

# Comparison of *in situ* time-series of temperature with gridded sea surface temperature datasets in the North Atlantic

Sarah L. Hughes, N. Penny Holliday, Eugene Colbourne, Vladimir Ozhigin, Hedinn Valdimarsson, Svein Østerhus, and Karen Wiltshire

Hughes, S. L., Holliday, N. P., Colbourne, E., Ozhigin, V., Valdimarsson, H., Østerhus, S., and Wiltshire, K. 2009. Comparison of *in situ* time-series of temperature with gridded sea surface temperature datasets in the North Atlantic. – ICES Journal of Marine Science, 66: 1467–1479.

Analysis of the effects of climate variability and climate change on the marine ecosystem is difficult in regions where long-term observations of ocean temperature are sparse or unavailable. Gridded sea surface temperature (SST) products, based on a combination of satellite and *in situ* observations, can be used to examine variability and long-term trends because they provide better spatial coverage than the limited sets of long *in situ* time-series. SST data from three gridded products (Reynolds/NCEP OISST.v2., Reynolds ERSST.v3, and the Hadley Centre HadISST1) are compared with long time-series of *in situ* measurements from ICES standard sections in the North Atlantic and Nordic Seas. The variability and trends derived from the two data sources are examined, and the usefulness of the products as a proxy for subsurface conditions is discussed.

**Keywords:** climatology, North Atlantic, sea surface temperature, time-series.

Received 15 August 2008; accepted 25 January 2009; advance access publication 15 March 2009.

S. L. Hughes: Fisheries Research Services, Marine Laboratory, Aberdeen, UK. N. P. Holliday: National Oceanography Centre, Southampton, UK. E. Colbourne: Northwest Atlantic Fisheries Centre, St Johns, Canada. V. Ozhigin: Knipovich Polar Research Institute of Marine Fisheries and Oceanography, Murmansk, Russia. H. Valdimarsson: Marine Research Institute, Reykjavík, Iceland. S. Østerhus: Bjerknes Centre for Climate Research, Bergen, Norway. K. Wiltshire: Alfred Wegener Institute, Bremerhaven, Germany. Correspondence to S. L. Hughes: tel: +44 1224 876544; fax: +44 1224 295511; e-mail: s.hughes@marlab.ac.uk.

## Introduction

To study the effects of climate variability on the structure and function of marine ecosystems requires clear indicators of the changing physical nature of the ocean. Hydrographic conditions, and their spatial and temporal variations, are known to have significant influence on biological processes. Long time-series of directly measured ocean temperature, salinity, and currents are extremely sparse and access to these data is often limited. As a result, in studies of change at all trophic levels (from plankton to seabirds), indices derived from widely available datasets, such as those for sea surface temperature (SST), are frequently used as a measure of upper ocean conditions. A researcher wishing to investigate the influence of climate variability on marine ecosystems may then face the question: which of the available environmental indices are most appropriate?

In an attempt to answer this question, we have chosen to investigate data from three globally complete, interpolated SST products (Table 1), and compare them with *in situ* hydrographic data. The SST products are: HadISST1 developed at the UK Meteorological Office, Hadley Centre for Climate Prediction and Research (Rayner *et al.*, 2003), the Optimum Interpolation Sea Surface Temperature Analysis (OISST.v2) developed by the US National Oceanic and Atmospheric Administration (NOAA; Reynolds *et al.*, 2002), and the Improved Extended Reconstructed SST (ERSST.v3) dataset developed by NOAA (Smith *et al.*, 2008).

At the time of its publication, HadISST1 was the only globally complete SST product with a 1° grid extending back for more than 100 years to 1870. The ERSST.v3, extending back to 1854, offers an alternative to HadISST1, but is only available on a 2° grid. Despite its more limited time-span, one clear advantage of the OISST.v2 product is the weekly output. A daily version at 1/4° resolution (Reynolds *et al.*, 2007) is also available, but has not been investigated here.

Because oceans cover the majority (70%) of the earth's surface, understanding the long-term variability in SST has been a key issue in climate change research (Smith *et al.*, 2008), and the development of gridded SST products has been driven in part by the need to provide data input to general atmospheric circulation models (Hurrell *et al.*, 2008). At the time of writing, a global comparison of gridded SST products was continuing, with the aim of improving the long-term climate data records (<http://ghrsst.nodc.noaa.gov/intercomp.html>).

In the published literature, comparison and analysis of existing SST products have focused on how well the products represent the long-term trends and variability on spatial scales generally ranging from global to ocean basin scale (Reynolds *et al.*, 2002; Rayner *et al.*, 2003). These, and other comparisons (Barton and Casey, 2005; Vecchi and Soden, 2007; Hurrell *et al.*, 2008), offer some insight as to where and why the gridded SST products differ. However, it is difficult to conclude from the comparisons which, if any, of the products are thought to be the best. Uppala

**Table 1.** Summary of key features of the globally complete, gridded SST datasets.

Name	Spatial resolution	Start date	Updated	Underlying datasets	Data formats
ERSST.v3	2° lat. × 2° long.	1854	Monthly	I-COADS dataset to 1997, bias corrected AVHRR, real-time ship, and buoy observations available to the US National Meteorological Centre via GTS	ASCII and NetCDF <sup>a</sup>
OISST.v2	1° lat. × 1° long.	November 1981	Weekly	I-COADS dataset to 1997, bias corrected AVHRR, real-time ship, and buoy observations available to the US National Meteorological Centre via GTS	ASCII and NetCDF (via ftp or OPeNDAP) <sup>b</sup>
HadISST1	1° lat. × 1° long.	1870	Monthly	Met Office Marine Databank, via GTS since 1995, COADS dataset 1871–1995, bias corrected AVHRR since January 1982	ASCII and NetCDF (via ftp, LAS server or OPeNDAP) <sup>c</sup>

<sup>a</sup><http://www.ncdc.noaa.gov/oa/climate/research/sst/ersstv3.php>.<sup>b</sup><http://www.cdc.noaa.gov/cdc/data.noaa.oisst.v2.html>.<sup>c</sup><http://hadobs.metoffice.com/hadisst/>.

*et al.* (2005), in the ERA-40 reanalysis project, used HadISST1 for the period before 1981 with OISST.v2 thereafter. Hurrell *et al.* (2008) followed a similar procedure and state that for their particular application, OISST.v2 is currently the best available SST product.

Although Barton and Casey (2005) suggest that the gridded SST products are underutilized in this manner, they have been used for marine ecosystem/climate variability studies, e.g. a study of climate variability and seabirds in the North Atlantic using OISST.v2 (Sandvik *et al.*, 2005), investigation of climate effects on the growth of the shell *Arctica islandica* in the North Sea using ERSST.v3 (Schone *et al.*, 2005), and assessments of the variability of plankton in North Atlantic and European waters related to HadISST1 (Richardson and Schoeman, 2004; Bonnet *et al.*, 2005).

In this study, we aim to test the following hypotheses: (i) variations in global gridded SST products reflect variations in subsurface conditions, and (ii) global gridded SST products reflect true variations on monthly to decadal time-scales. We posit that these are implicit assumptions made by researchers who make use of gridded SST products as proxies for subsurface conditions.

We will test these hypotheses from a regional perspective (as opposed to global fields), in the context of examining the effects of changes in temperature on biological time-series. The simple approach we take is to compare *in situ* CTD-based subsurface

and near-surface temperature time-series with SST time-series extracted from three different global gridded SST products. Reynolds *et al.* (2002) encourage the use of independent datasets to help quantify the difference between the gridded SST products. Although our study falls short of the rigorous testing methodology suggested by Reynolds *et al.* (2002), we hope that these results can offer some guidance to others on the most appropriate SST product to use for their analysis. We examine the longest time-series from stations and sections in the ICES region of interest, using only data since 1950 to be confident of better data quality (Table 2). All the *in situ* time-series we examine are reported annually in the ICES Report on Ocean Climate (IROC; Hughes and Holliday, 2007) and are increasingly being used in the study of long-term variations in fisheries and ecosystem processes.

## Methods

### Gridded SST products

The three gridded SST products we selected [HadISST1, OISST.v2, and ERSST.v3; Rayner *et al.* (2003), Reynolds *et al.* (2002), and Smith *et al.* (2008)] provide interpolated SST data over the global ocean and include information on sea ice at higher latitudes. Table 1 summarizes the main structural differences between the datasets, illustrating the level of information (i.e. spatial resolution, update frequency, available data formats) likely to be used

**Table 2.** Details of *in situ* datasets, footnotes illustrating contact details of the data provider for each dataset.

Dataset	Depth	Station details	Lat.	Long.	T	σ	Occupation frequency
Kola <sup>a</sup>	0–200 m	3–7	70.5–72.5	33.5	3.92	0.49	12–15 per year before 1980, 7–9 in recent years
Kola <sub>U</sub>	0–50 m	–	–	–	4.71	1.66	
Mike <sup>b</sup>	50 m	Single station	66.0	2.0	7.41	0.33	Weekly over all 12 months
Nlce <sup>c</sup>	50–150 m	2–4	67.0–67.6	–18.8	3.34	1.09	4 occupations per year since 1970, 1 per year before that
FShet <sup>d</sup>	Upper layer (0–200 m)	1–2	61.0	–3.0	9.57	0.15	6–9 occupations per year
Helgo <sup>e</sup>	Surface	Single station	54.2	7.9	10.1	0.72	Almost daily
Newf <sup>f</sup>	0–175 m	Single station	47.5	–52.6	0.27	0.33	40–50 occupations per year in recent years
Newf <sub>U</sub>	0–5 m	–	–	–	4.81	0.68	

The average temperature (*T*) and standard deviation (*σ*) are calculated for the period 1971–2000. The subscript U refers to additional datasets derived from the same station, but calculated for the near-surface layer.

<sup>a</sup>Oleg V. Titov, PINRO—Knipovich Polar Research Institute of Marine Fisheries and Oceanography, Russia.

<sup>b</sup>Svein Østerhus (svein.osterhus@gfi.uib.no), Geophysical Institute, University of Bergen, Norway.

<sup>c</sup>Hedinn Valdimarsson (hv@hafro.is), Hafrannsóknastofnunin, Iceland Marine Research Institute, Iceland.

<sup>d</sup>Sarah Hughes (s.hughes@marlab.ac.uk), FRS—Fisheries Research Services, Aberdeen, UK.

<sup>e</sup>Karen Wiltshire (kwiltshire@awi-bremerhaven.de), AWI/BAH (Alfred-Wegener-Institut/Biologische Anstalt Helgoland), Germany.

<sup>f</sup>Eugene Colbourne (eugene.colbourne@dfo-mpo.gc.ca), Northwest Atlantic Fisheries Centre, Canada.

by researchers to determine which product may be suitable for their purposes.

The OISST product, adding the optimal interpolation methodology, superseded earlier products developed by Reynolds and Marsico (1993), with OISST.v2 updating and improving v1, using better algorithms for satellite bias correction and sea ice. The extended Reynolds SST dataset (ERSST) has been developed through versions 1 and 2, with version 3 released in 2007 (Smith *et al.*, 2008); most relevant to this study are improvements to tuning that improve the analysis in data-sparse regions. At the time of writing, an update to the Hadley Centre SST product was being prepared and for this reason we refer to the dataset investigated here as HadISST1.

The data products are essentially similar, derived from a combination of *in situ* and satellite observations. As well as data from real-time observations in recent years, each of the three products incorporates *in situ* data from the International Comprehensive Atmosphere Ocean Dataset (ICOADS), 1871–1995 (Woodruff *et al.*, 1998) and satellite-based SST observations from the Advanced Very High Resolution Radiometer (AVHRR). Although the HadISST1 dataset incorporates COADS data up to 1995, its main database is the UK Meteorological Office Marine Databank (Rayner *et al.*, 2003). To provide regular updates to the products, new observations for all datasets are obtained in near real time via the Global Telecommunications System (GTS), although the GTS data are collated through different data centres: the US National Meteorological Center for OISST.v2 and ERSST.v3 and the UK Meteorological Office for HadISST1.

A key difference between the datasets is the interpolation technique used to fill gaps in spatial and temporal coverage. The two longer datasets (ERSST.v3 and HadISST1) also have to use reconstruction techniques to fill the data-sparse periods before 1948. These techniques, based on the method of empirical orthogonal functions, assess the pattern of spatial variability from recent periods with good data coverage and use these patterns to fill gaps in earlier periods of sparse observation. Again, the detail of the reconstruction technique differs between the SST products.

### **In situ time-series data**

*In situ* time-series data from the six longest available time-series in the North Atlantic have been selected. The time-series data are presented annually in the IROC (Hughes and Holliday, 2007). The averaging depths have been determined by each data provider as those most suitable to represent the hydrographic conditions of the region of the upper water column (Table 2). Because these time-series are regularly published and already widely used as indicators in their particular region, we chose to use these time-series as published for comparison with SST data, accepting that the analysis methods used to prepare each time-series may not be consistent. A brief description of six time-series is presented below.

The first observations along the Kola Section (Kola, Figure 1) were made starting in 1900, and to date this site has been occupied more than 950 times, making it one of the longest oceanographic time-series in the world. The data used in this study are monthly and yearly (1951–2006) temperature values from the upper 200 m layer, at stations 3–7 (70.5°–72.5°N). This time-series, which samples the Murmansk Current, is widely used in fisheries oceanography and studies of regional climate variability, because it reflects variability in hydrographic conditions in the entire

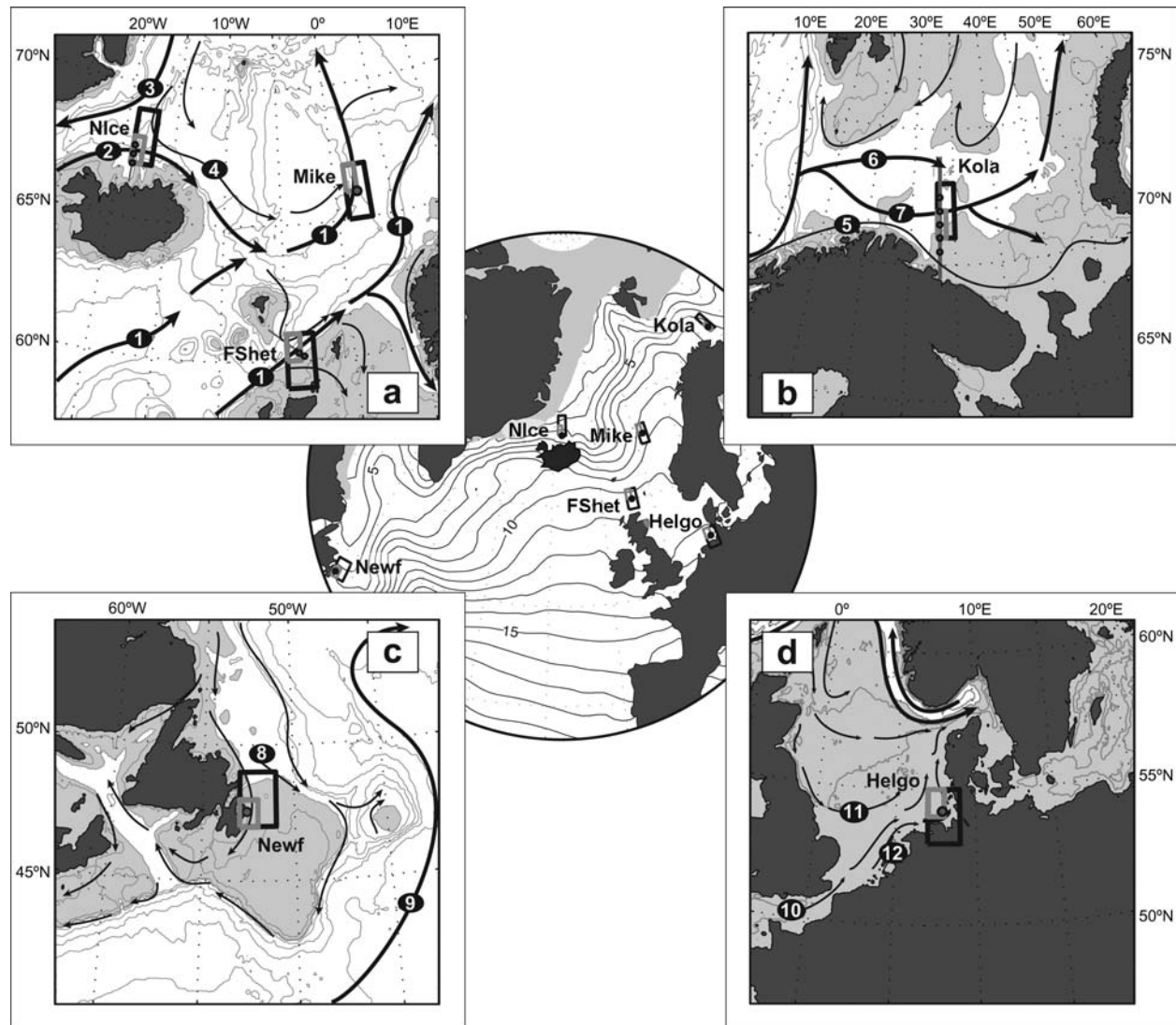
southern Barents Sea (Yndestad *et al.*, 2008). The Murmansk Current, a continuation of the North Cape Current entering the Barents Sea through its western boundary, is related to the haline/density frontal zone located between 71.0°N and 72.0°N, which separates the warmer and more saline Atlantic Water to the north and cooler fresher coastal waters to the south (Figure 1b). There were two gaps in observations during 1907–1920 and 1941–1944. The gaps were filled by Bochkov (1982) with the use of multiple regression models fitted on the data for 1946–1974. The section was surveyed most frequently from the 1950s to the 1980s, with 12–15 occupations a year. In recent years (1990s to present), it has been occupied around 7–9 times per year. The temperature time-series contains annual and quarterly values from 1900 to 2007, whereas monthly values are available from the period 1921–2007.

Ocean Weather Station Mike (Mike, Figure 1) is a single-station hydrographic time-series situated in 2200 m of water in the eastern margin of the Norwegian deep basin (Gammelsrød *et al.*, 1992; Østerhus and Gammelsrød, 1999). Temperature and salinity profiles have been collected daily to 1000 m, and weekly to 2200 m, since 1948, making it a remarkable and uniquely homogeneous ocean time-series. The station monitors the inflowing Atlantic Water, as well as conditions in the deep Norwegian Sea. The proximity to the Polar Front means that the upper ocean properties can also be influenced by Arctic Waters (Figure 1a).

Situated to the north of Iceland, the Siglunes hydrographic section (Nice, Figure 1) runs north–south at longitude 18.8°W, from the deep Iceland Sea towards the Icelandic coast (68°N–66.26°N) and has been extensively analysed to demonstrate seasonal and interannual variability in north Icelandic waters (Malmberg and Valdimarsson, 2003). The time-series from 50- to 150-m depth at stations 2–4 (67°N–67.6°N) indicates variability in the North Icelandic Irminger Current, a branch of the North Atlantic Current, which flows northwards to the west of Iceland and turns eastwards to flow along the edge of the shelf (Figure 1a). The conditions in north Icelandic waters are largely warm and saline because of the Irminger Current, but periodically the shelf region is affected by incursions of cold, fresh Polar Water (e.g. in the late 1960s, late 1970s, 1988, and late 1990s), which have significant effects on local conditions and fish stocks (Malmberg and Valdimarsson, 2003).

The Faroe Shetland Channel (FShet, Figure 1) is the deepest gap in the shallow ridge that separates the eastern North Atlantic Ocean from the Nordic Seas. The easternmost branch of warm, saline Atlantic inflow travels northwards from the Rockall Trough into the Norwegian Sea in the upper layers of the FShet (Figure 1a). At depth, the cold dense overflow water flows southwards into the North Atlantic. Two sections across the channel, running from the Faroe Plateau to the Scottish continental shelf, have been sampled since 1897, at a rate of between one and five sections per year (Turrell *et al.*, 1999). At present, the sections are occupied 3–5 times per year, although sampling rates were lower before the 1980s. Although the net flow of the upper ocean is northwards, the region is known to have a high level of mesoscale variability (Sherwin *et al.*, 2006). There is lateral exchange between the Atlantic water, currents on the continental shelf, and a modified form of Atlantic water that has travelled from the Iceland Basin via the north edge of the Faroe Plateau. The *in situ* time-series we present has been derived to reduce the effect of mesoscale variability as much as possible and represents the properties North Atlantic water in the high-salinity core next to the continental shelf break.





**Figure 1.** Central: average SST for 1997–2007 calculated from OISST.v2 grid boxes used for SST data are illustrated in black for ERSST.v3 and grey for the OISST.v2 and HadISST1 data. The map projection is equal-area so accurately represents the different spatial extent of the grid boxes. Areas where sea ice (>50% ice cover) occurs for more than 3 months of the year are shaded grey. Inset maps: (a) Location of the Siglunes Section, North of Iceland, Ocean Weather Station Mike, and the Faroes-Shetland section, (b) the Kola Section, (c) Station 27 on the Newfoundland Shelf, and (d) the Helgoland Roads Station. Oceanographic sections (grey lines) and important stations (circles) are indicated. Grey boxes indicate the extent of the  $1^\circ$  grid (OISST.v2 and HadISST1), and black boxes indicate the extent of the  $2^\circ$  grid (ERSST.v3). Arrows indicate the pathways of Atlantic Water (a1), The North Icelandic Irminger Current (a2), East Greenland Current (a3), the East Icelandic Current (a4), warmer and fresher Coastal Water (b5), the Murmansk Current (b6), the Atlantic Water inflow (b7), the Labrador Current (c8), the warmer and more saline North Atlantic Current (c9), Channel Water entering the North Sea from the North Atlantic (d10), South North Sea Water (d11), and Continental Coastal Water (d12). Shaded area is <250-m depth, contours are 50 m (d only), 100, 250, 500 m, then every 1000 m.

It is made up of data from both standard sections. The mesoscale variability means that the location and depth of the high-salinity core changes from one occupation to another, but it is always shallower than 200-m depth. In the IROC (Hughes and Holliday, 2007), the time-series is presented with a 2-year running mean applied. Here, we have calculated annual mean values for better comparison with annual mean SST data.

Helgoland Roads (Helgo, Figure 1) is a shallow island station in the German Bight, in the southern North Sea. It has been sampled almost daily for surface temperature since 1873, with some gaps in the late 1940–1950s, and for other physical and biological parameters since 1967 (Wiltshire and Manly, 2004). The German

Bight is shallow (averaging around 25 m), with the circulation dominated by tides and windforcing. The mesoscale variability is strong, and the main water masses are continental coastal water and central southern North Sea water (Figure 1d), which includes oceanic and river run-off influences (Becker *et al.*, 1992). The waters around Helgoland can be influenced by either of these water masses, as well as by the lateral movement of fronts and eddies. Therefore, the time-series is highly variable. There is a strong seasonal signal, but it is well resolved by the near-daily sampling resolution.

Systematic hydrographic measurements have been made at Station 27 on the Newfoundland Shelf (Newf, Figure 1) since

about 1947 (Colbourne and Fitzpatrick, 1994). In the early years, the station was occupied 1–2 times per month, but in recent years, sampling increased to ~2–4 times per month, on average. Seasonally, the sampling reveals a bias towards a maximum in spring and early summer and a minimum during winter. Station 27 is close to the coast on the eastern Canadian continental shelf and is situated within the inshore branch of the Labrador Current. This current, which generally flows southeastwards along the continental shelf, is made up of a reasonably strong western boundary current along the shelf break, and a weaker component over the banks and inshore regions (Figure 1c). It is responsible for advecting cold, reasonably fresh, subpolar water, together with sea ice and icebergs from the Arctic to lower latitudes along the Labrador Coast. At Station 27, the cold ( $<0^{\circ}\text{C}$ ) water mass, which forms the cold intermediate layer on the continental shelf, is present year-round, so variations in water properties at this site are representative of conditions across a broad area of the Newfoundland Shelf (Colbourne *et al.*, 1997). The depth-integrated (0–176 m) time-series of temperature is a robust index of climate conditions on the shelf (Colbourne *et al.*, 1994).

Depending on the local hydrographic conditions, the *in situ* time-series are not necessarily expected to reflect the conditions at the surface. For this analysis, to examine the differences between the *in situ* time-series and the surface data at two sites (Kola and Newf), we have included an additional dataset, derived from the same underlying *in situ* data, but calculated for the near-surface layer (Table 2). Although some of the time-series extend back farther, we have chosen to analyse the data since 1950, which is the common period for each of the six datasets and has the advantage of limiting the time-series to later periods, where the data quality is better.

### Comparison of datasets

For comparison with *in situ* time-series data, we extracted gridded SST data from the grid cell whose centre was closest to the central location of the *in situ* observations (Figure 1). Each of the *in situ* data sites is in a very different hydrographic region, with markedly different temperature regimes (Figure 1, Table 3).

When comparing the gridded SST datasets with *in situ* time-series that are generally calculated as depth-averaged values, the absolute temperature values would not be expected to match. Ignoring absolute temperature values and calculating temperature anomalies allows the variability in temperature to be examined, with the assumption that the variability in the upper layers of the ocean will follow the patterns similar to those at the surface. This also allows us to compare variability within each SST dataset, and removes the problem of differing climatologies.

Annual mean temperature anomalies were calculated for selected grid cells from each of the three gridded SST datasets. Temperature anomalies were calculated by subtracting the monthly mean climatology, calculated for the period 1971–2000, from monthly values. Annual mean anomalies are annual averages of monthly values. Although the OISST.v2 dataset only extends to 1981, the product is offered with a climatology prepared by Xue *et al.* (2003), which allows anomalies to be calculated relative to a 1971–2000 reference period. To calculate monthly means from the weekly OISST.v2, the data were first interpolated onto a daily time base, then averaged over a calendar month, as described in Reynolds *et al.* (2002).

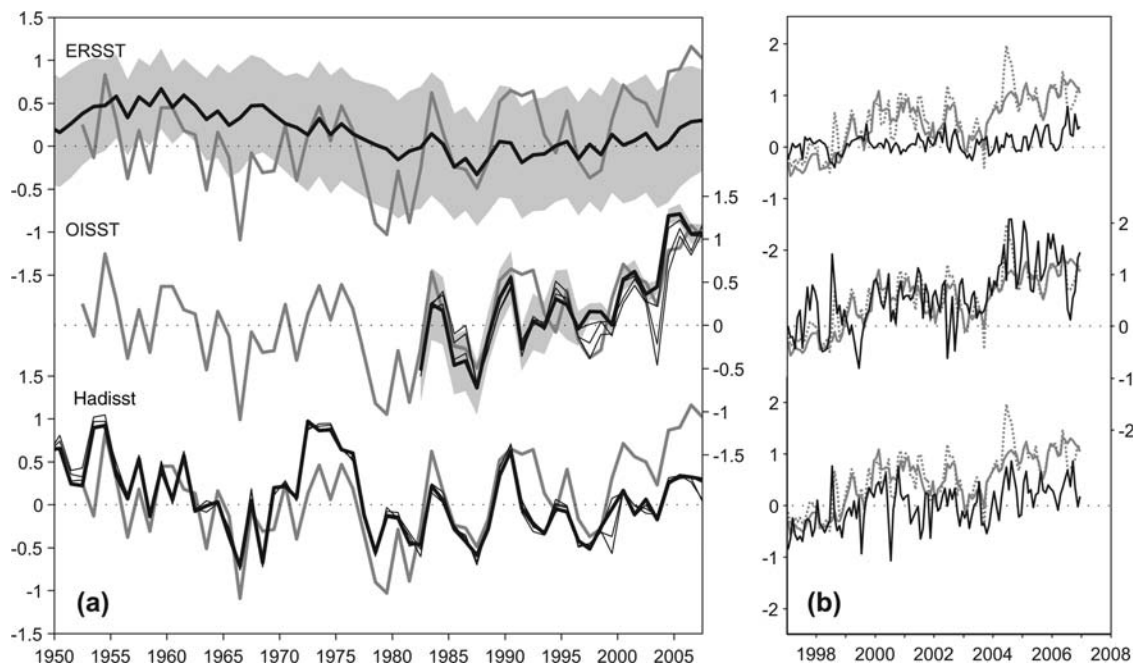
Anomalies were calculated in a similar way for the *in situ* time-series, referencing to the period 1970–2000. As well as the mean conditions, seasonal variability was removed from the data to calculate anomalies. It should be noted that the sampling frequency of the *in situ* time-series varies markedly between the datasets, from daily at Helgo to 2–3 times per year at NIce (Table 2). A sample here refers to a single occupation of each station or section. Where sampling was more than once per month, monthly averages were calculated for comparison with *in situ* data. Where sampling was less than once per month, data from the single occupation were used as a representative sample for that month.

For visual comparison, the annual mean temperature anomalies from the *in situ* time-series were plotted along with the same data calculated from each gridded SST product (Figures 2–7). However, for statistical analysis, a slightly different approach was followed. To limit the influence of the different sampling variability, the statistical analysis was undertaken not on the annual data, but with a 3-year running mean applied (because the OISST dataset begins at the end of 1981), so to compare all three products, statistics were calculated only for the common period 1982–2006 (Table 4).

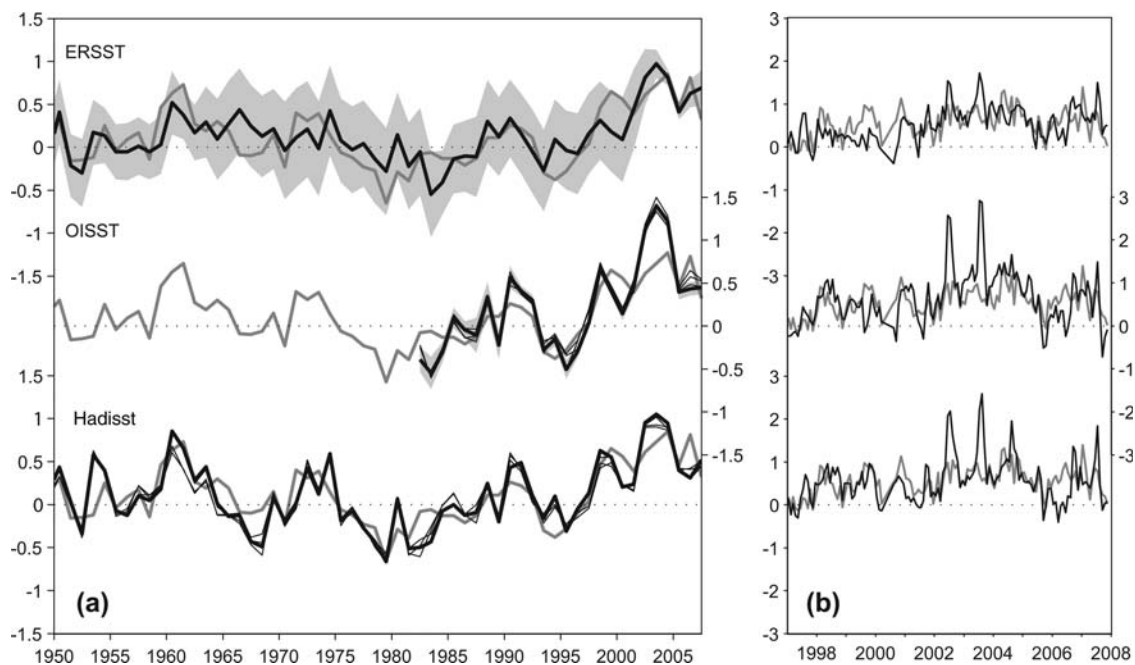
One advantage of the SST datasets over the *in situ* time-series is the enhanced temporal resolution, with data available every month and some of the products now also available at weekly (Reynolds *et al.*, 2002) and even daily resolution (Reynolds *et al.*, 2007). To determine the ability of the gridded datasets to reflect the monthly variability, we also calculated statistics from the monthly *in situ* time-series (Table 5). Here, it should be noted that because of the different sampling frequency of the datasets, the monthly value can be calculated as an average of a number of regular observations (Helgo—daily, Mike—weekly, and Newf—weekly or fortnightly). For those time-series sampled less frequently than monthly, or that have months for which no data are available, the SST dataset was subsampled to the same monthly sampling frequency as the *in situ* data. Shaded error bars on the time-series figures illustrate the error estimates in

**Table 3.** Statistics of SST (mean, minimum, and maximum) at each selected grid point for the period 1997–2006.

1997–2006	ERSST.v3			OISST.v2			HadISST1		
	Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.
Kola	4.72	2.49	8.31	5.37	1.94	11.25	4.92	2.36	9.63
Mike	8.40	5.89	12.46	8.25	4.92	14.24	8.17	5.33	13.62
NIce	2.85	0.02	7.59	4.09	0.64	9.64	3.86	0.55	9.08
FShet	10.17	7.54	13.74	10.19	7.82	14.22	9.95	7.35	14.36
Helgo	10.83	2.69	20.51	11.39	2.95	20.82	11.11	0.82	20.83
Newf	5.07	−0.91	14.74	5.86	−0.46	16.56	5.71	0.29	15.67



**Figure 2.** Kola Section: Eastern Barents Sea (Kola, 71.5°N 33.5°E). (a) Annual mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). For ERSST.v3 and OISST.v2, shaded areas provide estimate of error. For OISST.v2 and HadISST1, data from three adjacent grid squares are also plotted (fine black). (b) Monthly mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). *In situ* data from near-surface level are also illustrated (dashed grey).



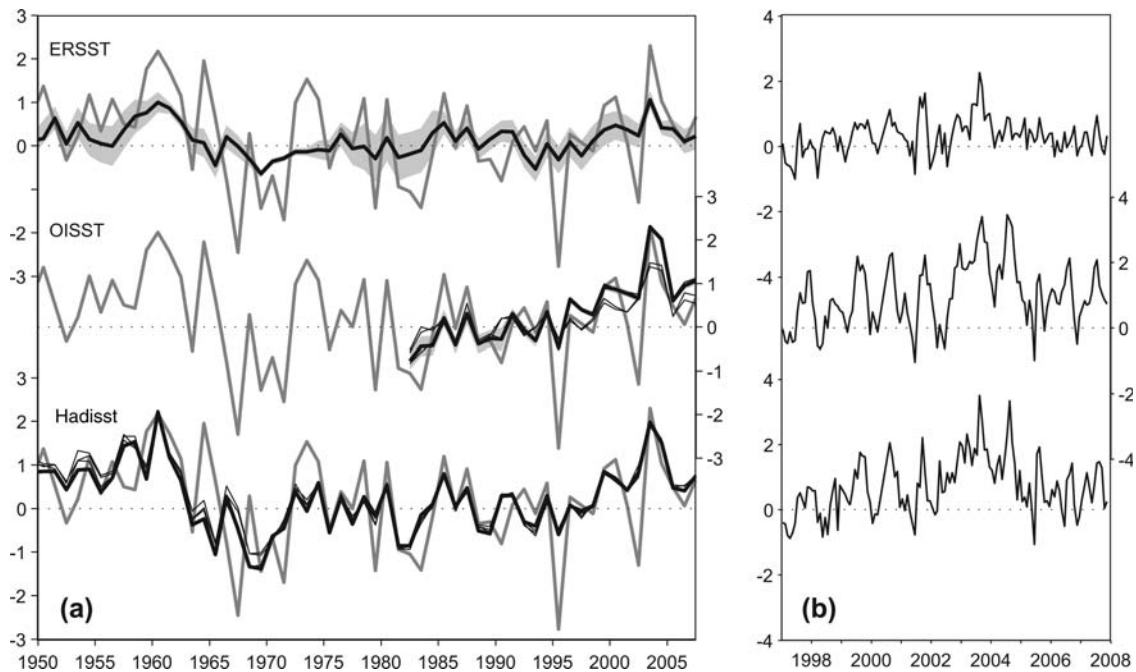
**Figure 3.** Ocean Weather Station Mike: Norwegian Sea (Mike, 66°N 2°E). (a) Annual mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). For ERSST.v3 and OISST.v2, shaded areas provide estimate of error. For OISST.v2 and HadISST1, data from three adjacent grid squares are also plotted (fine black). (b) Monthly mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black).

the ERSST.v3 and OISST.v2, using values included within the gridded SST datasets.

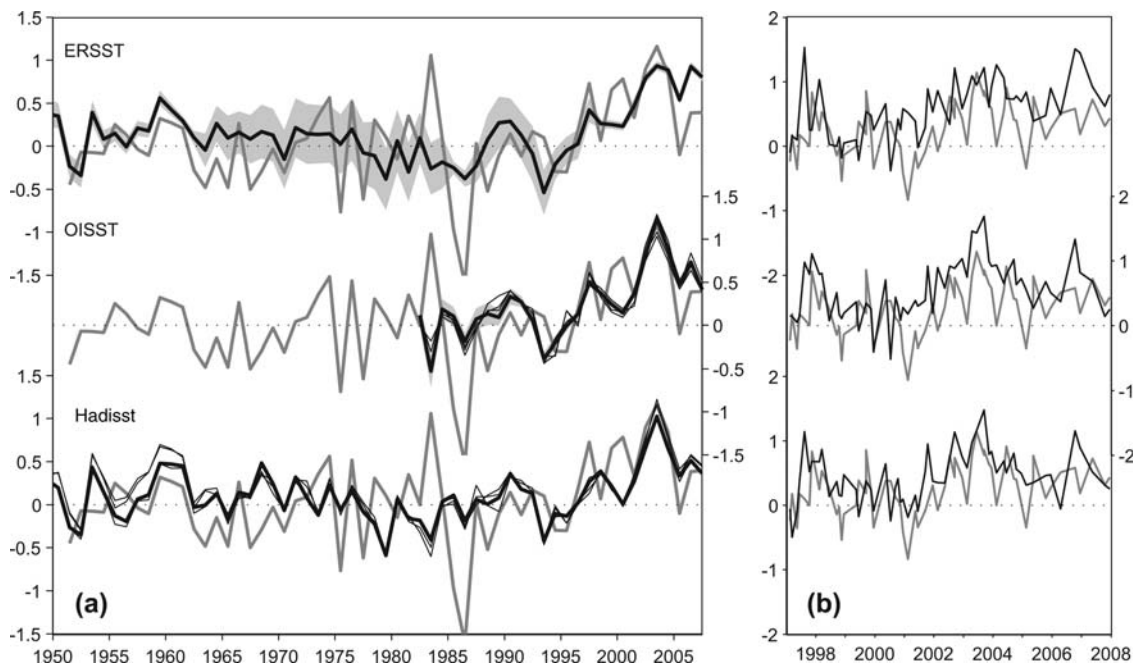
For each time-series (3-year annual mean and monthly subsample, Tables 4 and 5), we calculated the Pearson correlation

coefficient ( $r$ ) and the root-mean-square error (RMSE). These two parameters can be used to determine the degree of similarity between the two datasets, on the assumption that the SST dataset with the highest value of  $r$  and lowest RMSE represents the best fit





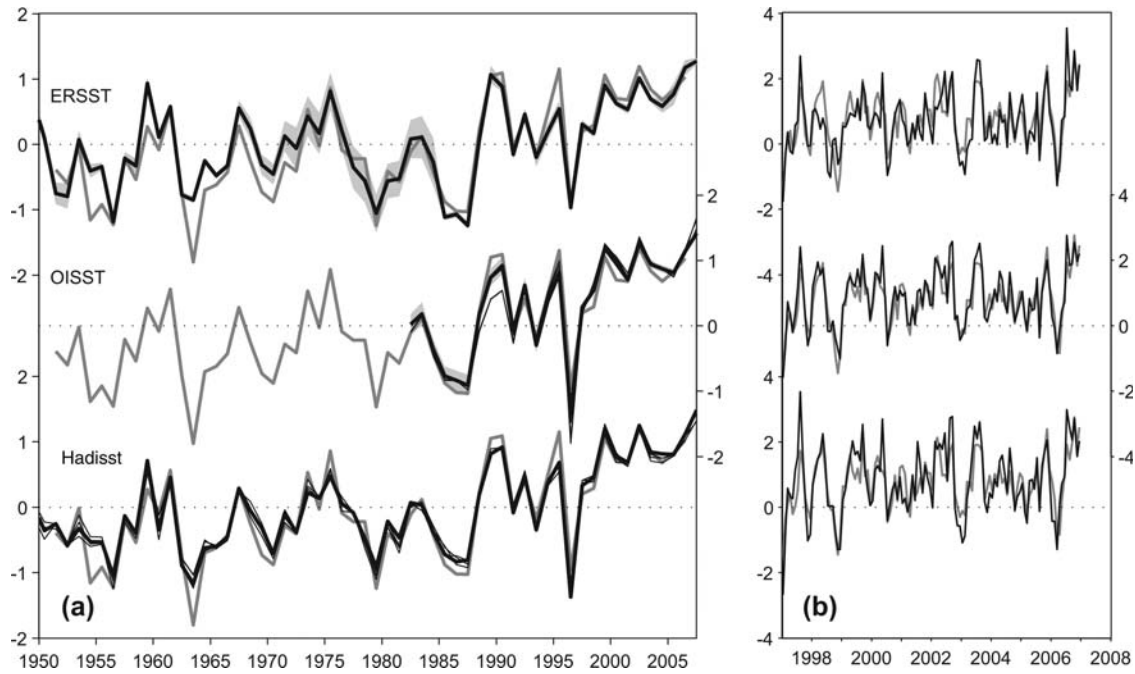
**Figure 4.** Siglunes: North Iceland Irminger Current (Nlce, 67.3°N 18.8°W). (a) Annual mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). For ERSST.v3 and OISST.v2, shaded areas provide estimate of error. For OISST.v2 and HadISST1, data from three adjacent grid squares are also plotted (fine black). (b) Note that monthly mean anomalies from *in situ* data are not available at this location. Monthly mean anomalies from gridded SST products (solid black) are illustrated.



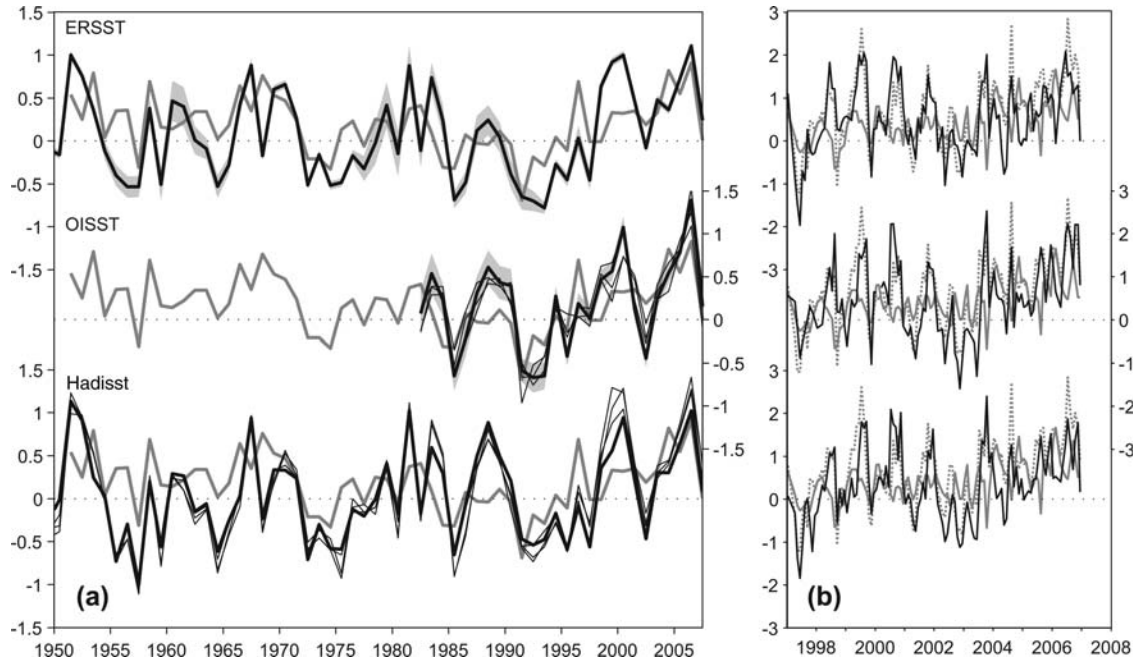
**Figure 5.** Faroe Shetland Channel: North Atlantic Water (FShet, 61°N 3°W). (a) Annual mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). For ERSST.v3 and OISST.v2, shaded areas provide estimate of error. For OISST.v2 and HadISST1, data from three adjacent grid squares are also plotted (fine black). (b) Monthly mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black).

to the data. Although the correlation coefficient can provide an estimate of how well the pattern of variability of the gridded products (g) matches the pattern of variability in the *in situ* data (t), it does

not measure how well the scale of variability is matched. To address this, we therefore also calculated a Skill score [S; Equation (1)], following the method suggested by Taylor (2001), using the correlation



**Figure 6.** Helgoland Roads: North Sea (Helgo, 54.2°N 7.9°E). (a) Annual mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). For ERSST.v3 and OISST.v2, shaded areas provide estimate of error. For OISST.v2 and HadISST1, data from three adjacent grid squares are also plotted (fine black). (b) Monthly mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black).



**Figure 7.** Station 27: Newfoundland Shelf (Newf, 47.5°N 52.6°W). (a) Annual mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). For ERSST.v3 and OISST.v2, shaded areas provide estimate of error. For OISST.v2 and HadISST1, data from three adjacent grid squares are also plotted (fine black). (b) Monthly mean anomalies (solid grey), compared with equivalent data from gridded SST products (solid black). *In situ* data from near-surface level also illustrated (dashed grey).

coefficient and standard deviation. A Skill score of 1 indicates a perfect match of the gridded datasets to the *in situ* time-series:

$$S = \frac{4(1+r)^4}{(\hat{\sigma}_g + 1/\hat{\sigma}_t)^2(1+r_o)^4}, \quad (1)$$

where  $\hat{\sigma}_g$  is the standard deviation normalized by the observations  $\sigma_g/\sigma_t$ , and  $r$  is the Pearson correlation coefficient, with  $r_o$  as the maximum attainable correlation, set here to 1.

Intercomparison of data products with a different spatial resolution must be approached with caution. For the 1° datasets, data



**Table 4.** Statistical comparison of annual mean anomalies from gridded SST and *in situ* time-series for the period 1982–2006.

1982 – 2006		ERSST.v3				OISST.v2				HadISST1			
Site	s.d. ( $\sigma_t$ )	corr. ( $r$ )	s.d. ( $\sigma_g$ )	RMSE	Skill	corr. ( $r$ )	s.d. ( $\sigma_g$ )	RMSE	Skill	corr. ( $r$ )	s.d. ( $\sigma_g$ )	RMSE	Skill
Kola	0.38	0.77	0.11	0.30	0.18	0.84	0.44	0.23	<b>0.70</b>	0.90	0.23	0.20	0.63
Mike	0.34	0.88	0.34	0.16	0.78	0.91	0.44	0.19	0.78	0.92	0.34	0.13	<b>0.85</b>
Nlce	0.55	0.69	0.25	0.41	0.28	0.68	0.63	0.46	<b>0.49</b>	0.77	0.51	0.35	0.61
FShet	0.48	0.76	0.39	0.31	<b>0.57</b>	0.70	0.33	0.34	0.45	0.67	0.27	0.35	0.36
Helgo	0.57	0.99	0.56	0.09	<b>0.97</b>	0.97	0.58	0.13	0.94	0.97	0.56	0.13	0.94
Newf	0.29	0.80	0.47	0.29	0.53	0.76	0.39	0.25	<b>0.55</b>	0.56	0.38	0.32	0.35

Columns indicate the standard deviation ( $\sigma$ ), the Pearson correlation coefficient ( $r$ ), root-mean-square error (RMSE), and skill level. Values emboldened illustrate the highest skill score for each *in situ* time-series.

**Table 5.** Statistical comparison of monthly mean anomalies from gridded SST and *in situ* time-series for the period 1997–2006.

1997–2006		ERSST.v3					OISST.v2				HadISST1			
Site	$n$	s.d. ( $\sigma_t$ )	corr. ( $r$ )	s.d. ( $\sigma_g$ )	RMSE	Skill	corr. ( $r$ )	s.d. ( $\sigma_g$ )	RMSE	Skill	corr. ( $r$ )	s.d. ( $\sigma_g$ )	RMSE	Skill
Kola	120	0.52	0.39	0.19	0.50	0.10	0.61	0.62	1.28	0.41	0.65	0.41	0.40	<b>0.44</b>
Kola <sub>U</sub>	120	0.52	0.33	0.19	0.52	0.08	0.67	0.62	1.34	<b>0.48</b>	0.63	0.41	0.40	0.42
Mike	120	0.33	0.40	0.41	1.11	0.23	0.50	0.63	1.39	0.22	0.50	0.53	1.21	<b>0.25</b>
Nlce	–	–	–	–	–	–	–	–	–	–	–	–	–	–
FShet	59	0.40	0.37	0.44	1.23	0.22	0.52	0.44	0.41	0.33	0.56	0.40	0.97	<b>0.37</b>
Helgo	120	0.80	0.82	0.91	1.45	0.68	0.92	0.87	1.78	<b>0.84</b>	0.85	1.07	1.75	0.68
Newf	112	0.44	0.16	0.81	1.33	0.08	0.14	0.88	1.32	0.07	0.16	0.84	1.10	<b>0.08</b>
Newf <sub>U</sub>	112	0.82	0.74	0.81	1.60	0.57	0.76	0.88	1.57	<b>0.60</b>	0.75	0.84	1.33	0.58

Note that for some time-series, data do not exist for every month. The number of samples available for each dataset is displayed ( $n$ ). Columns indicate the standard deviation ( $\sigma$ ), the Pearson correlation coefficient ( $r$ ), root-mean-square error (RMSE), and skill level. Values emboldened illustrate the highest skill score for each *in situ* time-series.

from the three adjacent grid cells that lie within the spatial extent of the 2° grid cell were also investigated in an effort to examine the local spatial variability and to establish whether this explains the difference between the SST products with different spatial scales. The time-series from adjacent grid cells of OISST.v2 and HadISST1 are plotted for comparison on the annual mean figures (Figures 2a–7a).

As mentioned previously, because the SST products are generated on a time basis (weekly, monthly), the recent data from these products can only incorporate observations submitted to the collating centres in real time or near real time. These real-time data are collected by satellites, as well as by buoys and vessels, which transmit their observations to national meteorological data centres via the GTS. As none of the data providers submits data in real time, the gridded SST products are independent of the IROC *in situ* time-series data, at least for the period 1997 onwards. Before 1997, the *in situ* time-series observations are likely to have been incorporated into the ICOADS dataset as part of the delayed mode archive within ICOADS, and as such, the two datasets cannot be considered truly independent before this date. Monthly anomalies (Figures 2b–7b) were compared for the past decade 1997–2007, corresponding to the period for which the *in situ* data are not informing the SST products.

## Results

### Kola Section: Eastern Barents Sea (Kola, 71.5°N 33.5°E)

At Kola, the OISST.v2 time-series displays the best fit with the data (Figure 2, Table 4), and this is most notable in the years since 2000, when the *in situ* temperatures exhibited a substantial increase not represented in either the ERSST.v3 or HadISST1 data. Despite

some underestimating of the long-term trend, HadISST1 represents the interannual to decadal variability quite closely, as reflected by the high correlation. The error bars in ERSST.v3 are large for this region, and the product is not representative of inter-annual or even decadal scale variability.

The 1° grid cells of the OISST.v2 and HadISST1 coincide with stations 4–6 on the Kola Section (Figure 1), and the 2° area covered by ERSST.v3 extends farther to the north, closer to the influence of Atlantic Waters. Station 3 of the Kola Section, which is included in the *in situ* data, is to the south of all the grid boxes. When comparing the annual mean data from adjacent 1° grid cells, there is evidence of some variability in the temperature anomalies, for example, in the OISST.v2 dataset; temperatures in the two more northern grid cells were lower during 2003 (Figure 2a). However, this variance is not large enough in itself to explain the difference between ERSST.v3 and OISST.v2.

To examine whether the difference between the SST products and the Kola *in situ* data might be a result of a real difference between surface conditions and the subsurface index, we can compare the near-surface (0–50 m) time-series with the depth-averaged data (0–200 m; Figure 2b). The ERSST.v3 and HadISST1 data have already been established to be a poor match at this site, specifically for the later period. Therefore, this comparison focuses on OISST.v2. The surface *in situ* data display a slightly larger amplitude of variability than the 0–200 m data, as might be expected, but reveal the same underlying trend. The RMSE and Skill values calculated from the SST products compared with the near-surface dataset increase slightly compared with those calculated from the deeper dataset (Tables 4 and 5). The warm event recorded in the near-surface *in situ* data during the late

summer of 2004 is captured in the OISST.v2 dataset, although there is some difference in the timing of this event.

### **Ocean Weather Station Mike: Norwegian Sea (Mike, 66°N 2°E)**

As this station is well sampled, we would expect to see some of the highest correlation scores at Mike. The time-series of annual anomalies demonstrate that all three SST products reflect the 50-m multiyear variability reasonably well, with no significant difference from the *in situ* data (Figure 3a, Tables 4 and 5), although the products reveal some clear differences. The error bars in ERSST.v3 are still large in this region.

The position of Mike is right in the centre of the 2° grid cell from the ERSST.v3 dataset. The 1° grid cell selected from HadISST1 and OISST.v2 for comparison is likely to reflect the influence from Arctic waters more than the adjacent three grid cells, which lie closer to the Atlantic water domain. However, examination of the time-series from adjacent cells (Figure 3a) reveals little variability in the annual mean anomalies across the grid cells.

HadISST1 has the highest Skill score, both in the annual mean and monthly time-series (Tables 4 and 5). However, at this site, on time-scales of 1–3 years, the SST data tend to have a greater amplitude of variability than the *in situ* time-series at 50 m. The monthly data in Figure 3b indicate that increases in the annual mean can occur in the SST products because of just 1 or 2 months in an individual year having a significantly higher value (e.g. 2002 and 2003). These anomalously high (or sometimes low) monthly values are not observed in the *in situ* data and this will affect the Skill scores.

### **Siglunes: North Icelandic Irminger Current (NICE, 67.3°N 18.8°W)**

The 1° grid cell of the HadISST1 and OISST.v2 datasets coincides closely with the position of the Siglunes stations 2–4 and should therefore be representative of conditions in the Irminger Current.

The ERSST.v3 dataset, with a 2° resolution cover a larger area, covers more of the region affected by waters of Arctic origin in the East Greenland and East Icelandic Currents (Figure 1). Although we might expected fairly high variability between adjacent grid cells, there is very little difference in those selected for comparison from HadISST1 (Figure 4), although there are some small differences in OISST.v2. The contrasting properties of the Atlantic and Polar Waters, combined with a reasonably low sampling frequency, result in an *in situ* time-series that is highly variable, with a large range of temperature anomalies.

The SST products do not reveal the same interannual extremes of temperature seen in the *in situ* time-series, though OISST.v2 and HadISST1 perform better than ERSST.v3 (Table 4). Because of the low sampling frequency in the Siglunes section (4 times per year), perhaps the extremes in the time-series are not a good representation of annual means. Although we do not have monthly data at Siglunes, we can examine the monthly time-series from the SST data products to establish whether this captures individual extreme similar events, for example, during 2002, when a cold event was observed in the *in situ* record. There is no evidence of a similar strong cold event in any of the gridded SST datasets, suggesting that it could have been a subsurface phenomenon, because the *in situ* time-series is averaged from 50 to 150 m.

OISST.v2 and HadISST1 appear to represent the *in situ* subsurface variability on longer time-scales (approximately >3 years) fairly well. In contrast to Ocean Weather Station Mike, the amplitude of the variability recorded by the SST products is rather less than the variability revealed by the *in situ* subsurface data, although again this may be strongly influenced by the low sampling frequency of the *in situ* data.

### **Faroe Shetland Channel: North Atlantic Water (FShet, 61°N 3°W)**

The location of the selected grid cells at FShet coincides fairly well with the location of the core of Atlantic Water and the position of the *in situ* data (Figure 1). The larger 2° grid cell used for the ERSST.v3 dataset extends farther south, into the region occupied by a mixture of oceanic waters with waters of coastal origin. Despite this, the data from adjacent grid cells display very little significant variance (Figure 5a).

The SST products at this location visually display similar patterns of variability to each other, and reflect the >3 year subsurface variability reasonably well, although statistically the correlations are rather low (Table 4). The low correlation is influenced by the period between 1975 and 1990, when the sampling frequency along this section was lower (only occupied 1–2 times per year during the period 1978–1988) and like NICE, there is strong interannual variability in the *in situ* data, which does not appear in the SST products. In the recent period, comparing data at original sampling frequency, the statistics place HadISST1 as the closest match to these *in situ* data (Table 5).

### **Helgoland Roads: North Sea (Helgo, 54.2°N 7.9°E)**

There is a very strong correlation between all the SST data products and the *in situ* time-series at Helgo. The ERSST.v3 has a slightly greater difference from the *in situ* data than the others (Table 4), particularly before the 1980s (Figure 6a). This may reflect the effect of the coarser grid in this region of strong mesoscale variability. The 2° grid of ERSST.v3 extends right up to the coast, compared with the 1° grid selected for the HadISST1 and OISST.v2 (Figure 1), which does not. However, examination of the time-series in adjacent boxes suggests that the same conclusions would be reached if data from the adjacent time-series had been used for analysis. The monthly anomalies illustrated in Figure 6b reveal that the SST products closely track the variability at shorter time-scales too, though ERSST.v3 and HadISST1 tend to have more variability than the *in situ* data (Table 5).

From daily SST measurements at Helgo, we calculated for the period 1997–2006 a mean temperature of 10.68°C, with a maximum of 18.85°C and a minimum of 3.13°C. Both the mean value and the temperature range are lower than the same statistics calculated from the gridded SST products, with ERSST.v3 offering the best match. The most likely explanation for this difference is the high mesoscale variability in the coastal waters of the German Bight.

Despite the very good correlation with all SST datasets at this station (Table 5), the scale and timing of the observed maximum temperature anomaly during the period 1997–2006 is different. The OISST.v2 dataset matches with the observations, giving a maximum temperature anomaly in July 2006. The ERSST.v3 differs slightly with the timing of the maximum in October 2006. However, in the HadISST1 dataset, the maximum temperature during this period was observed in August 1997.

### Station 27: Newfoundland Shelf (Newf, 47.5°N 52.6°W)

A comparison of the depth-integrated *in situ* time-series with the three SST products reveals similar decadal patterns, but they all exhibit greater variability than the *in situ* data on interannual time-scales (Figure 7a).

An examination of the monthly time-series from both levels (Figure 7b) indicates that the near-surface data are much more closely correlated with the SST products. The monthly surface anomaly *in situ* data at Newf have a strong and varying seasonal component, which resemble the SST monthly data closely, whereas the monthly depth-integrated *in situ* data have low correlations with all three SST products (Table 4). In other words, as expected, the Skill scores greatly improve when comparing the *in situ* surface data with the SST products. This result indicates that differences are at least to some extent caused by the variation between the surface and subsurface water properties.

There is little difference between the SST products, though OISST.v2 performed slightly better in recent years. This is most evident in the monthly time-series from 2003 to 2006 (Figure 7b). We conclude that the SST products are not good indicators of subsurface variability at monthly to interannual time-scales at this location, because of the real difference between the surface and subsurface variability. However, on longer time-scales (approximately >3 years), the SST products to some extent reflect the variability of the subsurface *in situ* data.

### Discussion

In this study, we wanted to examine whether variations in global gridded SST products reflect variations in subsurface conditions at our measurement sites, and if the global gridded SST products were able to reflect observed variations on monthly to decadal time-scales. We also hoped to determine which products, if any, offered the best representation of the known *in situ* variability at each location, and offer some guidance on which would be the most appropriate gridded SST dataset to use.

The analysis period was 1950–2007, and no attempt was made to investigate earlier data. However, significant differences in the gridded datasets do occur in earlier, data-sparse periods before 1940 (Smith *et al.*, 2008), and we would advise caution to anyone using them to investigate climate variability in an ecosystem context during this period.

Before using any dataset to investigate climate variability, it is important to understand its limitations. Although the gridded SST datasets are globally complete and can therefore provide a continuous time-series for any ocean region as far back as the end of the 19th century, at any particular time a large part of the dataset is reconstructed. Between 1982 and 2000, ship and buoy observations provided ~30% coverage of the world's oceans. Satellite coverage is wider, but even the best coverage between 1994 and 2000 missed ~10% of the available sea area (Reynolds *et al.*, 2002).

The *in situ* time-series also have their limitations, and some have a reasonably low sampling frequency, making it difficult to resolve temporal variability at scales of less than a year (Table 2). For this reason, we applied a 3-year smoother before making a statistical comparison of the longer term variability. For the monthly analysis (Table 5), the number of monthly samples at FShet is much lower than for the other sites, and we place less confidence in the monthly results at this station.

Both *in situ* and satellite observations were generally sparse in our study area, with the exception, perhaps, at Helgo. Reynolds *et al.* (2002) illustrated the distribution of ship and buoy data in 2000, a typical example of the data distribution in recent years, clearly demonstrating the large data-sparse areas north of 55°N in the North Atlantic. It should also be noted that three of the sites (Kola, Nice, and Newf) are reasonably close to the areas affected by sea ice (Figure 1), and Hurrell *et al.* (2008) note that some of the largest differences between OISST.v2 and HadISST1 occur in the marginal ice areas.

There will always be difficulties when trying to compare a time-series at a single point with the one derived from an area average. Our method of examining anomalies reduces the problem somewhat, because we would expect the temperature variability, on time-scales of greater than 1 month, to have broadly similar patterns over wide areas. In frontal regions, where water masses of different origins and characteristics meet, we would expect some of the variability to be lost when examining an area average. Examination of adjacent time-series within a 2° grid cell revealed that, at the sites we investigated, variability over this spatial scale is mostly small, and as such, there is little difference between the 1° and 2° gridded products. In fact, the size of the grid may not necessarily truly reflect the resolution of the data. Because of the methods used in data analysis, HadISST1 could not resolve “localized” SST variability, such as Gulf Stream meanders, whereas OISST.v2 could (Reynolds *et al.*, 2002; Hurrell *et al.*, 2008); hence, the resolution of OISST.v2 can be said to be better than that of HadISST1, although both are presented on a 1° grid.

The error bars provided with the gridded SST datasets indicate the degree of confidence that can be placed on the data, highlighting areas and periods when data are sparse and confidence is low. We concur with Rayner *et al.* (2003) and Barton and Casey (2005) that it is essential to take account of the error estimates when using these datasets. Error bars for the HadISST1 dataset are not as easily available as those for OISST.v2 and ERSST.v3. Rayner *et al.* (2003) suggest using the ungridded HadSST2 dataset to determine error estimate, using the procedure followed by Folland *et al.* (2001).

For researchers investigating climate variability and its effect on the marine ecosystem, variability at inter- and intra-annual time-scales is relevant, and the occurrence and persistence of extreme events may be of interest. Sicard *et al.* (2006), for example, attempted to use the ERSST data at a shallow coastal site to identify periods where SST was above levels that could be lethal to shellfish. They concluded that the ERSST dataset was not a valid substitute for *in situ* measurements at their study site, because the ERSST data underestimated the amplitude of observed temperature variability and revealed a different temporal variation.

From our limited comparisons of monthly time-series, we suggest that caution should be taken when using gridded SST products for this type of study. First, stemming from differences in the underlying climatology, the means of SSTs calculated from different datasets at the six locations differ by up to 1.0°C (Table 3), with differences greatest at the sites in marginal ice areas (Kola, Nice, and Newf). For the period 1997–2006, annual mean SSTs from OISST.v2 were warmer than HadISST1 and ERSST.v3 at five of the six sites. Second, we can see that even when examining temperature anomalies, none of the time-series reflects the exact timing of extreme events as recorded by the *in situ* time-series data. Even at Helgo, the value and date of the maximum temperature anomaly observed during the period



1997–2006 differs for each dataset, and especially between HadISST1 and the observations (Figure 6b). At the other stations, the timing of maximum and minimum temperature anomalies varies widely between SST datasets and between SST datasets and the observations.

In our comparison of surface and subsurface data, we assumed that the relationship between surface and subsurface conditions is constant throughout the year. It is clear from the analysis at Mike that the SST products contain greater monthly to interannual variability than evident in the *in situ* data at 50 m. However, it is likely that, because of thermal stratification during summer, there will be some monthly variability in this relationship. It might be expected that the relationship would be stronger in winter, when the water column is well mixed. We suggest that this is an area requiring further investigation.

Bearing in mind the seasonal influence (although we have only been able to examine two time-series), we have demonstrated that variability in SST does not always reflect conditions in upper layers of the ocean, and we advise careful consideration of the local hydrographic conditions before using the SST data as a proxy for subsurface variability. In particular, the scale of the temperature variability in the surface layers is likely to be much higher than that in deeper layers, although the long-term trends might be similar. At Mike, an earlier analysis revealed that variability in the upper 150 m was similar, particularly in the long-term variability (Gammelsrød and Holm, 1984). Where we examined near-surface time-series (Kola, Newf), these compared better with SST products, but only at Newf, where we examined data from the upper 5 m of the water column, were we able to increase the skill of the SST datasets.

As described earlier, the three datasets are based on large databases of *in situ* observations, including the ICOADS. Including more *in situ* data within the gridded SST products, specifically where other *in situ* observations are sparse, would logically improve estimates. However, as Reynolds *et al.* (2002) noted, there are problems associated with using data collected with profiling instruments to represent SST. Since October 2007, Fisheries Research Services (Marine Laboratory, Aberdeen, UK) have reported their *in situ* observations in the FShet in near real time (usually within 10 days). Data from other stations (Newf, Mike) are also reported in near real time, but it is uncertain if this is rapid enough to be incorporated into the SST products. Sea surface observations made by a research ship during a research cruise and transmitted in real time would ensure that SST data are provided at the same time and position, and more effort should perhaps be made to submit research vessel data in this way.

Although significant differences have been noted within the different gridded products, there appears to be an even spread throughout the literature for each product, with no apparent bias towards the source institute (i.e. European researchers do not necessarily use HadISST1). As they do not make it explicit in their research methods, it is difficult to know how and why researchers chose a particular product to use for their comparison. A review of the literature relating these products to ecosystem variability reveals no clues why researchers chose a particular product.

The frequency of update may be an important factor, although all the products are updated within a month or two of data collection. The regular weekly updates of OISST.v2 may make this an attractive product for those wishing to analyse very recent data. The weekly data, however, require careful analysis. For example, the day of the week when the data are reported changed in 1989.

In addition, the data are available directly in monthly format, if monthly resolution is adequate.

The length of the available time-series may also be a deciding factor, and those needing a long time-series (before 1982) are limited to only HadISST1 and ERSST.v3. The ease of availability of data products is an issue, and for these large gridded datasets, the size of the data files may also be a limiting factor, because specialist code and software may be required to read in the data. A summary of the data-access options is presented in Table 1. Although all the products are available in both ASCII and NetCDF format, for some researchers, the size of these globally complete datasets may be a barrier to their use. The OPeNDAP (<http://www.opendap.org/>) data server (previously known as DODS) allows researchers to access data over the Internet, and to import a subset directly into familiar data analysis and visualization packages (such as Excel and Matlab), without concerns about the underlying data formats. Once set up, this makes regular updates of data easy, and both OISST.v2 and ERSST.v3 are available via this data server. An alternative access route to the data is via the Live Access Server, which, for example, makes extracting subsets of ERSST.v3 data easy.

## Conclusions

Because we only investigated data at six locations in the North Atlantic, we cannot claim to have undertaken a thorough validation of each SST product. In addition, some of our *in situ* time-series are located in high-latitude and data-sparse regions. Although these traits may be a useful test for the products, they should be borne in mind when generalizing conclusions to less data-sparse regions.

From our analysis, we found a more consistent correlation between the OISST.v2 dataset and our *in situ* time-series data, closely followed by the HadISST1 dataset. The average skill score for OISST.v2 was 0.65, although this is not significantly different from the average score of 0.62 obtained from HadISST1. When comparing just the two long time-series (results not illustrated), HadISST1 was overall better than ERSST.v3 with mean Skill scores of 0.55 and 0.28, respectively.

For our particular applications, we found the OPeNDAP access to OISST.v2 convenient, and the availability of error estimates for download alongside the OISST.v2 and ERSST.v3 an advantage when determining the suitability of the data product to a particular region.

We have demonstrated that the SST products appear to match the observed *in situ* variability better at time-scales > 3 years. At monthly to 3-year time-scales, we concluded that different and inconsistent conclusions could be drawn from using different products.

Using the gridded SST data to identify the mean temperatures at a position, or to investigate the scale and timing of maxima and minima events, might give very different results, both between SST datasets and from those of the *in situ* data. Typically, data from the SST products revealed greater variation (higher amplitude) than the *in situ* data.

Surface *in situ* data may or may not have a close relationship with subsurface data. Therefore, it is important to consider the local hydrography if using SST data as a proxy for subsurface conditions. However, the agreement between surface and subsurface *in situ* with the SST products improves on time-scales > 3 years.

The results from our analysis match broadly with the conclusions of Barton and Casey (2005) that SST gridded data can provide useful indicators of long-term variability and may offer suitable proxies when *in situ* data are unavailable. We do,

however, advise caution in the application of these products, and suggest that it is essential to consider the quality of the underlying observations and that, as a minimum, the error estimates should be investigated along with the SST data, or the results from one or more gridded datasets should be computed.

## Acknowledgements

NOAA OISST.v2 data provided by the NOAA/OAR/ESRL PSD, Boulder, CO, USA, from their website at <http://www.cdc.noaa.gov>. ERSST.v3 data were provided from NCEP and National Climatic Data Center (NCDC) by Dick Reynolds and Tom Smith, <ftp://eclipse.ncdc.noaa.gov/pub/ersst/>. The UK Meteorological Office, Hadley Centre, HadISST 1.1–Global sea-Ice coverage and SST (1870–present), was obtained from <http://www.hadobs.org/>.

## References

- Barton, A. D., and Casey, K. S. 2005. Climatological context for large-scale coral bleaching. *Coral Reefs*, 24: 536–554.
- Becker, G. A., Dick, S., and Dippner, J. W. 1992. Hydrography of the German Bight. *Marine Ecology Progress Series*, 91: 9–18.
- Bochkov, Y. A. 1982. Water temperature in the 0–200 m layer in the Kola-Meridian section in the Barents Sea, 1900–1981. *Sbornik Nauchnykh Trudov Polyarno Nauchno-Issledovatel'skogo Instituta Rybnogo Khozyajstva I Okeanografii (PINRO)*, Murmansk, 46: 113–122 (in Russian).
- Bonnet, D., Richardson, A. J., Harris, R., Hirst, A., Beaugrand, G., Edwards, M., Ceballos, S., *et al.* 2005. An overview of *Calanus helgolandicus* ecology in European waters. *Progress in Oceanography*, 65: 1–53.
- Colbourne, E., de Young, B., Narayanan, S., and Helbig, J. 1997. A comparison of hydrography and circulation on the Newfoundland Shelf during 1990–1993 with the long-term mean. *Canadian Journal Fisheries and Aquatic Sciences*, 54: 68–80.
- Colbourne, E., and Fitzpatrick, C. 1994. Temperature, salinity and density at Station 27 from 1978 to 1993. *Canadian Technical Report of Hydrography and Ocean Sciences*, 159.
- Colbourne, E., Narayanan, S., and Prinsenberg, S. 1994. Climatic changes and environmental conditions in the Northwest Atlantic, 1970–1993. *ICES Marine Science Symposia*, 198: 311–322.
- Folland, C. K., Karl, T. R., Christy, J. R., Clarke, R. A., Gruza, G. V., Jouzel, J., Mann, M. E., *et al.* 2001. Observed climate variability and change. In *Climate Change 2001: the Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 99–181. Ed. by J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, *et al.* Cambridge University Press, New York. 881 pp.
- Gammelsrød, T. S., and Holm, A. 1984. Variations of temperature and salinity at Station M (66°N 02°E) since 1948. *Rapports et Procès-Verbaux des Réunions du Conseil International pour l'Exploration de la Mer*, 185: 188–200.
- Gammelsrød, T. S., Østerhus, S., and Godøy, Ø. 1992. Decadal variations of the ocean climate in the Norwegian Sea observed at ocean station “Mike” (66°N 2°E). *ICES Marine Science Symposia*, 195: 68–75.
- Hughes, S. L., and Holliday, N. P. (Eds.) 2007. *ICES Report on Ocean Climate 2006*. ICES Cooperative Research Report, 289. 55 pp.
- Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., and Rosinski, J. 2008. A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *Journal of Climate*, 21: 5145–5153.
- Malmberg, S.-A., and Valdimarsson, H. 2003. Hydrographic conditions in Icelandic waters, 1990–1999. *ICES Marine Science Symposia*, 219: 50–60.
- Østerhus, S., and Gammelsrød, T. 1999. The abyss of the Nordic Seas is warming. *Journal of Climate*, 12: 3297–3304.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., *et al.* 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysics Research*, 108: 4407. doi: 10.1029/2002JD002670.
- Reynolds, R. W., and Marsico, D. C. 1993. An improved real-time global sea surface temperature analysis. *Journal of Climate*, 6: 114–119.
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W. 2002. An improved *in situ* and satellite SST analysis for climate. *Journal of Climate*, 15: 1609–1625.
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. 2007. Daily high-resolution blended analyses for sea surface temperature. *Journal of Climate*, 20: 5473–5496.
- Richardson, A. J., and Schoeman, D. S. 2004. Climate impact on plankton ecosystems in the Northeast Atlantic. *Science*, 305: 1609–1612. doi: 10.1126/science.1100958.
- Sandvik, H., Erikstad, K. E., Barrett, R. T., and Yoccoz, N. G. 2005. The effect of climate on adult survival in five species of North Atlantic seabirds. *Journal of Animal Ecology*, 74: 817–831.
- Schone, B. R., Houk, S. D., Freyre Castro, A. D., Fiebig, J., Oschmann, W., Kroncke, I., Dreyer, W., *et al.* 2005. Daily growth rates in shells of *Arctica islandica*: assessing sub-seasonal environmental controls on a long-lived bivalve mollusk. *Palaos*, 20: 78–92.
- Sherwin, T. J., Williams, M. O., Turrell, W. R., Hughes, S. L., and Miller, P. I. 2006. A description and analysis of mesoscale variability in the Faroe-Shetland Channel. *Journal of Geophysical Research: Oceans*, 111: C03003. doi: 10.1029/2005JC002867.
- Sicard, M. T., Maeda-Martinez, A. N., Lluch-Cota, S. E., Lodeiros, C., Roldan-Carrillo, L. M., and Mendoza-Alfaro, R. 2006. Frequent monitoring of temperature: an essential requirement for site selection in bivalve aquaculture in tropical-temperate transition zones. *Aquaculture Research*, 37: 1040–1049.
- Smith, T. M., Reynolds, R. W., and Lawrimore, J. 2008. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *Journal of Climate*, 21: 2283–2296.
- Taylor, K. E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106: 7183–7192.
- Turrell, W. R., Slesser, G., Adams, R. D., Payne, R., and Gillibrand, P. A. 1999. Decadal variability in the composition of Faroe Shetland Channel bottom water. *Deep Sea Research I*, 46: 1–25.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J. K., *et al.* 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131: 2961–3012. doi: 10.1256/qj.04.176.
- Vecchi, G. A., and Soden, B. J. 2007. Effect of remote surface temperature change on tropical cyclone potential intensity. *Nature*, 450: 1066–1070. doi: 10.1038/nature06423.
- Wiltshire, K. H., and Manly, B. F. 2004. The warming trend at Helgoland Roads, North Sea: phytoplankton response. *Helgoland Marine Research*, 58: 269–273. doi: 10.1007/s10152-04-0196-0.
- Woodruff, S. D., Diaz, H. F., Elms, J. D., and Worley, S. J. 1998. COADS Release 2 data and metadata enhancements for improvements of marine surface flux fields. *Physics and Chemistry of the Earth*, 23: 517–526.
- Xue, Y., Smith, T. M., and Reynolds, R. W. 2003. Interdecadal changes of 30-yr SST normals during 1871–2000. *Journal of Climate*, 16: 1601–1612.
- Yndestad, H., Turrell, W. R., and Ozhigin, V. 2008. Lunar nodal tide effects on variability of sea level, temperature and salinity in the Faroe Shetland Channel and the Barents Sea. *Deep Sea Research I*, doi 10.1016/j.dsr.2008.06.003.