An investigation into the impact of using various techniques to estimate Arctic surface air temperature anomalies

Article

Published Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

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

EMMA M. A. DODD AND CHRISTOPHER J. MERCHANT

Department of Meteorology, University of Reading, Reading, United Kingdom

NICK A. RAYNER AND COLIN P. MORICE

Met Office Hadley Centre, Exeter, United Kingdom

ABSTRACT

Time series of global and regional mean surface air temperature (SAT) anomalies are a common metric used to estimate recent climate change. Various techniques can be used to create these time series from meteorological station data. 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 was investigated using reanalysis data as a test bed. Techniques that interpolated anomalies were found to result in smaller errors than noninterpolating techniques relative to the reanalysis reference. 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, root-mean-square errors (RMSEs) up to 0.55 K larger than the two kriging techniques tested. Noninterpolating techniques provided the least representative anomaly estimates. Nonetheless, they serve as useful checks for confirming whether estimates from interpolating techniques are reasonable. The interaction of meteorological station coverage with estimation techniques between 1850 and 2011 was simulated using an ensemble dataset comprising repeated individual years (1979–2011). All techniques were found to have larger RMSEs for earlier station coverages. This supports calls for increased data sharing and data rescue, especially in sparsely observed regions such as the Arctic.

1. Introduction

The Arctic is recognized as an important region in the study of climate change because of expected and observed changes in this region. Temperature changes are predicted to be more rapid in the Arctic compared to those predicted at lower latitudes as a result of climate amplification processes (Serreze and Barry 2011). Arctic average temperatures have already warmed more rapidly than those over the rest of the world (e.g., Bekryaev et al. 2010; Serreze et al. 2009; Screen and Simmonds

2010). Many other changes associated with climate change are also being recorded in the Arctic, including reduced, thinner and younger sea ice cover; retreating glaciers; increased ice sheet melt; thawing permafrost; changes in precipitation; more frequent wildfires; and shifting vegetation (ACIA 2005; Comiso et al. 2008; Comiso 2012; Maslanik et al. 2007; Parkinson and Comiso 2013). However, monitoring Arctic temperature change is challenging, particularly in areas covered by sea ice for all or part of the year.

Time series 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 (Fig. 1), and the records are often short. The sparseness of SAT 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., Brohan et al. 2006; Cowtan and Way 2014; Jones et al. 2012; Parker
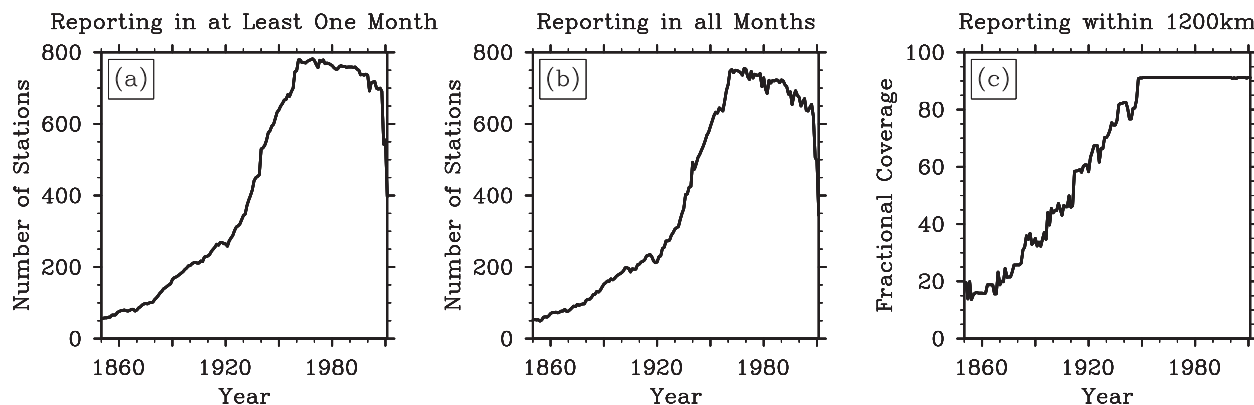

FIG. 1. The number of stations (a) reporting at least one temperature in each year and (b) reporting temperatures in all months of each year, and (c) the percentage of grid cells with at least one station reporting within 1200 km (fractional coverage).

et al. 2009; Pielke et al. 2007). Satellites provide consistent, continuous, and detailed observations of the radiometric temperature of ice and water surfaces in the Arctic (Merchant et al. 2013). These radiometric surface temperatures can be related to SATs and are available for the last three decades. However, satellite temperatures are not generally utilized at present for producing datasets of SAT anomalies.

There are many different groups that generate datasets of SAT anomalies using various techniques. These result in time series of temperature changes that generally produce similar trends in postindustrialization climate change (see Fig. 2). There are many different techniques and methods that can be used to quantify SAT changes over the Arctic from sparse in situ measurements.

One technique that can be employed is to use available in situ temperature measurements exclusively. This technique is used to create the Hadley Centre/Climatic Research Unit (CRU) global temperature anomaly dataset, version 4 (HadCRUT4; Morice et al. 2012), which is produced by the Met Office Hadley Centre from a combination of a land surface temperature anomaly dataset produced by the Climatic Research Unit of the University of East Anglia in conjunction with the Met Office Hadley Centre (CRUTEM4; Jones et al. 2012) and the Met Office Hadley Centre's sea surface temperature (SST) anomaly dataset, version 3 (HadSST3; Kennedy et al. 2011a,b). Since they do not spatially infill data, any grid boxes that 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). This means that large areas of the Arctic are unrepresented in these datasets as both land station records and SST records, especially in sea ice regions, are temporally and spatially sparse even with recent updates to the dataset (Kennedy et al. 2011a; Morice et al. 2012).

Some datasets, such as the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS) Surface Temperature Analysis, also known as GISTEMP, interpolate and extrapolate Arctic temperatures. GISTEMP is a combination of linearly interpolated and extrapolated SAT data over land and sea ice and the extended reconstructed sea surface temperature dataset version 3b (ERSSTv3b), which replaced the previous SST dataset in 2013, over the ocean (GISS 2014; Hansen et al. 2010; Smith et al. 2008).
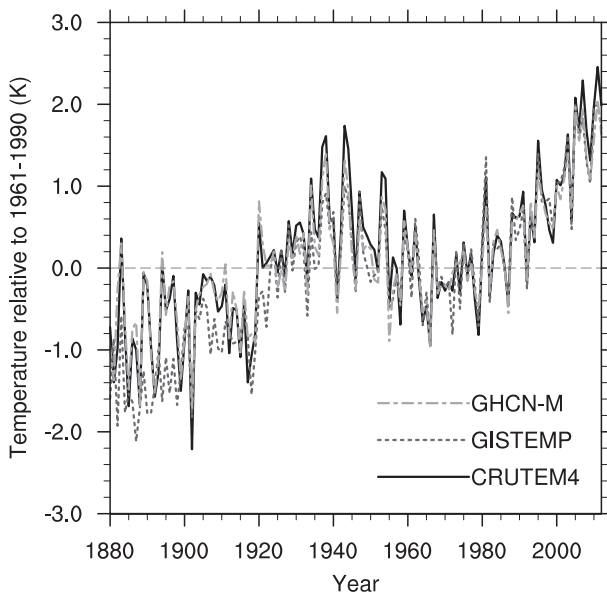
FIG. 2. The annual Arctic SAT anomaly (K) over land relative to 1961–90 for several temperature anomaly datasets: GHCN-M (Lawrimore et al. 2011), GISTEMP (Hansen et al. 2010), and CRUTEM4 (Jones et al. 2012). The time series are produced from the dataset grids using grid boxes north of 65°N over land and converted to be relative to 1961–90. The dataset versions used for this figure are GHCN-M.3.2.2.20140729, GISTEMP (downloaded on 29 Jul 2014), and CRUTEM4.2.0.0.

SAT anomalies from the Global Historical Climatology Network monthly temperature dataset, version 3 (GHCN-Mv3), along with some additional data, are interpolated between stations and extrapolated up to 1200 km into regions with no measurements (Hansen et al. 2010; Lawrimore et al. 2011). SAT anomalies are produced for areas of sea ice by extrapolating SAT data from nearby land stations into sea ice regions (Hansen et al. 2010). Sea ice areas are therefore treated as if they were land areas with a geographical area which changes over time with sea ice extent.

Kriging is another interpolating technique used to produce temperature anomaly datasets. The Berkeley Earth surface temperature dataset is an interpolated dataset of SAT anomalies that uses temperature records from many preexisting datasets and statistical techniques to produce a globally complete estimate of temperature anomalies over land and ocean areas (Berkeley Earth 2014a; Rohde et al. 2013a,b; Muller et al. 2013; Rohde et al. 2012). The Berkeley dataset method utilizes 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. The interpolating technique of simple kriging is used as part of the Berkeley method along with other statistical techniques (Rohde et al. 2013a,b; Muller et al. 2013; Rohde et al. 2012). A recent update to the dataset adds HadSST data to the Berkeley land-area dataset and extrapolates SAT data from the land-area dataset over sea ice to make the Berkeley dataset globally complete (Berkeley Earth 2014a,b). The Berkeley dataset is therefore spatially complete in the Arctic and sea ice areas are treated as if they were land areas.

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 HadCRUT4 method of estimation and linear interpolation in the style of 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, 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. The degree of difference arising from using these different estimation techniques to estimate Arctic SAT anomalies was explored using European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) data as a test bed. ERA-Interim has been found to be consistent with independent observations of Arctic SATs and provides realistic estimates of Arctic temperatures and temperature trends that outperform, or are comparable to, other currently available reanalyses for all areas of the Arctic so far investigated (Chung et al. 2013; Dee and Uppala 2009; Jakobson et al. 2012; Lindsay et al. 2014; Lüpkes et al. 2010; Screen and Simmonds 2011). Therefore ERA-Interim data were identified as a suitable test bed for this study. Two investigations were undertaken. First, 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. Second, 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 paper is as follows. Section 2 describes the data and techniques used in this study. Section 3 evaluates estimation technique performance in recent decades. Section 4 analyses the effect of the historical coverage of meteorological stations. The final section discusses the results and provides a summary and conclusions.

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 2c.

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.

The estimation techniques yield estimated anomalies—that is, 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 that 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 noninterpolating 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.

a. 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 2-m air temperature data between 1979 and 2011 relative to a 10-yr climatology (1990–99). A 10-yr climatology was used instead of the conventional 30-yr climatology in order to simulate the use of a conventional climatology in a longer dataset using ERA-Interim data covering only 33 years. This was necessary if the performance of the estimation techniques outside of the climatology 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–February (DJF)], spring [March–May (MAM)], summer [June–August (JJA)], and autumn [September–November (SON)]. 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.

b. 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 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 CRUTEM4 dataset version used in this study was 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 that 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. 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 visualization purposes, the locations of the identified, nonduplicated stations are shown in Fig. 3. Monthly anomalies from the ERA-Interim grid cell containing each identified meteorological station were extracted from the reference dataset. These station anomaly time series were masked and used to create input anomaly datasets for estimating anomalies in recent decades and using historical coverages.

· 53°N - 65°N, not reporting 1979-2011   • > 65°N, not reporting 1979-2011

+ 53°N - 65°N, reports 1979-2011   ■ > 65°N, reports 1979-2011
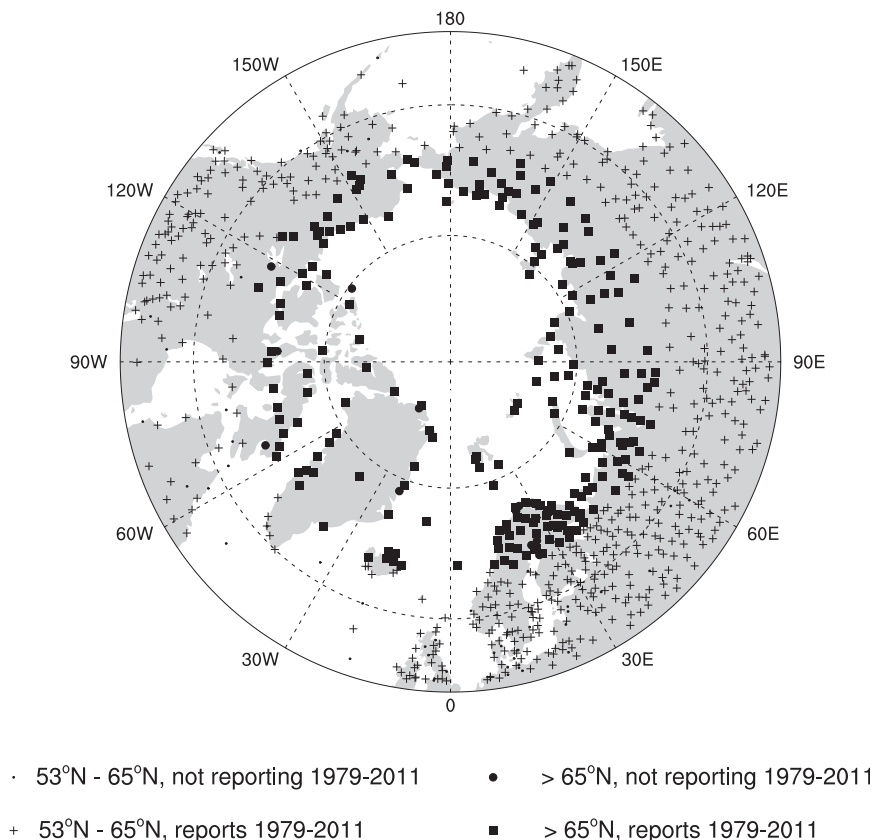
FIG. 3. The locations of all meteorological stations in the CRUTEM4 databank. Different markers are used to show the locations of the meteorological stations depending on whether they are above 53°N or above 65°N and whether or not they report a single temperature between 1979 and 2011.

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 that 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 Fig. 3.

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 (1850–2011) to the estimation of Arctic anomalies for all years of the ERA-Interim period (1979–2011). For each year's historical coverage this gives an ensemble of 33 results (ensemble members) whose statistical properties can be investigated. Each ensemble member comprises a set of repeated instances of each 12-month year between 1979 and 2011, therefore 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 Fig. 3.

c. Estimation techniques

1) LINEAR INTERPOLATION

LI (also known as kernel smoothing using a conical filter) 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 1200-km radius of that grid cell. The weight w decreases linearly from unity for input anomalies located at the grid cell being analyzed (distance d between the input anomaly and the grid point being analyzed is 0 km) to zero for input anomalies located more than or equal to the radius away (1200 km) [Eq. (1)]. This weighting method and the 1200-km radius is the same as used for GISTEMP:

$$w = 1 - \frac{d}{\text{Radius}}, \quad \text{when} \quad d \le \text{Radius}. \tag{1}$$

2) KRIGING

Kriging refers to a set of geostatistical methods of interpolation that 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 this section. A full description of the equations and variables used for both kriging methods is available in the appendix. A summary of the approach detailed in the appendix is as follows:

- A semivariance function was determined from the input anomalies. This function prescribed the weighting for this interpolation; stations up to 3585.9 km away can contribute to the kriged anomaly and the station weighting halves at about 1200 km.
- 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 analyzed.
- 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.

(i) Global ordinary kriging

The first kriging method investigated is 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.

(ii) 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 that includes a simple kriging technique. In the Berkeley method the temperature anomaly field over land is conceptualized 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) NONINTERPOLATING TECHNIQUES

Noninterpolating techniques were the final techniques applied in this study. In this study "noninterpolation" techniques are those that 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 HadSST3 use noninterpolating techniques. In this study two noninterpolating techniques were explored. Comparing the results from these two techniques allows us to explore the impact of spreading information through the use of 5° grid boxes.

(i) The binning technique

The first technique, the binning technique, is similar to the technique used by the CRUTEM4 dataset. The input anomalies were gridded to the 5° × 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).

(ii) Not interpolating

In the second technique, NI, input anomalies were treated as estimated anomalies and analyzed without modification.

d. 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. These monthly anomalies were used to produce estimates of spatially resolved Arctic anomalies and Arctic-average anomalies at monthly, seasonal, and annual time scales 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. Estimated monthly, seasonal, and annual Arctic-average anomalies were produced by area-weighting the estimated anomalies at these time scales 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.

e. 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 time scales. Two error metrics were calculated: the root-mean-square error (RMSE) and compound relative error (CRE). RMSE is a metric that measures the absolute error. CRE measures the relative error (i.e., the error variance as a fraction of the expected variance) [Eq. (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:

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

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. The performance of estimation techniques in recent decades

The performance of all estimation techniques, both interpolating and noninterpolating, 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 analyzed by comparing the estimated anomalies to the reference anomalies.

a. Arctic-average anomalies

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

The time series for annual anomalies are shown in Fig. 4. NI was the least accurate technique for annual Arctic-average anomalies. NI produced estimated annual anomalies with errors of up to nearly 1 K, whereas the errors for all other estimation techniques were below 0.60 K. Interpolating techniques were more accurate than noninterpolating techniques; their errors were below 0.30 K. The errors produced by the different interpolating techniques were very similar to each other. Figure 5 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 noninterpolating techniques, with errors generally below 1 K, 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 4 K 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 (Fig. 5). This seasonality arises from seasonality in
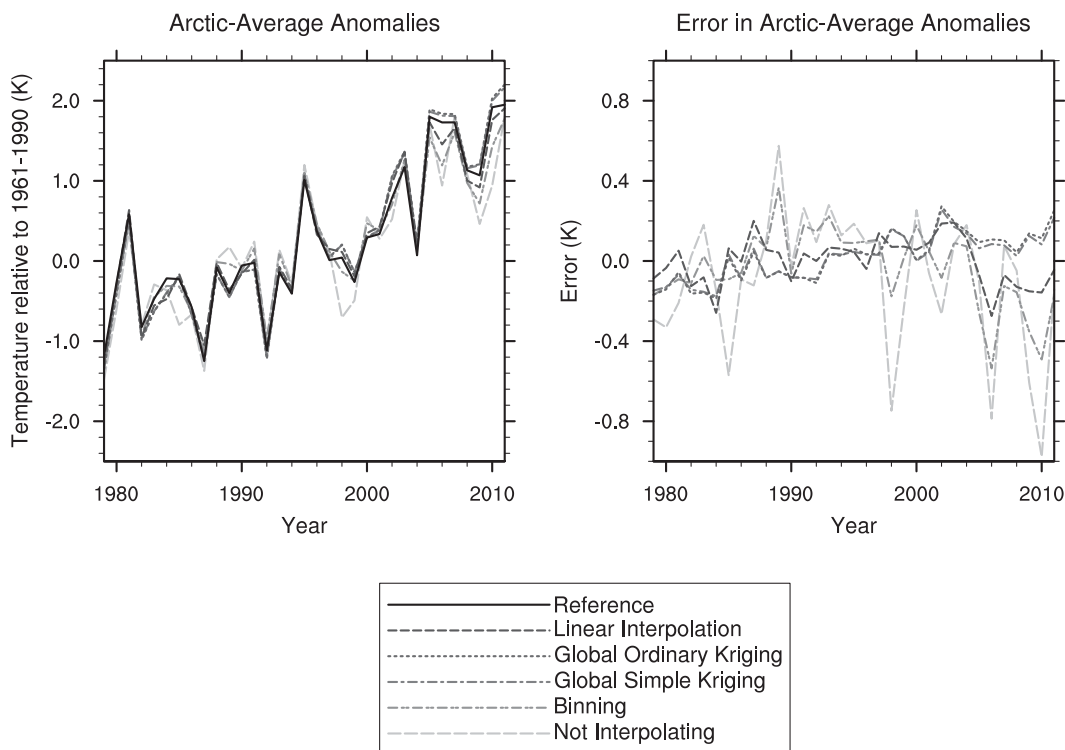
FIG. 4. (left) Time series of annual Arctic-average anomalies between 1979 and 2011 produced using each estimation technique investigated in this study and from reference anomalies. (right) The errors in estimated anomalies relative to the reference anomalies.

Arctic temperatures. Figure 6 is a box-and-whisker plot of monthly area-weighted Arctic SATs over land and sea ice areas from ERA-Interim. As shown in Fig. 6, 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 2003). 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 2003). 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 by producing a Taylor diagram from the errors in estimated annual and monthly Arctic-average anomalies (Fig. 7). Taylor diagrams are a way of graphically summarizing how well estimated variables match a reference dataset. The Taylor diagram confirms the results mentioned above

for estimated 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 Fig. 4. 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). But as both of the interpolation
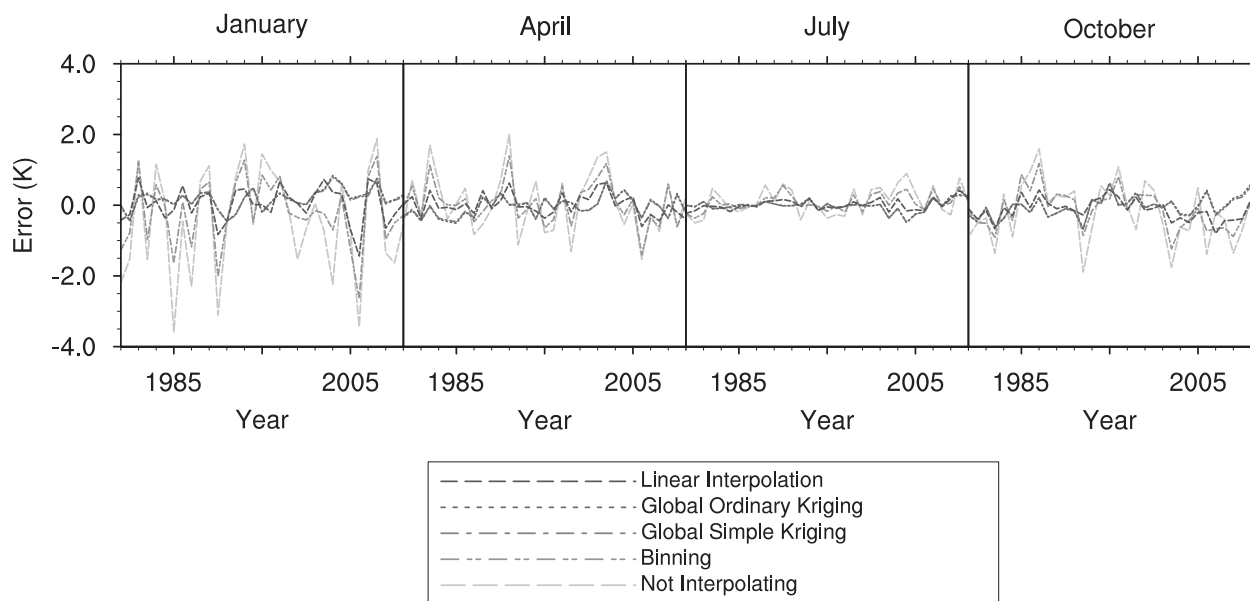
FIG. 5. Time series of the errors in estimated monthly Arctic-average anomalies relative to the reference anomalies between 1979 and 2011 for each estimation technique investigated in this study. One representative month for each season is shown.

techniques investigated here use fixed variance functions for all grid cells and for all time steps, the changes in Arctic sea ice areas resulting from 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 Fig. 1. LI may be more sensitive to reductions in temperature record coverage than kriging techniques. This is explored more fully in section 4b.

To summarize, the interpolating techniques used in this study provided a more accurate estimate of Arctic-average anomalies than noninterpolating 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.

b. 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 8 shows the mapped RMSE and CRE for estimated annual anomalies. All interpolating techniques investigated produced estimates of annual anomalies with RMSEs below 2 K for more than 99% of grid cells. For monthly anomalies (not shown), RMSEs were below 2 K for 47%–99% of grid cells depending on the month of

the year; on average 83% of grid cells had RMSEs below 2 K. Therefore all interpolating techniques investigated produced estimates of monthly and annual anomalies that were, for the majority of grid cells, within 2 K of the reference. The area-weighted average of the RMSE and CRE across the Arctic was calculated for estimated
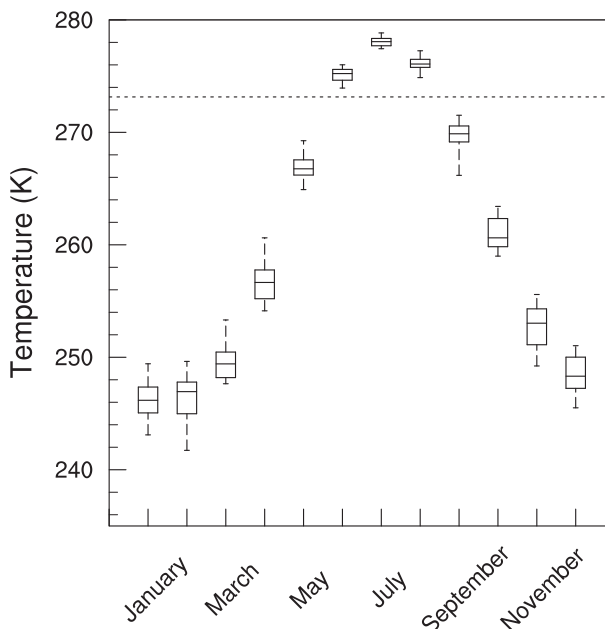

FIG. 6. A box-and-whisker plot of the range, median, and lower and upper quartiles of monthly area-weighted Arctic SAT averaged over land and sea ice from ERA-Interim between 1979 and 2011. A reference line is included at 273.13 K.
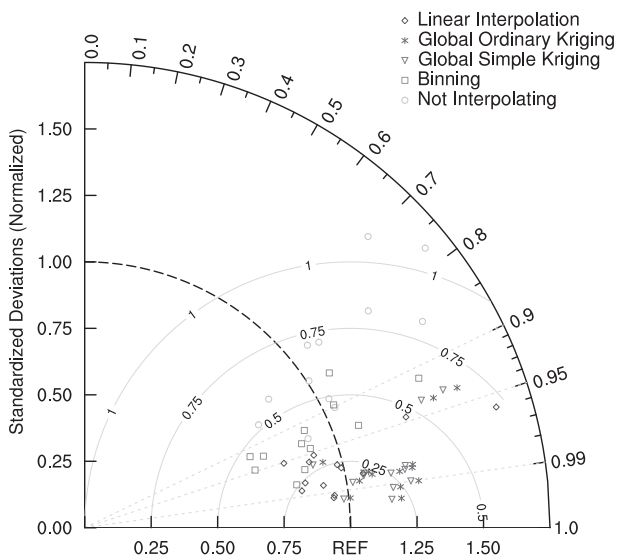
FIG. 7. A Taylor diagram comparing estimated Arctic-average monthly and annual anomalies produced by each estimation technique investigated in this study to the reference anomalies. Each symbol plotted represents a month of the year or the annual value. Cross correlation is shown by the angle with respect to the x axis. The standard deviations (normalized with respect to the reference standard deviation) can be read from the y axis. The RMSE (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 the binning technique are off the scale of this diagram.

annual and monthly anomalies using a cosine of latitude weighting. Figure 9 shows the monthly and annual area-weighted averages for RMSE and CRE. All interpolating techniques produced estimated anomalies that were, on average, within 2 K of the ERA-Interim reference as shown in Fig. 9. 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 3a. 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 Fig. 9. The annual and monthly anomaly error metrics for LI were up to 0.15 and 0.77 K larger 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 Fig. 8. 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 0.10 K. In addition, at the grid cell scale, GSK errors were, for the majority of grid cells in the Arctic, within 0.10 K of GOK errors (Fig. 8). However, when the maps of RMSE and CRE in Fig. 8 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 subregions of the Arctic cancel each other out for Arctic-average anomalies estimated using interpolation techniques. Comparing the errors spatially avoids this cancellation of errors and therefore emphasizes differences in the performance of the interpolation techniques.

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 analyzed by comparing the
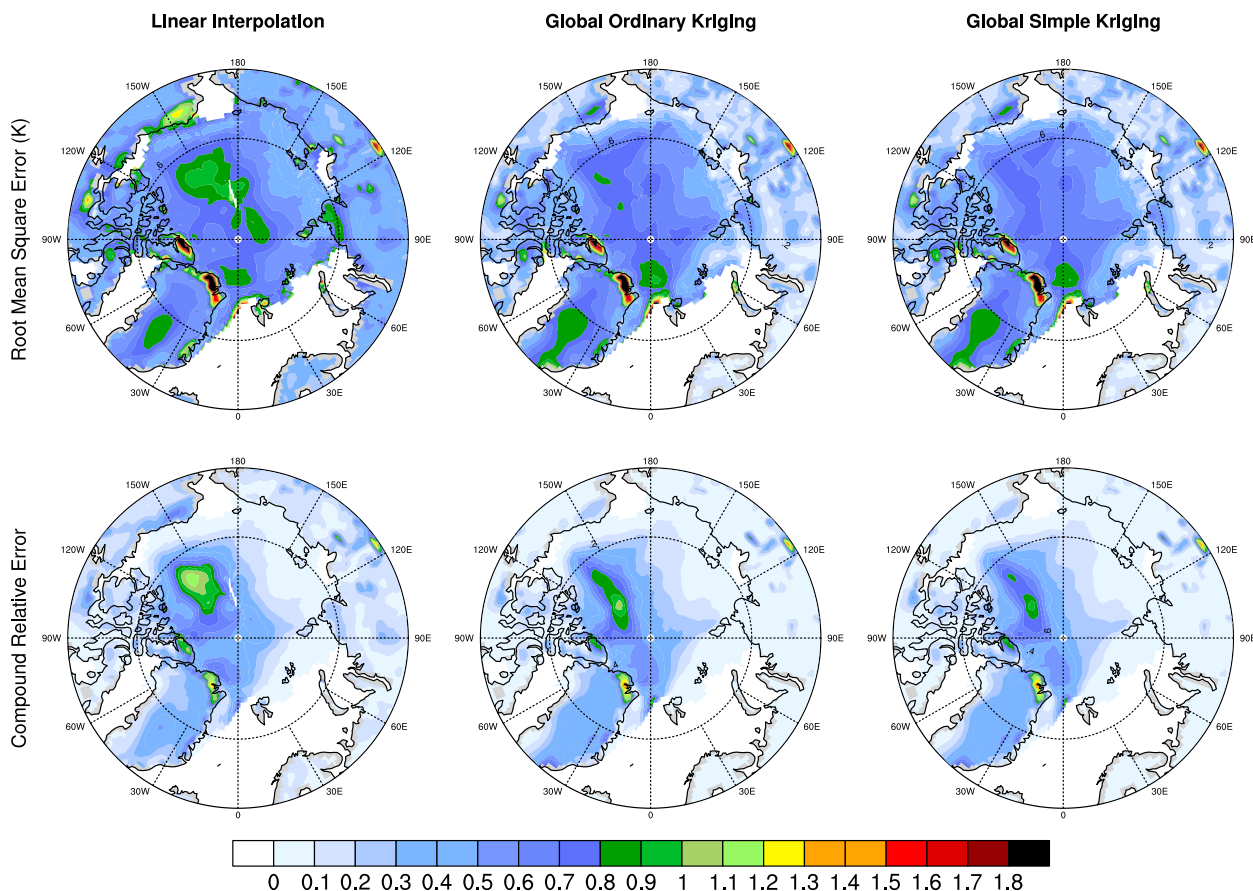
FIG. 8. The RMSE and CRE between spatially resolved annual Arctic anomalies estimated using the investigated interpolating techniques and reference anomalies in recent decades (1979–2011). CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

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 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.

a. Relative performance of estimation techniques

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 Fig. 10. The RMSE and CRE across ensemble members for each year of historical coverage are shown in Fig. 11.

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 noninterpolating methods. For historical coverages before 1930 the RMSE and CRE were generally smallest for kriging methods; kriging error metrics were smaller than 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
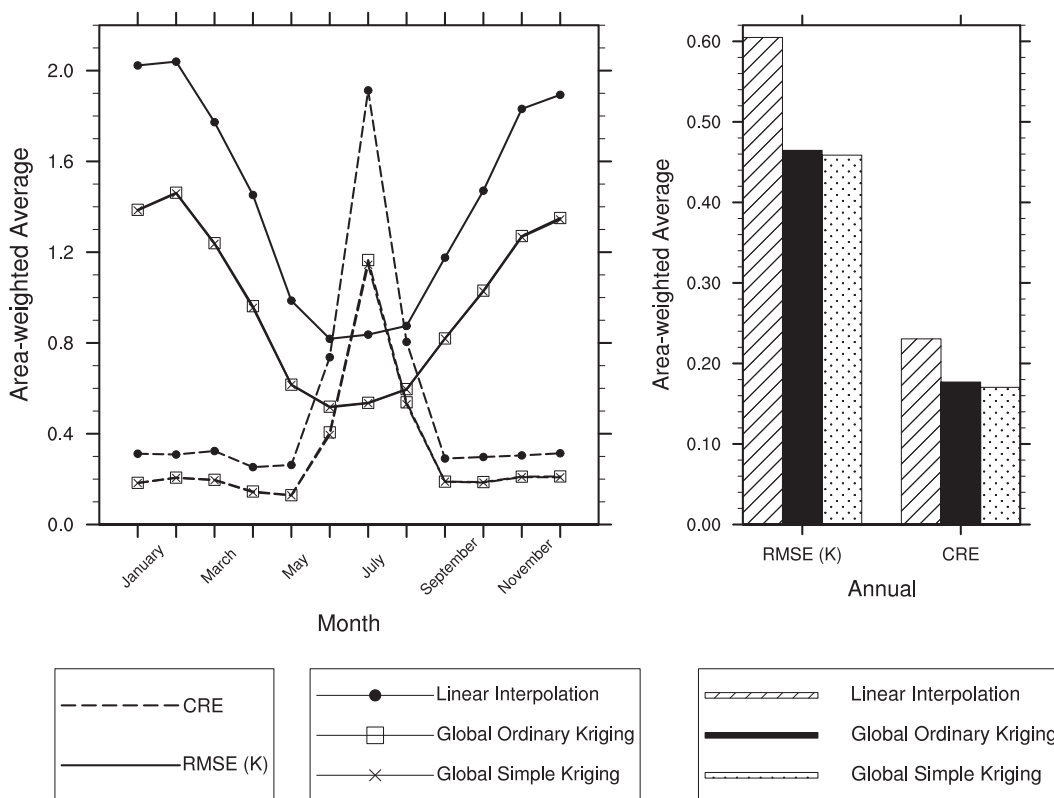
FIG. 9. The area-weighted average of the RMSE and CRE between estimated spatially resolved Arctic anomalies (estimated using the investigated interpolating techniques) and reference anomalies in recent decades (1979–2011): (left) monthly anomalies and (right) annual anomalies. CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

estimates which were more representative (Fig. 11). 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 Fig. 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 3a. Nevertheless, the errors for the techniques are, for the majority of years, within 0.20 K 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.

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 Fig. 12.

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 0.2–0.4 K. 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 0.20 K 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. But when maps of the RMSE were examined, the performance of the kriging
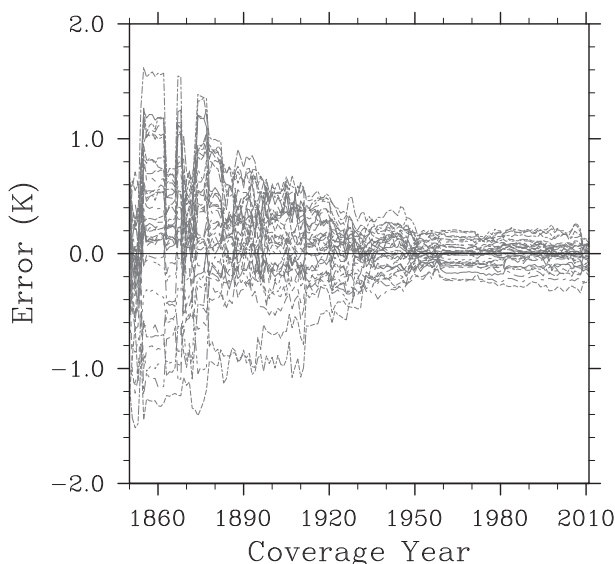
FIG. 10. The error in annual Arctic-average anomalies estimated by linearly interpolating each year of ERA-Interim anomalies (1979–2011, each year is shown by one line) masked using historical station coverages (1850–2011). (Similar graphs for all estimation techniques and seasons are provided in the supplemental material.)

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 Fig. 13. After this GSK was more representative of SATs over Arctic sea ice.

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.

b. 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 (Fig. 11) 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 2 K 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 (Fig. 12) 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 2.5 K 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. 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.

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 noninterpolating 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
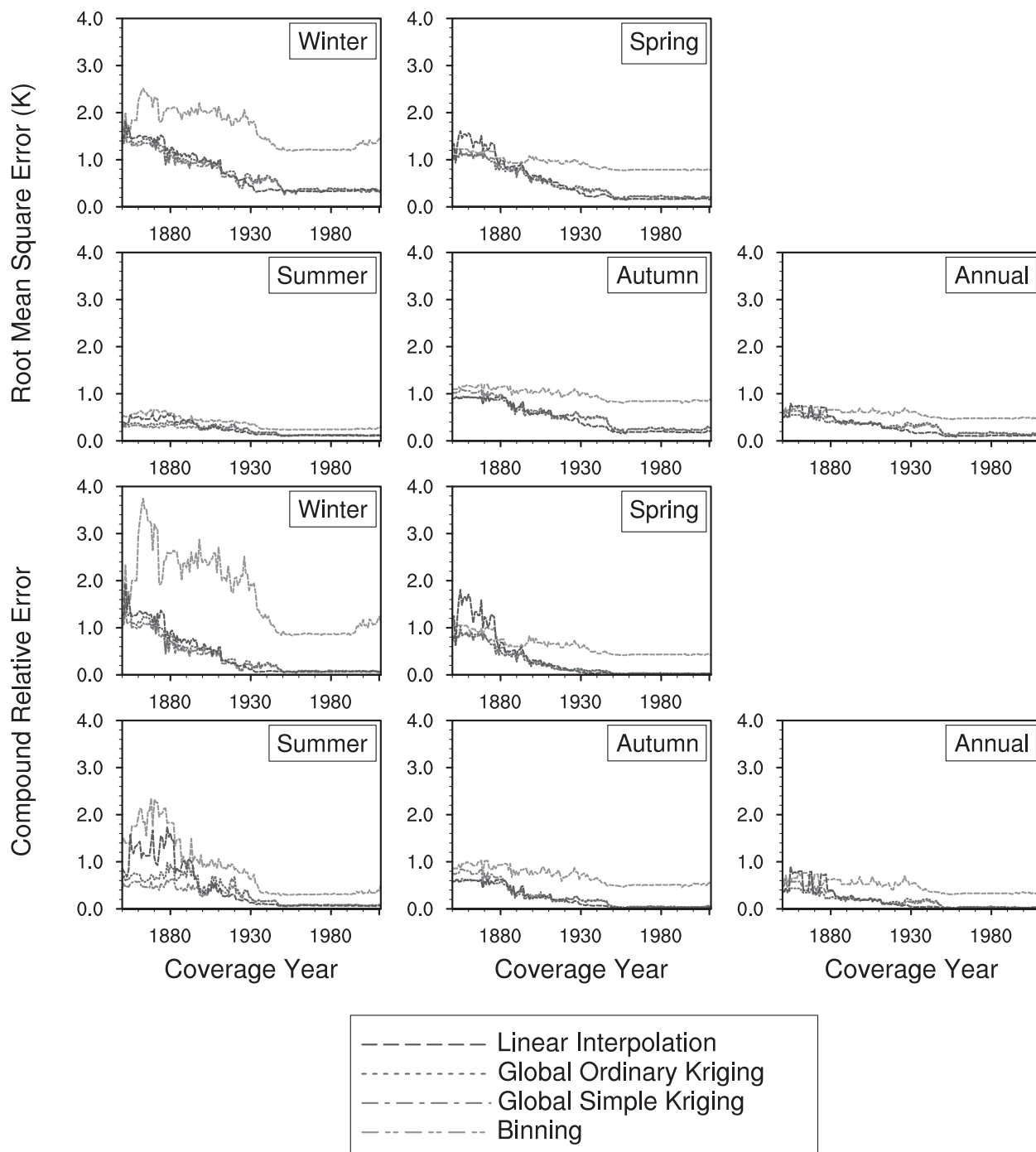
FIG. 11. The RMSE (K) and CRE across ensemble members (each year of ERA-Interim anomalies 1979–2011) in each historical coverage year for estimated seasonal and annual Arctic-average anomalies. CRE is a unitless metric where 0 is the best result and higher numbers represent a higher relative error.

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
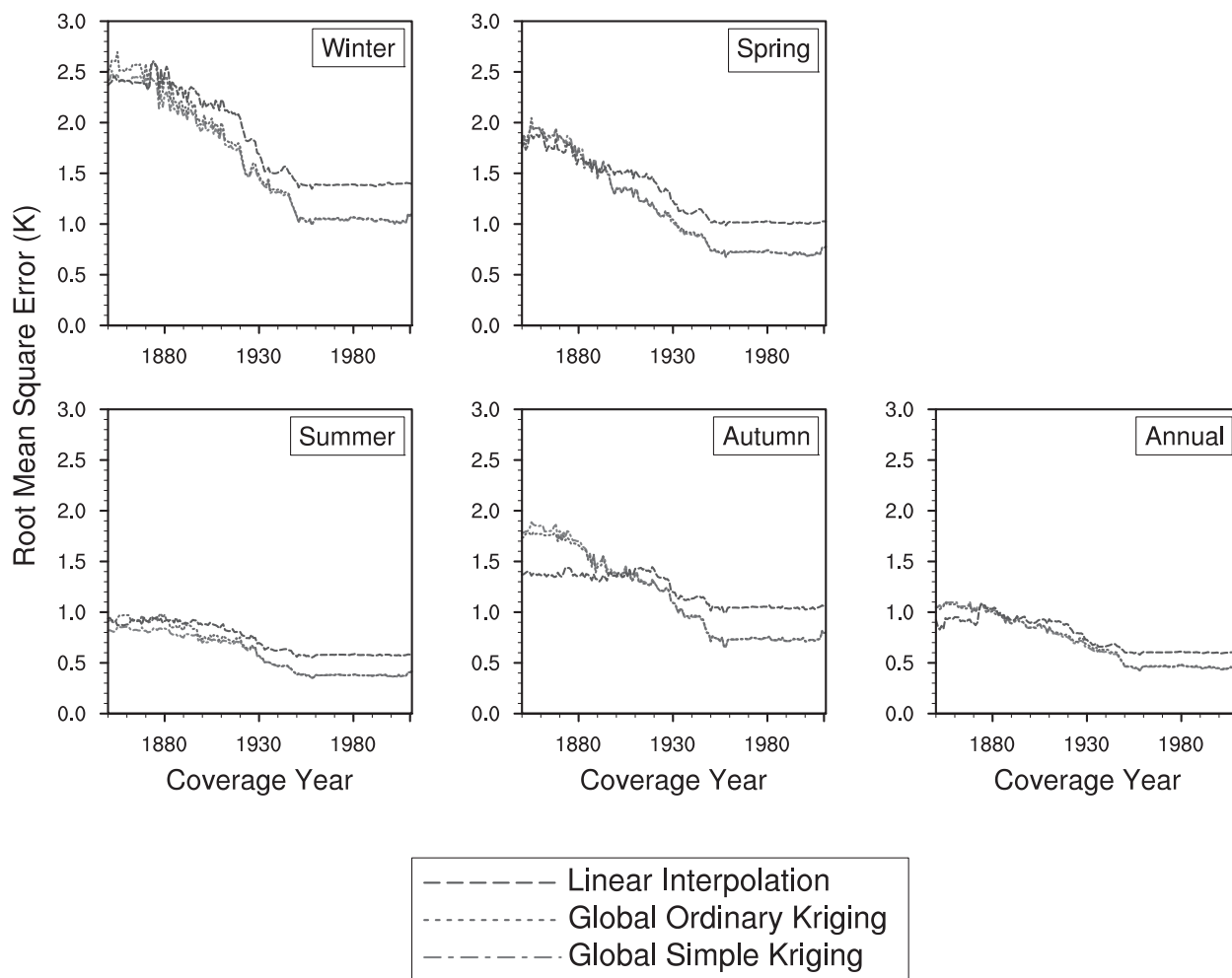
FIG. 12. The area-weighted RMSE (K) for spatially resolved Arctic anomalies across ensemble members (each year of ERA-Interim anomalies 1979–2011) produced by the investigated interpolating techniques.

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. First, 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. Second, in sparsely observed regions where the distance to the nearest station is large, kriging techniques relax toward a prior value, here either 0 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 toward a prior value and this increases the influence of a small

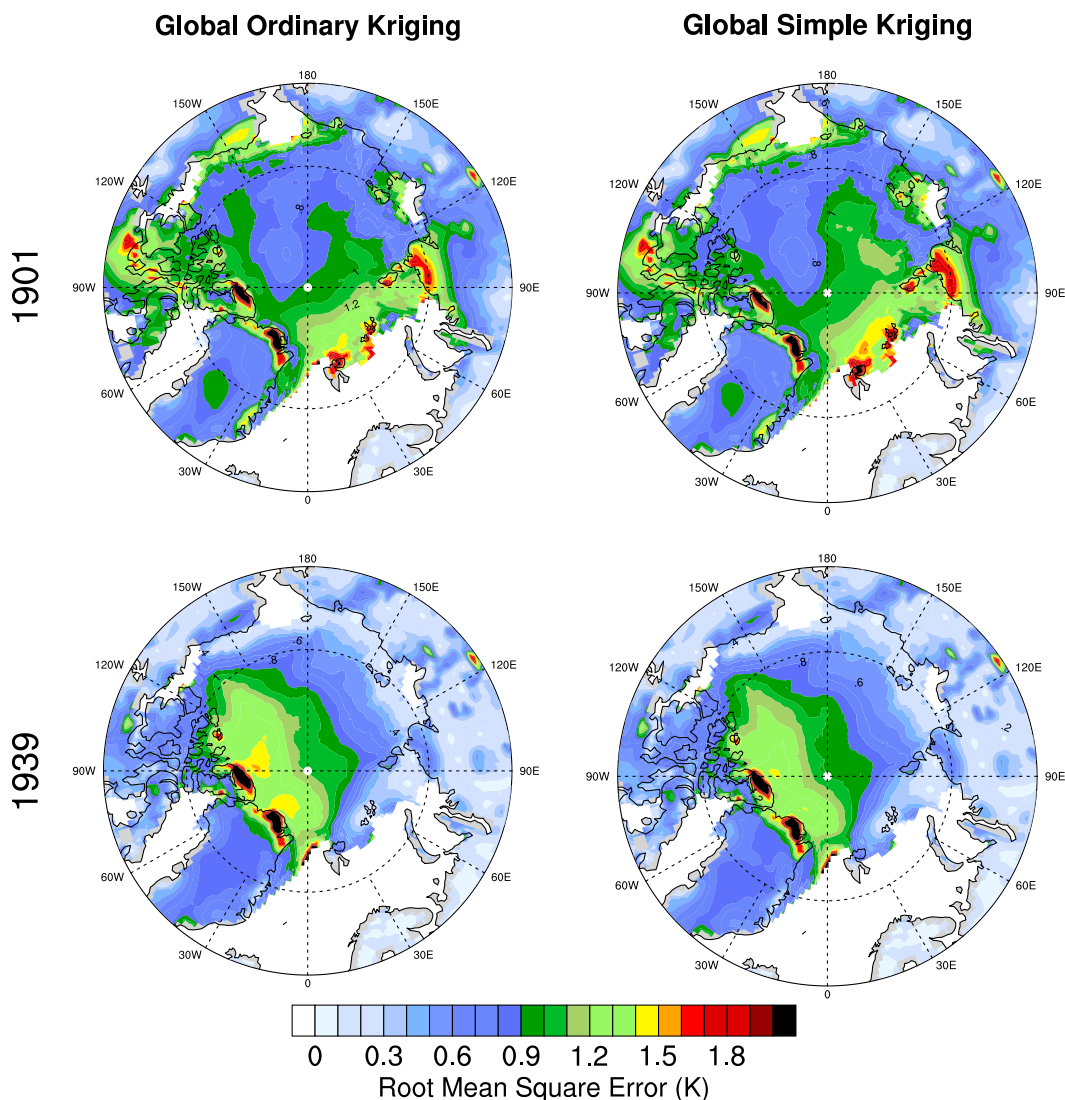**Global Ordinary Kriging**   **Global Simple Kriging**

FIG. 13. The mapped RMSE (K) across an ensemble of 33 different years of ERA-Interim data (1979–2011) for spatially resolved annual Arctic anomalies produced using GOK and GSK for two example historical coverage years: (top) 1901 and (bottom) 1939.

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-yr climatology used in this study but may be less representative for 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 noninterpolating 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 3585.9 km (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 that 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 consists of 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, such as 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. 2011; 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. Moreover, the purpose of this work was to investigate the impact of using certain estimation techniques, related to current temperature anomaly datasets, to estimate temperatures over Arctic land and sea ice areas. Investigating other estimation techniques, which may produce more accurate reconstructions, and the estimation of temperatures over other Arctic areas were beyond the scope of this study. Before conclusions are drawn

about the performance of estimation techniques, and the effects of changing station coverage, for Arctic temperatures and temperature changes in the real world further work is needed. The gap between the reanalysis and the reality in terms of the results presented here should be assessed; parametric, observational, and structural uncertainty should be further investigated; the impact of using monthly or seasonal covariance values should be examined; and the reconstruction of anomalies in areas of open ocean in the Arctic, as well as marginal ice zones, should be studied.

6. Conclusions

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 test bed. 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. Noninterpolating 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.

APPENDIX

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 for $i = 1, \ldots, n$ [calculated using Eq. (A1)], where $\mathbf{w}$ is a vector containing the optimal weights $w(x_i)$ for each observation of variable t at coordinates x_i for $i = 1, \ldots, n$, and $\mathbf{t}$ is a vector containing the observations of variable t at coordinates x_i for $i = 1, \ldots, n$:

$$\hat{t}(x_o) = \mathbf{w} \cdot \mathbf{t}, \qquad (A1)$$

where

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

The optimal weights in vector $\mathbf{w}$ are calculated using Eq. (A2). In that equation, $\mathbf{A}$ is a matrix of the expected covariance between each pair of observations of variable t at coordinates x_i for $i = 1, \ldots, n$ 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 for $i = 1, \ldots, n$ and the output point x_o so that $b(x_o, x_n)$ is the expected covariance between $t(x_o)$ and $t(x_n)$:

$$\mathbf{w} = \mathbf{A}^{-1}\mathbf{B}. \qquad (A2)$$

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. Therefore, for ordinary kriging the optimal weights in vector $\mathbf{w}$ are calculated using Eq. (A2) with the vectors and matrices shown below. Here ϕ is a Lagrange multiplier required for solving the equations.

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

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 **t** of the observations of variable t to produce an estimate of $t(x_o)$. Note that, because of the use of a mean of 0 for global simple kriging in this study, the same equation [Eq. (A1)] is used for both kriging methods. Therefore the optimal weights in vector **w** are calculated using Eq. (A2) with the vectors and matrices shown below:

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

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. (A3) with fitted parameter values of 0 for s, 7.6418 for y, and 3585.9 for ϵ. Also, d is the distance in kilometers and $\gamma(d)$ is the model semivariance at distance d,

$$\gamma(d) = \begin{cases} s + y\left[\frac{3}{2}\left(\frac{d}{\epsilon}\right) - \frac{1}{2}\left(\frac{d}{\epsilon}\right)^3\right] & \text{for} \quad 0 < d < \epsilon \\ s + y & \text{for} \quad d \geq \epsilon \end{cases}. \tag{A3}$$

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. (A2) to calculate the weights required for kriging.

REFERENCES

ACIA, 2005: *Arctic Climate Impact Assessment.* Cambridge University Press, 1046 pp.

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. *Bull. Amer. Meteor. Soc.,* **92,** 1421–1425, doi:10.1175/2011BAMS3218.1.

Bekryaev, R. V., I. V. Polyakov, and V. A. Alexeev, 2010: Role of polar amplification in long-term surface air temperature variations and modern Arctic warming. *J. Climate,* **23,** 3888–3906, doi:10.1175/2010JCLI3297.1.

Berkeley Earth, 2014a: Land + ocean (1850–recent): Monthly global average temperature. Berkeley Earth. [Available online at http://berkeleyearth.lbl.gov/auto/Global/Land_and_Ocean_complete.txt.]

——, 2014b: Land + ocean data. Berkeley Earth. [Available online at http://berkeleyearth.org/land-and-ocean-data.]

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.,* **111,** D12106, doi:10.1029/2005JD006548.

Chung, C., 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. *Atmos. Chem. Phys.,* **13,** 11 209–11 219, doi:10.5194/acp-13-11209-2013.

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. *Environ. Res. Lett.,* **7,** 014007, doi:10.1088/1748-9326/7/1/014007.

Comiso, J. C., 2012: Large decadal decline of the Arctic multiyear ice cover. *J. Climate,* **25,** 1176–1193, doi:10.1175/JCLI-D-11-00113.1.

——, C. L. Parkinson, R. Gersten, and L. Stock, 2008: Accelerated decline in the Arctic sea ice cover. *Geophys. Res. Lett.,* **35,** L01703, doi:10.1029/2007GL031972.

Cowtan, K., and R. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.,* **140,** 1935–1944, doi:10.1002/qj.2297.

Cressie, N., 1990: The origins of kriging. *Math. Geol.,* **22,** 239–252, doi:10.1007/BF00889887.

Dee, D. P., and S. Uppala, 2009: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. Roy. Meteor. Soc.,* **135,** 1830–1841, doi:10.1002/qj.493.

——, and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.,* **137,** 553–597, doi:10.1002/qj.828.

GISS, 2014: GISS surface temperature analysis (GISTEMP): Updates to analysis. [Available online at http://data.giss.nasa.gov/gistemp/updates_v3/.]

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.,* **48,** RG4004, doi:10.1029/2010RG000345.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei, 2008: Comparison of six methods for the interpolation of daily, European climate data. *J. Geophys. Res.,* **113,** D21110, doi:10.1029/2008JD010100.

Inoue, J., T. Enomoto, T. Miyoshi, and S. Yamane, 2009: Impact of observations from Arctic drifting buoys on the reanalysis of surface fields. *Geophys. Res. Lett.,* **36,** L08501, doi:10.1029/2009GL037380.

Jakobson, E., T. Vihma, T. Palo, L. Jakobson, H. Keernik, and J. Jaagus, 2012: Validation of atmospheric reanalyses over the central Arctic Ocean. *Geophys. Res. Lett.,* **39,** L10802, doi:10.1029/2012GL051591.

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. *J. Geophys. Res.,* **117,** D05127, doi:10.1029/2011JD017139.

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. *J. Geophys. Res.,* **103,** 18 567–18 589, doi:10.1029/97JC01736.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.,* **116,** D14103, doi:10.1029/2010JD015218.

——, ——, ——, ——, and ——, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.,* **116,** D14104, doi:10.1029/2010JD015220.

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. *J. Geophys. Res.,* **116,** D19121, doi:10.1029/2011JD016187.

Lindsay, R., M. Wensnahan, A. Schweiger, and J. Zhang, 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. *J. Climate,* **27,** 2588–2606, doi:10.1175/JCLI-D-13-00014.1.

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. *J. Geophys. Res.,* **113,** D10105, doi:10.1029/2007JD009380.

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. *Geophys. Res. Lett.,* **37,** L09810, doi:10.1029/2010GL042724 .

Maslanik, J. A., 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. *Geophys. Res. Lett.,* **34,** L24501, doi:10.1029/2007GL032043.

Merchant, C. J., and Coauthors, 2013: The surface temperatures of earth: Steps towards integrated understanding of variability and change. *Geosci. Instrum. Methods Data Syst.,* **2,** 305–321, doi:10.5194/gi-2-305-2013.

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. *J. Geophys. Res.,* **117,** D08101, doi:10.1029/2011JD017187.

Muller, R., and Coauthors, 2013: Decadal variations in the global atmospheric land temperatures. *J. Geophys. Res. Atmos.,* **118,** 5280–5286, doi:10.1002/jgrd.50458.

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. *J. Climate,* **13,** 2217–2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2.

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. *J. Geophys. Res.,* **114,** D05104, doi:10.1029/2008JD010450.

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. *Geophys. Res. Lett.,* **40,** 1356–1361, doi:10.1002/grl.50349.

Pielke, R. A., and Coauthors, 2007: Unresolved issues with the assessment of multidecadal global land surface temperature trends. *J. Geophys. Res.,* **112,** D24S08, doi:10.1029/2006JD008229.

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. *Agric. For. Meteor.,* **101,** 81–94, doi:10.1016/S0168-1923(99)00169-0.

Przybylak, R., 2003: *The Climate of the Arctic.* Kluwer Academic, 272 pp.

Rohde, R., 2013: Comparison of Berkeley Earth, NASA GISS, and Hadley CRU averaging techniques on ideal synthetic data. Berkeley Earth Tech. Rep., 15 pp. [Available online at http://static.berkeleyearth.org/memos/robert-rohde-memo.pdf.]

——, and Coauthors, 2012: Appendix to Berkeley Earth temperature averaging process: Details of mathematical and statistical methods. Berkeley Earth Tech. Rep., 19 pp. [Available online at http://scitechnol.com/2327-4581/2327-4581-1-103a.pdf.]

——, and Coauthors, 2013a: Berkeley Earth temperature averaging process. *Geoinfo. Geostat. Overview,* **1** (2), doi:10.4172/2327-4581.1000103.

——, and Coauthors, 2013b: A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinfo. Geostat. Overview,* **1** (1), doi:10.4172/2327-4581.1000101.

Screen, J. A., and I. Simmonds, 2010: The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature,* **464,** 1334–1337, doi:10.1038/nature09051.

——, and ——, 2011: Erroneous Arctic temperature trends in the ERA-40 reanalysis: A closer look. *J. Climate,* **24,** 2620–2627, doi:10.1175/2010JCLI4054.1.

Serreze, M. C., and R. G. Barry, 2011: Processes and impacts of Arctic amplification: A research synthesis. *Global Planet. Change,* **77,** 85–96, doi:10.1016/j.gloplacha.2011.03.004.

——, A. Barrett, J. Stroeve, D. Kindig, and M. Holland, 2009: The emergence of surface-based Arctic amplification. *Cryosphere,* **3,** 11–19, doi:10.5194/tc-3-11-2009.

Slonosky, V., 2014: Canadian historical data typing project. [Available online at https://sites.google.com/site/historicalclimatedata/canadian-historical-data-typing-project.]

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperature analysis. *J. Climate,* **21,** 2283–2296, doi:10.1175/2007JCLI2100.1.

Stroeve, J. C., M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, 2012: The Arctic's rapidly shrinking sea ice cover: A research synthesis. *Climatic Change,* **110,** 1005–1027, doi:10.1007/s10584-011-0101-1.

Thorne, P. W., and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends are they really achievable? *Bull. Amer. Meteor. Soc.,* **91,** 353–361, doi:10.1175/2009BAMS2858.1.

——, and Coauthors, 2011: Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bull. Amer. Meteor. Soc.,* **92,** ES40–ES47, doi:10.1175/2011BAMS3124.1.

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 Res.,* **24,** 161–180, doi:10.3354/cr024161.