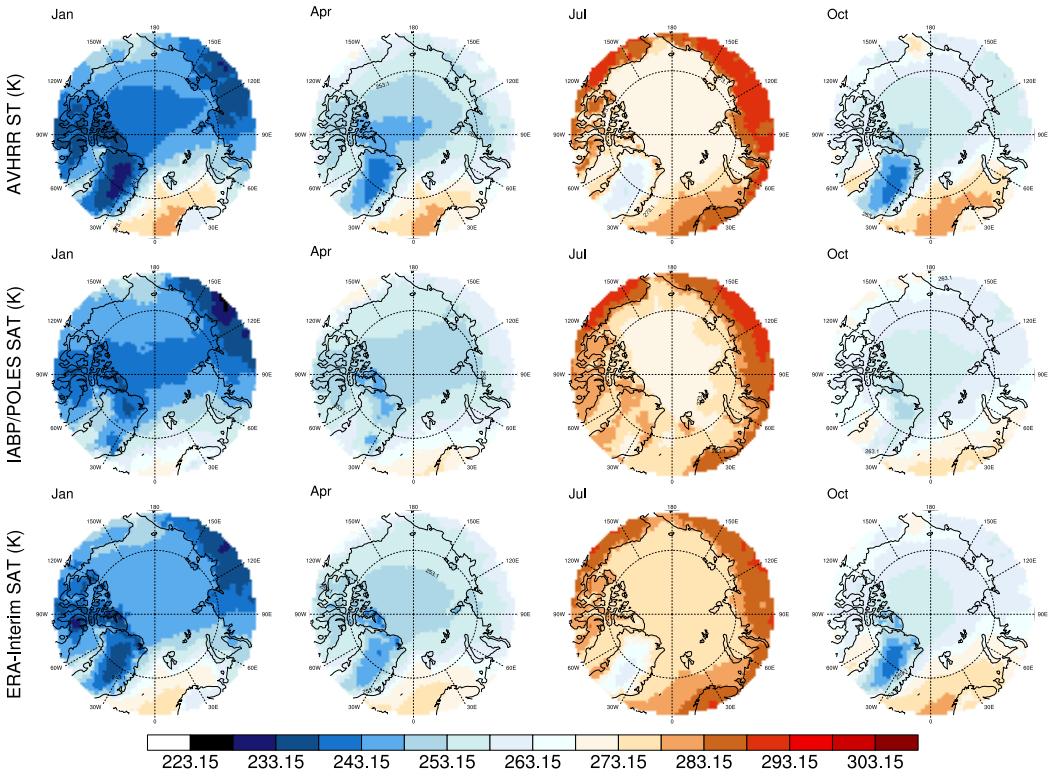

Figure 6.2: Spatially resolved calendar monthly average surface temperatures (K) from the AVHRR dataset of Comiso and Hall 2014, and monthly average surface air temperatures from IABP/POLES (Rigor et al., 2000) and ERA-Interim for the period 1981-2004.

RMSE values were 2.8K. Other studies comparing AVHRR ISTs to in situ data over sea ice found that there were few differences in the statistics of the temperatures and that AVHRR ISTs were within 3K of in situ SATs (Comiso, 2001; Key and Haefliger, 1992; Key et al., 1994, 1997; Yu et al., 1995). In these previous studies AVHRR ISTs had biases of +0.1K in summer and -0.4K for other months compared to in situ SATs (Comiso, 2001; Key and Haefliger, 1992; Key et al., 1994, 1997; Yu et al., 1995). The differences in statistics between these studies likely result from the use of different algorithms to derive surface temperatures from AVHRR, as well as the different techniques used to correct calibration, degradation and atmospheric effects.

Lastly, I compared monthly temperature timeseries from the AVHRR dataset and in situ data sources for two grid cells on Arctic sea ice to the temperature timeseries from a nearby Arctic land-based meteorological station (Figure 6.5) and a distant meteorological station (Figure 6.6). The temperature timeseries from all sources show a similar seasonal pattern. In addition, SATs and surface temperatures from all sources were generally within 3.5K of each other for most months. In boreal summer there are more notable differences between the timeseries; the re-normalised AVHRR surface temperatures vary around 271.15K, in situ SATs vary around 273.15K, and land station SATs are higher and more variable (Figures 6.5 and 6.6). Land station SATs are larger and more variable than SATs over sea ice in summer as they are not constrained by latent heat effects from

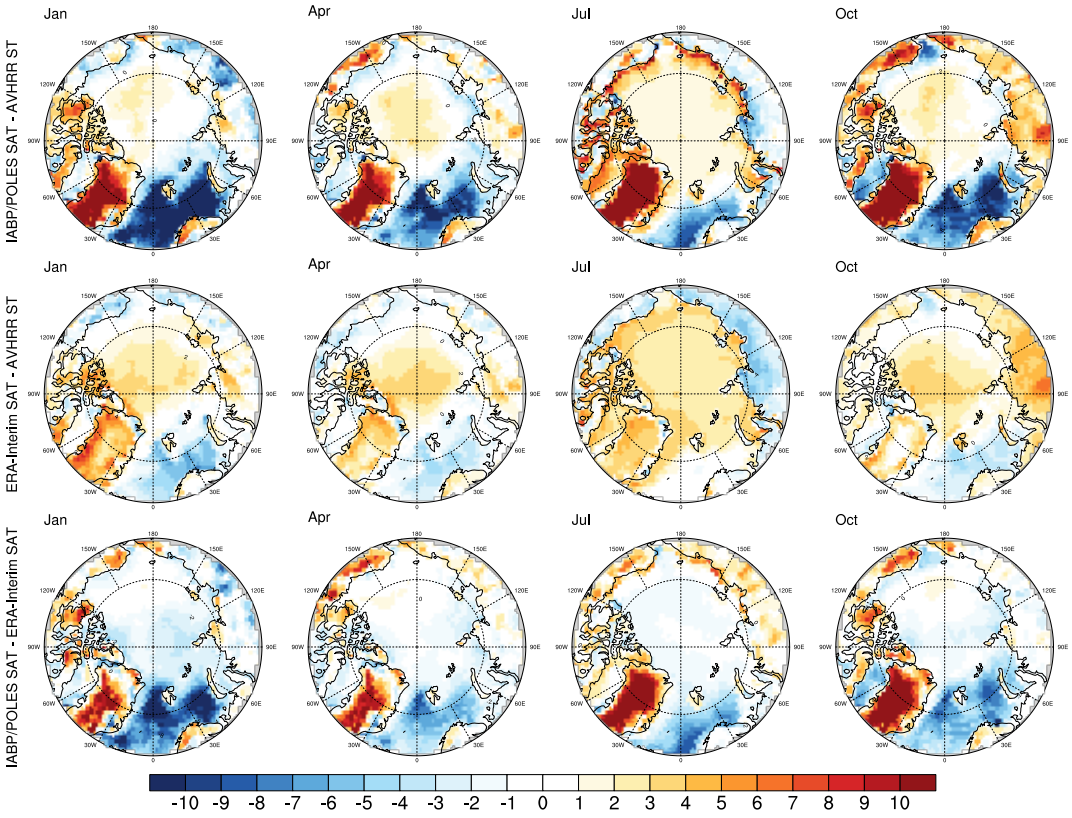

Figure 6.3: The Mean Difference (K) between spatially resolved monthly average surface temperatures from the AVHRR dataset of Comiso and Hall 2014, and monthly average surface air temperatures from IABP/POLES (Rigor et al., 2000) and ERA-Interim between 1981 and 2004.

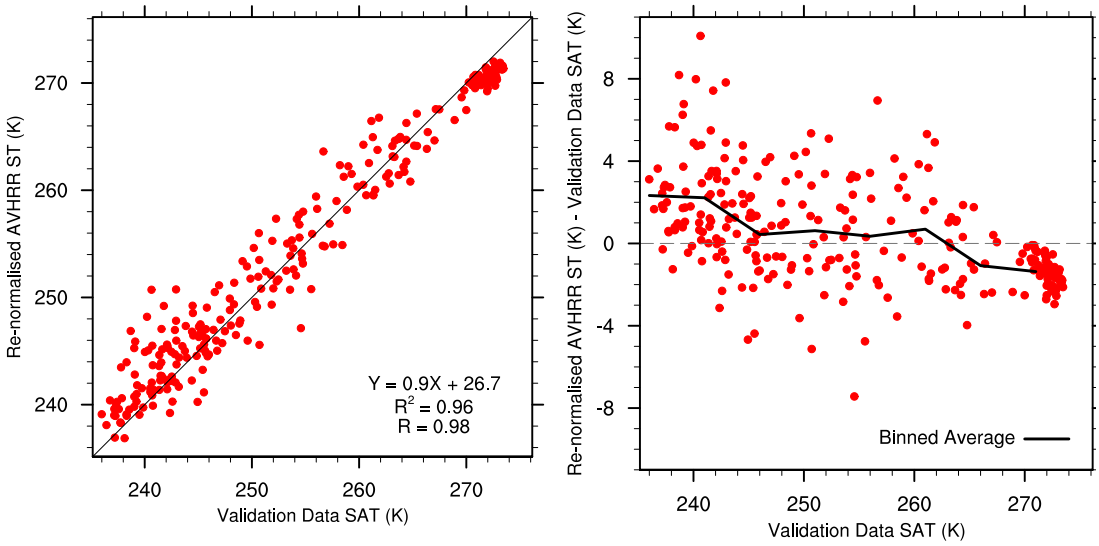

Figure 6.4: Monthly surface temperature data 1981-2008 from the AVHRR dataset of Comiso and Hall 2014 compared to the validation data used in this study (from NPDS and other reliable data sources over Arctic sea ice).

melting sea ice (Przybylak, 2003a). The AVHRR surface temperatures were closer than those from land stations to temperatures from North Pole Drifting Stations (NPDS), as would be expected (Przybylak, 2003a). However, the temperatures from NPDS, other in situ data sources, and previous studies (e.g. Chen et al., 2002; Wiese et al., 2015) suggest that surface temperatures over sea ice should be around 273.15K during the melt season. This relatively consistent cold bias could point to cloud masking issues in this dataset, or an inappropriate emissivity of sea ice in the summer. AVHRR data are known to have some serious cloud masking deficiencies resulting from instrument limitations (Comiso, 2014b).

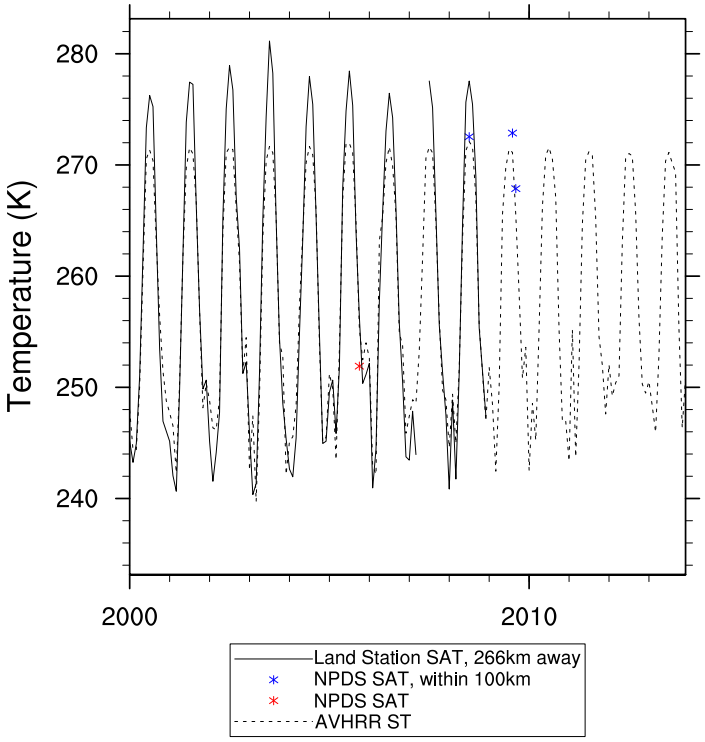

Figure 6.5: Monthly temperature timeseries from the AVHRR dataset of Comiso and Hall 2014 and NPDS observations (both those in the grid cell and those up to 100km away) for the EASE grid cell located at 83.62N, 8.13W compared to the temperature timeseries from a nearby Arctic land-based meteorological station (within 300km of the grid cell).

Anomalies were also produced from the temperature timeseries investigated above and the anomaly timeseries were compared (Figures 6.7 and 6.8). Land station and AVHRR anomalies were produced with respect to normals produced from these datasets (see Section 6.2.1.1 and 6.2.1.2.3). NPDS anomalies were produced with respect to the IABP/POLES analysis (see Section 6.2.2). It was found that the anomalies were fairly similar between datasets (Figures 6.7 and 6.8), with RMSEs often less than 3.5K. For AVHRR compared to NPDS observations the RMSE was less than 2.5K for all but one timeseries comparison.

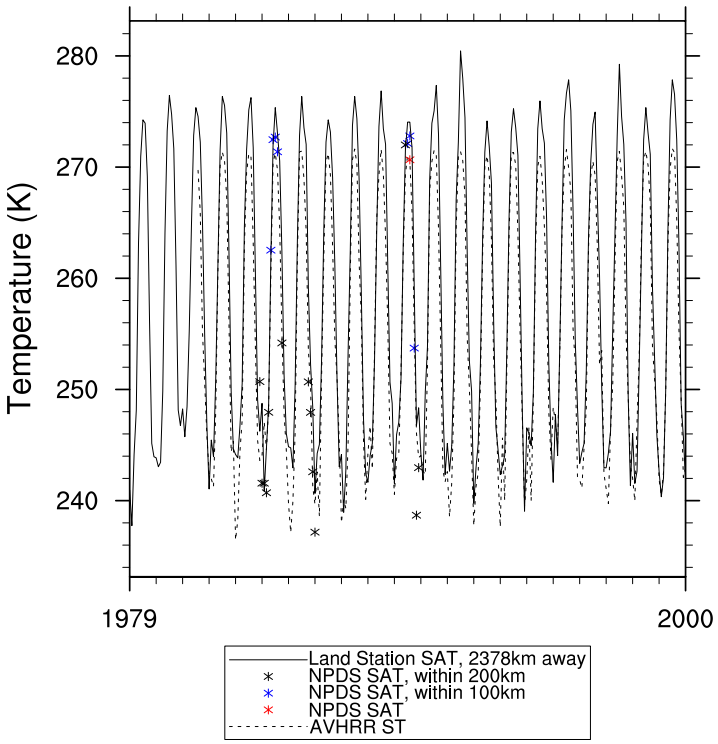

Figure 6.6: Monthly temperature timeseries from the AVHRR dataset of Comiso and Hall 2014 and NPDS observations (both those in the grid cell and those up to 200km away) for the EASE grid cell located at 81.83N, 73.66W compared to the temperature timeseries from a distant Arctic land-based meteorological station (more than 1200km distance from the grid cell).

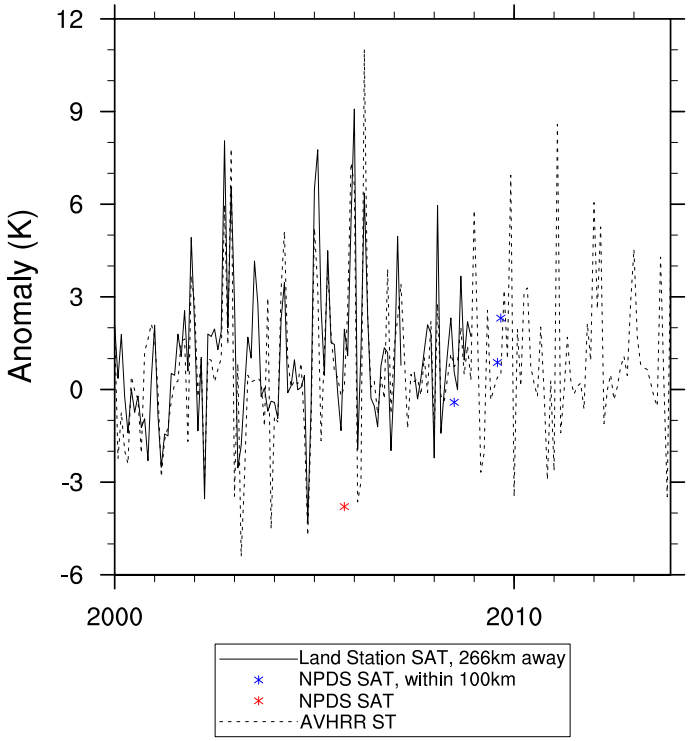

Figure 6.7: Monthly temperature anomaly timeseries from the AVHRR dataset of Comiso and Hall 2014 and NPDS observations (both those in the grid cell and those up to 100km away) for the EASE grid cell located at 83.62N, 8.13W compared to the temperature timeseries from a nearby Arctic land-based meteorological station (within 300km of the grid cell).

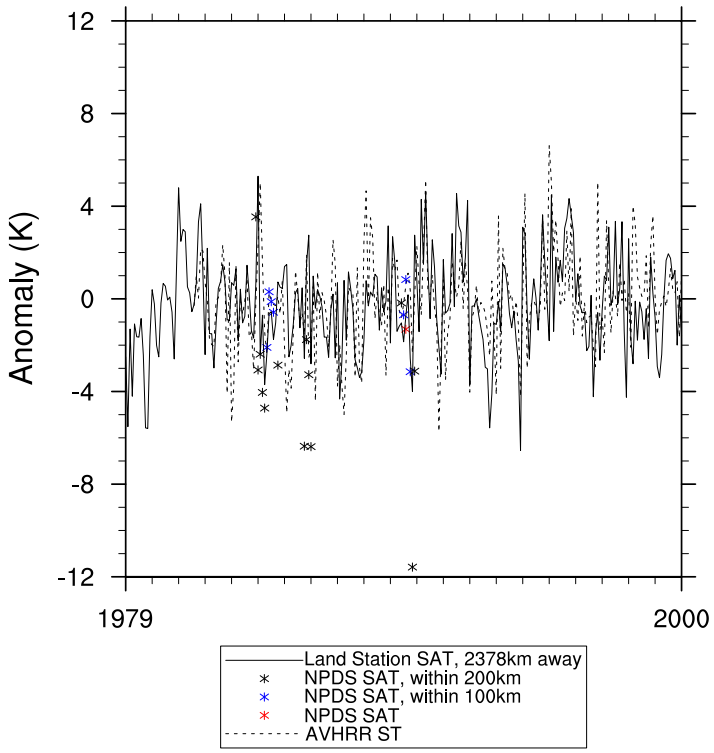

Figure 6.8: Monthly temperature timeseries from the AVHRR dataset of Comiso and Hall 2014 and NPDS observations (both those in the grid cell and those up to 200km away) for the EASE grid cell located at 81.83N, 73.66W compared to the temperature timeseries from a distant Arctic land-based meteorological station (more than 1200km distance from the grid cell).

Therefore, the re-normalised AVHRR surface temperature dataset used to produce input anomalies in this additional satellite study are similar to the validation data, although some biases were noted. In addition, the spatial distribution of temperatures and anomalies in the Arctic are similar to those of the ERA-Interim reanalysis, which is physically plausible in the Arctic (Chapter 2), and the IABP/POLES analysis, which is produced by optimally interpolating in situ data (Rigor et al., 2000).

6.2.1.2.3 AVHRR Derived Monthly Surface Temperature Anomalies The monthly re-normalised AVHRR surface temperatures were regridded to a 100km EASE grid (see Chapter 4). Surface temperatures between 1981 and 2004 were used as a climatology to produce monthly AVHRR surface temperature anomalies. Sea ice concentration information from ERA-Interim was gridded to the same EASE-grid and grid cells with at least 15% sea ice in all months between 1981 and 2013 were identified. AVHRR surface temperature anomaly timeseries from these 311 grid cells were used as input anomalies in this study and their locations are shown in Figure 6.9.

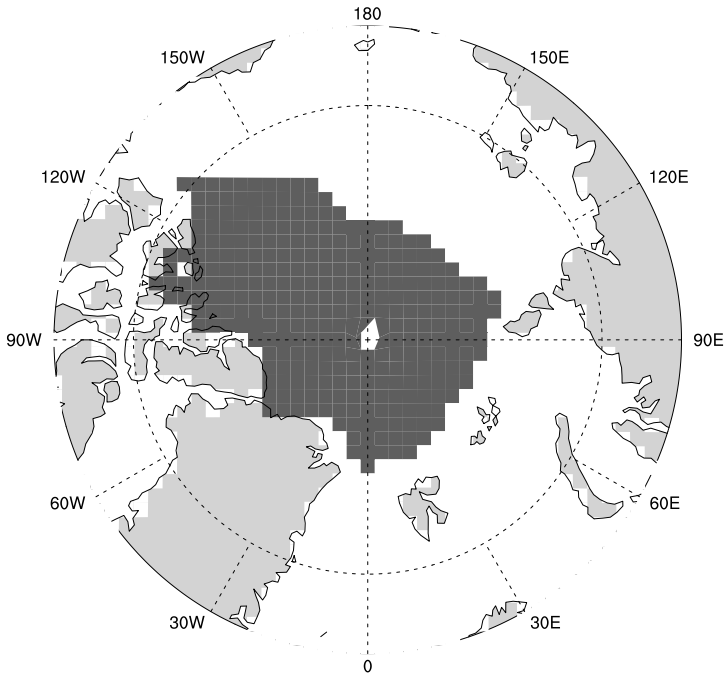

Figure 6.9: The location of AVHRR input surface temperature anomalies (dark grey). The AVHRR input anomalies are located in grid cells which contain permanent sea ice (at least 15% sea ice in all months) in ERA-Interim reanalysis data (gridded to a 100km EASE grid) in the time period 1981-2013.

6.2.2 Validation Anomalies

6.2.2.1 Validation Data

In situ sources of SAT data over Arctic sea ice, which reported radiation shielded SATs at between 1.25m and 2m height (the WMO definition of SAT) between 1981 and 2013 and for which data was available online, were identified from Appendix D. The selected data sources were: NPDS, the Arctic Ocean Experiment-2001 (AOE-2001, Tjernström 2005; Tjernström et al. 2005), the Arctic Summer Cloud-Ocean Study (ASCOS, Tjernström et al. 2014), the Applied Physics Laboratory Ice Station (APLIS, IARC 2012) 2007, National Oceanic and Atmospheric Administration (NOAA) Pacific Marine Environmental Laboratory (PMEL) weather stations (2004-2009), the Surface Heat Budget of the Arctic Ocean (SHEBA, Uttal et al. 2002) Ice Station, and Ice Station Tara (Vihma et al., 2008).

Sub-daily SAT data from the identified data sources were averaged into daily SAT records and gridded to the EASE grid as in Chapter 4. A full month of data records from drifting platforms is not generally available for a grid cell in a given year. The required sampling for a monthly average SAT in this study was determined using a similar method to that of Chapter 4. Based on this investigation it was decided to use a required sampling of 15 days for a monthly average. This was a compromise between the increase in apparent anomaly error arising from undersampling of the validation data and having enough data available to calculate difference metrics for the estimated anomalies. It should be noted here that the investigated grid cells are not required to have observations in more than one year, due to the use of a climatology to calculate anomalies rather than comparing actual temperatures using a double difference statistic. Validation anomalies were produced from these monthly average SATs, with respect to 1981-2004, using IABP/POLES data as a climatology (see Section 6.2.2.2). These monthly SAT anomalies constitute the validation anomalies in this study.

6.2.2.2 An Arctic Surface Air Temperature Climatology from IABP/POLES

The climatology used to produce validation anomalies from in situ sources of SAT data in this chapter was the IABP/POLES SAT analysis (Rigor et al., 2000). IABP/POLES provides 12-hourly gridded fields of Arctic SATs between 1979 and 2004 (Rigor et al., 2000; IABP, 2014). The analysis was produced by interpolating SAT data from various sources to produce a complete analysis of Arctic SATs (see Section 6.2.2.2.1, Rigor et al. 2000).

The ARCSS Surface Air Temperature (ARCSS-SAT) Analysis is a preliminary update to this dataset which extends IABP/POLES to 2008 and improves the analysis (Rigor et al., 2000; EOL, 2014). However, an initial investigation of ARCSS-SAT in this study noted that there may be issues with the format of the dataset version available at the time of this additional satellite data study. The data file had constant monthly temperature values for each grid cell when 12 hourly data were expected. Furthermore, these data

do not correspond with monthly average temperatures from other sources of temperature data in the Arctic such as NPDS. It is likely that these values are repeated instances of the first 12 hourly temperature in each month. These issues were not present in the IABP/POLES analysis. As a result the suitability of the IABP/POLES analysis, rather than the updated ARCSS-SAT analysis, for providing a climatology for the validation data was investigated.

6.2.2.2.1 The IABP/POLES Surface Air Temperature Analysis The IABP/POLES SAT analysis was produced primarily from in situ SAT observations from manned drifting stations and quality controlled buoy observations (Rigor et al., 2000). These SAT observations were optimally interpolated using a monthly first guess field, and monthly correlation length scales derived from the correlations between different observation types (ocean, coastal or land interior) for observation weighting (Rigor et al., 2000). The first guess field was produced by spatially binning and optimally interpolating in situ SAT observations supplemented with NCEP/NCAR Reanalysis 1 (NCEP R1) data in sparsely observed areas such as the North Atlantic (Rigor et al., 2000). The monthly correlation length scales were also utilised in producing these first guess fields (Rigor et al., 2000).

6.2.2.2.2 The Suitability of IABP/POLES as a Climatology I compared monthly average SATs produced from the IABP/POLES analysis to the in situ validation data, as well as other datasets, in order to try and examine the suitability of this dataset as a climatology. In Figure 6.2 the spatial distribution of SATs in IABP/POLES were compared to the re-normalised AVHRR dataset used for input anomalies, as well as to ERA-Interim reanalysis data. As noted in Section 6.2.1.2.2, the spatial distribution IABP/POLES SATs were consistent with those in the AVHRR dataset. In addition, the temperature distribution was found to be consistent with expectations and similar to the distribution of ERA-Interim SAT data which is physically plausible for Arctic SATs, although some large differences between the datasets were noted.

I also compared calendar monthly average SATs from IABP/POLES to in situ validation data from NPDS (an input to IABP/POLES) as well as to other validation data sources which are independent of the IABP/POLES analysis (AOE-2001, ASCOS, SEDNA, NOAA PMEL weather stations, Ice Station SHEBA and Ice Station Tara). The IABP/POLES monthly SAT values are largely within 2.5K of both the NPDS and independent in situ SAT data as shown in Figure 6.10. The Root Mean Square Error (RMSE) for the entire time period is 2.3K when compared to NPDS SATs and 2.4K when compared to independent data. For boreal summer (June to August) and winter (November to March) the RMSE is 0.6K and 2.6K respectively for NPDS data, and 0.8K and 2.8K for independent data. There is not a large discrepancy in the differences compared to NPDS and independent data. This corresponds with the findings of Rigor et al. 2000 who found that the IABP/POLES dataset had a Root Mean Square Error (RMSE) of 2.5K

(1.1K in summer and 2.8K in winter) compared to SAT observations from NP-28, which was excluded as an input to the analysis for that test.

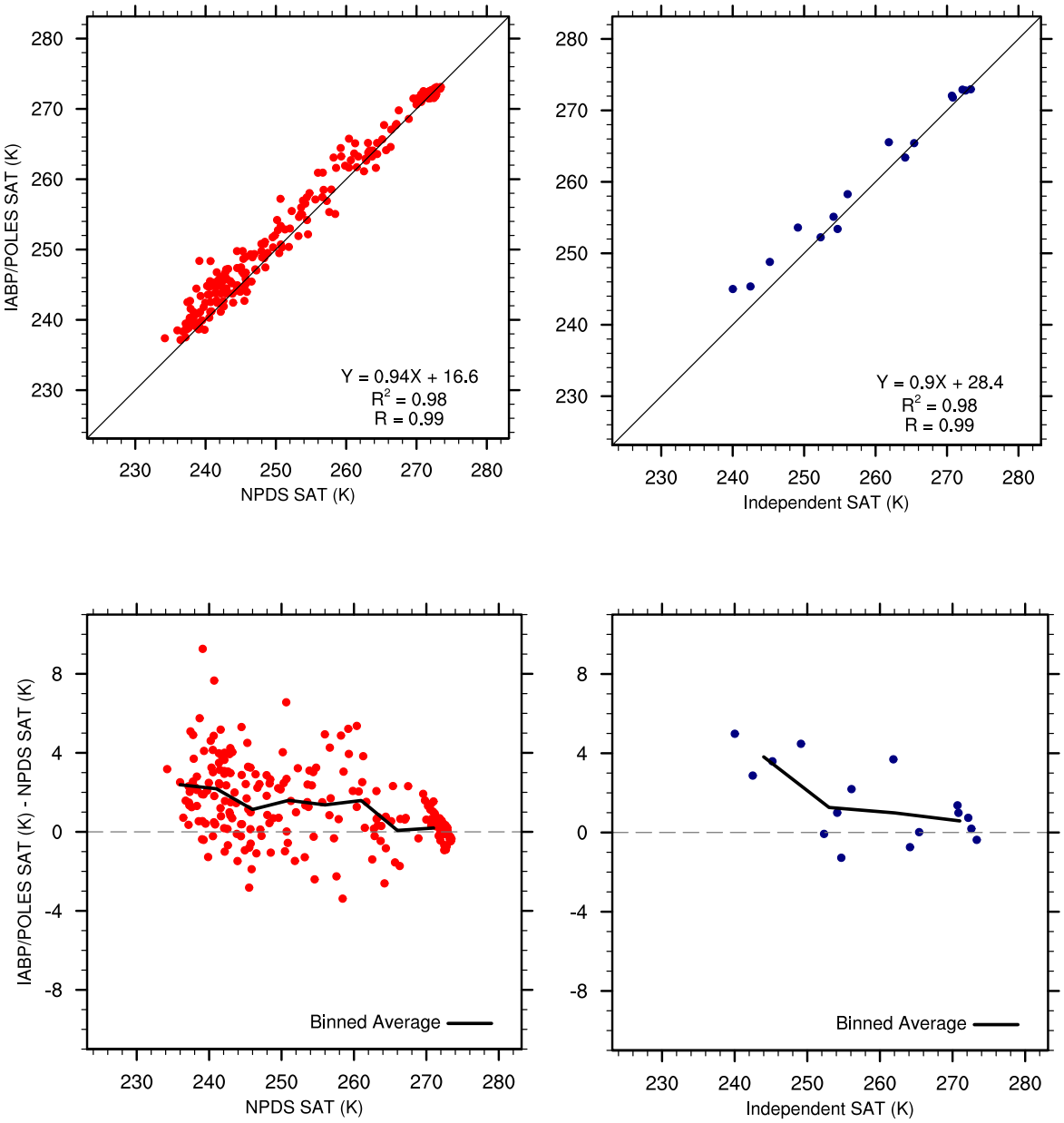

Figure 6.10: Monthly Surface Air Temperature data (K) 1981-2004 from the IABP/POLES dataset compared to the validation data used in this study, both NPDS and other data sources which are not used in the IABP/POLES analysis.

The bias in IABP/POLES data compared to the validation data is generally positive with larger biases for lower temperatures as shown in Figure 6.10. The average warm bias compared to NPDS data is 1-2K for temperatures of 260K or lower and is of a similar size compared to independent validation data (Figure 6.10). This positive bias of around 1-2K over sea ice can also be seen in Figure 6.3 when IABP/POLES SATs are compared to AVHRR surface temperatures. This bias has been noted in previous studies (Chen

et al., 2002) and biases are a known issue for this SAT analysis (EOL, 2014; Rigor, 2014). Buoy data, used as an input for IABP/POLES, have warm biases, especially early in the record, resulting from absorption of solar radiation and snow cover insulating the buoy and temperature sensor (Chen et al., 2002). In addition NCEP R1 data, which is also an input, has warm biases in the Arctic (see Chapter 2).

However, there are few datasets which can provide a climatology for Arctic sea ice. Datasets which could potentially provide a climatology in this region are: IABP/POLES and ARCSS-SAT (interpolated and extrapolated from primarily in situ data); satellite datasets with a long record (such as AVHRR); and reanalysis datasets. ARCSS-SAT, as mentioned before, was found to have data format issues at the time this study was conducted. Satellite instruments which provide surface temperatures in the Arctic (see Appendix D) either had short temperature records during the time period of interest for this study (MODIS, TOVS) or datasets/data products could not be identified. Using AVHRR data as a climatology would mean that the validation anomalies would not provide an independent validation of the AVHRR interpolation and this would affect the results of this study. Finally, reanalysis datasets can have biases and errors compared to in situ data and therefore may not be suitable as a climatology. For example, ERA-Interim, which was noted in Chapter 2 as physically plausible in the Arctic and a suitable testbed for investigating SATs in this region has a persistent warm bias compared to in situ SAT observations. When compared to the validation data used in this chapter the average warm bias is up to 4K, with higher biases for lower temperatures. Therefore, it was decided to use IABP/POLES as a climatology for the validation data used in this study.

6.2.2.2.3 An IABP/POLES Climatology IABP/POLES was gridded to the common 100km EASE Grid and calendar monthly average SATs were produced from the 12 hourly data between 1981 and 2004. These monthly average SATs were then bias corrected; the bias relative to the validation data was estimated using a linear regression and the estimated bias was removed. The estimated bias is given by $y = -0.06x + 16.8$, where x is the uncorrected IABP/POLES SAT (K). These bias corrected monthly average IABP/POLES SATs constitute the climatology used to produce validation anomalies from the in situ validation data sources over sea ice.

6.2.3 Comparison of Estimated Anomalies to the Validation Data

The estimated anomalies from both the reference interpolation and the AVHRR interpolation, produced by applying GSK to the input anomalies for this study, were compared to monthly validation anomalies over Arctic sea ice. Monthly differences were calculated by comparing the estimated anomalies of both interpolations to the validation anomalies for all months and grid cells where validation data are available. Difference metrics were

then produced from the monthly difference. As in other Chapters of this thesis the Root Mean Square Difference (RMSD), Mean Difference (MD), Standard Deviation of Difference (SDD), and Compound Relative Difference (CRD) metrics were calculated. These metrics are described, and the equation used to produce the CRD metric is given, in Chapter 3 of this thesis. As in Chapter 4, these metrics are calculated using the difference from the validation data, rather than the errors used in Chapter 3 and Chapter 5 where the 'true' temperature value was known due to the use of a reanalysis testbed. An additional metric, which measures the absolute difference, was also calculated in this study, but was found to be fairly similar to the RMSD, and therefore not reported here.

6.3 The Impact of Including Additional, AVHRR Derived Surface Temperature Data on the Performance of Estimation Techniques

The anomalies from both interpolations were compared to the validation anomalies for all months and grid cells where validation anomalies are present. Scatter plots of the monthly anomalies 1981-2013, from both the reference and AVHRR interpolations, compared to validation anomalies are given in Figure 6.11. The estimated anomalies from both interpolations were generally within 2K of the validation anomalies and have a slightly positive bias in most months. Overall the MD and RMSD is 0.5K and 2.4K for the reference, and 0.4K and 2.3K for the AVHRR interpolation. The RMSD was lower for AVHRR estimated anomalies on average, and slightly lower or the same for the majority of months. The bias in the AVHRR estimated anomalies was also lower on average, but the bias was lower in the reference estimated anomalies for the majority (7 out of 12) of months.

To investigate the differences between interpolations further, monthly difference metrics were produced (Figure 6.12). 95% confidence interval estimates for the metrics were calculated using the SDD. The sample size of the validation data was relatively small, around 30 ($21 \geq n \geq 33$), so I decided to use the Student's t-distribution to estimate the confidence intervals. The AVHRR interpolation produces anomalies with smaller difference metric values than the reference interpolation for half of the months. For these months the RMSD and CRD values are around 0.2K and 0.1 (CRD is a unitless metric) smaller respectively. Similar changes in accuracy were seen for months where the reference interpolation is more representative.

The AVHRR interpolation shows a more consistent improvement in summer anomalies (Figure 6.12), despite a relatively consistent cold bias of the re-normalised AVHRR derived surface temperatures in this season (Section 6.2.1.2.2). AVHRR surface temperatures show a similar size and variation in summer compared to the in situ SATs over sea ice; both are influenced by the latent heat effects from melting sea ice. In comparison, SATs from land stations are larger and show greater variation in the summer (Figures 6.5 and

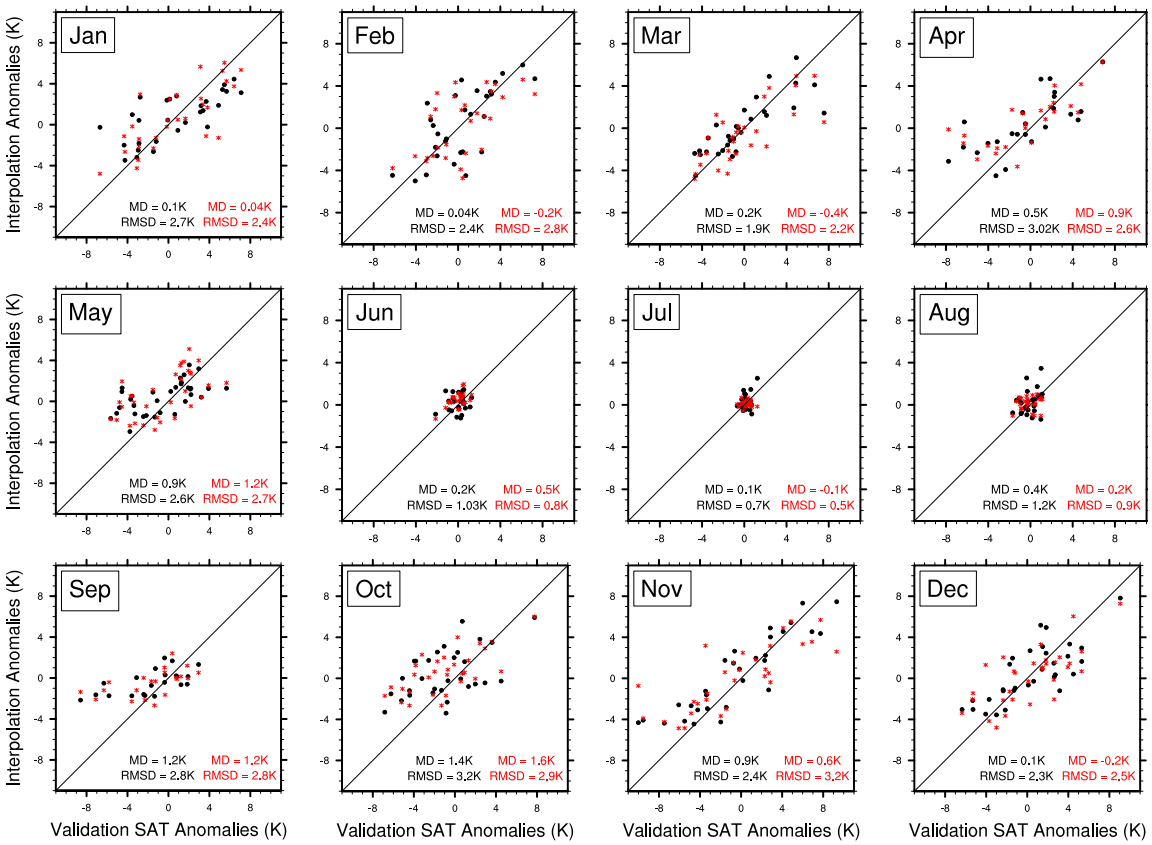

Figure 6.11: Monthly scatter plots of monthly anomalies 1981-2013 from the reference and AVHRR interpolations compared to validation anomalies. Black markers represent the reference interpolation, red markers represent the AVHRR interpolation.

6.6). This may explain the relatively large decrease in CRD values when supplementary surface temperatures derived from AVHRR are added over Arctic sea ice areas. But, this improvement could also be a feature of the small size of the anomalies in summer and the calculation of the CRD value. A small improvement in the absolute difference leads to a large improvement in the CRD if the variation in the reference, in this case the variance in the validation anomalies, is small, as is noted over Arctic sea ice in the summer. However, as shown in Figure 6.12, the confidence interval estimates for both interpolations overlap in all months and therefore they are not significantly different at the 95% confidence interval.

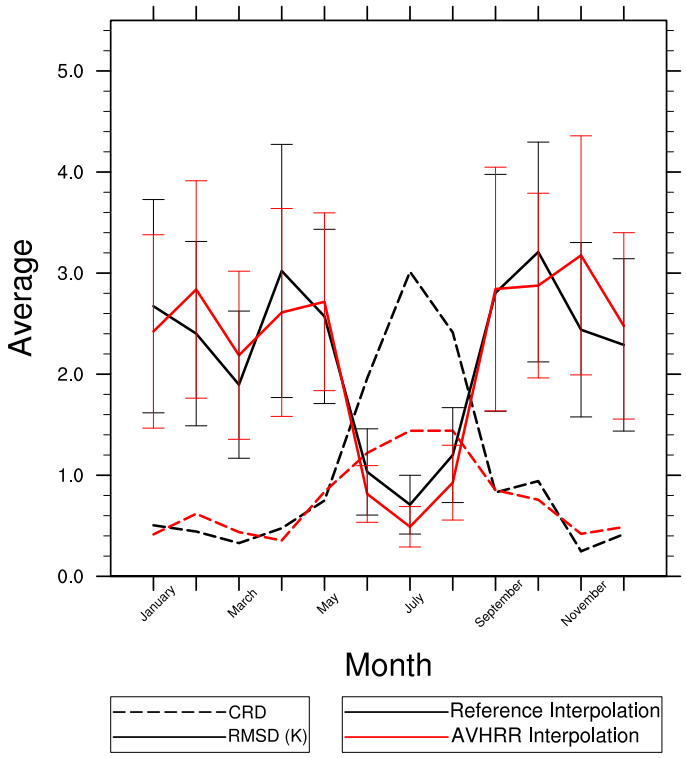

Figure 6.12: The Root Mean Square Difference (K) and Compound Relative Difference between estimated Arctic anomalies, from the reference and the AVHRR interpolations, and validation data 1981-2013. 95% confidence intervals are given for the RMSE from both interpolations. CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

The change in RMSD between the two interpolations was mapped and is shown for several example months in Figure 6.13. As noted for the scatter plots and metric graphs there is often a decrease in the difference metrics when AVHRR data are added to the input anomalies, but increases in the metrics are also often noted. 50% of the 323 grid cells for which an RMSD could be calculated show an increase in accuracy for the AVHRR interpolation (Figure 6.13). For all of the spatial comparisons mentioned here there was no obvious spatial pattern to the change in RMSD values. It was difficult to ascertain spatial differences in errors in this study as a result of the relatively temporally and spatially sparse sample of validation data.

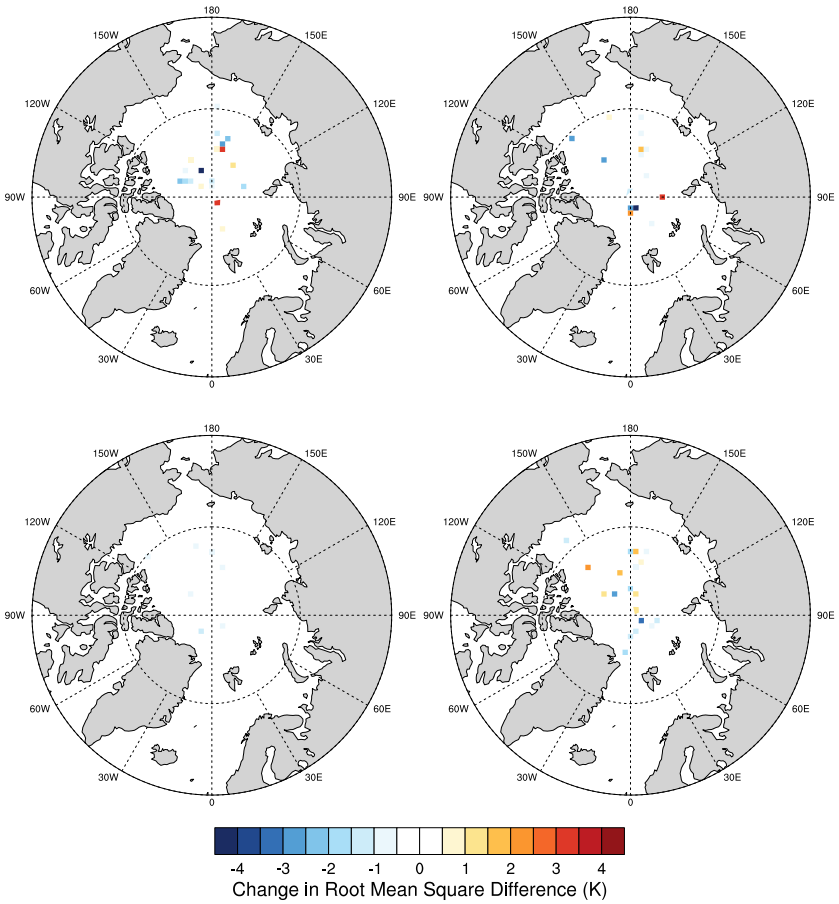

Figure 6.13: Change in Root Mean Square Difference (K) for several example months between AVHRR interpolation and reference interpolation monthly estimated anomalies for produced.

The decadal accuracy of estimated anomalies from the reference and AVHRR interpolations was investigated by producing decadal scatter plots (Figure 6.14). The chosen decades start from the beginning of the time period of interest (1981-2013). The sparseness of data for the second decade investigated (1991-2000) is due to the gap in the NPDS data between 1991-2003 and a paucity of suitable validation data prior to 2001. The reference and AVHRR estimated anomalies have similar biases and RMSD values in each decade; the MD is between -0.2 and 0.8K and the RMSD values are between 2.1 and 2.5K. Both interpolations show slight increases in accuracy for later decades. For the AVHRR interpolation, the decades chosen happen to be almost concurrent with changes in the AVHRR sensor version deployed on satellites which are relevant to this study (AVHRR/2 1981-2007, AVHRR/3 1998-present) so the increases in accuracy for later decades could be a result of changes in sensor accuracy. For the reference interpolation, this increase in accuracy is difficult to explain. As shown in Figure 3.1 of this thesis, the number of meteorological station records in the CRUTEM4 databank version used for this study has decreased (although fractional coverage is fairly static), and has likely decreased in the input anomalies used for this study. Therefore the increase in accuracy is not due to an increase in input anomaly coverage. It could be that uncertainties are larger in the validation anomalies for earlier years or that there are time-varying biases in the data contributing to the validation anomalies.

The metric values from the two interpolations are very similar for 1991-2000 and 2001-2010 as shown in Figure 6.14. Between 1981 and 1990 the AVHRR interpolation MD and RMSD are slightly higher than those from the reference interpolation (Figure 6.14); in this decade there is a decrease in accuracy when AVHRR data are added to the input anomalies. However, it must be noted that the differences in decadal accuracy between the estimated anomalies produced by the two interpolations, and within the same interpolation, are small and, as there is a relatively small sample of validation data, the differences may not be meaningful.

6.4 Discussion

The impact of the addition of TIR satellite sensor derived surface temperatures over Arctic sea ice to the estimation of SAT anomalies in this region using GSK was investigated. Often there was a slight increase in accuracy when surface temperatures derived from the AVHRR instruments were utilised. AVHRR interpolation estimated anomalies have smaller error metric values than the reference interpolation for half of calendar months. For these months the difference metrics were generally 0.2-0.1(K) smaller than those produced from the reference interpolation with consistently better anomaly accuracy in boreal summer months; RMSE and CRE values are around 0.2K and about 1.6 smaller respectively. However, decreases in accuracy were noted nearly as often as increases in accuracy. Furthermore the differences were generally small, for example the differences

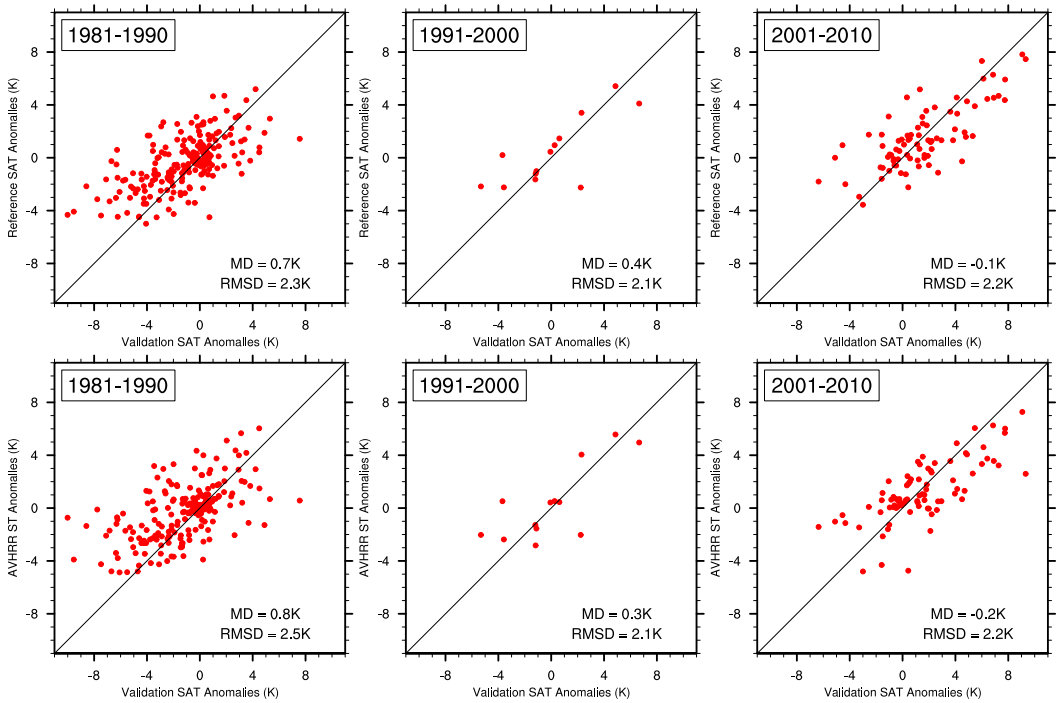

Figure 6.14: Decadal scatter plots of monthly anomalies (K) from the two interpolations compared to validation anomalies. A reference line is given showing x=y. The Mean Error and Root Mean Square Error values are given for each decade and for each interpolation.

were within the 95% confidence intervals for monthly average difference metrics. So, in general there is not a notable improvement in the accuracy of estimated anomalies over Arctic sea ice from adding AVHRR derived surface temperatures. But equally the results do show that the addition of this data equally did not lead to large decreases in accuracy.

The metric values for estimated anomalies from both interpolations are similar in size and seasonal pattern to those produced using a reanalysis testbed in Chapter 3. In contrast, the RMSD in anomalies estimated from in situ data using GSK (Chapter 4) were larger when compared to validation data over sea ice using very similar, or the same, data sources as used for validation in this study. These larger metric values were related to the use of a double difference statistic as noted in Chapter 4. The results of this additional satellite data study show that datasets such as IABP/POLES can be used as a climatology over Arctic sea ice areas, to allow the use of the Climate Anomaly Method. In addition, use of IABP/POLES introduced fewer errors into the calculation of difference metrics compared to the double difference statistic, even though notable biases are present.

The results of the study detailed in this chapter also show similarities with those from Chapter 5, which investigated the impact of adding supplemental input anomalies in a reanalysis testbed. Adding supplemental data often leads to an increase in the accuracy of estimated anomalies. Error metrics decreased by around 0.2(K) overall when additional data are added while larger decreases can occur for some areas and grid cells; up to around 1K for some grid cells in Chapter 5 while in this additional satellite data study some grid cell RMSEs decreased by up to 6.2K. Both studies utilised a relatively small sample of additional data sources for a limited time period so these increases in accuracy are

impressive. However, decreases in accuracy also occurred in these two studies. Additional data do not always increase accuracy as discussed in Chapter 5, however, this is generally noted for techniques such as Linear Interpolation which do not account for observation clustering, unlike GSK. Therefore decreases in accuracy in this chapter are more likely to result from uncertainties in the data sources used. But, the similarity of the changes in accuracy between the two additional data studies gives some confidence to the results detailed here, even though the monthly average metrics for both interpolations were not significantly different.

As mentioned previously, there are various uncertainties present in the data sources used for this study. The AVHRR derived data were re-normalised in order to produce monthly surface temperatures more consistent with in situ measurements (Comiso, 2014b). However, these data have some serious cloud masking problems, resulting from instrument limitations, which the re-normalisation technique will not correct and which may degrade the normalisation. Residual calibration and degradation problems may also remain after this re-normalisation. In addition, the AVHRR data were re-normalised using in situ data from Arctic meteorological stations from the WMO. Quality control was undertaken on these station records, but there may still be biases and inhomogeneities associated with these in situ records. Furthermore, these land-based stations may be located quite a distance from the AVHRR observations re-normalised over sea ice areas. This may impact the representativeness of these AVHRR surface temperatures over sea ice areas. Other uncertainties and issues may also be present in this dataset. The use of other satellite sources of TIR which are known to be relatively stable (e.g. Along Track Scanning Radiometers, Veal et al. 2013) or those which have fewer instrument limitations for cloud masking (e.g. Moderate Resolution Imaging Spectroradiometers, Comiso 2014b) should be investigated. The addition of data from other satellite sensors could lead to further, or larger, improvements in anomaly accuracy.

The in situ data sources used to provide validation data in this study may suffer from biases, inhomogeneities, noise and local effects. There are also uncertainties resulting from producing monthly average temperatures from less than a month of observations from these data sources. In addition, the climatology used for this study was IABP/POLES data between 1981 and 2004. The ARCSS-SAT analysis, which is preliminary and was not utilised in this study, is noted as extending and improving the IABP/POLES dataset (EOL, 2014). A positive bias was noted in the IABP/POLES dataset compared to the validation data used here and has been noted in previous studies. This bias may result from absorption of solar radiation, snow cover insulating the buoy and temperature sensor (Chen et al., 2002), and/or the incorporation of NCEP reanalysis data which is known to have a positive bias compared to in situ observations. Work is still ongoing to identify and account for all biases, inhomogeneities and other issues in the input data for these analyses (Rigor, 2014). There may also be biases and uncertainties resulting from the reanalysis dataset used as a first guess field for the IABP/POLES analysis as well as the optimal

interpolation method used. Other potential climatology datasets, for example reanalysis or satellite datasets and the finalised version of the ARCSS-SAT analysis, should be considered.

6.5 Summary

The addition of surface temperatures derived from the AVHRR instruments often increased the accuracy of anomalies estimated using GSK relative to independent validation data. The difference metrics were 0.2-0.1(K) smaller than those produced from the reference interpolation where increases in accuracy were observed. For some grid cells the accuracy increased by up to 6.2K. However, decreases in accuracy were also often noted and these could be up to 4.9K for some grid cells. Furthermore the differences were generally small.

These results show similarities with those from Chapter 5, a similar study using reanalysis data as a testbed. Adding supplemental data generally leads to an increase in accuracy although decreases in accuracy also occurred in these two studies. However, it must be noted that there are numerous uncertainties and limitations associated with the data utilised in this satellite data study.

Chapter 7

Summary and Conclusions

7.1 Importance of the Work

The Arctic is recognised as an important region in the study of climate change. However, monitoring Arctic temperature change is challenging, particularly in areas covered by sea ice for all or part of the year. How should the sea ice regions of the Arctic be treated in global and regional averages of surface temperature? This question guided the aims and objectives of this thesis.

7.2 Summary of the Work

Currently, the main temperature anomaly datasets estimate Arctic temperature change from Surface Air Temperature (SAT) and Sea Surface Temperature (SST) measurements using various techniques. In order to ascertain whether we can improve estimates of recent climate change, by altering how Arctic temperatures are monitored, we first need to quantify the uncertainties associated with these estimation techniques. Accordingly the first aim of this thesis was to:

1. Quantify the uncertainties in extrapolation and interpolation of information over the Arctic.

As the Arctic is a very data sparse region I decided to investigate the above aim initially using a reanalysis dataset as a testbed. The first objective of the first aim was:

(a) Examine how well the ECMWF (European Centre for Medium-Range Weather Forecasts) Re-Analysis (ERA) project interim reanalysis, or ERA-Interim, represents surface temperatures in the Arctic in order to determine whether ERA-Interim data are a suitable testbed for studying Surface Air Temperatures in this region.

I produced a review of the literature on reanalyses and their representation of Arctic temperatures (Section 2.2). According to the investigated literature, ERA-Interim

produces generally realistic estimates of Arctic temperature and variability trends (Section 2.2.6.3). This was despite a consistent warm bias of around 2K near the surface. ERA-Interim was also found to outperform other reanalyses available at the time of this study, in terms of biases and Root Mean Square Errors (RMSEs), for Arctic SATs (Section 2.2.7). Further to this I conducted an investigation into the representation of Arctic SAT Correlation Length Scales (CLS) in ERA-Interim (Section 2.3). I compared CLS derived from ERA-Interim SATs to reference CLS, determined in previous studies by Hansen and Lebedeff 1987 and Rigor et al. 2000. The annual and monthly ERA-Interim CLS were found to be close to, but slightly longer than, the reference annual and seasonal CLS (Sections 2.3.3 and 2.3.4). These longer CLS were anticipated; ERA-Interim data are gridded (unlike meteorological station data) and reanalysis data do not contain the inhomogeneities, uncertainties, noise and local effects associated with meteorological station records. On the basis of the literature review and the CLS investigation, I concluded that ERA-Interim is physically plausible for Arctic SAT anomalies and is a suitable reanalysis testbed to be utilised for the second objective:

(b) To investigate the impact of using various techniques to estimate Arctic surface air temperature anomalies.

I investigated the degree of difference arising from using five different techniques, based on existing temperature anomaly dataset techniques, to estimate Arctic SAT anomalies over land and sea ice using reanalysis data as a testbed (Chapter 3). The estimation techniques were applied to monthly SAT anomalies from ERA-Interim sampled at Arctic meteorological station locations. This yielded estimates of Arctic-average and spatially resolved SAT anomalies at several temporal resolutions. I compared these estimated anomalies to monthly SAT anomalies from ERA-Interim, which was the reference in this study. Techniques which interpolated anomalies were found to result in smaller errors than non-interpolating techniques relative to the reanalysis reference (Section 3.3). Kriging techniques provided the smallest errors in estimates of Arctic anomalies and Simple Kriging was often the best kriging method in this study, especially over sea ice. A linear interpolation technique had, on average, RMSEs up to 0.55 K larger than the two kriging techniques tested. Non-interpolating techniques provided the least representative anomaly estimates. Nonetheless, they serve as useful checks for confirming whether estimates from interpolating techniques are reasonable. I simulated the interaction of meteorological station coverage with estimation techniques between 1850 and 2011 using an ensemble dataset comprising repeated individual years (1979-2011). All techniques were found to have larger RMSEs for earlier station coverages (Section 3.4). This supports calls for increased data sharing and data rescue, especially in sparsely observed regions such as the Arctic. However, do we see similar results outside of a reanalysis testbed? Therefore, the third objective was to:

(c) Determine whether we can assess the impact of estimating Arctic surface air temperature anomalies with Global Simple Kriging using in situ data and whether it

corresponds with the results of objective (b).

I applied Global Simple Kriging (GSK), which was the most representative technique in the reanalysis study, to monthly SAT anomalies from Arctic meteorological stations (Chapter 4). I assessed the accuracy of the estimated anomalies produced using independent validation data. Independent meteorological stations were used over land and drifting platforms over sea ice. A 'double difference' statistic was used to estimate the accuracy of anomalies over sea ice. The accuracy of estimated anomalies over land were found in Section 4.3 to be very close to those from the reanalysis, both in seasonal pattern and size. Over sea ice the difference from the validation data from the two studies had a similar seasonal pattern, and size once the impacts of sample size and the double difference statistic were accounted for (Section 4.4). However, it should be noted that the confidence intervals are large and the sample size is very small so caution is needed in drawing quantitative conclusions from these data. There were also various uncertainties associated with the use of in situ validation data over sea ice, such as undersampling of monthly average temperatures, as noted in Section 4.5. Despite this the seasonal patterns and metric sizes from the two studies were similar and the two independent validation data sources over sea ice produced very similar results, despite providing a relatively small sample of data. This gives us confidence in using the results from the reanalysis testbed to quantify the uncertainties in extrapolation and interpolation of information over the Arctic, especially over land areas.

The second aim of this thesis was to:

2. Explore whether additional data sources would improve estimates of surface temperature change over the Arctic.

I investigated the use of supplemental data within a reanalysis testbed initially (Chapter 5). The additional input anomalies were monthly SAT anomalies from ERA-Interim sampled at supplementary data source locations over Arctic land, or sea ice, areas. Overall, adding data over land and sea ice led to an increase in accuracy within the reanalysis testbed. For Arctic-average anomalies the RMSE can decrease by up to 0.01K when additional data is added over land (Section 5.3.1), and by up to 0.1K when additional data is added over sea ice (Section 5.4.1). For spatially resolved anomalies with additional land data the RMSE decreases by around in general 0.1-0.2K, with decreases in RMSE of up to 0.6K (Section 5.3.2). For sea ice the decrease in RMSE is generally 0.2-0.4K, but can be up to 1K (Section 5.4.2). Substantial improvements in accuracy were associated with additional data sources located in the Arctic Ocean, such as over sea ice or on islands in sparsely observed regions. Furthermore, techniques which accounted for station clustering showed a more consistent improvement in accuracy. The results of this study and Chapter 4 suggest that the use of additional data sources would improve the estimation of Arctic SAT anomalies over land and sea ice from in situ data sources, assuming the additional data are of sufficient quality. As such the third aim of this thesis is to:

3. Identify appropriate data sources which could be used to monitor Arctic surface temperature change over Arctic sea ice.

In Appendix D I identified many surface temperature data sources over Arctic sea ice areas from in situ data sources, satellite instruments and reanalyses which could be employed for monitoring temperature changes in this region. Following this, and also in accordance with the second aim of this thesis, in Chapter 6 I explored whether the inclusion of surface temperatures over Arctic sea ice derived from the Advanced Very High Resolution Radiometer (AVHRR) satellite instruments would improve the accuracy of estimated SAT anomalies in this region. I also examined whether a gridded dataset of Arctic SATs could be used as a climatology, which would allow surface temperature anomalies to be calculated using the commonly employed Climate Anomaly Method in that chapter. This was prompted by the increase in error associated with the use of the double difference statistic in Chapter 4.

I applied GSK to in situ monthly SAT anomalies either from the CRUTEM4 databank exclusively, or combined with surface temperature timeseries derived from AVHRR. I compared the estimated anomalies from both interpolations to validation anomalies over sea ice areas. The validation anomalies were produced using a selection of in situ data sources identified in Appendix D, and the International Arctic Buoy Programme and NASA Polar Exchange at the Sea Surface program (IABP/POLES) analysis as a reference climatology. Often there was a slight increase in accuracy when surface temperatures derived from the AVHRR instruments were utilised. Overall the AVHRR interpolation estimated anomalies have slightly smaller biases and difference metrics than those from the reference interpolation (Section 6.3). Furthermore, the monthly difference metrics are 0.2-0.1(K) smaller for half of calendar months with consistently better anomaly accuracy in boreal summer months. But, decreases in accuracy were noted nearly as often as increases in accuracy and differences between the estimated anomalies were small. There is not a notable improvement in estimated anomalies over Arctic sea ice from adding AVHRR derived surface temperatures, but likewise it did not result in large decreases in accuracy. It is likely that uncertainties in AVHRR data, such as resulting from cloud masking limitations, impacted the results of this study. More notable improvements in accuracy may be accomplished by using temperature data derived from other satellite sensors. However, these results show similarities with those from Chapter 5 outside of the reanalysis testbed which gives some confidence to the results of these studies.

The final aim of this thesis is to:

4. Provide some recommendations on how to monitor temperature changes in the Arctic, especially over sea ice areas.

These recommendations are given in Section 7.4.

7.3 Research Limitations

As with all research, limitations exist in the data and methods used for this thesis. Some limitations are chapter specific, and are noted within the relevant chapter, whereas others are more general. Here I outline and discuss the general limitations of this thesis.

7.3.1 Data Sparseness

The challenge of representing the Arctic in global surface temperature time-series given that in situ measurements of Arctic temperatures are both spatially and temporally sparse, and the records are often short, was the catalyst for this thesis. It was also a limitation in exploring how the sea ice regions of the Arctic should most appropriately be treated in global and regional averages of surface temperature. In some chapters this limitation was overcome. In Chapters 3 and Chapter 5 I utilised spatially and temporally complete ERA-Interim reanalysis data as a testbed due to the sparseness of in situ observations. Comparing these with similar studies utilising in situ data assessed the gap between the reanalysis and the reality and gave us more confidence in the results of those chapters.

For some chapters the limitation of data sparseness could be mitigated, but not overcome. In Chapter 4 I used a longer time period of interest (compared to Chapter 3) so that sufficient sea ice data were available for the validation. In addition, I employed an equal area grid, instead of using the same grid as ERA-Interim, as equal angle grids exacerbate data sparseness. However, the validation data was still temporally and spatially sparse and monthly average temperatures were generally undersampled. In Chapter 5 these mitigation techniques could not be employed. So, in order to include enough additional data to see an impact in the estimated anomalies, I decided that if there was at least one NPDS record in a grid cell, month and year then it would be assumed that a monthly average could be computed. This is, however, a limitation as it is a compromise between the need to use data and techniques that are as close as possible to a previous chapter and having enough extra input data to investigate the impact of adding additional data.

The limitation of surface temperature data sparseness for the Arctic may be further mitigated or overcome in the future through data sharing and rescue efforts, the use of satellite sensor derived temperatures and/or the use of reanalysis.

7.3.2 Data Accessibility and Quality

A further limitation was data accessibility and quality. In Chapters which used in situ data sources over sea ice (Chapter 4 and 6) data accessibility and quality limitations were mitigated though alternative data choices. In Chapter 4 I considered using International Arctic Buoy Programme (IABP) data, but individual observations from these buoys are no longer made available on the IABP website. This is a response to data quality limitations identified in IABP records. Therefore I decided to use NPDS and Ice Mass Balance

Buoy (IMB) data from the U.S Army Cold Regions Research and Engineering Laboratory (CRREL) as validation data sources. In addition to being accessible these data are radiation shielded, which is important for temperature data quality; temperature sensors may suffer from solar heating effects if not radiation shielded, leading to errors in their observations. Automatic sensors, such as IMBs, may also be subject to bias problems resulting from rime ice and snow cover as they are not subject to routine maintenance. For this reason, I selected only manned drifting platforms as validation data in Chapter 6. However, data accessibility and quality limitations remain. For example, much of the CRREL IMB data is preliminary and may not be quality controlled.

Looking forward, the trend in science is to make data more widely accessible. There are currently efforts by many initiatives (see Section 7.4.2) to increase sharing, and rescue, of temperature data. This allows more people to work on the data and will help to mitigate and overcome data accessibility limitations in the future. However, increasing data sharing and accessibility can be difficult for many reasons and often people will want to see the benefits before sharing data. Efforts are also being made to quantify and document data quality, which is also very important for research and a particular concern in the Arctic. Conditions in this region are harsh and observation platforms are difficult to maintain. Furthermore, the wide variety of types of sensors deployed in the Arctic make the question of data quality for in situ observations in this region more complex. Hall et al. 2015 have highlighted a need for discussions and the development of protocols for measuring air and surface temperatures in the Arctic. Increased data accessibility would enable intercomparison of data from various sensors and platforms observing temperatures over sea ice and allow biases to be assessed.

7.3.3 The Use of a Single Reanalysis Dataset as a Testbed

ERA-Interim data were used as a testbed for Chapters 3 and 5. However, as reanalysis data can contain biases, uncertainties and errors, repeating the analyses in the above chapters would strengthen the analysis and conclusions of these chapters. Using multiple reanalysis testbeds would ascertain whether the results are sensitive or insensitive to the choice of reanalysis. In addition, the use of other reanalyses would improve the Historical Coverage investigation detailed in Section 3.4 and help provide a more robust investigation. Using other reanalyses as testbeds would strengthen the analysis by ascertaining the sensitivity of results to the reanalysis chosen. Also, these data could be used in combination with ERA-Interim to produce a larger ensemble of anomaly errors. Furthermore, using a 20^{th} century reanalysis, such as ERA-20C would allow us to investigate the impact of changing station coverage for actual 1850-1950 temperature anomaly patterns rather than simulating the impact using 1979-2011 temperature anomalies. These steps could be used to overcome this limitation and produce a more robust investigation.

7.4 Recommendations for Monitoring Temperature Changes in the Arctic over Sea Ice and Future Work

In this section I provide some recommendations and ideas for how to monitor temperature changes over Arctic sea ice areas, as well as identifying areas of further research. The focus is on sea ice areas with a sea ice concentration of more than 15%. This is due to the focus of this thesis and because data sources in these regions, which can cover a large fraction of the Arctic, are not generally employed in temperature anomaly datasets. This does not, however, mean that other Arctic areas are observed well enough not to be considered. Surface temperature data are temporally and spatially sparse throughout the Arctic. As a result I support data sharing and data rescue efforts for this region. In particular, monitoring temperature changes in areas of open ocean, seasonal sea ice areas, and marginal ice zones warrants further study. In common with areas of sea ice they are very sparsely observed, often treated differently between temperature anomaly datasets, and are experiencing rapid changes as a result of climate change.

7.4.1 Estimation Techniques

It was found in Chapter 3 that the choice of estimation techniques had an impact on the accuracy of estimated Arctic SAT anomalies over land and sea ice areas. Some techniques, such as kriging techniques, were found to result in smaller errors than other techniques, such as non-interpolating techniques. Furthermore, the estimation techniques are impacted differently when additional data are added; techniques which accounted for clustering were more robust and showed more consistent improvements in accuracy (Chapter 5). Similar results were noted when in situ data were used (Chapters 4 and 6). These studies give us an estimate of the uncertainties in estimating Arctic SAT anomalies using the investigated techniques. However, despite these results, I do not recommend the sole use of any one interpolation technique for generating timeseries of Arctic surface temperature change. Using different techniques to estimate anomalies allows us to investigate structural uncertainty in estimates of climate change.

These estimation technique studies could be extended to look at other estimation techniques; parametric, observational and structural uncertainty; and the impact of using monthly or seasonal covariance values. These areas of future work would help identify other valuable estimation techniques, further quantify uncertainties in estimation techniques, and may lead to improvements in the representation of the Arctic in global surface temperature timeseries. The inclusion of monthly or seasonal weighting functions is of particular interest as CLS in Arctic surface temperatures are shorter in boreal summer than in other seasons (see Section 2.3) and this is not accounted for in current temperature anomaly dataset methods. Furthermore the CLS also differ between surface types in

the Arctic (Rigor et al., 2000).

7.4.2 Surface Temperature Data in Sea Ice Areas

In situ data sources, from both manned and unmanned platforms, provide the longest records of surface temperatures over Arctic sea ice compared to other potential data sources. These data sources have been included in Arctic only datasets, but are not included in global temperature anomaly datasets. This is a result of the complexities involved in calculating SAT change from sparse, short, data records measured on drifting platforms. Yet Chapter 5 and 6 show that the addition of data sources over Arctic sea ice can improve the accuracy of estimated anomalies. The increases in accuracy noted in those studies were small compared to variability and trends in Arctic temperatures in recent decades. However, those chapters only used relatively small amounts of data so the estimates of increases in accuracy are likely conservative. These studies begin to assess the impact of using additional sources of temperature data in these areas, but further investigation is needed to quantify the influence of these data as well as how they could be included in temperature anomaly datasets.

Anomalies could be produced from in situ observations of surface temperatures using the Climate Anomaly Method by employing a gridded climatology. Existing datasets could be used as a climatology, for example: interpolated datasets such as IABP/POLES (Rigor et al., 2000) and ARCSS-SAT (EOL, 2014); reanalysis datasets such as ERA-Interim; or climate quality records of satellite derived surface temperatures. Further research should be conducted into the utilisation of existing datasets as a climatology over Arctic sea ice areas. A new climatology dataset could also be created by combining various data sources. One way this could be accomplished is by combining in situ observations from various data sources with climate quality satellite datasets and/or physically plausible reanalysis datasets using a technique which appropriately weights the contribution of each data source, such as optimal estimation. Once a climatology is chosen and an analysis is created it can then be improved and iterated upon. The climatology and analysis can be compared and contrasted with other datasets, and this can guide the incorporation of improved methods, further data sources and the identification of any data quality problems. In addition to the Climate Anomaly Method, further research could explore whether other methods could be utilised for producing estimates of temperature change in the Arctic, such as the Reference Station Method (Hansen and Lebedeff 1987, perhaps using satellite derived or reanalysis data as a 'reference station') or the First Difference Method (Peterson et al., 1998). Given the reliability of some in situ observations (e.g. NPDS, Rigor et al. 2000) and the potential improvement in accuracy of estimated anomalies in the Arctic, I recommend efforts to include in situ data sources over Arctic sea ice in global temperature anomaly datasets.

To increase the amount of in situ data available, as mentioned previously, I encourage data sharing and data rescue for the Arctic through initiatives such as the International

Surface Temperature Initiative (ISTI), the international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, and the Canadian historical data typing project (Allan et al., 2011; Thorne et al., 2011; Slonosky, 2014). To facilitate this I believe that there is the need for production of an Arctic sea ice SAT databank. This is within the scope of the ISTI (Thorne et al., 2011), which is a databank for in situ observations of surface meteorological variables and associated quality control information and metadata at several temporal resolutions. ISTI already includes some SAT data over Arctic sea ice and it is an established initiative, which would help facilitate Arctic data exchange. Ideally efforts would also be made to produce uncertainty estimates for these data. Further to this, I echo the conclusions of Hall et al. 2015; there is a need for discussions within the international scientific community about developing protocols for measuring air and surface temperatures in the Arctic, as has happened for land- and ocean-based observing systems used for climate monitoring. Also of benefit would be the extension of the work in Chapter 3 and Chapter 5 to create a map which identifies the most important regions in terms of including additional data. This could be accomplished by identifying each grid cell which is not used for input anomalies in Chapter 3, adding each of these grid cells in turn into one of the kriging estimation techniques (as these are most representative and not influenced by station clustering), and analysing the results. This would produce a map showing the change in RMSE resulting from adding each grid cell to the interpolation. This would identify the most important data sources currently available, help tailor data sharing and rescue initiatives, and identify key areas for future observing networks and in situ sensor deployment.

However, even with data sharing and data rescue efforts, in situ measurements in the Arctic will be sparse. Therefore I recommend using observations from other sources, such as satellites and reanalysis, both in combination with in situ data in this region and independently. Satellite-derived surface temperature data or reanalysis temperature estimates could be used to infill gaps in temperature records from in situ data, as a first guess field such as employed for ARCSS-SAT (EOL, 2014), or, as mentioned before, as a gridded climatology for in situ data over sea ice. I also recommend using reanalysis and satellite datasets independently of in situ observations in order to provide additional, and in some cases independent, datasets of temperature anomalies over the Arctic. Comparing and contrasting datasets produced using different methods and data sources in this region will be very informative. Finally, it must be noted that the use of satellite and reanalysis data is likely to become increasingly important as the reduction in sea ice extent and thickness both increases the difficulty of deploying manned and unmanned measuring platforms and shortens the length of time these sensors can operate on ice.

7.5 Concluding Remarks

Monitoring surface temperature change in the Arctic is an important, yet complex, area of research. Temperature changes in this region are accelerated compared to those at lower latitudes and will have local, regional and global impacts on the atmosphere, cryosphere, hydrosphere and biosphere. But the observing network in this area - in particular over the Arctic Ocean - is sparse, records are short, and conditions for instruments deployed here are harsh. In this thesis I have shown that using certain estimation techniques and additional data sources can improve our estimation of temperature changes in this region. If these methods and data are incorporated into global and Arctic temperature anomaly datasets they will help to quantify temperature changes in, and improve our understanding of, this dynamic and rapidly changing polar region.

Appendix A

Acronyms

Acronym	Meaning
20CR	NOAA-CIRES 20^{th} Century Reanalysis
3D-Var	Three-dimensional Variational analysis
4D-Var	Four-dimensional Variational analysis
AARI	Arctic and Antarctic Research Institute
AATSR	Advanced Along-Track Scanning Radiometer
ACRE	Atmospheric Circulation Reconstructions over the Earth initiative
ADEOS	Advanced Earth Observing Satellite
AIDJEX	Arctic Ice Dynamics Joint Experiment
AIRS	Atmospheric Infra-Red Sounder
AMSR(-E)	Advanced Microwave Scanning Radiometer (for EOS)
AMSU(-A/B)	Advanced Microwave Sounding Unit (-A/B)
AOE	Arctic Ocean Experiment
APLIS	Applied Physics Laboratory Ice Station
AR5	IPCC Fifth Assessment Report
ARCSS	Arctic System Science program
ARCSS-SAT	ARCSS Surface Air Temperature analysis
ARLIS	Arctic Research Laboratory Ice Station
ASCOS	Arctic Summer Cloud-Ocean Study
ASR	Arctic System Reanalysis
AT	Atmospheric Temperature
ATMS(-E)	Advanced Technology Microwave Sounder (for EOS)
ATS	Application Technology Satellite
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BROMEX	BRomine, Ozone, and Mercury EXperiment
CAM	Climate Anomaly Method
CEAREX	Coordinated Eastern Arctic Research Experiment
COADS	Comprehensive Ocean-Atmosphere Data Set

COBE-SST	Centennial in situ Observation-Based Estimates of the variability of SSTs and marine meteorological variables dataset
CRD	Compound Relative Difference
CRD*dd*	Compound Relative *dd* (Double Difference)
CRE	Compound Relative Error
CrIS	Cross-track Infrared Sounder
CRREL	U.S Army Cold Regions Research and Engineering Laboratory
CRU	Climatic Research Unit of the University of East Anglia
CRUTEM	CRU and Met Office Hadley Centre land surface temperature anomaly dataset
ECMWF	European Centre for Medium-Range Weather Forecasts
DARMS	Drifting Automatic Radiometeorological Stations
DMI	Danish Meteorological Institute
DOE	U.S Department of Energy
EASE-Grid	Equal-Area Scalable Earth grid
EOL	Earth Observing Laboratory
EOS	Earth Observing System
ERA-15	ECMWF 15 year Reanalysis
ERA-20C	ECMWF 20^{th} Century atmospheric Reanalysis
ERA-40	ECMWF 40 year Reanalysis
ERA-Clim	ECMWF Reanalysis of Global Climate Observations project
ERA-Interim	ECMWF Interim Reanalysis
ERS	European Remote-sensing Satellite
ERSST	Extended Reconstructed Sea Surface Temperature dataset
ETM+	Enhanced Thematic Mapper Plus
FDM	First Difference Method
FGGE	First Global Atmospheric Research Program Global Experiment
FMI	Finnish Meteorological Institute
GCM	General Circulation Model
GCOM	Global Change Observation Mission for water
GEOS	Goddard Earth Observing System
GHCNCAMS	Global Historical Climatology Network and the Climate Anomaly Monitoring System
GHCN-M	Global Historical Climatology Network-Monthly temperature dataset
GISS	Goddard Institute for Space Studies
GISTEMP	NASA GISS Surface Temperature Analysis
GLI	Global Imager
GOES	Geostationary Operational Environmental Satellite
GOK	Global Ordinary Kriging

GSK	Global Simple Kriging
HadAT	Met Office Hadley Centre Atmospheric Temperature dataset
HadCET	Met Office Hadley Centre Central England Temperature dataset
HadCRUT	Met Office Hadley Centre and CRU global temperature anomaly dataset
HadISST	Met Office Hadley Centre's interpolated SST and sea ice concentration dataset
HadMAT	Met Office Hadley Centre Marine Air Temperature dataset
HadNMAT	Met Office Hadley Centre Nighttime only Marine Air Temperature dataset
HadSST	Met Office Hadley Centre SST dataset
HIRS	High-resolution Infrared Sounder
HRIR	High Resolution Infrared Radiometer
IABP	International Arctic Buoy Programme
IABP/POLES	IABP and NASA Polar Exchange at the Sea Surface program
IAOE	International Arctic Ocean Expedition
IARC	International Arctic Research Center
IASI	Infrared Atmospheric Sounding Interferometer
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IKFS	Infrared Fourier Spectrometer
IMB	Ice Mass Balance Buoy
IMG	Interferometric Monitor for Greenhouse gases
INSAT	Indian National Satellite
IOEB	Ice-Ocean Environmental Buoy
IPCC	Intergovernmental Panel on Climate Change
IRAS	Infra Red Atmospheric Sounder
IRIS	InfraRed Interferometer Spectrometer
ISPD	International Surface Pressure Databank
IST	Ice Surface Temperature
ISTI	International Surface Temperature Initiative
ITPR	Infrared Temperature Profile Radiometer
IUK	Iterative Universal Kriging radiosonde analysis project
JAMI	Japanese Advanced Meteorological Imager
JAMSTEC	Japan Agency for Marine-Earth Science and Technology
J-CADs	JAMSTEC Compact Arctic Drifters
JDCAS	JMA Climate Assimilation System reanalysis
JMA	Japanese Meteorological Agency
JPSS	Joint Polar Satellite System
JRA-25	Japanese 25 year Reanalysis
JRA-55	Japanese 55 year Reanalysis

LeadEx	Lead Experiment
LI	Linear Interpolation
LST	Land Surface Temperature
LSWT	Lake Surface Water Temperature
MAD	Mean Absolute Difference
MAD*dd*	Mean Absolute *dd* (Double Difference)
MAT	Marine Air Temperature
ME	Mean Error
MERRA	NASA Modern Era-Retrospective Analysis for Research and Applications dataset
MLOST	NOAA NCDC Merged Land-Ocean Surface Temperature Analysis
MODIS	MODerate-resolution Imaging Spectroradiometer
MRIR	Medium Resolution Infrared Radiometer
MSU	Microwave Sounding Unit
MSU-MR	Multispectral scanning imager-radiometer
MTS-1/2	Micro-Wave Temperature Sounder - 1/2
MVISR	Multichannel Visible Infrared Scanning Radiometer
MW	Microwave
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCDC	NOAA National Climatic Data Center
NCEP	National Centers for Environmental Prediction
NCEP CFSR	NCEP Climate Forecast System Reanalysis
NCEP R1	NCEP-NCAR Reanalysis
NCEP R2	NCEP/DOE Reanalysis
NI	Not Interpolating
NMAT	Nighttime only Marine Air Temperature
NOAA	National Oceanic and Atmospheric Administration
NOAA-CIRES	National Oceanic and Atmospheric Administration - Cooperative Institute for Research in Environmental Sciences
NOCv2.0	National Oceanography Centre Surface Flux and Meteorological Data set v2.0
NPDS	AARI North Pole Drifting Stations
NPEO	North Pole Environmental Observatory
NSIDC	National Snow and Ice Date Center
OCTS	Ocean Color and temperature Scanner
OSTIA	Operational Sea Surface Temperature and Sea-Ice Analysis
PATMOS-X	AVHRR Pathfinder Atmospheres - Extended
PAWS	Polar Automatic Weather Station
PMEL	Pacific Marine Environmental Laboratory

PPE	Percentage of ensemble members with Positive Errors
RAOBCORE	RAdiosonde OBservation COrrection using REanalyses
RATPAC	Radiosonde Atmospheric Temperature Products for Assessing Climate
RICH	Radiosonde Innovation Composite Homogenization
RMSD	Root Mean Square Difference
RMSD*dd*	Root Mean Square *dd* (Double Difference)
RMSE	Root Mean Square Error
RSM	Reference Station Method
RSS	Remote Sensing Systems
SAT	Surface Air Temperature
SDD	Standard Deviation of Difference
SDD*dd*	Standard Deviation of *dd* (Double Difference)
SDE	Standard Deviation of Error
SEVIRI	Spinning Enhanced Visible Infra-Red Imager
SHEBA	Surface Heat Budget of the Arctic Ocean
SI-GDR	Spectrometer/Interferometer - German Democratic Republic
SIMBA	Scottish Association for Marine Science Ice Mass Balance Array Buoys
SIRS-A/B	Satellite Infra-Red Spectrometer - A/B
SLSTR	Sea and Land Surface Temperature Radiometer
SMMR	Scanning Multichannel Microwave Radiometer
SMS	Synchronous Meteorological Satellite
SR	Scanning Radiometer
SSM/I	Special Sensor Microwave - Imager
SSM/T	Special Sensor Microwave - Temperature
SSMIS	Special Sensor Microwave - Imager/Sounder
SST	Sea Surface Temperature
SSU	Stratospheric Sounding Unit
ST	Surface Temperature
S-VISSR	Stretched Visible-Infrared Spin Scan Radiometer
TES-Nadir	Tropospheric Emission Spectrometer - Nadir
THIR	Temperature-Humidity Infrared Radiometer
TIR	Thermal Infrared
TIROS	Television and Infra-Red Observation Satellite
TLT	Lower Tropospheric Temperature
TM	Thematic Mapper
TOVS	TIROS Operational Vertical Sounder
TRMM	Tropical Rainfall Measuring Mission
UAH	University of Alabama in Huntsville

UAV	Unmanned Aerial Vehicle
VAS	Visible-Infrared Spin Scan Radiometer Atmospheric Sounder
VHRR	Very High Resolution Radiometer
VIIRS	Visible/Infrared Imager Radiometer Suite
VIRS	Visible and Infra Red Scanner
VTPR	Vertical Temperature Profile Radiometer
WMO	World Meteorological Organization
WRF	Weather Research and Forecasting model

Appendix B

Kriging

Kriging is used to estimate the value of a variable at an unsampled point where the covariance or semivariance function as a function of distance is known. This appendix describes the equations and variables used for both kriging methods in this study.

The value of the variable of interest t at an unsampled coordinate x_o is estimated using a linear combination of observed values of t at coordinates x_i, $i = 1...n$ calculated using Eq. (B.1). $\mathbf{w}$ is a vector containing the optimal weights $w(x_i)$ for each observation of variable t at coordinates x_i, $i = 1...n$. $\mathbf{t}$ is a vector containing the observations of variable t at coordinates x_i, $i = 1...n$.

$$\hat{t}(x_o) = \mathbf{w} \cdot \mathbf{t} \tag{B.1}$$

$$\text{where: } \mathbf{w} = \begin{bmatrix} w(x_1) \\ ... \\ w(x_n) \\ \phi \end{bmatrix} \text{ and } \mathbf{t} = \begin{bmatrix} t(x_1) \\ ... \\ t(x_n) \\ \phi \end{bmatrix}$$

The optimal weights in vector $\mathbf{w}$ are calculated using Eq. (B.2). $\mathbf{A}$ is a matrix of the expected covariance between each pair of observations of variable t at coordinates x_i, $i = 1...n$ so that $a(x_i, x_n)$ is the expected covariance between $t(x_i)$ and $t(x_n)$. $\mathbf{B}$ is a matrix of the expected covariance between observations of variable t at coordinates x_i, $i = 1...n$ and the output point x_0 so that $b(x_o, x_n)$ is the expected covariance between $t(x_o)$ and $t(x_n)$.

$$\mathbf{w} = \mathbf{A}^{-1} \cdot \mathbf{B} \tag{B.2}$$

Two methods of kriging were investigated in this study: Global Ordinary Kriging and Global Simple Kriging. The optimal weights in vector $\mathbf{w}$ were calculated slightly differently depending on the method of kriging used.

For 'Ordinary' kriging, where the mean is unknown a priori, the mean is calculated during interpolation by constraining the optimal weights so that they sum to 1. For

Ordinary kriging the optimal weights in vector **w** are therefore calculated using Eq. (B.2) with the vectors and matrices shown below. ϕ is a Lagrange multiplier required for solving the equations.

$$\begin{bmatrix} w(x_1) \\ ... \\ w(x_n) \\ \phi \end{bmatrix} = \begin{bmatrix} a(x_1,x_1) & ... & a(x_1,x_n) & 1 \\ ... & ... & ... & 1 \\ a(x_n,x_1) & ... & a(x_n,x_n) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} b(x_o,x_1) \\ ... \\ b(x_o,x_n) \\ 1 \end{bmatrix}$$

For 'Simple' kriging which assumes that the mean is known a priori the optimal weights do not need to be constrained to sum to 1. Instead the mean is added to the dot product of the vector of optimal weights **w** and the vector of the observations of variable t, **T** to produce an estimate of $t(x_o)$. Note that, due to the use of a mean of 0 for Global Simple Kriging in this study, the same equation (Eq. (B.1)) is used for both kriging methods. Therefore the optimal weights in vector **w** are calculated using Eq. (B.2) with the vectors and matrices shown below.

$$\begin{bmatrix} w(x_1) \\ ... \\ w(x_n) \end{bmatrix} = \begin{bmatrix} a(x_1,x_1) & ... & a(x_1,x_n) \\ ... & ... & ... \\ a(x_n,x_1) & ... & a(x_n,x_n) \end{bmatrix}^{-1} \cdot \begin{bmatrix} b(x_o,x_1) \\ ... \\ b(x_o,x_n) \end{bmatrix}$$

The expected covariance is calculated using a model function, which approximates the covariance as a function of distance, determined from available observations as follows. An experimental semivariogram was produced by calculating the semivariance from the average difference between each pair of this study's input anomalies separated by binned distances. 25 km bins were used. The experimental semivariogram was plotted and it was determined that a spherical model would provide the best model fit to the experimental semivariogram. The model function is given in Eq. (B.3) with fitted parameter values of 0 for s, 7.6 for y and 3585.9 for ϵ . d is the distance in km and $\gamma(d)$ is the model semivariance at distance d.

$$\gamma(d) = \begin{cases} s + y[\frac{3}{2}(\frac{d}{\epsilon}) - \frac{1}{2}(\frac{d}{\epsilon})^3] & \text{for } 0 < d < \epsilon \\ s + y & \text{for } d \geq \epsilon \end{cases} \tag{B.3}$$

This model function fitted to the semivariogram was used to calculate the expected covariances in this study by calculating the expected semivariances first and then calculating the expected covariances from the expected semivariances. The expected covariances were then used in Eq. (B.2) to calculate the weights required for kriging.

Appendix C

Supplemental Material from Chapter 3

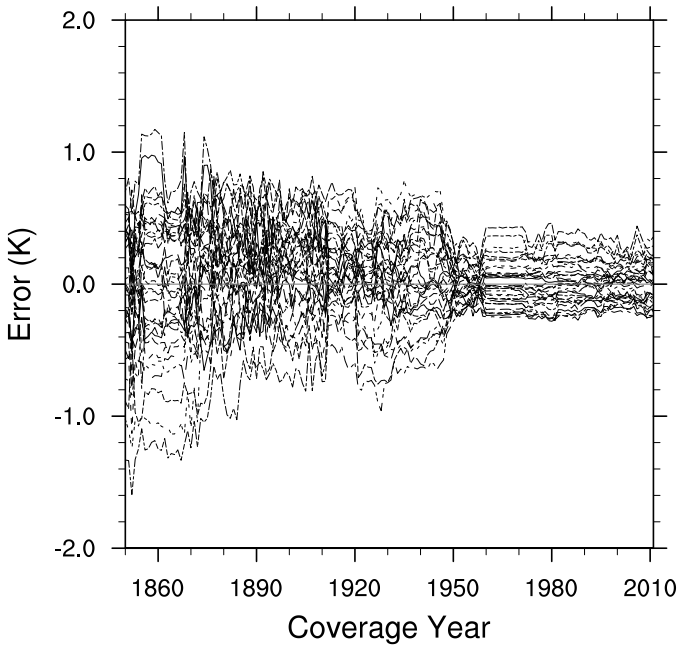

Figure C.1: The error in annual Arctic-average anomalies estimated by using Global Ordinary Kriging on each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

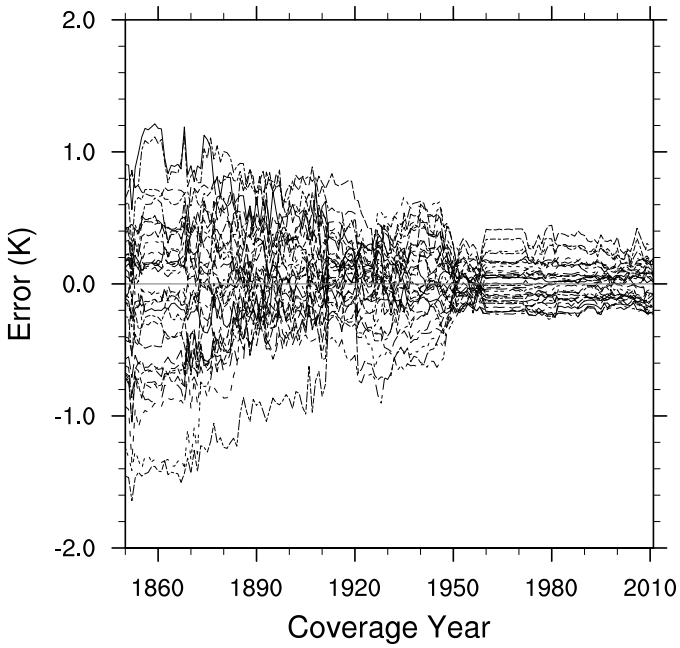

Figure C.2: The error in annual Arctic-average anomalies estimated by using Global Simple Kriging on each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

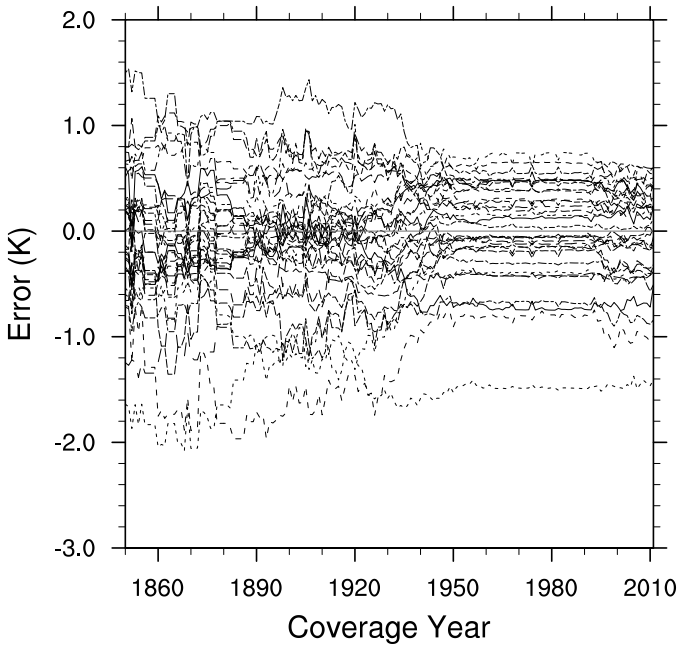

Figure C.3: The error in annual Arctic-average anomalies estimated by using the Binning technique on each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

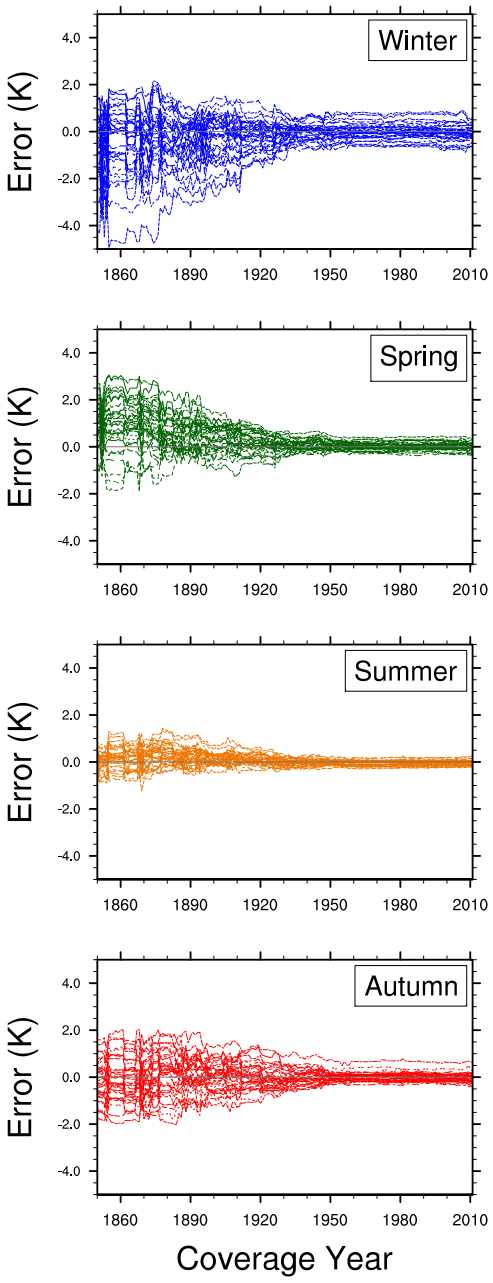

Figure C.4: The error in seasonal Arctic-average anomalies estimated by Linearly Interpolating each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

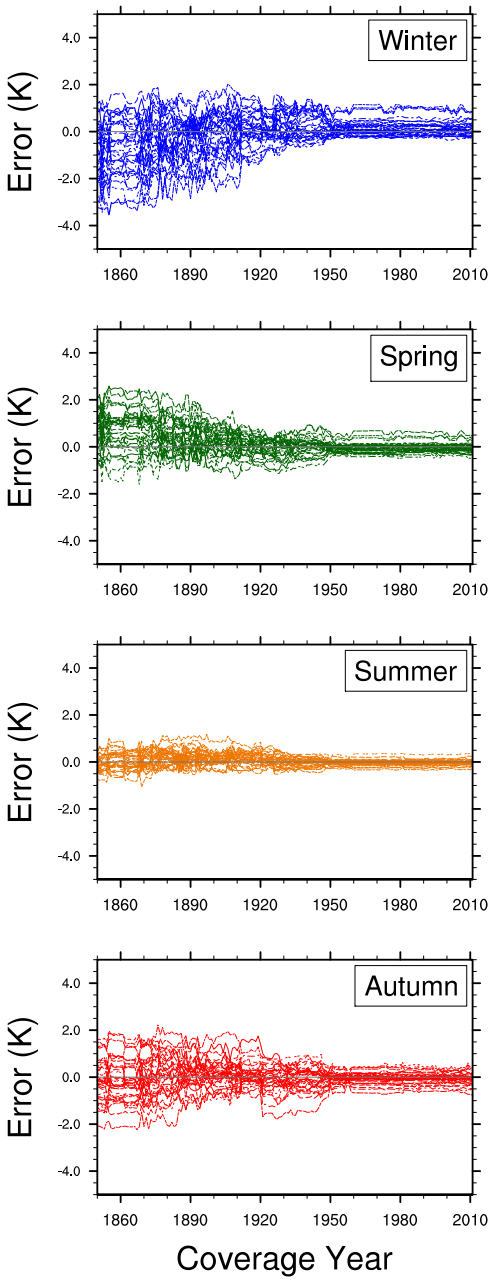

Figure C.5: The error in seasonal Arctic-average anomalies estimated by using Global Ordinary Kriging on each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

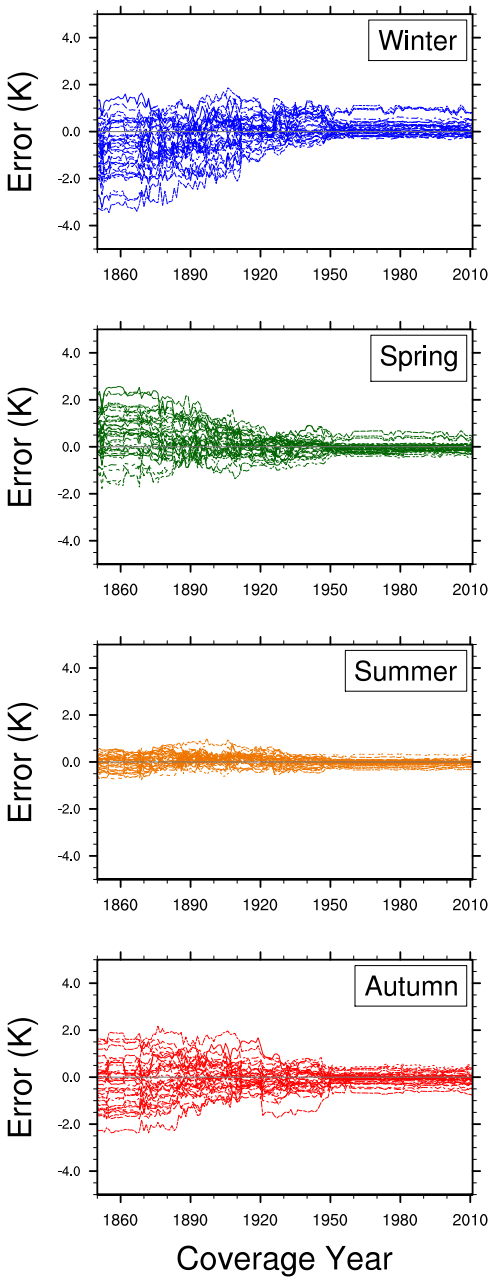

Figure C.6: The error in seasonal Arctic-average anomalies estimated by using Global Simple Kriging on each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

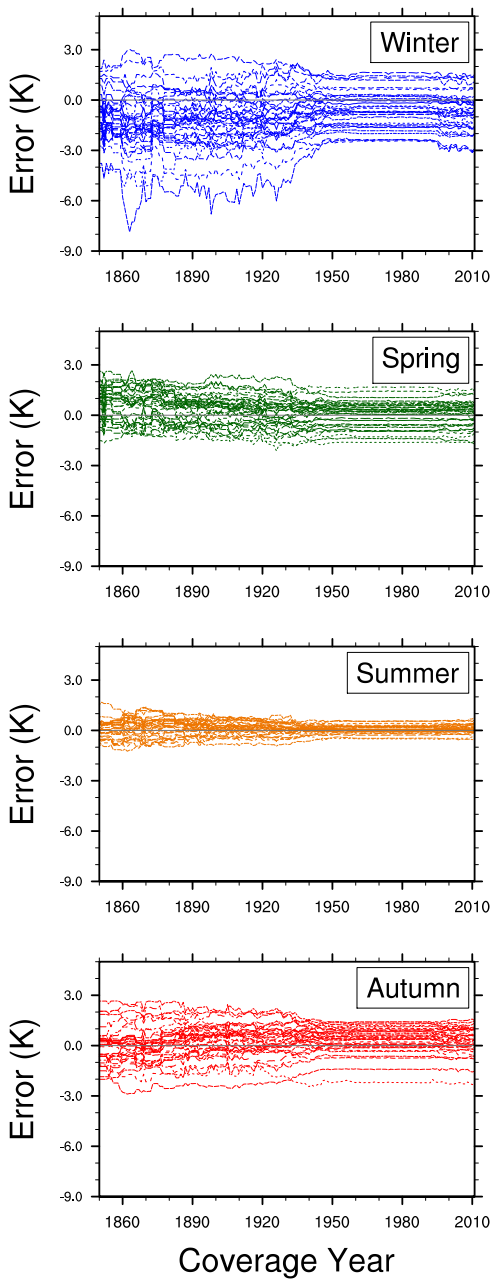

Figure C.7: The error in seasonal Arctic-average anomalies estimated by using the Binning technique on each year of ERA-Interim anomalies (1979-2011, each year is shown by one line) using historical station coverages (1850-2011).

Appendix D

Potential Sources of Surface Air Temperature Data in the Arctic

Table D.1: Potentially suitable sources of further SAT data over sea ice areas in the Arctic from in situ observations, satellite sensors and reanalysis datasets.

In Situ						
Data Source	Time Period	Variable/s	Name/Organisation	Dataset/s?	Issues	Notes
Ships and Buoys	1850–	MAT, SST	Collated in ICOADS (Woodruff et al., 2011) and the Met Office Marine Data Bank (Parker et al., 1995)	HadMAT1, HadISST and HadISST2 (Rayner et al., 2003); HadNMAT2 (Kent et al., 2013); HadSST3 (Kennedy et al., 2011b,c); ERSSTv3b (Smith et al., 2008); COBE-SST (Ishii et al., 2005); NOCv2.0 (Berry and Kent, 2009).	Daytime solar heating issues (MAT). Observations measured over ocean areas not sea ice.	
Ice Bound Ships and Icebreakers	1879-1881	SAT	USS Jeanette	Will be included in ICOADS.		Being digitised by oldWeather (Brohan, 2013).
	1893-1896		Fram	Arctic meteorology and climate atlas (Environmental Working Group, 2000).		
	1922-1924		Maud			
	1937-1940		Georgiy Sedov	May be in ship and buoy datasets.		(Marchenko, 2012; Zubov, 1939).
	1979–		CCGS Amundsen			
	1984–		RV Polarstern			
	1988–		Oden			
	1999–		USCGC Healy			
	There may be additional, similar ships and icebreakers not identified here					
Manned Drifting Ice Stations	1937-1991, 2003-2013	SAT	AARI NPDS	Included in IABP/POLES, ARCSS-SAT (Rigor et al., 2000; EOL, 2014), and the EWG 2000.	Hiatus after the dissolution of the Soviet Union.	Reliable in situ SATs (Rigor et al., 2000)
	1951-1965		US Drifting Stations (Polar Ice Pack Station, T-3, Alpha, Bravo, Charlie, ARLIS I to IV)	Arctic meteorology and climate atlas (EWG, 2000).		(Smith, 1971)

	1970-1978		AIDJEX Station			(Untersteiner et al., 2007)
	1978–		APLIS			(IARC, 2012).
	1991, 1996, 2001, 2008		IAOE 1991; AOE 1996, 2001; ASCOS		SAT at 1.3m height (Leck et al., 2001, 1996; Tjernström et al., 2005).	(Tjernström, 2005; Tjernström et al., 2014)
	1997-1998		Ice Station SHEBA			(Uttal et al., 2002)
	2002–		NOAA PMEL, at Ice Station Borneo/Barneo (NOAA, 2000)			2010-2014 were IABP PAWS buoys.
	2006-2007		Ice Station Tara			(Vihma et al., 2008)
	2012		BROMEX			(Hall et al., 2015)
	There may be additional, similar expeditions not identified here					
DARMS	1957-1976	SAT	AARI	Available in the Arctic meteorology and climate atlas (EWG, 2000).	Limited range and accuracy of location triangulation (Untersteiner and Thorndike, 1982).	(Olenicoff 1968, cited in Untersteiner and Thorndike 1982).
Drifting Buoys	1970-1978	SAT	AIDJEX Buoys			(Untersteiner et al., 2007)
	1979–		IABP	IABP/POLES 1979-2004 (Rigor et al., 2000), ARCSS-SAT 1979-2008 (EOL, 2014)	Trying to quantify biases (Rigor, 2014). Snow cover on buoys (Chen et al., 2002).	Raw IABP data not available online. ARCSS-SAT is preliminary.
	1992-1998		IOEB			(Honjo et al., 1995)
	2000-2009		NPEO, J-CADS and PMEL Buoys			(Honjo et al., 1995)
	2008-2009		Hamburg Arctic Buoy Drift Experiment			
	2009–		O-Buoys		Primarily a chemical sensor (Knepp et al., 2010).	
	There may be additional, similar buoy programs not identified here					

IMBs	1993–	SAT, IST	CRREL		SAT at 1m (Elder, 2013). Data is often preliminary. Primarily for ice mass balance measurements.	Sometimes deployed with ice stations e.g. NPDS.
	2011–		SIMBA		Not radiation shielded. Primarily for ice mass balance measurements.	Can also provide ISTs. SIMBA instruments are deployed by other agencies e.g. FMI and DMI (SRSL, 2014; Hoyer, 2014).
Thermochrons	2012	IST	BROMEX		Snow cover an issue, different colours used for handles, not shielded (Hall et al., 2015).	
Radiosondes/ Rawinsondes	≤1947–	AT, SAT	CEAREX, Lead Experiment (LeadEx) and various other sources	HadAT (Thorne et al., 2005); RAOBCORE and RICH (Haimberger et al., 2012); RATPAC (Free et al., 2005); IUK (Sherwood, 2014) and the NCEP/NCAR Arctic marine rawinsonde archive (NCEP/NCAR, 1997).	Structural uncertainty in datasets, heterogeneities and biases (McCarthy et al., 2008; Titchner et al., 2009).	CEAREX and LeadEx used to validate reanalysis data (Bromwich and Wang, 2005).
Satellite - Thermal Infrared						
Sensor	Time Period	Variable/s	Satellite/s	Dataset/s or Data Products?	Issues	Notes
HRIR, THIR, MRIR, IRIS	1960-1994	AT, ST	Nimbus, TIROS			THIR IST investigated for Arctic sea ice (Comiso, 1994; Hall et al., 2004)
SIRS-A/B	1969-1980	AT, ST	Nimbus-3/4		Coarse vertical resolution.	

VHRR, AVHRR	1972–	ST	TIROS-M/N, NOAA, Metop	Pathfinder AVHRR (1981–), PATMOS-X (1981–) (Heidinger et al., 2014)	Issues with cloud masking for AVHRR and well as differences between AVHRR instruments (Comiso, 2014b).	AVHRR IST investigated for Arctic sea ice (e.g. Comiso, 2001, 2003; Key and Haefliger, 1992; Lindsay and Rothrock, 1994; Wang and Key, 2005; Yu et al., 1995). Good coverage, long record (Rayner et al., 2003).
ITPR, VTPR, HIRS(/2,3,4)	1972-1993	AT, ST	Nimbus, NOAA, TIROS-N, Metop	TOVS (MSU, HIRS, and SSU) Pathfinder Path-P (Francis and Schweiger, 1999).	Coarse spatial and vertical resolution.	TOVS IST investigated for Arctic sea ice (Chen et al., 2002; Schweiger et al., 2002).
SI-GDR	1976-1981	AT, ST	Meteor-Priroda			
TM, ETM+	1982–	ST	Landsat			Used for IST measurements (Hall et al., 2004)
SM, MSU-MR	1985–	ST	Meteor		Coarse vertical resolution.	
MVISR, IRAS, VIRR	1988–	ST	Feng-Yun		Some sensors have coarse spatial and vertical resolution.	
ATSR(-2), AATSR, SLSTR	1991-2012	ST	ERS-1/2, Envisat			SLSTR to be deployed on Sentinel missions, 2015.
IMG, OCTS, GLI	1996-2003	AT, ST	ADEOS			
VIRS	1997-2014	ST	TRMM			

MODIS	2000–	AT, ST	EOS-Terra/Aqua	Daily 0.25km IST product.		IST investigated for Arctic sea ice (Hall et al., 2004)
AIRS, CrIS	2002–	AT, ST	EOS-Aqua, JPSS, Soumi-NPP		Coarse spatial resolution.	
WindSat	2003–	ST	Coriolis			
TES-nadir	2004–	AT, ST	EOS-Aura			
IASI	2006–	AT, ST	Metop		Coarse spatial resolution.	
VIIRS	2011–	ST	Soumi-NPP	VIIRS IST product in production (NASA, 2014)	Cloud mask and other aspects.	
IKFS-2	2014–	AT, ST	Meteor			
Satellite - Microwave						
Sensor	Time Period	Variable/s	Satellite/s	Dataset/s or Data Products?	Issues	Notes
SMMR, SSM/I , SSM/T, SSMIS	1972–	AT (NSIDC, 2010), ST	Seasat, Nimbus, Block5D-2/3		May be large differences (SMMR) compared to in situ data (Gloersen et al., 1992).	SMMR investigated for Arctic sea ice temperatures (Cavalieri et al., 1984; Gloersen et al., 1992).
MSU, AMSU, ATMS	1978–	TLT	TIROS-N, NOAA, EOS-Aqua, Metop, Suomi-NPP	UAH Lower Troposphere Temperatures (Christy et al., 2003), used in the Cowtan and Way dataset (Cowtan and Way, 2014). TOVS (MSU, HIRS, and SSU) Pathfinder Path-P (Francis and Schweiger, 1999).	Many adjustments needed to produce the datasets. Coarse vertical resolution.	Used to validate reanalyses in the Arctic (Screen and Simmonds, 2011). TOVS IST investigated for Arctic sea ice (Chen et al., 2002; Schweiger et al., 2002).
IRAS	1988–	ST	Feng-Yun		Coarse spatial and vertical resolution.	

AMSR/2, AMSR-E	2002–	ST	ADEOS, GCOM, EOS-Aqua		Coarse spatial resolution.	
MWTS-1/2	2008–	AT	Feng-Yun 3 satellites			
Reanalyses						
Institution	Time Period	Reanalysis	Other Organisations	Summary of Chapter 2		Notes
				Advantages	Issues	
NOAA-CIRES	1871–	20CR		Very long reanalysis record. Timeseries are very similar to those from in situ observations.	Least representative reanalysis for Arctic SATs over land. Bad sea ice specification (Lindsay et al., 2014)?	Produced using atmospheric pressure data only.
NCEP	1948–	NCEP R1	NCAR	Realistic SATs and SAT patterns.	Some errors and biases. Overestimates Arctic warming.	Used as a first guess field for some areas in the IABP/POLES and ARCSS-SAT datasets (Rigor et al., 2000; EOL, 2014).
	1979–	NCEP R2	DOE	Realistic SATs and SAT patterns. Outperforms R1 when compared to IABP data (Alexeev et al., 2012)	Some biases, overestimates Arctic warming.	Corrected version of R1, but only minor differences (Kanamitsu et al., 2002).
	1979–	NCEP CFSR		Higher resolution reanalysis. Most representative NCEP reanalysis for the Arctic.	Some biases.	
ECMWF	1979 to 1993	ERA-15		Outperforms NCEP R1 for Arctic temperatures.	Overestimated Arctic temperatures by up to 2-3°C (Bromwich and Wang, 2005)	

	1957 to 2003	ERA-40		Realistic SATs and SAT patterns.	Consistent warm bias. Discontinuity in 1997 (Screen and Simmonds, 2011).	
	1979–	ERA-Interim		Realistic SATs and SAT patterns. Outperforms, or is comparable to, other reanalyses in terms of SAT patterns, anomalies, and SAT trends.	Warm bias in the atmosphere below 400 m of up to 2°C (Jakobson et al., 2012).	Explicitly reconstructs SAT rather than interpolating between model layers (Dee et al., 2011b; Lindsay et al., 2014).
	1900–	ERA-20C				(ERA-CLIM, 2014; Reanalyses.org, 2014).
JMA	1979 to 2004	JRA-25				
	1979–	JDCAS		Realistic Arctic SATs over land and realistic temperature trends compared to MSU.	Large RMSEs. Outperformed by other reanalyses in the Arctic ocean.	
	1958–	JRA-55		SAT timeseries are similar to those from other reanalyses and from datasets utilising in situ observations (Simmons and Poli, 2014).		Data for JRA-55 was made available in January 2014 (Ebita et al., 2011; JRA Project, 2014; Kobayashi et al., 2015).
NASA	1979–	MERRA		Realistic for Arctic SATs over land and above 200m.	Warm bias of about 2°C for air temperatures below 200 m (Jakobson et al., 2012).	
	1980–	MERRA2				In production (Reanalyses.org, 2014).

ASR	2000–	ASR		Realistic SATs and SAT patterns.	Some biases are noted by Hines et al. 2011; Wilson et al. 2011.	

Note: Satellite and sensor data was generally informed by the WMO OSCAR (Observing Systems Capability Analysis and Review Tool).

Bibliography

ACIA, 2005: *Executive Summary*, chap. Executive Summary, 8–21. Cambridge University Press, Cambridge.

Alexeev, V. A., I. Esau, I. V. Polyakov, S. J. Byam, and S. Sorokina, 2012: Vertical structure of recent arctic warming from observed data and reanalysis products. *Climatic Change*, **111 (2)**, 215–239.

Allan, R., P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, and S. Brönnimann, 2011: The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bulletin of the American Meteorological Society*, **92 (11)**, 1421–1425.

Anisimov, O. A., D. G. Vaughan, T. V. Callaghan, C. Furgal, H. Marchant, T. D. Prowse, V. H., and J. E. Walsh, 2007: *Polar regions (Arctic and Antarctic)*, chap. 15, 653–685. Cambridge University Press, Cambridge.

Beesley, J. A., C. S. Bretherton, C. Jakob, E. L. Andreas, J. M. Intrieri, and T. A. Uttal, 2000: A comparison of cloud and boundary layer variables in the ECMWF forecast model with observations at Surface Heat Budget of the Arctic Ocean (SHEBA) ice camp. *Journal of Geophysical Research-Atmospheres*, **105 (D10)**, 12 337–12 349.

Belousov, S. L., L. S. Gandin, and S. A. Mashovich, 1971: 210. Israel Program for Scientific Translations.

Berkeley Earth, 2014: Land + ocean (1850 - recent): Monthly global average temperature. Berkeley Earth, *http* : *//berkeleyearth.lbl.gov/auto/Global/Land_and_Ocean_complete.txt*.

Berry, D. I. and E. C. Kent, 2009: A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bulletin of the American Meteorological Society*, **90 (5)**, 645–656.

Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Monthly Weather Review*, **124 (6)**, 1256–1271.

Brigham, L., 2011: Marine protection in the arctic cannot wait. Nature, *http* : *//www.nature.com/news/*2011/111012/*full*/478157*a.html*.

Brohan, P., 2013: Ice station jeannette. oldWeather, *http* : *//blog.oldweather.org*/2013/08/08/*ice* − *station* − *jeannette/*.

Brohan, P., J. J. Kennedy, I. Harris, S. F. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research: Atmospheres*, **111 (D12)**.

Bromwich, D., Y.-H. Kuo, M. Serreze, J. Walsh, L.-S. Bai, M. Barlage, K. Hines, and A. Slater, 2010: Arctic system reanalysis: Call for community involvement. *Eos, Transactions American Geophysical Union*, **91 (2)**, 13.

Bromwich, D. H., R. L. Fogt, K. I. Hodges, and J. E. Walsh, 2007: A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *Journal of Geophysical Research-Atmospheres*, **112 (D10)**.

Bromwich, D. H., K. M. Hines, and L.-S. Bai, 2009: Development and testing of polar weather research and forecasting model: 2. arctic ocean. *Journal of Geophysical Research-Atmospheres*, **114 (D8)**.

Bromwich, D. H. and S.-H. Wang, 2005: Evaluation of the NCEP-NCAR and ECMWF 15-and 40-Yr reanalyses using rawinsonde data from two independent arctic field experiments. *Monthly Weather Review*, **133 (12)**, 3562–3578.

Bromwich, D. H., A. B. Wilson, L. Bai, G. E. K. Moore, and P. Bauer, 2014: Contrasting the regional Arctic System Reanalysis with the global ERA-Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **In Review**.

Callaghan, T. V., et al., 2011: The changing face of arctic snow cover: A synthesis of observed and projected changes. *Ambio*, **40 (1)**, 17–31.

Callendar, G. S., 1938: The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society*, **64 (275)**, 223–240.

Casey, K. S., T. B. Brandon, P. Cornillon, and R. Evans, 2010: The past, present, and future of the AVHRR Pathfinder SST program. *Oceanography from Space*, Springer, chap. 16, 273–287.

Cavalieri, D. J., P. Gloersen, and W. J. Campbell, 1984: Determination of sea ice parameters with the nimbus 7 smmr. *Journal of Geophysical Research: Atmospheres*, **89 (D4)**, 5355–5369.

Chen, Y., J. A. Francis, and J. R. Miller, 2002: Surface temperature of the arctic: Comparison of TOVS satellite retrievals with surface observations. *Journal of Climate*, **15 (24)**, 3698–3708.

Christy, J. R., R. W. Spencer, W. B. Norris, W. D. Braswell, and D. E. Parker, 2003: Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures. *Journal of Atmospheric and Oceanic Technology*, **20 (5)**, 613–629.

Chung, C. E., H. Cha, T. Vihma, P. Räisänen, and D. Decremer, 2013: On the possibilities to use atmospheric reanalyses to evaluate the warming structure in the arctic. *Atmospheric Chemistry and Physics*, **13 (22)**, 11 209–11 219.

Church, J. A., et al., 2013: *Sea Level Change*, chap. 13, 1152–1153. Cambridge University Press, Cambridge.

Cohen, J. L., J. C. Furtado, M. A. Barlow, V. A. Alexeev, and J. E. Cherry, 2012: Arctic warming, increasing snow cover and widespread boreal winter cooling. *Environmental Research Letters*, **7 (1)**.

Comiso, J. C., 1994: Surface temperatures in the polar regions from Nimbus 7 temperature humidity infrared radiometer. *Journal of Geophysical Research-Oceans*, **99 (C3)**, 5181–5200.

Comiso, J. C., 2000: Variability and trends in antarctic surface temperatures from in situ and satellite infrared measurements. *Journal of Climate*, **13 (10)**, 1674–1696.

Comiso, J. C., 2001: Satellite-observed variability and trend in sea-ice extent, surface temperature, albedo and clouds in the arctic. *Annals of Glaciology*, **33**, 457–473.

Comiso, J. C., 2003: Warming trends in the arctic from clear sky satellite observations. *Journal of Climate*, **16 (21)**, 3498–3510.

Comiso, J. C., 2006: Arctic warming signals from satellite observations. *Weather*, **61 (3)**, 70–76.

Comiso, J. C., 2012: Large decadal decline of the arctic multiyear ice cover. *Journal of Climate*, **25 (4)**, 1176–1193.

Comiso, J. C., 2014a: Questions about how you renormalised the avhrr data. Personal Communication.

Comiso, J. C., 2014b: Re: Avhrr monthly surface temperature data. Personal Communication.

Comiso, J. C., D. J. Cavalieri, and T. Markus, 2003: Sea ice concentration, ice temperature, and snow depth using AMSR-E data. *IEEE Transactions on Geoscience and Remote Sensing*, **41 (2)**, 243–252.

Comiso, J. C. and D. K. Hall, 2014: Climate trends in the arctic as observed from space. *Wiley Interdisciplinary Reviews: Climate Change*, **5 (3)**, 389–409.

Compo, G. P., P. D. Sardeshmukh, J. S. Whitaker, P. Brohan, P. D. Jones, and C. McColl, 2013: Independent confirmation of global land warming without the use of station temperatures. *Geophysical Research Letters*, **40 (12)**, 3170–3174.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society*, **87 (2)**, 175–190.

Compo, G. P., et al., 2011: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137 (654)**, 1–28.

Cowtan, K. and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*.

Cressey, D., 2011: Scientific challenges in the arctic: Open water. Nature, *www.nature.com/news/*2011/111012/*full*/478174*a.html*.

Cressie, N., 1990: The origins of kriging. *Mathematical Geology*, **22 (3)**, 239–252.

Curry, J. A., J. L. Schramm, A. Alam, R. Reeder, T. E. Arbetter, and P. Guest, 2002: Evaluation of data sets used to force sea ice models in the arctic ocean. *Journal of Geophysical Research-Oceans*, **107 (C8)**.

Dee, D. P., E. Källén, A. J. Simmons, and L. Haimberger, 2011a: Comments on reanalyses suitable for characterizing long-term trends. *Bulletin of the American Meteorological Society*, **92 (1)**, 65–70.

Dee, D. P. and S. Uppala, 2009: Variational bias correction of satellite radiance data in the era-interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **135 (644)**, 1830–1841.

Dee, D. P., et al., 2011b: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597.

Dodd, E. M. A., C. J. Merchant, N. A. Rayner, and C. P. Morice, 2015: An investigation into the impact of using various techniques to estimate arctic surface air temperature anomalies. *Journal of Climate*, **28 (5)**, 1743–1763, doi: http://dx.doi.org/10.1175/JCLI-D-14-00250.1.

Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, **116**, 140–158.

Douville, H., 1998: *The ECMWF surface analysis: diagnostics and prospects*. European Centre for Medium-Range Weather Forecasts.

Ebita, A., et al., 2011: The Japanese 55-year Reanalysis "JRA-55": An interim report. *SOLA*, **7 (0)**, 149–152.

ECMWF, 2007: IFS documentation CY31R1. Part II: Data assimilation. Tech. rep., European Centre for Medium-Range Weather Forecasts, Shinfield, Reading, UK.

Elder, B., 2013: Reliability of imb air temperature sensors. personal communication.

Embury, O., 2015: AVHRR calibration issues. Personal Communication.

Embury, O. and C. J. Merchant, 2012: A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. *Remote Sensing of Environment*, **116**, 47–61.

EOL, 2014: ARCSS Surface Air Temperature (ARCSS-SAT) Analysis. EOL, *http* : $//data.eol.ucar.edu/codiac/dss/id = 106.308$.

ERA-CLIM, 2014: European reanalysis of global climate observations (ERA-CLIM). ECMWF, $www.era - clim.eu/about$.

ESA, 2014: About SST CCI: Overview. ESA, $http : //www.esa - sst - cci.org/?q = overview$.

EWG, 2000: Arctic meteorology and climate atlas. National Snow and Ice Data Center, Boulder, CO.

Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. R. Christy, and R. A. Pielke, 2011: Analysis of the impacts of station exposure on the US Historical Climatology Network temperatures and temperature trends. *Journal of Geophysical Research-Atmospheres*, **116 (D14)**.

Fan, Y. and H. Van den Dool, 2008: A global monthly land surface air temperature analysis for 1948–present. *Journal of Geophysical Research: Atmospheres*, **113 (D1)**.

Folland, C. K. and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, **121 (522)**, 319–367.

Francis, J. A. and A. J. Schweiger, 1999: TOVS Pathfinder Path-P daily and monthly polar gridded atmospheric parameters. National Snow and Ice Data Center, Boulder, CO.

Free, M., D. J. Seidel, J. K. Angell, J. Lanzante, I. Durre, and T. C. Peterson, 2005: Radiosonde atmospheric temperature products for assessing climate (RATPAC): A new data set of large-area anomaly time series. *Journal of Geophysical Research: Atmospheres*, **110 (D22)**.

Gibson, J. K., P. Kallberg, and S. Uppala, 1999: 1. ERA-15 description (version 2 - january 1999). Tech. rep., European Centre for Medium-Range Weather Forecasts, Shinfield, Reading, UK. *http* : $//www.ecmwf.int/publications/$.

GISS, 2014: GISS surface temperature analysis (GISTEMP) - updates to analysis. GISS, *http* : $//data.giss.nasa.gov/gistemp/updates\_v3/$.

Gloersen, P., W. J. Campbell, D. J. Cavalieri, J. C. Comiso, C. L. Parkinson, and H. J. Zwally, 1992: Arctic and antarctic sea ice, 1978-1987: Satellite passive-microwave observations and analysis. *Annals of Glaciology*, **17**, 149–154.

Haimberger, L., C. Tavolato, and S. Sperka, 2012: Homogenization of the global radiosonde temperature dataset through combined comparison with reanalysis background series and neighboring stations. *Journal of Climate*, **25 (23)**, 8108–8131.

Hall, D. K., J. R. Key, K. A. Casey, G. A. Riggs, and D. J. Cavalieri, 2004: Sea ice surface temperature product from MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, **42 (5)**, 1076–1087.

Hall, D. K., S. V. Nghiem, I. G. Rigor, and J. A. Miller, 2015: Uncertainties of temperature measurements on snow-covered land and sea ice from in-situ and MODIS data during BROMEX. *Journal of Applied Meteorology and Climatology*.

Hanna, E., et al., 2013: Ice-sheet mass balance and climate change. *Nature*, **498 (7452)**, 51–59.

Hansen, J. and S. Lebedeff, 1987: Global trends of measured surface air temperature. *Journal of Geophysical Research: Atmospheres*, **92 (D11)**, 13 345–13 372.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Reviews of Geophysics*, **48**.

Hartmann, D. L., et al., 2013: *Observations: Atmosphere and Surface*, chap. 2, 187201. Cambridge University Press, Cambridge.

Hawkins, E. and P. Jones, 2013: On increasing global temperatures: 75 years after callendar. *Quarterly Journal of the Royal Meteorological Society*, **139 (677)**, 1961–1963.

Heidinger, A. K., M. J. Foster, A. Walther, and X. Zhao, 2014: The pathfinder atmospheres extended (PATMOS-x) AVHRR climate data set. *Bulletin of the American Meteorological Society*, **95 (6)**.

Hines, K. M. and D. H. Bromwich, 2008: Development and testing of polar weather research and forecasting (WRF) model. part i: Greenland ice sheet meteorology. *Monthly Weather Review*, **136 (6)**, 1971–1989.

Hines, K. M., D. H. Bromwich, L.-S. Bai, M. Barlage, and A. G. Slater, 2011: Development and testing of polar WRF. part iii: Arctic land. *Journal of Climate*, **24 (1)**, 26–48.

Hinzman, L. D., et al., 2005: Evidence and implications of recent climate change in northern alaska and other arctic regions. *Climatic Change*, **72 (3)**, 251–298.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei, 2008: Comparison of six methods for the interpolation of daily, european climate data. *Journal of Geophysical Research-Atmospheres*, **113 (D21)**, 19.

Honjo, S., R. Krishfield, J. Kemp, T. Takezawa, and K. Hatakeyama, 1995: Drifting buoys make discoveries about interactive processes in the arctic ocean. *EOS, Transactions American Geophysical Union*, **76 (21)**, 209–219.

Hoyer, J., 2014: Ice mass balance buoy data. personal communication.

IABP, 2014: Surface air temperature. IABP, *http* : *//iabp.apl.washington.edu/data*$_s$*atemp.html*.

IARC, 2012: APLIS07: Welcome. International Arctic Research Center, *http* : *//aplis*07*.iarc.uaf.edu/en/research/projects/aplis*07.

Inoue, J., T. Enomoto, T. Miyoshi, and S. Yamane, 2009: Impact of observations from arctic drifting buoys on the reanalysis of surface fields. *Geophysical Research Letters*, **36 (8)**.

IPCC, 2013: chap. Summary for Policymakers. Cambridge University Press, Cambridge.

Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. *International Journal of Climatology*, **25 (7)**, 865–879.

Jakobson, E., T. Vihma, T. Palo, L. Jakobson, H. Keernik, and J. Jaagus, 2012: Validation of atmospheric reanalyses over the central arctic ocean. *Geophysical Research Letters*, **39 (10)**.

Jin, M. and R. E. Treadon, 2003: Correcting the orbit drift effect on AVHRR land surface skin temperature measurements. *International Journal of Remote Sensing*, **24 (22)**, 4543–4558.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research-Atmospheres*, **117**, 29.

JRA Project, 2014: JRA-55 the Japanese 55-year Reanalysis. JRA Project, *http* : $//jra.kishou.go.jp/JRA-55/index_en.html$.

Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77 (3)**, 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*, **83 (11)**, 1631–1643.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research-Oceans*, **103 (C9)**, 18 567–18 589.

Kennedy, J., C. Morice, and D. Parker, 2011a: Global and regional climate in 2010. *Weather*, **66 (7)**, 188–194.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. measurement and sampling uncertainties. *Journal of Geophysical Research-Atmospheres*, **116**.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011c: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. biases and homogenization. *Journal of Geophysical Research-Atmospheres*, **116**.

Kent, E. C. and P. . Challenor, 2006: Toward estimating climatic trends in SST. Part II: Random errors. *Journal of Atmospheric and Oceanic Technology*, **23 (3)**, 476–486.

Kent, E. C. and A. Kaplan, 2006: Toward estimating climatic trends in SST. Part III: Systematic biases. *Journal of Atmospheric and Oceanic Technology*, **23 (3)**, 487–500.

Kent, E. C., N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, and D. E. Parker, 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *Journal of Geophysical Research-Atmospheres*, **118 (3)**, 1281–1298.

Kent, E. C. and P. K. Taylor, 2006: Toward estimating climatic trends in SST. Part I: Methods of measurement. *Journal of Atmospheric and Oceanic Technology*, **23 (3)**, 464–475.

Key, J. and M. Haefliger, 1992: Arctic ice surface temperature retrieval from AVHRR thermal channels. *Journal of Geophysical Research-Atmospheres*, **97 (D5)**, 5885–5893.

Key, J., J. A. Maslanik, T. Papakyriakou, M. C. Serreze, and A. J. Schweiger, 1994: On the validation of satellite-derived sea-ice surface-temperature. *Arctic*, **47 (3)**, 280–287.

Key, J. R., A. J. Schweiger, and R. S. Stone, 1997: Expected uncertainty in satellite-derived estimates of the surface radiation budget at high latitudes. *Journal of Geophysical Research-Oceans*, **102 (C7)**, 15 837–15 847.

Knepp, T. N., et al., 2010: Development of an autonomous sea ice tethered buoy for the study of ocean-atmosphere-sea ice-snow pack interactions: the O-buoy. *Atmospheric Measurement Techniques*, **3 (1)**, 249–261.

Kobayashi, S., et al., 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, **93 (1)**.

Kullerud, L., 2011: Sustainability: A green arctic. Nature, *http* : *//www.nature.com/nature/journal/v*478/*n*7368/*full*/478179*a.html*.

Kwok, R., 2007: Near zero replenishment of the arctic multiyear sea ice cover at the end of 2005 summer. *Geophysical Research Letters*, **34 (5)**.

Kwok, R., G. F. Cunningham, M. Wensnahan, I. Rigor, H. J. Zwally, and D. Yi, 2009: Thinning and volume loss of the arctic ocean sea ice cover: 2003-2008. *Journal of Geophysical Research-Oceans*, **114 (C7)**.

Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *Journal of Geophysical Research-Atmospheres*, **116 (D19)**.

Leck, C., E. K. Bigg, D. S. Covert, J. Heintzenberg, W. Maenhaut, E. D. Nilsson, and A. Wiedensohler, 1996: Overview of the atmospheric research program during the International Arctic Ocean Expedition of 1991 (IAOE-91) and its scientific results. *Tellus B*, **48 (2)**, 136–155.

Leck, C., E. D. Nilsson, E. K. Bigg, and L. Bäcklin, 2001: Atmospheric program on the Arctic Ocean Expedition 1996 (AOE-96): An overview of scientific goals, experimental approach, and instruments. *Journal of Geophysical Research: Atmospheres*, **106 (D23)**, 32 051–32 067.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the arctic. *Journal of Climate*, **27 (7)**, 2588–2606.

Lindsay, R. W. and D. A. Rothrock, 1994: Arctic sea ice surface temperature from AVHRR. *Journal of Climate*, **7 (1)**, 174–183.

Liu, J., Z. Zhang, Y. Hu, L. Chen, Y. Dai, and X. Ren, 2008: Assessment of surface air temperature over the arctic ocean in reanalysis and IPCC AR4 model simulations with IABP/POLES observations. *Journal of Geophysical Research-Atmospheres*, **113 (D10)**.

Lorenc, A. C. and F. Rawlins, 2005: Why does 4D-Var beat 3D-Var? *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3247–3257.

Lüpkes, C., T. Vihma, E. Jakobson, G. König-Langlo, and A. Tetzlaff, 2010: Meteorological observations from ship cruises during summer to the central arctic: A comparison with reanalysis data. *Geophysical Research Letters*, **37 (9)**.

Makshtas, A., D. Atkinson, M. Kulakov, S. Shutilin, R. Krishfield, and A. Proshutinsky, 2007: Atmospheric forcing validation for modeling the central arctic. *Geophysical Research Letters*, **34 (20)**.

Manley, G., 1953: The mean temperature of central england, 1698-1952. *Quarterly Journal of the Royal Meteorological Society*, **79 (340)**, 242–261.

Marchenko, N., 2012: *The Laptev Sea*, chap. 4, 120–123. Springer, Heidelberg.

Markwardt, C. B., 2009: Non-linear least squares fitting in IDL with MPFIT. *Proc. Astronomical Data Analysis Software and Systems XVIII*.

Maslanik, J., C. Fowler, J. Stroeve, S. Drobot, J. Zwally, D. Yi, and W. Emery, 2007: A younger, thinner arctic ice cover: Increased potential for rapid, extensive sea-ice loss. *Geophysical Research Letters*, **34 (24)**.

McCarthy, M. P., H. A. Titchner, P. W. Thorne, S. F. B. Tett, L. Haimberger, and D. E. Parker, 2008: Assessing bias and uncertainty in the HadAT-adjusted radiosonde climate record. *Journal of Climate*, **21 (4)**, 817–832.

Mears, C. A. and F. J. Wentz, 2009: Construction of the Remote Sensing Systems v3. 2 atmospheric temperature records from the MSU and AMSU microwave sounders. *Journal of Atmospheric and Oceanic Technology*, **26 (6)**, 1040–1056.

Menne, M. J., C. N. Williams, and R. S. Vose, 2009: The US Historical Climatology Network Monthly Temperature Data, Version 2. *Bulletin of the American Meteorological Society*, **90 (7)**, 993–1007.

Merchant, C. J., P. Le Borgne, A. Marsouin, and H. Roquet, 2008: Optimal estimation of sea surface temperature from split-window observations. *Remote Sensing of Environment*, **112 (5)**, 2469–2484.

Merchant, C. J., et al., 2012: A twenty-year independent record of sea surface temperature for climate from Along Track Scanning Radiometers. *Journal of Geophysical Research-Oceans*, **117 (C12)**.

Merchant, C. J., et al., 2013: The surface temperatures of earth: steps towards integrated understanding of variability and change. *Geoscientific Instrumentation, Methods and Data Systems*, **2 (2)**, 305–321.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research-Atmospheres*, **117 (D8)**.

Muller, R. A., et al., 2013: Decadal variations in the global atmospheric land temperatures. *Journal of Geophysical Research-Atmospheres*, **118 (11)**, 5280–5286.

NASA, 2014: VIIRS ice surface temperature product status. NASA, *http* : $//viirsland.gsfc.nasa.gov/Products/ISTStatus.html$.

NCEP/NCAR, 1997: NCEP/NCAR arctic marine rawinsonde archive. National Snow and Ice Data Center, Boulder, CO, dataset.

New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate*, **13 (13)**, 2217–2238.

Nghiem, S. V., I. G. Rigor, D. K. Perovich, P. Clemente-Colón, J. W. Weatherly, and G. Neumann, 2007: Rapid reduction of arctic perennial sea ice. *Geophysical Research Letters*, **34 (19)**.

NOAA, 2000: North pole weather data from 2004 deployment. NOAA, *http* : $//www.arctic.noaa.gov/np2004/gallery_np_weatherdata.html$.

NSIDC, 2010: Special Sensor Microwave Imager/Sounder (SSMIS). NSIDC, *http* : *//nsidc.org/data/docs/daac/ssmis*$_i$*nstrument/*.

NSIDC, 2012a: Arctic sea ice extent settles at record seasonal minimum. NSIDC, *http* : *//nsidc.org/arcticseaicenews/*2012/09/*arctic* − *sea* − *ice* − *extent* − *settles* − *at* − *record* − *seasonal* − *minimum/*.

NSIDC, 2012b: Arctic sea ice falls below 4 million square kilometers. NSIDC, *http* : *//nsidc.org/arcticseaicenews/*2012/09/*arctic* − *sea* − *ice* − *falls* − *below* − 4 − *million* − *square* − *kilometers/*.

NSIDC, 2013: All about sea ice: Thermodynamics : Albedo. NSIDC, *http* : *//nsidc.org/cryosphere/seaice/processes/albedo.html*.

NSIDC, 2014: Arctic sea ice reaches minimum extent for 2014. NSIDC, *http* : *//nsidc.org/arcticseaicenews/*2014/09/*arctic* − *minimum* − *reached/*.

Olenicoff, S. M., 1968: Soviet development and use of drifting automatic radiometeorological stations for arctic research.

Onogi, K., et al., 2007: The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan*, **85 (3)**, 369–432.

Overland, J. E. and M. Wang, 2013: When will the summer arctic be nearly sea ice free? *Geophysical Research Letters*, **40 (10)**, 2097–2101.

Parker, D. E., 2010: Urban heat island effects on estimates of observed climate change. *Wiley Interdisciplinary Reviews: Climate Change*, **1 (1)**, 123–133.

Parker, D. E., C. K. Folland, and M. Jackson, 1995: Marine surface temperature: observed variations and data requirements. *Climatic Change*, **31 (2-4)**, 559–600.

Parker, D. E., P. Jones, T. C. Peterson, and J. Kennedy, 2009: Comment on "unresolved issues with the assessment of multidecadal global land surface temperature trends" by Roger A. Pielke Sr. et al. *Journal of Geophysical Research-Atmospheres*, **114 (D5)**.

Parkinson, C. L. and J. C. Comiso, 2013: On the 2012 record low arctic sea ice cover: Combined impact of preconditioning and an august storm. *Geophysical Research Letters*, **40 (7)**.

Perovich, D., S. Gerland, S. Hendricks, W. Meier, M. Nicolaus, and M. Tschudi, 2014: Sea ice [in arctic report card 2014]. NOAA, *http* : *//www.arctic.noaa.gov/reportcard*.

Peterson, T. C., T. R. Karl, P. F. Jamason, R. Knight, and D. R. Easterling, 1998: First difference method: Maximizing station density for the calculation of long-term global temperature change. *Journal of Geophysical Research: Atmospheres*, **103 (D20)**, 25 967–25 974.

Pielke, S., R. A., et al., 2007: Unresolved issues with the assessment of multidecadal global land surface temperature trends. *Journal of Geophysical Research-Atmospheres*, **112 (D24)**.

Polashenski, C., D. Perovich, J. Richter-Menge, and B. Elder, 2011: Seasonal ice mass-balance buoys: adapting tools to the changing arctic. *Annals of Glaciology*, **52 (57)**, 18–25.

Polyakov, I. V., J. E. Walsh, and R. Kwok, 2012: Recent changes of arctic multiyear sea ice coverage and the likely causes. *Bulletin of the American Meteorological Society*, **93 (2)**, 145–151.

Price, D. T., D. W. McKenney, I. A. Nalder, M. F. Hutchinson, and J. L. Kesteven, 2000: A comparison of two statistical methods for spatial interpolation of canadian monthly mean climate data. *Agricultural and Forest Meteorology*, **101 (2-3)**, 81–94.

Price, J. C., 1991: Timing of NOAA afternoon passes. *International Journal of Remote Sensing*, **12 (1)**, 193–198.

Przybylak, R., 1997: Spatial and temporal changes in extreme air temperatures in the arctic over the period 1951-1990. *International Journal of Climatology*, **17**, 615–634.

Przybylak, R., 2003a: *Air Temperature*, chap. 4, 78–79. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Przybylak, R., 2003b: *Atmospheric Circulation*, chap. 2, 25–28. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Przybylak, R., 2003c: *Introduction*, chap. 1, 3. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Przybylak, R., P. Wyszyński, Z. Vízi, and J. Jankowska, 2013: Atmospheric pressure changes in the arctic from 1801 to 1920. *International Journal of Climatology*, **33 (7)**, 1730–1760.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. 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 dataset. *Journal of Climate*, **19 (3)**, 446–469.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research-Atmospheres*, **108 (D14)**.

Reanalyses.org, 2014: Overview of current atmospheric reanalyses. Reanalyses.org, *http* : $//reanalyses.org/atmosphere/overview - current - reanalyses$.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *Journal of climate*, **15 (13)**, 1609–1625.

Richter-Menge, J. A., D. K. Perovich, B. C. Elder, K. Claffey, I. Rigor, and M. Ortmeyer, 2006: Ice mass-balance buoys: a tool for measuring and attributing changes in the thickness of the arctic sea-ice cover. *Annals of Glaciology*, **44 (1)**, 205–210.

Rienecker, M. M., et al., 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, **24 (14)**, 3624–3648.

Rigor, I. G., 2014: Info on IABP buoydata. Personal Communication.

Rigor, I. G., R. L. Colony, and S. Martin, 2000: Variations in surface air temperature observations in the arctic, 1979-97. *Journal of Climate*, **13 (5)**, 896–914.

Rohde, R., 2013: Comparison of Berkeley Earth, NASA GISS, and Hadley CRU averaging techniques on ideal synthetic data. Tech. Rep. 28^{th} August, Berkeley Earth. $http://berkeleyearth.org/pdf/robert-rohde-memo.pdf$.

Rohde, R., et al., 2012: Berkeley Earth temperature averaging process: A supplement on statistical and mathematical methods. Berkeley Earth, $http://berkeleyearth.org/pdf/methods-paper-supplement$.

Rohde, R., et al., 2013a: Berkeley Earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview*, **1 (1)**.

Rohde, R., et al., 2013b: A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*, **1 (1)**.

Saha, S., et al., 2006: The NCEP climate forecast system. *Journal of Climate*, **19 (15)**, 3483–3517.

Saha, S., et al., 2010: The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, **91 (8)**, 1015–1057.

Scambos, T. A., T. M. Haran, and R. Massom, 2006: Validation of AVHRR and MODIS ice surface temperature products using in situ radiometers. *Annals of Glaciology*, **44 (1)**, 345–351.

Schweiger, A. J., R. W. Lindsay, J. A. Francis, J. Key, J. M. Intrieri, and M. D. Shupe, 2002: Validation of TOVS Path-P data during SHEBA. *Journal of Geophysical Research: Oceans (1978-2012)*, **107 (C10)**, SHE–17.

Screen, J. A. and I. Simmonds, 2011: Erroneous arctic temperature trends in the ERA-40 reanalysis: A closer look. *Journal of Climate*, **24 (10)**, 2620–2627.

Serreze, M. C. and R. G. Barry, 2011: Processes and impacts of arctic amplification: A research synthesis. *Global and Planetary Change*, **77 (1)**, 85–96.

Serreze, M. C., et al., 2000: Observational evidence of recent change in the northern high-latitude environment. *Climatic Change*, **46**, 159–207.

Sherwood, S., 2014: IUK radiosonde analysis project – new v2 now available! University of New South Wales, *http : //web.maths.unsw.edu.au/ stevensherwood/radproj/index.html*.

Simmons, A. J. and P. Poli, 2014: Arctic warming in ERA-Interim and other analyses. *Quarterly Journal of the Royal Meteorological Society*.

Simmons, A. J., et al., 2004: Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research*, **109 (D24)**.

Slonosky, V., 2014: Canadian historical data typing project. *https : //sites.google.com/site/historicalclimatedata/canadian − historical − data − typing − project*.

Smith, C. L., 1971: A comparison of soviet and american drifting ice stations. *The Polar Record*, **15 (99)**, 877–885.

Smith, T. M. and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on coads data (1854-1997). *Journal of Climate*, **16 (10)**, 1495–1510.

Smith, T. M. and R. W. Reynolds, 2004: Improved extended reconstruction of sst (1854-1997). *Journal of Climate*, **17 (12)**, 2466–2477.

Smith, T. M. and R. W. Reynolds, 2005: A global merged land-air-sea surface temperature reconstruction based on historical observations (1880-1997). *Journal of Climate*, **18 (12)**, 2021–2036.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and L. J., 2008: Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis. *Journal of Climate*, **21 (10)**, 2283–2296.

SRSL, 2014: Sea-ice mass balance buoys. SRSL, *http : //www.samsrsl.co.uk/marine − services/sea − ice − mass − balance − buoys*.

Stroeve, J. C., M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012: The arctics rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, **110 (3-4)**, 1005–1027.

Tastula, E.-M. and T. Vihma, 2011: WRF model experiments on the antarctic atmosphere in winter. *Monthly Weather Review*, **139 (4)**, 1279–1291.

Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan, 2005: Revisiting radiosonde upper air temperatures from 1958 to 2002. *Journal of Geophysical Research-Atmosphere*, **110 (D18)**.

Thorne, P. W. and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends: Are they really achievable? *Bulletin of the American Meteorological Society*, **91 (3)**, 353–361.

Thorne, P. W., et al., 2011: Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bulletin of the American Meteorological Society*, **92 (11)**, ES40–ES47.

Titchner, H. A., P. W. Thorne, M. P. McCarthy, S. F. B. Tett, L. Haimberger, and D. E. Parker, 2009: Critically reassessing tropospheric temperature trends from radiosondes using realistic validation experiments. *Journal of Climate*, **22 (3)**, 465–485.

Tjernström, M., 2005: The summer arctic boundary layer during the arctic ocean experiment 2001 (AOE-2001). *Boundary-Layer Meteorology*, **117 (1)**, 5–36.

Tjernström, M., C. Leck, P. O. G. Persson, M. L. Jensen, S. P. Oncley, and A. Targino, 2005: Experimental equipment; a supplement to "the summer arctic boundary layer during the arctic ocean experiment 2001 (AOE-2001)". *Boundary-Layer Meteorology*, **117 (1)**, 5–36.

Tjernström, M., et al., 2014: The Arctic Summer Cloud Ocean Study (ASCOS): overview and experimental design. *Atmospheric Chemistry and Physics*, **14 (6)**, 2823–2869.

Trewin, B., 2010: Exposure, instrumentation, and observing practice effects on land temperature measurements. *Wiley Interdisciplinary Reviews: Climate Change*, **1 (4)**, 490–506.

Untersteiner, N. and A. S. Thorndike, 1982: Arctic data buoy program. *Polar Record*, **21 (131)**, 127–135.

Untersteiner, N., A. S. Thorndike, D. A. Rothrock, and K. L. Hunkins, 2007: AIDJEX revisited: A look back at the us-canadian arctic ice dynamics joint experiment 1970-78. *Arctic*, **80 (3)**, 327–336.

Uppala, S. M., et al., 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131 (612)**, 2961–3012.

Uttal, T., et al., 2002: Surface heat budget of the arctic ocean. *Bulletin of the American Meteorological Society*, **83 (2)**, 255–275.

Veal, K. L., G. K. Corlett, D. Ghent, D. T. Llewellyn-Jones, and J. J. Remedios, 2013: A time series of mean global skin SST anomaly using data from ATSR-2 and AATSR. *Remote Sensing of Environment*, **135**, 64–76.

Vicente-Serrano, S. M., M. A. Saz-Sanchez, and J. M. Cuadrat, 2003: Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research*, **24 (2)**, 161–180.

Vihma, T., J. Jaagus, E. Jakobson, and T. Palo, 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)**.

Vihma, T., J. Uotila, B. Cheng, and J. Launiainen, 2002: Surface heat budget over the weddell sea: Buoy results and model comparisons. *Journal of Geophysical Research-Oceans*, **107 (C2)**.

Vincent, L. A., X. L. Wang, E. J. Milewska, H. Wan, F. Yang, and V. Swail, 2012: A second generation of homogenized canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research-Atmospheres*, **117 (D18)**.

Vose, R. S., et al., 2012: NOAA's Merged Land-Ocean Surface Temperature analysis. *Bulletin of the American Meteorological Society*, **93 (11)**, 1677–1685.

Wang, M. and J. E. Overland, 2009: A sea ice free summer arctic within 30 years? *Geophysical Research Letters*, **36 (7)**.

Wang, W., P. Xie, S.-H. Yoo, Y. Xue, A. Kumar, and X. Wu, 2011: An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Climate Dynamics*, **37 (7-8)**, 1601–1620.

Wang, X. J. and J. R. Key, 2005: Arctic surface, cloud, and radiation properties based on the AVHRR polar pathfinder dataset. part i: Spatial and temporal characteristics. *Journal of Climate*, **18 (14)**, 2558–2574.

Wentz, F. J., C. Gentemann, D. Smith, and D. Chelton, 2000: Satellite measurements of sea surface temperature through clouds. *Science*, **288 (5467)**, 847–850.

Wentz, F. J. and T. Meissner, 2000: AMSR ocean algorithm, version 2, algorithm theoretical basis document. Tech. Rep. 121599A-1, Remote Sensing Systems, Santa Rosa, California.

Wentz, F. J. and T. Meissner, 2007: Supplement 1 algorithm theoretical basis document for AMSR-E ocean algorithms. Tech. Rep. 051707, Remote Sensing Systems, Santa Rosa, California.

Wiese, M., P. Griewank, and D. Notz, 2015: On the thermodynamics of melting sea ice versus melting freshwater ice. *Annals of Glaciology*, **56 (69)**, 191.

Wilson, A. B., D. H. Bromwich, and K. M. Hines, 2011: Evaluation of polar WRF forecasts on the arctic system reanalysis domain: Surface and upper air analysis. *Journal of Geophysical Research-Atmospheres*, **116 (D11)**.

Woodruff, S. D., et al., 2011: ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, **31 (7)**, 951–967.

Yu, Y., D. A. Rothrock, and R. W. Lindsay, 1995: Accuracy of sea ice temperature derived from the advanced very high resolution radiometer. *Journal of Geophysical Research-Oceans*, **100 (C3)**.

Zubov, N. N., 1939: The drift of the Sedov. *Nature*, **143**, 837–840.