EVALUATING CLIMATE VARIABILITY AND CHANGE FROM MODERN AND HISTORICAL SST OBSERVATIONS

Nick A. Rayner [(1)], Alexey Kaplan [(2)], Elizabeth C. Kent [(3)], Richard W. Reynolds [(4)], Philip Brohan [(1)], Kenneth S. Casey [(5)], John J. Kennedy [(1)], Scott D. Woodruff [(6)], Thomas M. Smith [(7)], Craig Donlon [(8)], Lars-Anders Breivik [(9)], Steinar Eastwood [(9)], Masayoshi Ishii [(10)] and Tess Brandon [(5)]

[(1)] *Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, U.K., Email: nick.rayner@metoffice.gov.uk, philip.brohan@metoffice.gov.uk, john.kennedy@metoffice.gov.uk*

[(2)] *Lamont-Doherty Earth Observatory of Columbia University, P.O. Box 1000, Palisades, NY 10964-8000, U.S.A., Email: alexeyk@ldeo.columbia.edu*

[(3)] *National Oceanography Centre, Waterfront Campus, European Way, Southampton SO14 3ZH, United Kingdom, Email: eck@noc.soton.ac.uk*

[(4)] *NOAA (National Oceanic and Atmospheric Administration)/National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801-5001, U.S.A., Email: Richard.W.Reynolds@noaa.gov*

[(5)] *NOAA (National Oceanic and Atmospheric Administration)/National Oceanographic Data Center, 1315 East-West Highway, Silver Spring, MD 20910, U.S.A., Email: Kenneth.Casey@noaa.gov, Tess.Brandon@noaa.gov*

[(6)] *NOAA/ESRL (National Oceanic and Atmospheric Administration/Earth System Research Laboratory) (R/PSD3), 325 Broadway, Boulder, CO 80305, U.S.A., Email: Scott.D.Woodruff@noaa.gov*

[(7)] *NOAA NESDIS STAR (National Oceanic and Atmospheric Administration/National Environmental Satellite-Data and Information-Service Satellite Applications and Research) and CICS/ESSIC (Cooperative Institute for Climate Studies /Earth Systems Science Interdisciplinary Center), 5825 Univ. Research Ct., Suite 4001, College Park, MD 20740, U.S.A., Email: Tom.Smith@noaa.gov*

[(8)] *ESA/ESTEC (European Space Agency/European Space Research and Technology Centre) (EOP-SME (Environment Operational Programme-Small and Medium-sized Enterprises)), Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands, Email: Craig.Donlon@esa.int*

[(9)] *Norwegian Meteorological Institute, P.O.BOX 43, Blindern, N-0313 Oslo, Norway, Email: s.eastwood@met.no, lars.anders.breivik@met.no*

[(10)] *Meteorological Research Institute, Japan Meteorological Agency, 1-1 Nagamine, Tsukuba, Ibaraki, 305-0052, Japan, Email: maish@mri-jma.go.jp*

1. ABSTRACT

There are many challenges in blending historical and modern observations of sea surface temperature (SST) into homogenous gridded data sets suitable for use in climate research. Many of these problems can be avoided if proper choices are made during design and deployment phases of new instrumentation and observing systems, as specified in the Global Climate Observing System (GCOS) climate monitoring principles (http://www.wmo.int/pages/prog/gcos/documents/GCOS_Climate_Monitoring_Principles.pdf).

2. INTRODUCTION

Sea surface temperature (SST) is one of the most important physical variables in the arena of climate change. The large thermal inertia of the surface ocean (as compared to the land or the atmosphere) makes SST a particularly suitable parameter for monitoring changes in the state of the climate system, for predicting seasonal-to-interannual climate variability, and for verifying projections of longer-term climate change. In order to project climate trends into the future, we need to know and understand their history. This task, in turn, requires observational samples long enough for secular trends and multi-decadal variability to be identifiable in the presence of the very energetic background of shorter period variability. Historical (century or longer) data sets of SST are needed for training coupled models for seasonal-to-interannual climate forecasts [1] and for verifying or constraining century-long climate projections [2]. Historical SST data sets are also essential for atmospheric reanalyses (e.g. [3] and [4]) and model-based studies of climate variability. They also play a major role in the calibration of paleoceanographic records, thus contributing to multi-century and millennial climate reconstructions (e.g. [5] and [6]).

For these tasks, the quality of information about past SST variations is almost as important as the quality of modern observations. The historical-SST community strives to make the best use of modern SST observations to improve analyses of the past. Here we summarise the progress and problems in this area in the last decade, since the OceanObs'99 review of SST analyses [7].

Most historical SST observations were extracted from the logbooks of ships, with buoy measurements and

satellite data becoming a significant portion of the data set in the 1980s. The ultimate collection of *in situ* surface ocean data is the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, [8, 9 and 10]). Data coverage prior to the second half of the nineteenth century is extremely poor (with little or no SST data prior to 1800), so only the period after 1850 has been used for climate analyses so far.

Most climate research uses require spatially complete representation of SST on regular grids. These "gridded analyses" of historical SST depend as much on their "analysis" method, i.e. how missing data are inferred and how existing data are smoothed, as on the underlying set of observations, their quality control (QC) and bias correction procedures. The much-improved sampling of SST in the last ten years by buoys and satellites has resulted not only in very good data for the last decade, but also in some qualitative improvements for the analysis of earlier data. Data abundance has made it possible to better identify and correct platform-specific systematic biases. It has also become possible to refine estimates of observational error covariance. This process was helped tremendously by the synergy created by the Group for High Resolution SST (formerly the GODAE (Global Ocean Data Assimilation Experiment) High Resolution Pilot Project, GHRSST-PP, see [11]).

In the following sections, we outline the main problems and progress in the recent development of century-scale gridded SST analyses, discuss ways to avoid these problems in the future and make recommendations that, if implemented, would result in a greatly improved *in situ* observing system.

3. PROBLEMS AND PROGRESS

3.1. Systematic biases and their correction

The historical database of SST observations contains both slowly and rapidly varying relative biases arising from changes in instrumentation and data sources. These relative biases are of comparable size to climate variations, so they must be corrected. The need for bias adjustments is well known and a number of adjustments have been developed (e.g. [12]). Uncertainties in SST bias adjustments have also been estimated in recent years [13], [14] and [15]. This is a large component of the total uncertainty in historical SST variations. Reducing this uncertainty is particularly important, as it tends to be correlated in space and time, and therefore is often not reduced by averaging. On a global average, the bias adjustment uncertainty peaks in the late 1930s at around 0.08°C, with local values of roughly 0.05-0.2°C. Reducing this uncertainty would greatly improve historical gridded analyses. However, this is hampered by inadequate metadata.

In the mid-nineteenth century, semi-insulated wooden buckets were used on ships to collect water samples, but by the early twentieth century, the majority of SST measurements were made using uninsulated canvas buckets, which tended to lose heat via evaporation. These were gradually phased out and typically replaced by measurements of the temperature of engine room intake (ERI) cooling water, or observations using better-insulated rubber buckets (Fig. 1, bottom panel). ERI measurements tend to be warm relative to measurements made using canvas buckets, as do those from insulated buckets. Each observing country switched from using canvas buckets to ERI or insulated buckets at different times, introducing temporal and geographical variations into the bias in the SST record [16]. Reference [17] gives more detail about Voluntary Observing Ship (VOS) measurements.
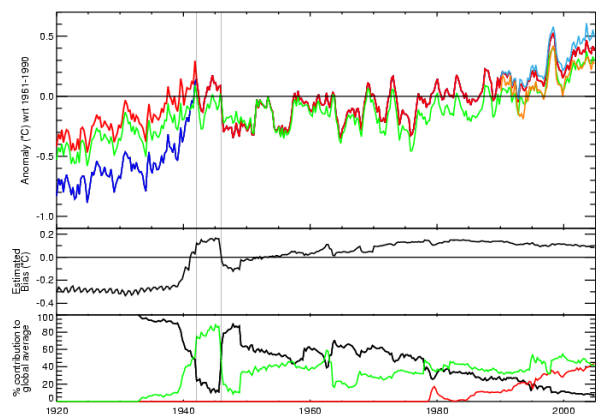

Figure 1. Top: global SST anomaly series (°C, relative to 1961-90 average), 1920-2006, from: uncorrected data (blue); HadSST2 (Hadley Sea Surface Temperature) (red,[14]); a provisional series corrected for all known inhomogeneities throughout the record (green, [16]); drifting buoys (orange) and ships (cyan). Middle: estimated bias in global average SST (°C, [16]). Bottom: percentage contribution to the global average of bucket (black), engine room intake (green) and buoy (red) measurements.

More recently, manual shipboard observations have been augmented by automatic measurements made by drifting and moored buoys (red curve in the bottom panel of Fig. 1). Drifting buoy design was standardised around 1993 and their collective measurements appear to have been stable since then. However, relative to shipboard observations, drifting buoy SST measurements tend to be cooler (Fig. 1, top panel). Since the mid 1980s, the mean ship-buoy SST difference is roughly 0.14°C. The increase in reliance on drifting buoy measurements of SST and the coincident decline in the contribution to the database of ship-borne

observations has led to a further temporally and geographically varying bias in the SST record. Accounting for this new bias is also important. However, since they provide a stable subset of data from 1993 onwards, drifting buoys can be used to better understand and correct historical ship observations.

Moored buoys from different nations have relative biases, but geographical coverage is limited. Other sources of *in situ* SST data make a relatively minor contribution to the record and are not considered in this section (but see Sect. 4.1 for further discussion).

Since the early 1980s, satellites have been retrieving a wealth of SST data. Like *in situ* data, satellite observations can suffer from biases geographically, temporally, and both within the lifetime of an individual sensor and between sensors. Combining *in situ* and satellite observations is additionally complicated both by the complex nature of the surface ocean and that every sensor, satellite and *in situ*, is measuring a different component of it. (More about satellite retrievals of SST can be found in [11]). However, stable sets of retrievals, such as from the Along Track Scanning Radiometer (ATSR) series or the Pathfinder reanalysis of Advanced Very High Resolution Radiometer (AVHRR), help us to better interpret the *in situ* record. The reverse is also true: extensive use is made of drifting buoy observations to calibrate or validate satellite retrievals. In addition, combining SST retrievals from different satellite instruments can further reduce biases (see e.g. [18]).

The AVHRR Pathfinder program ([19] and [20]) uses a multi-step method to minimize spurious trends and maximize cross-satellite homogeneity. Radiances are converted to brightness temperatures based on each sensor's operating temperature. The calculation of SST from these brightness temperatures is based on regression to *in situ* observations, which further ensures consistency across the transition from one satellite to the next.

3.2. Sampling and measurement uncertainties

In addition to systematic biases, all data types are affected by uncertainties due to measurement errors and under-sampling of variability. Measurement uncertainties in ship SST observations have been estimated by comparing pairs of nearby observations (Fig. 2a) [21]. The total variance of observations within each bin (Fig. 2b) should exceed measurement and sampling error variance by the real variance of the field there. High resolution satellite data (e.g. Pathfinder AVHRR, version 5 [20]) is useful for estimating this component (Fig. 2c) [22]. A combined sampling and measurement error for a single observation can be derived (Fig. 2d), but note that in the absence of an observation, sampling error equals the local SST

anomaly standard deviation. Assuming the errors in a grid box are uncorrelated, the error in grid box averages is then obtained by dividing by the square root of the number of observations in each grid box. These values can exceed 1°C locally. As Figs. 2e and f demonstrate, the resulting error model successfully captures the major spatial features of Pathfinder-ICOADS difference.

The distinction between random and bias errors can sometimes be blurred when observations from a random mixture of platform types with unequal biases are combined. For regions and periods where a single or small number of platforms contribute to the SST, estimate the common bias may remain even when random errors are reduced by averaging [23] and [24].

The availability of independent SST estimates from different platforms also allows estimates of uncertainties pertaining to each individual platform to be determined (e.g. [25]); note here another need for independent data.
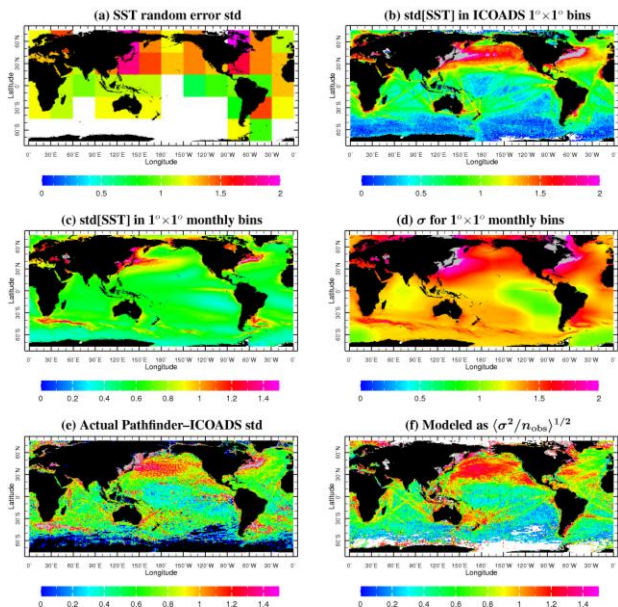

Figure 2. Effective observational error in in situ SST and its components: (a) random error estimate for one ship observation [21]; (b) actual standard deviation in ICOADS 1° x 1° monthly bins, 1960-2005; (c) SST variability within 1° x 1° monthly bins, 4 km Pathfinder v5 daily SST [20], 1985-2004; (d) standard error for a single observation in 1° x 1° monthly bins, estimated by combining (a) and (c); (e) standard deviation of SST difference between 1° x 1° monthly Pathfinder (night) and ICOADS, 2000-2004; (f) average ICOADS SST error in 1° x 1° monthly grid boxes during 2000-2004, estimated from a single observation error estimate and actual number of observations in ICOADS bins in each month.

3.3. Sea ice and marginal ice zone SST

For many applications, where information on fluxes of heat is needed, globally complete fields of SST and sea ice concentration (SIC) are required. Sea ice concentrations can be used to estimate SST values at high latitudes where there are few direct observations. Providing homogeneous fields for the last century or so is more challenging for SIC than for SST: the information content of data from different periods is more variable; data are less centrally organised; metadata are poor and independent data for validation are scarce.

Since 1978, the SMMR/SSM/I (Scanning Multifrequency Microwave Radiometer/Special Sensor Microwave Image) series has provided detailed and continuous information on SIC in both hemispheres (see [26]). Prior to this, information on sea ice extents, much less on concentration, is sketchy in the Southern Hemisphere. Homogeneity is an essential goal because, for example, an atmosphere-only model which uses a globally complete analysis of SST and sea ice as a lower boundary condition will suffer spurious discontinuities in heat fluxes should the characteristics of the SIC field change suddenly. For this reason, removal of biases in SIC retrievals from passive microwave instruments is also essential. Biases arise from, for example, surface melt ponding in the summer months, when ponds of fresh water form on top of the ice, causing underestimation of ice concentration.

In grid boxes with partial ice cover, the SST attributed to the open water in the grid box is also important. The water might be assumed to be near to its freezing point, but during the summer months large excursions of a few degrees can occur. A potential boost to our understanding of marginal ice zone SST has come with the POLEWARD project, which so far has deployed more than 100 drifting buoys in the Norwegian Sea and Barents Sea.

A reanalysis of the passive microwave record for the last thirty years is currently underway at the EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) Ocean and Sea Ice Satellite Application Facility. This work will ascribe each pixel with an error estimate for the first time. Significant work has also been undertaken with the historical record: a collection of digitised operational charts from various nations has been assembled in the Global Digital Sea Ice Data Bank. Some of these extend back to the 1930s. Reconciling recent and historical observations and assessing the resultant uncertainties is an area of active research.

3.4. Climate variability and change estimation and the effects of analysis methodology

As previously stated, gridded analyses are produced by inferring missing data and smoothing available observations. The analysis methods used make assumptions about the statistical properties of the observations. These assumptions can differ significantly from the properties of real-world observations, e.g. the climate of the last 150 years is not stationary and observational distribution is sometimes sparse and not random. These deviations of the real from the ideal world result in errors in the gridded analyses, e.g. reduced trends, over-fitting of data or lack of variance.
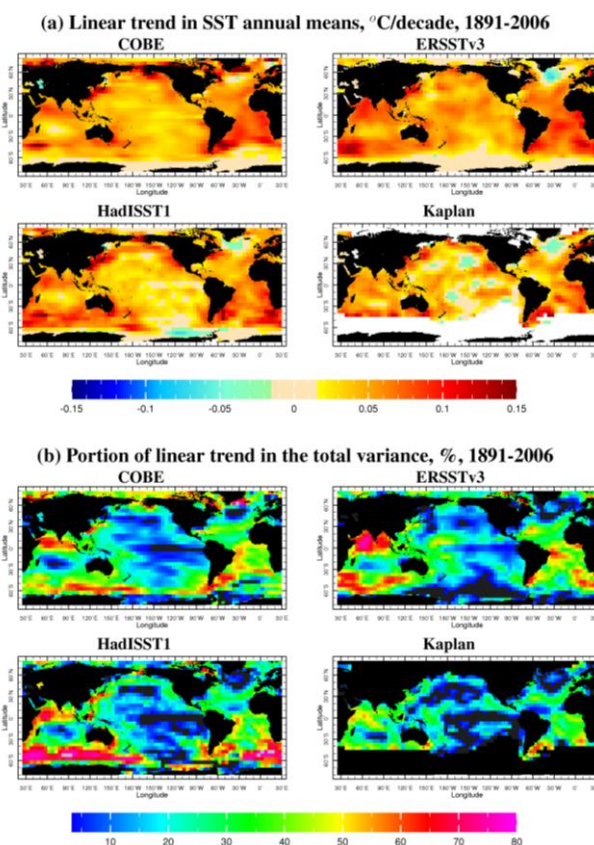

Figure 3. Comparison of four SST analyses, 1891-2006. a) Linear trends in annual mean SST anomaly and b) proportion of total variance explained by the trend. Blacked-out areas: trends are not significant at 5% level.

One of the most straightforward uses of gridded SST analyses (e.g. COBE (Centennial in-situ Observation-Based Estimates) SST [27], ERSST (Extended Reconstructed Sea Surface Temperatures) version 3 [15], HadISST1 [28] and Kaplan SST [29]) is computing and interpreting their long-term trend patterns. Such patterns for the 116-year period 1891-2006 are shown in Fig. 3a. Fig. 3b shows the percentage

of SST variance in this period due to the local trend. Blacked-out areas show where trends are not significant at the 5% level.

There are similarities and differences between the trend patterns in each analysis, due to differences in the underlying observational databases, their QC and bias correction procedures and in the analysis methods used. HadISST1 and Kaplan are the most similar in terms of the pattern shape, if not mean value: the Kaplan analysis was produced under the conservative assumption of stationarity of the mean climate for the entire analysis period, while HadISST1 includes a pre-processing, which explicitly derives the modes of long-term variability, capturing global warming. Therefore Kaplan has uniformly less warming than HadISST1. These two products use slightly different observational databases, with similar bias corrections and similar analysis methods. ERSSTv3 uses a different observational database, different set of bias corrections and a somewhat different analysis method. It shows stronger and more uniform warming than HadISST1 and Kaplan. The COBE analysis also shows large and uniform warming. The differences between trend patterns are driven mostly by the observational data, QC and bias correction differences in the first half of the trend period (1891-1948): the trend patterns are more different then (not shown). However, in the period of better and more abundant data (1949-2006) the four trend patterns are more similar to each other (not shown).

4. FUTURE OBSERVING SYSTEM NEEDS

4.1. Adequacy and stability of the SST observing system

As previously discussed, SST observations are collected from many different *in situ* platforms, including surface observations from ships (both VOS and research vessels), and moored and drifting buoys. In addition, measurements of the temperature of the ocean skin and sub-skin are made by a range of infrared and microwave satellite instruments, and near surface temperatures are obtained from oceanographic profiles. Near-surface temperatures (within the first few metres) from profiles measured by mechanical and expendable bathythermographs (MBTs and XBTs), conductivity-temperature-depth instruments (CTDs), Argo floats and gliders, are beginning to be augmented by higher vertical resolution SST measurements from a subset of Argo floats. This information on near surface temperature structure will allow us to better reconcile SST from *in situ* and satellite platforms. We note though, that while Argo is essential for studies of ocean heat content, understanding sea level changes and the vertical profile of the near-surface ocean, it could not be a substitute for other *in situ* SST measurements with its current, relatively low, sampling frequency. A combination of VOS, drifting buoys (with and without thermistor chains) and Argo floats is needed.

Whilst this multitude of observational sources makes SST one of the best observed ocean parameters, the diversity and evolution of the balance and quality of those observational sources means that the potential for data inhomogeneity is great (see Section 2.1). *In situ* SST is measured at many different depths, which are often unknown. **Metadata describing observing method and depth is needed to understand and reconcile those differences**, which are due to physical effects, such as diurnal variability (warming and subsequent cooling of the ocean surface controlled by solar heating and wind), and depth differences, from those which are biases in an individual platform or instrument. Satellite instruments produce a much higher density of observations than *in situ* platforms. However, the number of satellite instruments is much smaller than the number of *in situ* platforms. Thus, biases in satellite retrievals due to cloud and aerosol contamination (for infrared retrievals) and land and rainfall contamination (for microwave retrievals) can have a large, systematic impact on both the retrievals and the gridded analyses that use them. Furthermore, satellite instruments may drift from their prescribed orbit or calibration, degrading their accuracy. Reference [18] addresses many of these satellite-bias problems using comparisons to *in situ* data. However, such adjustments cannot be substantially improved without reducing the *in situ* bias, which requires more *in situ* metadata.

To remove systematic biases from the SST observing system, it is important to compare observations from different sources, both within and between platform types. Nearby observations are needed to ensure high data quality. This apparent **redundancy in the observing system is necessary** for the generation of accurate global gridded datasets and analyses (coupled importantly with redundancy in data management and archival systems). Routine comparisons between SST measured by surface platforms to that from oceanographic profiles would also help to determine relative biases in these different data types and could prevent future high profile problems with new instrumentation, such as those uncovered by [30].

Confidence in observed SST is also increased by its consistency with an independent quantity: marine surface air temperature from VOS. In general, **co-incident information on other atmospheric variables is helpful** for understanding and removing biases, which depend on environmental parameters. Such information is available in ICOADS, which brings together data from a diversity of platform types and limited available historical metadata (e.g. [31]), and in various GHRSST data sets.

Accuracy requirements for SST (and other variables, see http://www.wmo.int/pages/prog/sat/Databases.html) vary widely both within and between user communities, and between different assessments. For climate purposes, an accuracy of between 0.1 and 0.5°C on grid resolutions of 1-500km and 1-30 days is specified between the different assessments. This wide range of stated needs makes assessing the adequacy of the observing system rather challenging. **We recommend that databases of accuracy requirements are rationalised.**

Here we use fields of random synthetic observations with statistical characteristics equal to those of VOS and drifting buoys during the period August 2002-December 2007 to evaluate how many platforms and observations are required in any grid box to satisfy certain accuracy requirements. We use a general grid box, without defining its resolution in space and time. However, as sampling error (for one observation) varies from location to location with the variability of the field, we calculate results for different sampling errors. Fig. 4 summarises the results and provides a graphical "look up table" for various accuracy requirements.

As an example, if the requirement on a gridded average were "better than 0.1°C" and the sampling error in that grid box were 0.8°C for one observation, then 5 buoys making at least 30 observations each is one way to meet that requirement (as is 10 buoys making 15 observations, etc.). The other panels are used to inform less stringent requirements. The information is displayed in this way, because the results are strongly dependent on the estimate of sampling error used. Also, the shaded areas give a number of possible deployments that might be used to achieve the same objective, so displaying in map form is not straightforward. Note that the errors displayed in Fig. 2 are a combination of both sampling and measurement error, so would give a pessimistic view if combined directly with Fig. 4, which requires that we know the sampling error only (the measurement errors are included already following the method of [24]).

Note that it is not possible to meet the 0.1°C accuracy with VOS observations with error characteristics typical of the VOS fleet between 2002 and 2007. However, high accuracy can be achieved in VOS SST measurements without high costs [32] and exemplary deployments should be replicated.
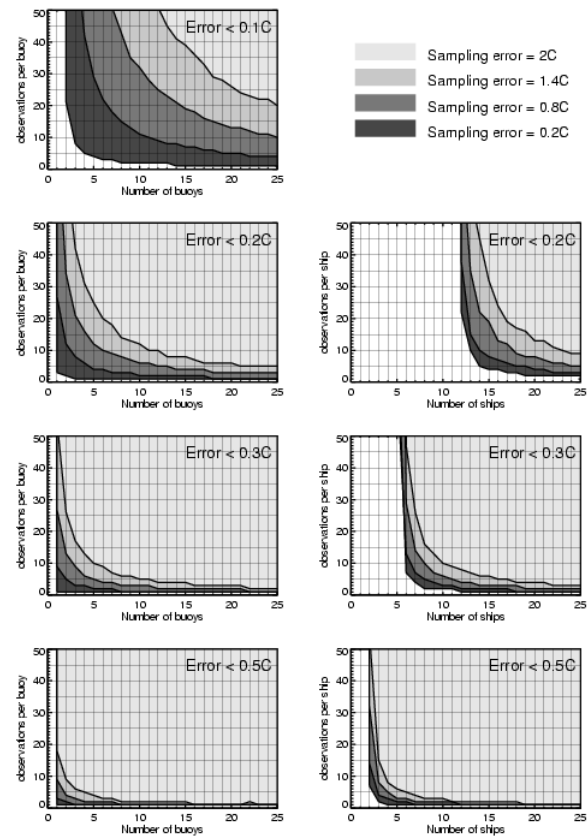

Figure 4. Number of platforms and observations needed within any grid box in order to satisfy different accuracy requirements. The top panel shows the number of buoys and number of observations required to ensure that the probability that the grid box error exceeds 0.1°C is less than one third. Shading indicates how the number of buoys and observations varies depending on the sampling error of the particular grid box - more highly variable regions will have larger sampling errors. See text for an example. Accuracy requirements are: better than 0.1°C (top); better than 0.2°C (second row); better than 0.3°C (third row) and better than 0.5°C (bottom). Left: requirements for buoys; right: requirements for ships.

The **desirable features of a surface marine observing system for SST include a large number of observations with known (preferably high) data quality and adequate sampling in space and time**. In addition, it is desirable to have **a large number of identifiable platforms with documented measurement methods and sufficient metadata;**

ongoing monitoring and analysis of data quality (see GCOS climate monitoring principle 4)**; a wide range of information on co-incident environmental parameters; consistency with (or quantifiable differences from) past observations; and feedback of monitoring and analysis results to observing system operators**. Data that include this range of features can be more fully analysed and adjusted for bias, and will lead to improved gridded SST data sets and analyses for climate studies.

4.2. The need for both in situ and satellite observing systems

Accurate global satellite infrared SST measurements began in 1981 while global microwave measurements began in 2002. The latter have better coverage while the former have better resolution. Both have far better coverage than *in situ* observations.

In situ observations allow historical gridded analyses of large spatial (typically $\geq 2°$) and temporal scales (typically ≥ 1 month). In addition, other variables (e.g., winds, air temperature, humidity) are often observed simultaneously from *in situ* platforms (mostly ships). Satellite observations allow much higher spatial and temporal resolution gridded analyses. In gridded analyses, especially those using multiple satellite instruments, the direct contribution of *in situ* observations is usually overwhelmed. Furthermore, some analysis schemes have chosen one satellite instrument as a standard and eliminated the direct use of *in situ* observations. However, in most cases *in situ* **data play a critical role in the ongoing calibration and validation of satellite algorithms for climate-quality gridded SST analysis products**.

The availability of high temporal resolution SST observations from *in situ* and satellite sources has also enabled better understanding of high resolution physical processes. An example is the diurnal variability and vertical profile of temperature in the top few metres of the ocean [33], although here better integration of surface and sub-surface observations is required.

4.3. The need to maintain long enough overlap between satellite missions

GCOS monitoring principle number 12 says: "A suitable period of overlap for new and old satellite systems should be ensured for a period adequate to determine inter-satellite biases and maintain the homogeneity and consistency of time-series observations." Evidence of this need abounds in the literature, perhaps most famously for observations of tropospheric temperatures from the series of (Advanced) Microwave Sounding Units ((A)MSU) on the NOAA (National Oceanic and Atmospheric Administration) polar orbiting satellites (e.g. [34]). Because of

differences between satellite instruments, their trends took a long time to reconcile to better-known surface temperature trends.

Each time a new satellite replaces an old satellite, a step discontinuity is introduced into the record. Over a number of missions these discontinuities, exacerbated by drifts in the satellite calibration, constitute a random walk that adds red noise to the merged record. A long overlap between sensors minimises this and allows a self-consistent, independent record to be created. Short overlaps make it difficult to estimate the relative biases. The length of suitable overlap depends on the characteristics of each sensor. Appropriate overlap periods for new sensors should be calculated as part of mission design.

Such problems can be corrected using other instruments – *in situ* or satellite – at the cost of their independence. Independent records are desirable because they decrease uncertainty, thereby increasing confidence in all products and making failures easy to trace; they contribute to a robust network that is not overly reliant on any single component.

4.4. Requirements for future sea ice observations

There are many data sets of passive microwave retrievals of SIC, but none are globally validated. For users to have an understanding of retrieval accuracy in different locations, **passive microwave data sets need global validation and retrieval uncertainties for each pixel**.

Passive microwave retrievals of SIC constitute a 30-year record, albeit with serious shortcomings, e.g. under-estimation due to melt ponds on the surface of the ice and due to thin ice. Synthetic Aperture Radar (SAR) imagery is widely thought to be the most accurate source of sea ice information and is used by operational ice services in the drafting of their charts. Yet for studies of climate variability and change, a record of sea ice changes using SAR data is not available. This is partly because of data cost and partly because no automatic retrieval of sea ice concentration from SAR imagery is yet possible. The best we have at present is an analyst's subjective interpretation of SAR imagery in the form of an ice chart. While we have very high-resolution images of SST, we do not have free and easy access to interpreted high-resolution sea ice images from SAR. **It should be possible for SAR data providers to derive a product for climate use.**

4.5. Need for further recovery and interpretation of historical observations and metadata

The present-day network of VOS and buoys provides frequent and widespread coverage of the world's oceans. But the coverage of the historical archive, on which we

rely for estimating, understanding and predicting climate change, should be improved to the maximum extent possible. The number of observations in ICOADS, and the fraction of ocean surface from which observations are available, both decrease steadily as we go back in time (with additional temporary reductions during the two world wars). Many early historical VOS sources were digitised decades ago. Due to the limitations of early computer technology they often lack adequate metadata (including ship identification) and may have been subject to undetected data processing errors [35]. With the existing archive, we cannot reconstruct even global average SST changes before about 1850, and reliable regional reconstructions are even shorter.

Platform and instrumental metadata for VOS have been regularly collected by WMO since 1955 [31]. Historical metadata from drifting and moored buoys, and other automated platform types, starting in the 1970s may still exist at widely dispersed data centres, possibly at risk of media degradation or other loss. International efforts to systematically consolidate and archive these important metadata are only just beginning through the Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM).

Improving our historical reconstructions of surface marine climate is a high priority. Recently released data for 1908-1958 from the Twentieth Century Reanalysis Project ([4] and [36]) extend the global atmospheric record, reanalyzed by state-of-the-art General Circulation Models, to the century scale. This product and its ongoing extension for 1891-2002 would not have been possible without a recent significant international effort on historical data inventory, digitization, and exchange. However, an extension of this activity for 1840s-2011, the NOAA-CIRES (Cooperative Institute for Research in. Environmental Sciences) Surface Input Reanalysis for Climate Applications, is still significantly restricted in its quality and scope by the shortage of historical observations.

There is much potential for improving the number and coverage of historical observations and metadata, but this requires international collaboration on data recovery and QC. Recent and ongoing initiatives [37] include: recovery of U.K. Royal Navy observations for 1939-47; Norwegian and international polar exploration results for 1867-1930; the German Maury collection for 1845-67; English East-India Company records for 1790-1834 and the Global Oceanographic Data Archaeology and Rescue project [38]. **Continuing work to digitise historical observations and metadata (from ships, buoys and oceanographic profiles), and blend them (using standard formats and improved quality controls) into ICOADS will result in much more**

detailed and reliable estimates of marine climate change and variability.

4.6. Further developments in analysis methodology

The GCOS SST/sea ice Working Group are conducting a test experiment, applying all individual analysis and QC methodologies to common observational databases. This effort will result in a better understanding of the source of the differences between gridded analyses and specific recommendations for the improvement of gridded analyses, in addition to those made here.

Different analysis techniques are currently applied to sparsely sampled ship and buoy databases (e.g. [28]) from those applied to much more abundant satellite data (e.g. [18]). Reduced-space analysis methods, using covariance estimates encapsulated by Empirical Orthogonal Functions (EOFs), based on *in situ* data produce relatively low resolution fields, leave permanent "holes" in the locations where poor data sampling makes this method of covariance estimation unfeasible, and do not explicitly account for the long-term variability (trends) in the data [39]. Because of users' demand, the HadISST, ERSST, and COBE products use additional heuristic approaches in addition to the reduced-space analysis to get around these problems. **More holistic methodological approaches should be developed, which cope better with real-world deviations from theoretically ideal observations.** This would result in better gridded analyses and better understanding of their features. They should also explicitly address the uncertainty in covariance estimates used (as in [40]).

As applications of gridded SST analyses become more and more sophisticated, **more reliable and user-friendly representations of uncertainty should be provided**. For example, this could be achieved by representing analysis uncertainty via an ensemble of equally probable realizations of the analysis, whose mean would be equal to the best-estimate analyzed fields currently produced and whose spread would correspond to the uncertainty of the analysis, simulating covariances of the analysis error as well [41]. This is a more user-friendly option than providing analysis standard errors.

4.7. Further developments in bias corrections

In the past, the evaluation of bias in historical SSTs has typically been done using aggregated data sets, i.e. not separated according to measurement type, with the bias adjustment then applied to the aggregated data. This creates problems when new historical data are digitized and added to the database since the new data may have been collected using different methods. To minimize this problem in the future, **multiple independent sets of**

bias adjustments and their uncertainties should be computed for individual homogeneous components of the historical (and contemporary) observing array (e.g. [16]) and an adjustment assigned to each individual observation. When new data are digitized, their bias could be separately evaluated using available metadata.

Without good metadata, an understanding of the evolution of the observing array is just educated guesswork. Whenever new instrumentation is deployed, therefore, **one of the most important considerations is the provision of comprehensive, accurate and accessible metadata** (as required by GCOS Climate Monitoring Principle number 3). For example, drifting buoys currently dominate the SST observing array, but the accessibility of good metadata on these instruments is poor. Metadata are currently accessible for just a subset of buoys and do not include important information such as measurement depth, which is used to understand the way different platforms observe diurnal variability in SST. Metadata for ships and moored buoys is also quite patchy (see Section 3.5 and [42]). While metadata including ship identification may be lost forever for some early historical sources, withholding ship call signs owing to commercial and security concerns is a recent development (~2007, see also [43]). This makes proper estimation of bias and uncertainty impossible, because without the call sign the link to detailed observational metadata cannot be made. Therefore, **withholding ship call signs must stop or be circumvented**.

The calibration of expendable instrumentation such as drifting buoys is also not checked or monitored directly. **Drifting buoys and their future equivalents should be randomly captured to allow any shift in their calibration to be quantified** in order to provide better understanding of the evolution of their accuracy through their lifetime. More diagnostic information should also be transmitted in real time.

5. REFERENCES

1. Chen, D., M.A. Cane, A. Kaplan, S.E. Zebiak & D.J. Huang (2004). Predictability of El Niño over the past 148 years. *Nature*, **428**(6984), 733-736.

2. Intergovernmental Panel on Climate Change (IPCC) (2007). *Climate Change 2007 – The Physical Science Basis*, Cambridge University Press, Cambridge, U.K.

3. Uppala, S.M. & coauthors (2005). The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**(612), 2961-3012.

4. Compo, G.P., J.S. Whitaker & P.D. Sardeshmukh (2006). Feasibility of a 100 year reanalysis using only surface pressure data. *Bull. Amer. Met. Soc.* **87**, 175-190.

5. Linsley, B.K., P.P.Zhang, A.Kaplan, S.S.Howe, G.M. Wellington (2008). Interdecadal-decadal climate variability from multicoral oxygen isotope records in the South Pacific Convergence Zone region since 1650 A.D. *Paleoceanography*, **23**, doi**:**10.1029/2007PA001539.

6. Black, D.E., M.A. Abahazi, R.C. Thunell, A. Kaplan, E.J. Tappa & L.C. Peterson (2007). An 8-century tropical Atlantic SST record from the Cariaco Basin: Baseline variability, twentieth-century warming, and Atlantic hurricane frequency, *Paleoceanography*, **22**, doi:10.1029/2007PA001427.

7. Reynolds, R.W., D.E. Harrison & D.C. Stokes (2001). Specific contributions to the observing system: Sea surface temperatures. In: *Observing the Oceans in the 21st Century, C.J. Koblinsky and N.R. Smith (eds)*, GODAE Project Office and Bureau of Meteorology, Melbourne, pp. 87-101.

8. Woodruff, S.D., H.F. Diaz, E.C. Kent, R.W. Reynolds & S.J. Worley (2008). The evolving SST record from ICOADS. In *Climate Variability and Extremes during the Past 100 Years* (S. Brönnimann, J. Luterbacher, T. Ewen, H.F. Diaz, R.S. Stolarski, and U. Neu, Eds.), Advances in Global Change Research, **33**, Springer, 65-83.

9. Worley, S. & Co-Authors (2010). "The Role of the International Comprehensive Ocean-Atmosphere Data Set in the Sustained Ocean Observing System" in these proceedings (Vol. 2), doi:10.5270/OceanObs09.cwp.94.

10. Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Eric Freeman, J., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R. and Wilkinson, C. (2011), ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, **31**, 7, 951–967,. doi:10.1002/joc.2103.

11. Donlon, C. & Co-Authors (2010). "Successes and Challenges for the Modern Sea Surface Temperature Observing System" in these proceedings (Vol. 2), doi:10.5270/OceanObs09.cwp.24.

12. Folland, C.K. & D.E. Parker (1995). Correction of instrumental biases in historical sea surface temperature data. *Q.J.R. Meteorol. Soc.* **121**, 319-367.

13. Folland, C.K., N.A. Rayner, S.J. Brown, T.M. Smith, S.S.P. Shen, D.E. Parker, I. Macadam, P.D. Jones, N.R. Jones, N. Nicholls & D.M.H. Sexton (2001). Global temperature change and its uncertainties since 1861. *Geophys. Res. Letts.* **28**, 2621-2624.

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

15. Smith, T.M., R.W. Reynolds, T.C. Peterson & J. Lawrimore (2008). Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *J. Climate*, **21**, 2283-2296.

16. Kennedy, J.J., R.O. Smith, N.A. Rayner, M. Saunby & D.E. Parker (2009). Reassessing biases and other uncertainties in sea surface temperature observations

since 1850. *In preparation*

17. Kent, E. & Co-Authors (2010). "The Voluntary Observing Ship (VOS) Scheme" in these proceedings (Vol. 2), doi:10.5270/OceanObs09.cwp.48.

18. Reynolds, R.W., Smith T.M., Liu C, Chelton, D.B., Casey, K.C. & Schlax, M.G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *J. Climate*, **20**, 5473-5496.

19. Kilpatrick, K., Podestá, G. & Evans, R.H. (2001). Overview of the NOAA/NASA AVHRR Pathfinder algorithm for sea. *Journal of Geophysical Research*, **106**(C5), 9179-9197.

20. Casey, K.S, R. Evans, K. Kilpatrick, E. J. Kearns, T. B. Brandon, R. W. Reynolds & J. Vazquez-Cuervo (2009). Toward the Sea Surface Temperature Climate Data Record from Space: An Improved Advanced Very High Resolution Radiometer Pathfinder Dataset, in preparation for *Journal of Climate*.

21. Kent, E.C. & Challenor, P.G. (2006). Toward estimating climatic trends in SST. Part II: random errors. *J. Atmos. Oceanic Technol.*, **23**(3), 476-486, doi:10.1175/JTECH1844.1.

22. Kaplan, A. (2009). Modeling observational error of binned in situ SST observations, *Int. J. of Climatology*, in prep.

23. Kent, E.C. & D.I. Berry (2008). Assessment of the Marine Observing System (ASMOS): Final Report, NOCS Research and Consultancy Report No. 32, 55pp.

24. Kennedy, J.J., R.O. Smith & N.A. Rayner (2009). Using AATSR data to assess the quality of *in situ* sea-surface temperature observations for climate studies, *submitted to Rem. Sens. Env*

25. O'Carroll, A.G, J. R. Eyre & R.W. Saunders (2008). Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations. *J. Atmos. Oceanic Technol.*, **25**(7), 1197-1207, doi:10.1175/2007JTECHO542.1

26. Breivik, L. & Co-Authors (2010). "Remote Sensing of Sea Ice" in these proceedings (Vol. 2), doi:10.5270/OceanObs09.cwp.11.

27. Ishii, M., A. Shouji, S. Sugimoto & T. Matsumoto (2005). Objective Analyses of SST and Marine Meteorological Variables for the 20th Century using ICOADS and the Kobe collection. *Int. J. Climatology*, **25**, 865-879.

28. Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell & A. Kaplan (2003). Globally complete analyses of SST, sea ice, and night marine air temperature, 1871-2000. *J. Geophysical Res.* **108**, doi:10.1029/2002JD002670.

29. Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal & B. Rajagopalan (1998). Analyses of global sea surface temperature 1856-1991. *J. Geophys. Res.*, **103**, 18567-18589.

30. Willis, J. K., J. M. Lyman, G. C. Johnson, and J. Gilson (2007), Correction to "Recent cooling of the upper ocean", *Geophys. Res. Lett.*, **34**, L16601, doi:10.1029/2007GL030323.

31. Kent, E.C., S.D. Woodruff & D.I. Berry (2007). WMO Publication No. 47 metadata and an assessment of observation heights in ICOADS. *J. Atmos. Oceanic Technol.*, **24**, 214-234.

32. Beggs, H., R. Verein, H. Kippo, M. Underwood, I. Barton, C. Steinberg, E. Schulz, R. Hibbins, A. Thomas & G. Ball (2009). Enhancing ship of opportunity sea surface temperature observations in the Australian region, GHRSST User Symposium abstract (http://www.ghrsst-pp.org/International-GHRSST-Data-Users-Symposium-Event.htm)

33. Donlon, C. & Co-Authors (2007). The Global Ocean Data Assimilation Experiment High-Resolution Sea Surface Temperature Pilot Project, *Bull. Amer. Meteorological Soc.*, **88**, 1197-1213.

34. Thorne P.W., Parker D.E., Christy J.R. & Mears C.A. (2005). Uncertainties in climate trends - Lessons from upper-air temperature records. *Bull. Amer. Meteorological Soc.*, 86, 1437-1442.

35. Woodruff, S.D., H.F. Diaz, S.J. Worley, R.W. Reynolds & S.J. Lubker (2005). Early ship observational data and ICOADS. *Climatic Change*, **73**, 169-194.

36. Compo, G.P. & Co-Authors (2009). The Twentieth Century Reanalysis Project, *Q. J. R. Meteorol. Soc.*, in preparation.

37. Wilkinson, C., Woodruff, S. D., Brohan, P., Claesson, S., Freeman, E., Koek, F., Lubker, S. J., Marzin, C. and Wheeler, D. (2011), Recovery of logbooks and international marine data: the RECLAIM project. *International Journal of Climatology*, **31**, 7, 968–979, doi: 10.1002/joc.2102

38. Levitus, S.; S. Sato, C. Maillard, N. Mikhailov, P. Caldwell & H. Dooley (2005). Building ocean profile-plankton databases for climate and ecosystem research. NOAA Tech. Report 117, 29 pp. (Available from NODC/NOAA, E/OC5, 1315 East West Highway, Silver Spring MD, 20910, U.S.A.)

39. Kaplan A., M.A. Cane & Y. Kushnir (2003). Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects, in *Advances in the Applications of Marine Climatology: The Dynamic Part of the WMO Guide to the Applications of Marine Climatology, WMO/TD-1081*, World Meteorological Organization, Geneva, Switzerland, pp. 199-216.

40. Ilin, A., and A.Kaplan, 2009: Bayesian PCA for reconstruction of historical sea surface temperatures, In: *Proceedings of International Joint Conference on Neural Networks (IJCNN 2009)*, Atlanta, GA, USA, June 14-19, 2009, paper 0246, pp.1322-1327.

41. Karspeck, A.R., S.R. Sain & A.Kaplan (2009). Modeling and prediction of non-stationary mid-scale spatial variability in historic Northern Hemisphere Atlantic sea surface temperatures, in prep.

42. Snowden, D. & Co-Authors (2010). "Metadata Management in Global Distributed Ocean Observation Networks" in these proceedings (Vol. 2), doi:10.5270/OceanObs09.cwp.8.

43. Woodruff, S. & Co-Authors (2010). "Surface In Situ Datasets for Marine Climatological Applications" in these proceedings (Vol. 2), doi:10.5270/OceanObs09.cwp.93.