INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4914

RMetS
Royal Meteorological Society

Assessing the health of the *in situ* global surface marine climate observing system

David I. Berry* and Elizabeth C. Kent

Marine Physics and Ocean Climate, National Oceanography Centre, Southampton, UK

ABSTRACT: The *in situ* surface marine climate observing system includes contributions from several different types of observing platforms. Most observations come from mobile platforms, e.g. ships or surface drifting buoys. Climate applications using marine observations often require fields of environmental parameters to be constructed on regular spatiotemporal grids. User requirements are therefore typically presented in terms of parameter uncertainty at particular space and timescales. It is therefore important to relate the characteristics of marine observations, in terms of their expected quality and sampling distribution, to these requirements. A simple method to estimate the instrumental uncertainty in fields derived from a mixture of observation types is presented. This method enables preliminary assessment of the extent to which the available observations meet the stated user requirements.

Example observing system adequacy assessments are presented for two climate variables, sea surface temperature (SST) and marine air temperature (MAT) using *in situ* data. The method is also applicable to gridded data sets constructed from combined *in situ* and satellite data. While the global metrics for SST show an improvement in observing system adequacy over time, the adequacy for MAT is declining. The assessments can determine the most efficient approach to improving observing system adequacy. For *in situ* SST the best approach would be to increase the number of different platforms making observations. For MAT, increasing the number of observations overall, regardless of platform and increasing the geographical coverage is required to reduce the uncertainty.

The assessments would be improved by more extensive evaluation of uncertainties associated with each different variable for each platform type. It would also be beneficial to review the completeness of the user requirements: e.g. to include user requirements relating to the stability of averages on large space and timescales required for climate monitoring, or for constructing estimates of air–sea exchange.

KEY WORDS observing system assessment; instrumental uncertainty; ICOADS; sea surface temperature; air temperature; Global Climate Observing System; marine observations

Received 31 March 2015; Revised 8 September 2016; Accepted 10 September 2016

1. Introduction

1.1. The Global Climate Observing System

The climate observing system underpins many activities beyond the monitoring of climate change. These include the detection and attribution of climate change (e.g. IPCC, 2013; Blunden and Arndt, 2014); short term and seasonal prediction (e.g. Balmaseda and Anderson, 2009); decadal prediction (e.g. Meehl *et al.*, 2014); hindcasting and re-analyses (e.g. Saha *et al.*, 2010; Compo *et al.*, 2011; Dee *et al.*, 2011; Rienecker *et al.*, 2011; Cardone *et al.*, 2015; Kobayashi *et al.*, 2015); and the impact of climate change on ecosystems (e.g. Reid and Valdés, 2011). Other research applications relying on an effective *in situ* surface marine climate observing system include satellite calibration and validation (e.g. Jackson *et al.*, 2009; Roberts *et al.*, 2010; Prytherch *et al.*, 2015) and climate model validation (IPCC, 2013). The design and implementation of the climate observing system is overseen and coordinated by the Global Climate Observing System (GCOS). It provides advice and guidance at national and international levels including through the publication of high-level documents on the progress and implementation of the GCOS (e.g. GCOS, 2009, 2010). Coordination through GCOS-affiliated expert bodies and panels, such as the World Meteorological Organisation (WMO)/Intergovernmental Oceanographic Commission (IOC), Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM), Ship Observations Team (SOT), the Data Buoy Cooperation Panel (DBCP), the Ocean Observations Panel for Climate (OOPC) and Atmospheric Observation Panel for Climate (AOPC).

Within GCOS, and associated expert panels, the observations are considered by domain (atmospheric, oceanic and terrestrial) and selected variables are designated as essential climate variables (ECVs). ECVs are defined as variables that are technically and economically feasible to observe globally and that will have a high impact on

* Correspondence to: D. I. Berry, Marine Physics and Ocean Climate, National Oceanography Centre, European Way, Southampton SO14 3ZH, UK. E-mail: dyb@noc.ac.uk

Table 1. List of surface ECVs for the marine atmospheric and oceanic domains from the 2010 GCOS implementation plan (GCOS, 2010).

Domain	ECV	Note
Atmospheric	Air temperature Wind speed and direction Water vapour Pressure Precipitation Surface radiation budget	Atmospheric surface includes measurements at standardized but globally varying measurements in close proximity to the surface.
Oceanic	Sea surface temperature Sea surface salinity Sea level Sea state Sea ice Surface current Ocean colour Carbon dioxide partial pressure Ocean acidity Phytoplankton	Including measurements within the surface mixed layer, usually within the upper 15 m.

the United Nations Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC). There are currently 50 defined ECVs (GCOS, 2010; Bojinski *et al.*, 2014), some spanning multiple domains and others specific to a single domain. Table 1 lists the current ECVs for the lower atmosphere over the oceans and the ocean surface.

The marine climate observing system is made up from many individual observing systems, both national and international systems (Lampitt *et al.*, 2010; Roemmich *et al.*, 2010) and satellite-based components (e.g. Merchant *et al.*, 2014). Examples include observations from Voluntary Observing Ships (VOS), Earth Observation satellites, moored and drifting buoys, and hereafter we refer to these different observing system components as platform types. Although these observing systems are coordinated internationally, they require individual nations, or groups of nations, to allocate resources for the observations, e.g. through the purchase and deployment of drifting buoys and floats, through the maintenance of arrays of moored buoys or contributions to the VOS Scheme. These systems have typically been designed to meet a range of requirements, and some have a focus on contributing to the operational observing system (e.g. through the WMO World Weather Watch). Some systems started out as research programmes and transition to becoming a sustained network once they have demonstrated sufficient maturity and utility (GCOS, 2010).

Elements of the observing system are known to suffer from pervasive systematic biases as discussed, e.g. by Kent *et al.* (2010) for sea surface temperature (SST) observations. These systematic biases may be correlated across observations from many different platforms under similar environmental conditions and for similar instrumentation. While possible to detect and, on average, remove the effect of these biases, residual errors remain that contribute to the uncertainty. Examples of such adjustments using models of bias include for bucket-measured SST (Folland and Parker, 1995), for daytime heating bias in MAT (Berry *et al.*, 2004) and to account for changes in measurement height in MAT (Kent *et al.*, 2013). In this article, we assume that these systematic biases, correlated across large components of the observing system, can be corrected and that any residual uncertainty due to these biases can be estimated. At the grid box level this residual uncertainty is typically small compared with the other terms but can be significant for the global mean.

1.2. User requirements for the observing system

In parallel to the coordination within GCOS, a series of requirements for the different components of the observing system for different uses have been specified. These are defined via the WMO Rolling Review of Requirements (http://www.wmo.int/pages/prog/www/OSY/GOS-RRR.html, accessed 21 October 2015) and listed in the WMO Observing Requirements database for different application areas and by ECV. For climate applications the requirements, integrated across all sensors (*in situ* and satellite), are developed by the OOPC and AOPC. For each variable and application area three levels of requirements are given (http://www.wmo-sat.info/oscar/observingrequirements, accessed 21 October 2015):

- *goal*: the level beyond which further improvements are not necessary;
- *breakthrough*: an intermediate level where, if achieved, significant improvement would be made for target application; and
- *threshold*: the minimum requirement to be met for the data to be useful.

The breakthrough level is seen as the optimal level for the design and planning of observing systems balancing cost against benefit. Examples of the requirements for SST and air temperature for climate applications are given in Table 2. These requirements undergo regular review. For example, as new research highlights deficiencies in the

Table 2. Climate requirements for SST and air temperature from the OOPC and AOPC as specified in the WMO OSCAR database http://www.wmo-sat.info/oscar/observingrequirements, last accessed 21 October 2015. The quoted uncertainties are for the 68% (1σ) confidence interval.

Variable	Application area	Level	Uncertainty (K)	Horizontal resolution (km)	Observing cycle (h)
SST	Climate-AOPC	Goal	0.250	10	3
		Breakthrough	0.400	50	6
		Threshold	1.000	500	24
SST	Climate-OOPC	Goal	0.100	1	1
		Breakthrough	0.126	8	3
		Threshold	0.200	500	24
Air temperature (at surface)	Climate-AOPC	Goal	0.100	25	3
		Breakthrough	0.150	50	6
		Threshold	0.300	100	12

targets and new measurement methods become available, more stringent targets can be set. As a result, the particular values used in this article are for illustrative purposes.

The user requirements presented in the WMO Rolling Review do not currently specify the stability requirement over time of large-scale averages, one of the most important characteristics of a climate observing system. This includes stability of individual sensors/instruments as well as the observing system as a whole. For the well-sampled case, the stability will be limited by changes over time to any unidentified or unadjusted pervasive biases (Jones and Wigley, 2010). This has been addressed partially in the satellite supplement to the GCOS implementation plan (GCOS, 2011), which lists stability requirements for decadal timescales. These stability requirements are extremely demanding: for SST the requirement is for the maximum change in systematic error to be less than 0.03 K per decade over 100 km scales. In the marine context such stability assessments have rarely been performed due to a lack of suitable reference measurements, and have so far only been possible for large-scale averages (Merchant *et al.*, 2012).

1.3. Assessment of the adequacy of the observing system

Early studies of the design of the global observing system compared observing requirements with instrument characteristics and estimates of natural variability to quantify the number of observations required to meet those requirements in different regions (e.g. Legler, 1991; Weller and Taylor, 1993). However, due to the composite nature of the GCOS and the diverse range of user requirements, assessments of the observing system are often qualitative and fragmented, with metrics defined by platform type. For example, the completeness of the oceanic component of the global observing system has been previously measured in terms of the number of platforms (e.g. of individual moored buoy, drifting buoy, VOS) making observations compared with a target number (e.g. GCOS, 2009). Using this metric, the *in situ* global ocean observing system was assessed as having reached 61% of initial goals (GCOS, 2009). Such metrics defined by platform or observation counts cannot be simply related to user requirements (such as those given in Table 2) and can only give a partial picture of the state of the observing system. This makes it difficult to assess whether requirements are being met. Observations of some ECVs have been in decline for many years (Kent *et al.*, 2006) and recently there has been concern over low data return from the tropical moored arrays (Legler and Hill, 2014).

A more detailed observation requirement assessment has been made for SST for satellite applications (Reynolds *et al.*, 2005). Here the buoy density (i.e. number of buoys within a grid box) required to reduce biases in satellite-derived SST to below 0.5 °C was determined. The *in situ* observing system was assessed, with drifting buoys, moored buoys and ships weighted according to their error characteristics to give an 'equivalent buoy density' (EBD). Regions where the EBD fell below the target threshold were then identified.

In this paper we will describe a method and metrics that could be used to make an integrated assessment of the observing system. The application of this method is demonstrated using *in situ* observations of MAT and SST but is applicable to other variables and to combined *in situ* and satellite observations. Application of the method highlights the improvement (for SST) and decline (MAT) in the climate observing system for these ECVs. In Section 2 we derive the uncertainty in grid box means calculated using observations from marine platforms and describe the method. In Section 3 we use this approach to estimate the uncertainty due to instrumental errors in the surface marine climate observing system. Sections 4 and 5 present and discuss the results.

2. Estimation of grid box uncertainty for mobile platforms

Observations over the oceans can contain correlated errors due to, e.g. multiple observations from the same platform containing a calibration error or for observations clustered together in space and time. Basing uncertainty estimates on the observed frequency distribution of those observations (e.g. using the standard error of the mean) is therefore likely to be unreliable, especially when the sample size is small. Similarly, uncertainty estimates based on the decomposition of the observed variance (e.g. Lindau, 2003) are likely to be inaccurate. Following others (e.g. Kent and Berry, 2008; Kennedy *et al.*, 2011) we can, however, perform a type B evaluation of the uncertainty (e.g. BIPM *et al.*, 2008).

Point observations within a grid box can be expressed as:

$$T_{\text{obs},l} = T_{\text{true},l} + e_{\text{cor},l} + e_{\text{uncor},l} \quad (1)$$

where $T_{\text{obs},l}$ is the value of observation l at a discrete point in time and space, $T_{\text{true},l}$ the true value at that location, $e_{\text{cor},l}$ the error in the observation due to correlated effects and $e_{\text{uncor},l}$ the error in the observation due to uncorrelated effects. Hereafter we refer to these as correlated and uncorrelated errors, respectively.

Examples of uncorrelated errors include instrumental noise and any other errors arising during observation that are expected to average to zero over many measurements. Correlated errors can be correlated between observations made by the same platform, for observations made using the same method on different platforms, for observations taken under particular environmental conditions, or for any other group of observations with a common characteristic that can lead to biased measurements. Examples of errors correlated for an individual platform include effects due to poor sitting of instruments, calibration, or parallax errors. These effects may be random when considered across many different platforms. Examples of errors correlated across observations made using a common method include bias in bucket SST measurements (Folland and Parker, 1995). We assume that such pervasive errors correlated across significant elements of the networks can be detected, the mean effect removed and the residual uncertainty estimated. This approach is commonly used (e.g. Kennedy *et al.*, 2011; Hirahara *et al.*, 2014; Huang *et al.*, 2015), including in the construction of data sets used in IPCC assessments (IPCC, 2013).

In order to estimate the grid box mean, and its uncertainty, we need to link the point values with the grid box mean. The true point value, corresponding to observation l, can be decomposed into the grid box mean ($\overline{T}$) and a deviation from the mean ($\Delta_{xyt,l}$) and expressed as:

$$T_{\text{true},l} = \overline{T} + \Delta_{xyt,l} \quad (2)$$

Conversely, the grid box arithmetic mean would be given by:

$$\overline{T} = \frac{1}{N}\sum_{l=1}^{N}\left(T_{\text{true},l} - \Delta_{xyt,l}\right) \quad (3)$$

For a well-sampled grid box the index l would traverse all discrete points in time and space within the grid box and $N \to \infty$. Replacing the true value with the observed we then have:

$$\overline{T} = \frac{1}{N}\sum_{l=1}^{N}\left(T_{\text{obs},l} - e_{\text{cor},l} - e_{\text{uncor},l} - \Delta_{xyt,l}\right) \quad (4)$$

The error in our estimate of the mean based on equally weighted observations is then given by:

$$\left(\frac{1}{N}\sum_{l=1}^{N}T_{\text{obs},l}\right) - \overline{T} = \frac{1}{N}\sum_{l=1}^{N}\left(e_{\text{cor},l} + e_{\text{uncor},l} + \Delta_{xyt,l}\right) \quad (5)$$

The mean effect of the correlated errors, $e_{\text{cor},l}$, gives rise to a bias in our grid box mean. By definition the mean effect of the uncorrelated errors, $e_{\text{uncor},l}$, will tend to zero for normally distributed errors and large sample sizes. The mean effect of the $\Delta_{xyt,l}$ gives rise to the sampling error, the difference between our sample mean and the mean of a perfectly sampled grid box (the population mean). For large sample sizes this will tend to zero. For small sample sizes, neither the sampling error nor the mean effect of the uncorrelated errors will necessarily be zero.

Assuming that the mean effect of correlated errors has been estimated and removed (i.e. bias adjusted) and that the terms on the right-hand side of Equation (5) have zero expectation, the error variance in our mean, the uncertainty, is then given by the expected value of the square of the right-hand side. This assumption is discussed further in Section 3. Squaring Equation (5) and taking an ensemble mean we have:

$$\sigma^2_{\text{mean}} = E\left\{\left(\frac{1}{N}\sum_{l=1}^{N}\left(e_{\text{cor},l} + e_{\text{uncor},l} + \Delta_{xyt,l}\right)\right)^2\right\} \quad (6)$$

where $E\{\ \}$ indicates an expected value. Assuming that each term in the summation is independent of the others Equation (6) becomes:

$$\sigma^2_{\text{mean}} = E\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}e_{\text{cor},i}e_{\text{cor},j} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}e_{\text{uncor},i}e_{\text{uncor},j} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Delta_{xyt,i}\Delta_{xyt,j}\right\} \quad (7)$$

where the indices i and j represent all pairs of observations within the grid box. Recognizing that Equation (7) contains covariance terms and replacing these with the product of the appropriate correlations and standard deviations we have:

$$\sigma^2_{\text{mean}} = \overbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{\text{cor},ij}\sigma_{\text{cor},i}\sigma_{\text{cor},j}}^{\text{uncertainty due to correlated errors}} + \overbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{\text{uncor},ij}\sigma_{\text{uncor},i}\sigma_{\text{uncor},j}}^{\text{uncertainty due to uncorrelated errors}} + \overbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{xyt,ij}\sigma_{xyt,i}\sigma_{xyt,j}}^{\text{uncertainty due to sampling effects}} \quad (8)$$

where $\rho_{\mathrm{cor},ij}$ is the correlation between the errors due to correlated effects for observations i and j; $\rho_{\mathrm{uncor},ij}$ the correlation between the errors due to uncorrelated effects for observations i and j (by definition 0 but left in for completeness); and $\rho_{xyt,ij}$ is the correlation between the errors due to sampling effects for observations i and j. For environmental data $\rho_{xyt,ij}$ will depend on the natural variability and the distribution of the observations within the grid box in space and time. Equation (8) is equivalent to expressions given by, *inter alia*: Kent and Berry (2008); Kennedy *et al.* (2011); and Morrissey and Greene (2009).

In Equation (3) and subsequent equations we have used the arithmetic mean of the observations. Using a weighted mean instead, taking into account the different error terms and their sizes proportional to each other, may give a better estimate of the mean value. For example, moored buoy observations could be given higher weight than VOS observations because of their higher quality data. Due to the nature of the marine climate observing system weighting by the observational error terms will have little impact on the grid box uncertainty. Where we have few observations (<10) these are likely to come from a single platform. Where we have many the improvements are marginal. Weighting according to sampling, taking into account the correlation, leads to the kriging equations (e.g. Cressie, 1993) and is beyond the scope of this paper.

3. Application to the surface marine climate observing system

As an example we have applied Equation (8) to the marine component of the climate observing system, specifically *in situ* SST and MAT observations. Over the oceans *in situ* observations of the SST and MAT are routinely made by the different marine platform types (VOS, moored buoys and drifting buoys). These have been collated within the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) (Woodruff *et al.*, 2011). For this study we have used all observations of SST and MAT from ICOADS Release 2.5. This includes observations from non ship or buoy sources, such as oceanographic measurements and US Coastal-Marine Automated Network (C-MAN) stations. For the purposes of this study these have been treated as ship observations. This will have little impact on the results as these platforms are geographically limited. Equation (8) requires the clustering of observations into groups from the same platform in order to set $\rho_{\mathrm{cor},ij}$. While some platform identifier information is available in ICOADS prior to 1960 (e.g. Carella *et al.*, 2015), we have limited our analysis to the period 1960–2007. We end our analysis in 2007 due to the loss of call sign information for the VOS within ICOADS after this date (Woodruff *et al.*, 2011; JCOMM, 2013), noting that many missing call signs for this period will be reinstated in the next ICOADS release (E. Freeman, 2015; personal communication).

In order to apply Equation (8) we have considered only errors correlated by individual platform and not those correlated across different platforms equivalent to assuming that pervasive bias can be quantified and minimized. Ideally, we would also correct for biases specific to an individual platform but this is not always possible. Instead we can estimate the distribution of relative biases for individual platforms through comparison with another source of data and centre the observations so as a whole they are unbiased compared with that source (e.g. Kent and Berry, 2008). For centred observations we have $e_{\mathrm{cor}} \sim N\left(0, \sigma^2_{\mathrm{cor}}\right)$ across all platforms and $\rho_{\mathrm{cor},ij} = 1$ when the observations i and j are from the same platform and zero otherwise. By definition the uncorrelated errors are independent and $\rho_{\mathrm{uncor},ij} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. We have assumed $e_{\mathrm{uncor}} \sim N\left(0, \sigma^2_{\mathrm{uncor}}\right)$.

While potentially large, the uncertainty due to sampling errors (e.g. Gulev *et al.*, 2007a) will follow broadly similar patterns to the uncertainty due to instrumental errors. For example, where we have a large number of observations we would expect both the sampling uncertainty and instrumental uncertainty to be small. Similarly, where we have few observations we would expect both to be large. In high variability regions the uncertainty due to sampling errors will usually be dominant. In regions with high variability that are well-sampled (e.g. the Gulf Stream and Kuroshio current region), both contributions to the uncertainty are expected to be small. In high variability regions that are poorly sampled (e.g. the Southern Ocean) both contributions to the uncertainty would be expected to be large. As discussed by Kent and Berry (2005), in highly variable regions it is more important to increase the number of observations than to improve the accuracy of the observations. Given that the sampling uncertainty is relatively complicated to calculate (e.g. Gulev *et al.*, 2007b; Morrissey and Greene, 2009) and it may not be easy to do so operationally we have not included it in this analysis. Due to its relationship with the instrumental uncertainties this should not impact on the main conclusions, but may underestimate total uncertainty in some regions and periods. More generally taking action to reduce measurement uncertainty by making more observations will also reduce sampling uncertainty.

With these assumptions we have:

$$\sigma^2_{\mathrm{mean}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{\mathrm{cor},ij}\sigma_{\mathrm{cor},i}\sigma_{\mathrm{cor},j} + \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2_{\mathrm{uncor},i} \quad (9)$$

Estimates of σ_{cor} and σ_{uncor} have been calculated separately for each *in situ* platform type through comparison with output from the UK Met Office NWP model (e.g. Met Office, 2002). $\rho_{\mathrm{cor},ij}$ has been set according to the ship call sign or WMO buoy number present within ICOADS, for observations from the same platform, or unknown platform, $\rho_{\mathrm{cor},ij} = 1$ and zero otherwise.

While we have only applied Equation (9) to *in situ* observations, it is possible to include satellite observations by treating each individual sensor as a different platform and using appropriate values for the uncertainties due to correlated and uncorrelated effects. For the purpose of this study, we focus on *in situ* data to highlight the contrasting

Table 3. Estimates of the uncertainty (68% (1σ)) due to correlated and uncorrelated effects for the sea surface temperature and air temperature observations (values from Kent and Berry, 2008).

Platform	Sea surface temperature		Air temperature	
	σ_{cor} (°C)	σ_{uncor} (°C)	σ_{cor} (°C)	σ_{uncor} (°C)
Ship	0.8	0.7	0.6	0.9
Moored buoy	0.2	0.3	0.5	0.3
Drifting buoy	0.3	0.6	0.9	1.0

development of the *in situ* component of the observing system.

For each individual identifiable ship or platform, the mean difference (platform mean difference) and variance (platform variance) from the model has been calculated (Kent and Berry, 2008). These statistics are then aggregated by platform type and half the variance of the platform mean differences used as an estimate of σ^2_{cor} for that platform type. In doing so we are assigning half of the uncertainty due to correlated errors to the observations and half to the model. Similarly, half of the mean platform variance has been used as an estimate of σ^2_{uncor} for the particular platform type (Table 3). Similarly Kennedy *et al.* (2011) compare *in situ* SST measurements with satellite SST measurements and decompose the observed differences into a systematic component (arising from correlated errors) and a random component (arising from uncorrelated errors). Kennedy *et al.* (2011) find similar values to those in Table 3, but a better estimate of the partition of uncertainty between observations and model would be beneficial.

The estimates of σ_{cor} and σ_{uncor} derived for ships, moored buoys and drifting buoys for SST and MAT (Table 3) have been used throughout the analysis period. Previous work (Berry, 2009) using the semi-variogram method (e.g. Lindau, 2003; Kent and Berry, 2005; Kent and Challenor, 2006) for estimating the uncertainty due to random errors, has shown that the uncertainty does not strongly vary over the period 1970–2007. Further confirmation that the assumption of constant uncertainty estimates is appropriate comes from a comparison of VOSClim (e.g. WMO, 2000), with higher quality instruments, and traditional VOS observations to model output (Ingleby, 2010). A small improvement in combined instrumental uncertainty is shown for VOSClim compared with VOS, with a reduction of order 0.1–0.2 °C for both SST and MAT. Similar results are found with the change from manual to automatic stations (Ingleby, 2010).

These results suggest that the sensor installation and environment have as much influence on the error characteristic as the quality of the sensor itself. Many factors beyond sensor quality can impact observation quality. Examples include bio-fouling (e.g. Delauney *et al.*, 2010); salt contamination (Weller *et al.*, 2008); air flow distortion (e.g. Moat *et al.*, 2005); solar heating (e.g. Anderson and Baumgartner, 1998; Berry *et al.*, 2004); and instrument exposure (Kent *et al.*, 1993 Berry and Kent, 2005). Considering the results of Ingleby (2010) and Berry (2009) and the low number of VOSClim and automatic weather stations on board VOS during the periods assessed, the use of constant values for the uncertainty due to correlated and uncorrelated errors is reasonable for this study. It should be noted that there are plans to upgrade the majority of VOS to VOSClim standard (e.g. JCOMM, 2013) and that ideally uncertainty estimates should be periodically re-evaluated to account for observing system changes.

4. Results

4.1. Sea surface temperature

Figure 1 shows the two components of the instrumental uncertainty, uncertainty due to correlated errors (top, (a) and (d)) and uncertainty due to uncorrelated errors (middle, (b) and (e)), calculated using Equation (9) for July 1987 (left) and July 2007 (right). Also shown is the total instrumental uncertainty (bottom, (c) and (f)). It should be noted that all uncertainties represent a 1σ (approximately 68%) confidence level.

For July 1987, during the period when VOS dominate the observing system, the uncertainty due to correlated errors (Figure 1(a)) is smallest over the major shipping routes where observations from many different platforms are concentrated. In contrast, the uncertainty due to uncorrelated errors (Figure 1(b)) tends to be low in most places, suggesting that errors due to uncorrelated effects have only a minor influence on the total instrumental error. This is confirmed by the similarity between the total instrumental uncertainty (Figure 1(c)) and the correlated component. In nearly all locations the most effective method for reducing the total instrumental uncertainty would be to increase the number of platforms making observations, or equivalently better quality observations with smaller uncertainty due to correlated errors.

The improvement in the observing system for SST by 2007 can be seen, both in terms of reduced uncertainty due to correlated errors (Figure 1(d)) and uncertainty due to uncorrelated errors (Figure 1(e)). The impact of the drifting buoy network is clear, with a marked decrease in the uncertainty due to correlated errors away from shipping lanes. As with July 1987, the uncertainty due to correlated errors is larger than the uncertainty due to uncorrelated errors, suggesting that increasing the number of platforms, rather than increased sampling from the same number of platforms, would have the most beneficial effect on the instrumental uncertainty.

4.2. Air temperature

Figure 2 shows an assessment of the uncertainty in the MAT observing system for July 1987 and July 2007. In 1987, as for SST, the uncertainty due to correlated errors (Figure 2(a)) is smallest over the shipping lanes. This component of uncertainty is smaller in 1987 for MAT than for SST. This is for two reasons. Firstly, the uncertainty due to correlated errors for each air temperature observation is itself smaller (Table 3). Secondly, during this period a greater number of VOS make air temperature observations

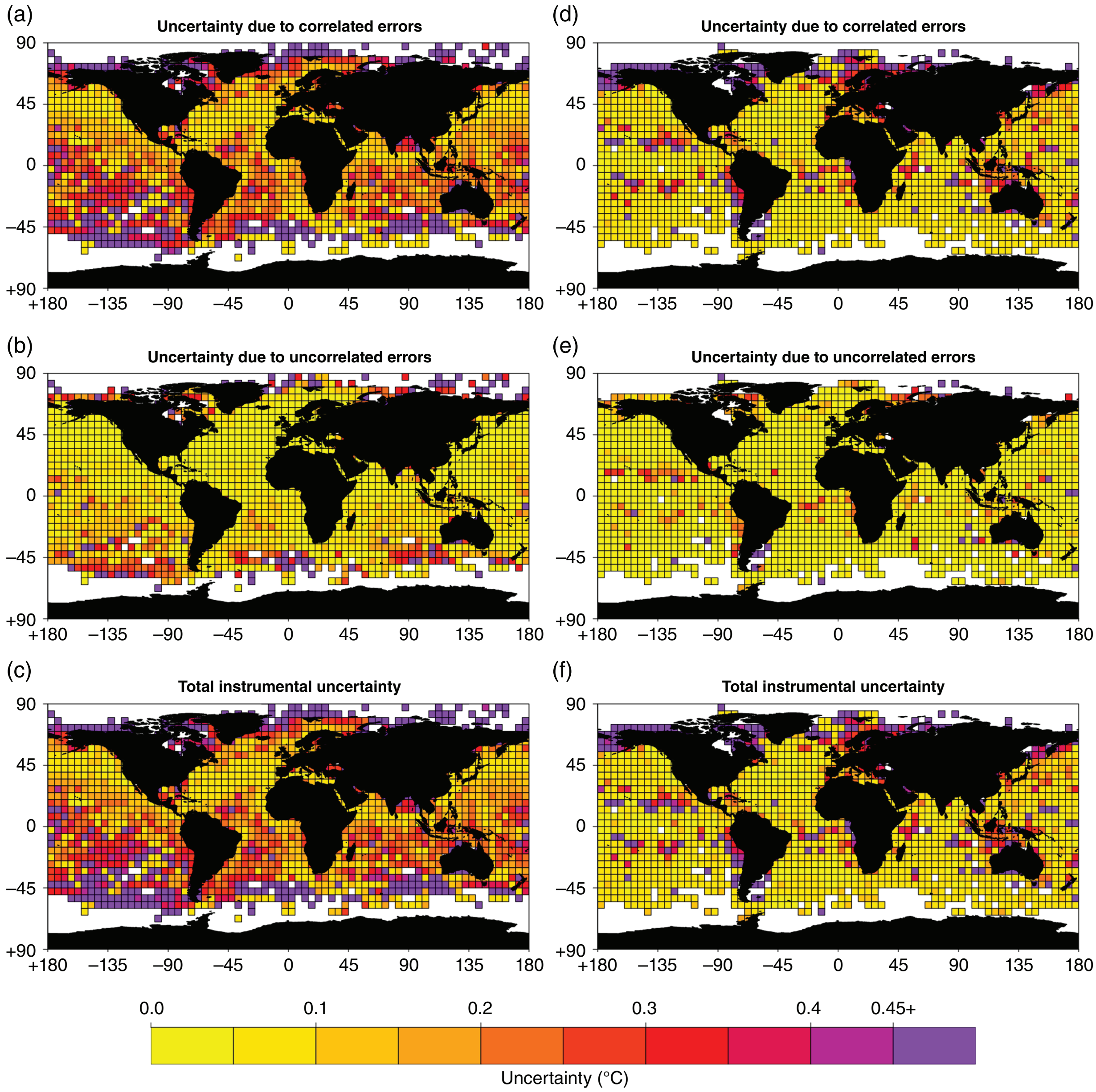

Figure 1. Components, uncertainty due to correlated errors (a and d) and uncorrelated errors (b and e), of the instrumental uncertainty (°C) in monthly mean SST for July 1987 (left) and July 2007 (right). Also shown is the total instrumental uncertainty (c and f). All values are at the 68% (1σ) confidence interval. Unsampled gridboxes are shown in white.

than SST observations. The uncertainty due to uncorrelated errors (Figure 2(b)) is also typically low between 50°S and 60°N. Both the uncertainty due to correlated errors and uncorrelated errors contribute similar amounts to the total instrumental uncertainty (Figure 2(c)), suggesting that having more observations, regardless of source, would be effective in reducing the instrumental uncertainty.

In contrast to SST the instrumental uncertainties in MAT have increased over the majority of the globe between 1987 and 2007. This increase is primarily due to the decline in the number of VOS over the period (e.g. Woodruff *et al.*, 2011), increasing both the uncertainty due to correlated errors (Figure 2(d)) and uncertainty due to uncorrelated errors (Figure 2(e)). The exception to this is in the tropics where the TAO (e.g. McPhaden *et al.*, 1998) and PIRATA arrays (e.g. Servain *et al.*, 1998) have had a positive impact on the observing system. As in 1987, more observations over the majority of the globe would improve the observing system.

4.3. Evolution and completeness of the observing system

To address whether the observing system requirements (e.g. Table 2) are being met we have shaded those 5° boxes where the threshold level for AOPC requirements for SST is met for differing proportions of the month during July 1987 (top) and July 2007 (bottom). This requirement is defined for a daily observing cycle (frequency of observations), so in this case we calculate the proportion of daily

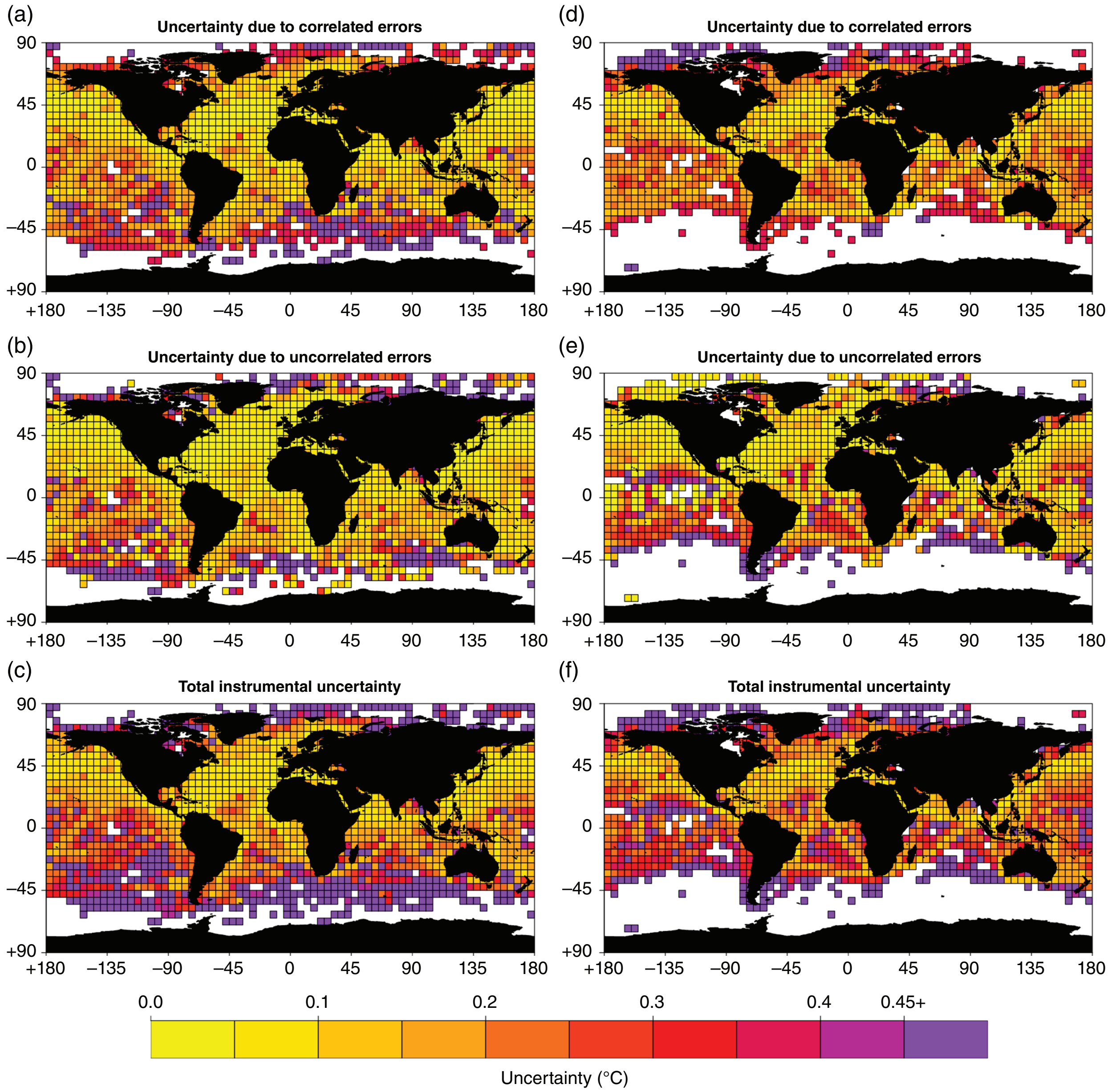

Figure 2. Same as Figure 1, but for marine air temperature.

averages with uncertainty at or below the limit. Regions meeting the requirements between 25 and 50% of the time are shown in red, between 50–75% in orange and ≥75% in green. In this study we present adequacy assessed using the AOPC SST requirement to enable comparison with the MAT assessment below and because the AOPC requirements are more reasonable for *in situ* observations. The requirements defined by the OOPC are more stringent but reasonable when both satellite and *in situ* observations are taken into consideration. Note that we have replaced the 500 km resolution with 5°, which will be similar in the tropics but at higher latitudes we are making the requirement stricter.

Figure 3(a) shows that the AOPC threshold requirement for SST is met at least 50% of the time over the majority of the northern extra-tropical oceans during July 1987. This includes at higher northern latitudes where our effective resolution is <500 km. The requirements are also being met in the shipping lanes of the southern hemisphere and tropics. The improvement in the observing system by 2007, with the threshold requirements met over the majority of the global ocean, can be seen in Figure 3(b). This improvement is due to the expansion of the drifting buoy network during the 1990s and 2000s in response to the needs of the satellite community (e.g. Reynolds *et al.*, 2005). For variables such as SST where the observing system includes satellites, these measurements will further improve the adequacy and completeness.

The AOPC threshold requirement for surface air temperature (0.3 K on a 100 km grid with an observing frequency/cycle of 12 h) is not achievable with the present marine surface observing system, requiring, e.g. three or more moored buoys per 100 km by 100 km area using current estimates of uncertainty. While algorithms for

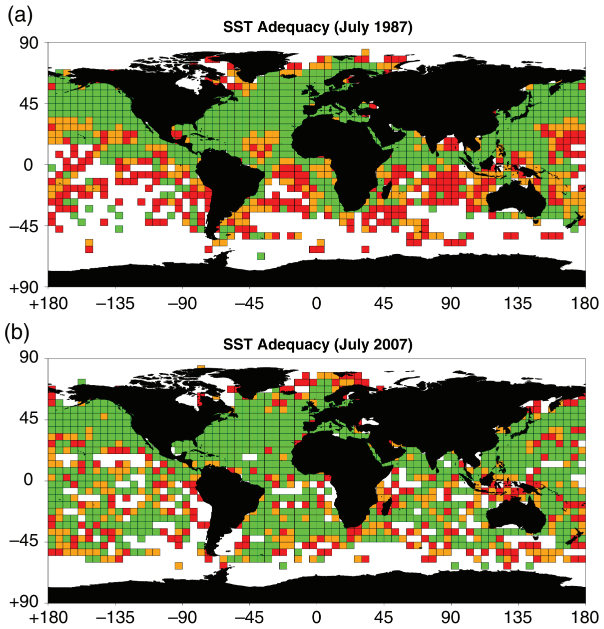

Figure 3. Adequacy of the observing system for SST during July 1987 (top) and July 2007 (bottom). Regions where the AOPC threshold requirements for SST (Table 2) are met for 25–50% (red), 50–75% (orange) and >75% (green) of the observing cycles (days) during the month are shown.

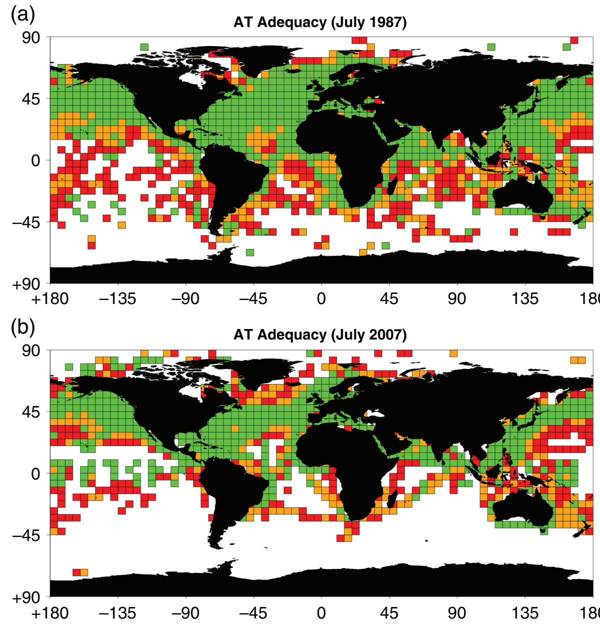

Figure 4. Same as Figure 3 but for air temperature and using the degraded requirements described in the text.

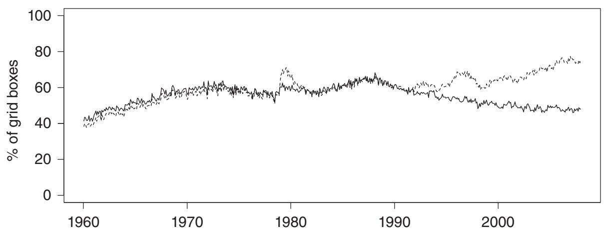

Figure 5. Percentage of ocean 5° grid boxes within 60°S–60°N meeting the threshold requirement for SST (dashed) and air temperature (solid) for at least 50% of the observing cycles in a month. A 12-month running mean filter has been applied.

estimating the air temperature from satellite radiometric observations exist (e.g. Jackson *et al.*, 2006) the algorithms and satellite network are not yet mature enough to routinely produce estimates of the MAT. Figure 4 therefore presents results for MAT assessed using the same requirements as for SST. During July 1987 the results for MAT (Figure 4(a)) are similar to SST (Figure 3(a)). The (reduced) threshold requirements are met over the majority of the northern oceans and over the shipping lanes in the tropics and southern hemisphere. However, by 2007 there has been a large retrenchment in the observing system for air temperature, especially in the Pacific and Indian Oceans. This is because no alternative source of global MAT observations has emerged to counter the decline in numbers of VOS.

Figure 5 shows a time series of the proportion of 5° ocean grid boxes (>25% water) between 60°S and 60°N meeting the AOPC threshold requirements for SST (dashed line) and reduced requirements described earlier for MAT (solid) for at least 50% of observing cycles (see Table 2) during each calendar month. It can be seen that for MAT the observing system was most complete, with respect to the reduced requirements described earlier, in the late 1980s, with almost 70% of the ocean sampled at the threshold (i.e. minimum usable) level during 1988. After this peak, there has been a steady decline in completeness, with only 48% of the ocean between 60°S and 60°N sampled at the threshold level during 2007. This level of completeness is comparable with that seen in the 1960s. The marine air temperature cannot be directly retrieved from satellite observations (e.g. Jin *et al.*, 2015) and the exclusion of satellite data will have no impact on this analysis. As noted earlier, indirect methods for estimating MAT from satellites do exist (e.g. Jackson *et al.*, 2009; Jackson and Wick, 2010; Roberts *et al.*, 2010). However, significant regional biases compared with *in situ* data exist in these estimates (e.g. Jackson and Wick, 2010).

In contrast to MAT, the completeness of the observing system for SST starts slightly lower than for air temperature before gradually rising over the last five decades to a maximum at the end of the period analysed. The cause of the peak around 1980 is unclear but could be due to issues with corrupted ship identifiers in the southern hemisphere leading to an overestimate of the number of different ships reporting and subsequent underestimate of the uncertainty due to correlated errors. As for MAT the completeness peaks at about 67% around 1988. SST then shows a short decline in coverage until the early 1990s when the development of the drifting buoy array increases coverage gradually to about 75% complete by 2007. Satellite SST measurements have not been include in this assessment but their inclusion would increase the proportion of grid cells meeting the requirements. For completeness, the

MAT observations have been compared with the AOPC threshold requirements (not shown) and these are met for <1% of grid boxes in 2007, largely driven by the moored buoy network.

5. Discussion

In this paper, building on previous work (e.g. Kent and Berry, 2008; Berry, 2009; Kennedy *et al.*, 2011), we have described a method for estimating instrumental uncertainty in grid box means (Section 2) to enable an integrated assessment of the adequacy of the marine climate observing system (Section 3). The aim is to develop a method that can be used operationally for monitoring the health of the observing system and identifying where gaps exist, as achieved by the EBD metric (Reynolds *et al.*, 2005).

To this end, Equation (9) can be used, given information on the uncertainty due to correlated errors and uncertainty due to uncorrelated errors for the different components of the observing system, to decompose the instrumental uncertainty into its component parts and to identify whether or not the observing systems meets the requirements for different application areas and ECVs. Maps, such as those shown in Figures 3 and 4, identify where we need to decrease uncertainty to meet a particular requirement. Uncertainty maps and their decomposition, such as Figures 1–2, then show the best way of achieving this reduction. For example, we can identify whether more observations are needed or whether better quality or more diverse platforms are needed in particular regions.

Assessments of observing system adequacy rely on the definition of appropriate requirements by a range of expert users. For the assessments presented in this paper the available user requirements were found to be problematic. For MAT the AOPC threshold requirement, defined as the minimum for the data to be useful, is almost impossible to achieve anywhere with the present marine observing system. The OOPC do not presently provide user requirements for MAT, despite MAT being required for turbulent air–sea flux estimation (Fairall *et al.*, 2010). For SST the threshold uncertainty is defined daily at 500 km resolution by both the OOPC and AOPC but with a much lower uncertainty required by OOPC (Table 2).

There are also no uncertainty requirements in the WMO Observing Requirements Database appropriate for very stable climate monitoring of large area, long-term changes near the ocean surface at monthly or longer observing cycles as required for climate assessments such as those of the IPCC. Similarly, the requirements do not address the problem of aliasing of the diurnal cycle, especially for those components where there is a daily, or longer, observing cycle. For example, any drift in the orbit of a satellite or between subsequent satellites (e.g. Ignatov *et al.*, 2004) or changes in observing practice (Kennedy, 2014) would risk aliasing changes in sampling of the diurnal cycle into estimates of long term trend.

Using Equation (9) we have performed a preliminary assessment of the health of the observing system for SST and MAT using the AOPC requirements for SST in both cases. We have shown that the completeness of the marine air temperature observing system peaked around 1988 at almost 70% completeness outside of the polar regions with a subsequent decline to less than 50% completeness by 2007. In contrast the SST component reached a maximum completeness (75%) in 2007. In performing this assessment we have used only *in situ* observations and the inclusion of satellite data would lead to an increase in the completeness for the SST component of the observing system but have no impact on the MAT component.

This allows us to make the following statements:

(1) In non-polar regions, the *in situ* marine climate observing system for SST was 75% complete at the 'threshold level' during 2007 for AOPC applications.
(2) In non-polar regions, the *in situ* marine climate observing system for SST was 54% complete at the 'threshold level' during 2007 for OOPC applications.
(3) In non-polar regions, the *in situ* marine climate observing system for air temperature was 1% complete at the 'threshold level' during 2007 for AOPC applications.
(4) In non-polar regions, the *in situ* marine climate observing system for air temperature was 48% complete during 2007 for an uncertainty of 1 °C, estimated daily at 5° spatial resolution.

Based on this paper a number of recommendations are made:

(1) The health of the observing system (e.g. GCOS, 2009) should be assessed using an integrated assessment based on the uncertainty characteristics of all observations contributing to an ECV and not on a platform-by-platform basis. This can be done using the method described in this paper.
(2) Evaluations of the completeness of the observing system should compare the uncertainty in the ECVs on a variety of space and timescales against pre-defined targets, such as those in the WMO Rolling Review of Requirements Database.
(3) Further work is required to evaluate the sampling uncertainty and its impact on the uncertainty in the grid box.
(4) Assessments, of the error characteristics for each different platform, for each ECV, should be made routinely and published as performance indicators alongside other metrics in assessments of progress towards implementation of the GCOS.
(5) An assessment of the completeness of user requirements should be made. Examples of additional requirements should include the stability of large-scale averages and requirements for the construction of observation-based estimates of air–sea interaction.

Acknowledgements

The authors acknowledge advice from Peter Challenor and Mark Bourassa and thank the reviewers for their constructive comments. This research was funded by NERC National Capability funding, NERC standard grant NE/J020788/1 and the UK Met Office through the National Centre for Ocean Forecasting.

References

Anderson SP, Baumgartner MF. 1998. Radiative heating errors in naturally ventilated air temperature measurements made from buoys. *J. Atmos. Ocean. Technol.* **15**: 157–173, doi: 10.1175/1520-0426(1998)015<0157:RHEINV>2.0.CO;2.

Balmaseda M, Anderson D. 2009. Impact of initialization strategies and observations on seasonal forecast skill. *Geophys. Res. Lett.* **36**: L01701, doi: 10.1029/2008GL035561.

Berry DI. 2009. *Surface Forcing of the North Atlantic: Accuracy and Variability*. Doctoral thesis, School of Ocean and Earth Science, University of Southampton, Southampton, 176 pp. http://eprints.soton.ac.uk/145001/1/Berry_2009_PhD.pdf (accessed 30 March 2015).

Berry DI, Kent EC. 2005. The effect of instrument exposure on marine air temperatures: an assessment using VOSClim data. *Int. J. Climatol.* **25**: 1007–1022, doi: 10.1002/joc.1178.

Berry DI, Kent EC, Taylor PK. 2004. An analytical model of heating errors in marine air temperatures from ships. *J. Atmos. Ocean. Technol.* **21**: 1198–1215, doi: 10.1175/1520-0426(2004)021<1198: AAMOHE>2.0.CO;2.

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML. 2008. *Guide to the Expression of Uncertainty in Measurement*. JCGM 100:2008, GUM 1995 with minor corrections. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf (accessed 30 March 2015).

Blunden J, Arndt DS. 2014. State of the climate in 2013. *Bull. Am. Meteorol. Soc.* **95**: S1–S238, doi: 10.1175/2014BAMSStateoftheClimate.1.

Bojinski S, Verstraete M, Peterson TC, Richter C, Simmons A, Zemp M. 2014. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **95**: 1431–1443, doi: 10.1175/BAMS-D-13-00047.1.

Cardone VJ, Callahan BT, Chen H, Cox AT, Morrone MA, Swail VR. 2015. Global distribution and risk to shipping of very extreme sea states (VESS). *Int. J. Climatol.* **35**: 69–84, doi: 10.1002/joc.3963.

Carella C, Kent EC, Berry DI. 2015. A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.*, doi: 10.1002/joc.4492.

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**: 1–28, doi: 10.1002/qj.776.

Cressie NAC. 1993. *Statistics for Spatial Data*, revised edition. John Wiley: New York, NY, 900 pp.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, BM M-S, Morcrette J-J, Park B-K, Peubey C, de Rosnay O, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597, doi: 10.1002/qj.828.

Delauney L, Compère C, Lehaitre M. 2010. Biofouling protection for marine environmental sensors. *Ocean Sci.* **6**: 503–511, doi: 10.5194/os-6-503-2010.

Fairall C, Barnier B, Berry DI, Bourassa MA, Bradley F, Clayson CA, de Leeuw G, Drennan WM, Gille ST, Gulev SK, Kent EC, McGillis WR, Quartly GD, Ryabinin V, Smith SR, Weller RA, Yelland MJ, Zhang H-M. 2010. Observations to quantify air-sea fluxes and their role in climate variability and predictability. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 2), Venice, Italy, 21–25 September 2009, Hall J, Harrison DE, Stammer D (eds). ESA Publication WPP-306, doi: 10.5270/OceanObs09.cwp.27.

Folland CK, Parker DE. 1995. Correction of instrumental biases in historical sea surface temperature data. *Q. J. R. Meteorol. Soc.* **121**: 319–367, doi: 10.1002/qj.49712152206.

GCOS. 2009. Progress report on the implementation of the Global Observing System for climate in support of the UNFCCC 2004–2008 (August 2009). GCOS Report No. 129, WMO, Geneva, 104 pp.

GCOS. 2010. Implementation plan for the Global Observing System for climate in support of the UNFCCC (2010 update). GCOS Report No. 138, WMO, Geneva, 180 pp.

GCOS. 2011. Systematic observation requirements for satellite-based products for climate. Supplemental details to the satellite-based component of the 'Implementation Plan for the Global Observing System for Climate in support of the UNFCC': 2011 update. GCOS Report No. 154, WMO, Geneva, 138 pp.

Gulev S, Jung T, Ruprecht R. 2007a. Estimation of the impact of sampling errors in the VOS observations on air-sea fluxes. Part I: Uncertainties in climate means. *J. Clim.* **20**: 279–301, doi: 10.1175/JCLI4010.1.

Gulev S, Jung T, Ruprecht R. 2007b. Estimation of the impact of sampling errors in the VOS observations on air-sea fluxes. Part II: Impact on trends and interannual variability. *J. Clim.* **20**: 302–315, doi: 10.1175/JCLI4008.1.

Hirahara S, Ishii M, Fukuda Y. 2014. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**: 57–75, doi: 10.1175/JCLI-D-12-00837.1.

Huang B, Banzon VF, Freeman E, Lawrimore J, Liu W, Peterson TC, Smith TM, Thorne PW, Woodruff SD, Zhang H. 2015. Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. *J. Clim.* **28**: 911–930, doi: 10.1175/JCLI-D-14-00006.1.

Ignatov A, Laszlo I, Harrod ED, Kidwell KB, Goodrum GP. 2004. Equator crossing times for NOAA, ERS and EOS sun-synchronous satellites. *Int. J. Remote Sens.* **10**: 5255–5266, doi: 10.1080/01431160410001712981.

Ingleby B. 2010. Factors affecting ship and buoy data quality: a data assimilation perspective. *J. Atmos. Ocean. Technol.* **27**: 1476–1489, doi: 10.1175/2010JTECHA1421.1.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, UK, and New York, NY, 1535 pp.

Jackson DL, Wick GA. 2010. Near-surface air temperature retrieval derived from AMSU-A and sea surface temperature observations. *J. Atmos. Ocean. Technol.* **27**: 1769–1776, doi: 10.1175/2010JTECHA1414.1.

Jackson DL, Wick GA, Bates JJ. 2006. Near-surface retrieval of air temperature and specific humidity using multisensory microwave satellite observations. *J. Geophys. Res.* **111**: D10306, doi: 10.1029/2005JD006431.

Jackson DL, Wick GA, Robertson FR. 2009. Improved multisensor approach to satellite-retrieved near-surface specific humidity observations. *J. Geophys. Res. Atmos.* **114**: D16303, doi: 10.1029/2008JD011341.

JCOMM. 2013. Ship observations team seventh session: final report. JCOMM Meeting Report No. 97, WMO, Geneva, 179 pp.

Jin X, Yu L, Jackson DL, Wick GA. 2015. An Improved Near-Surface Specific Humidity and Air Temperature Climatology for the SSM/I Satellite Period. *J. Atmos. Ocean. Technol.* **32**: 412–433, doi: 10.1175/JTECH-D-14-00080.1.

Jones PD, Wigley TML. 2010. Estimation of global temperature trends: what's important and what isn't. *Clim. Change* **100**: 59–69, doi: 10.1007/s10584-010-9836-3.

Kennedy JJ. 2014. A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* **52**(1): 1–32, doi: 10.1002/2013RG000434.

Kennedy JJ, Rayner NA, Smith RO, Parker DE, Saunby M. 2011. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850. Part 1: Measurement and sampling uncertainties. *J. Geophys. Res. Atmos.* **116**(D14): D14103, doi: 10.1029/2010JD015218.

Kent EC, Berry DI. 2005. Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *Int. J. Climatol.* **25**: 843–856, doi: 10.1002/joc.1167.

Kent EC, Berry DI. 2008. Assessment of the marine observing system (ASMOS): final report. NOCS Research and Consultancy Report No. 32, NOC, Southampton, 55 pp.

Kent EC, Challenor PG. 2006. Toward estimating climatic trends in SST. Part II: Random errors. *J. Atmos. Ocean. Technol.* **23**: 476–486, doi: 10.1175/JTECH1844.1.

Kent EC, Taylor PK, Truscott BS, Hopkins JS. 1993. The accuracy of voluntary observing ship's meteorological observations: results of the VSOP-NA. *J. Atmos. Ocean. Technol.* **10**: 591–608, doi: 10.1175/1520-0426(1993)010<0591:TAOVOS>2.0.CO;2.

Kent EC, Berry DI, Woodruff SD, Taylor PK. 2006. Voluntary observing ships: a vital marine observing system in decline. *CLIVAR Exch.* **38**(July): 20–21.

Kent EC, Kennedy JJ, Berry DI, Smith RO. 2010. Effects of instrumentation changes on ocean surface temperature measured *in situ*. *Wiley Interdiscip. Rev. Clim. Change* **1**: 718–728, doi: 10.1002/wcc.55.

Kent EC, Rayner NA, Berry DI, Saunby M, Moat BI, Kennedy JJ, Parker DE. 2013. Global analysis of night marine air temperature and its uncertainty since 1880: the HadNMAT2 data set. *J. Geophys. Res. Atmos.* **118**: 1281–1298, doi: 10.1002/jgrd.50152.

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K. 2015. The JRA-55 reanalysis: general specifications and basic characteristics. *J. Meteorol. Soc. Jpn.* **93**: 5–48, doi: 10.2151/jmsj.2015-001.

Lampitt R, Favali P, Barnes CR, Church MJ, Cronin MF, Hill KL, Kaneda Y, Karl DM, Knap AH, McPhaden MJ, Nittis KA, Priede IG, Rolin J-F, Send U, Teng C-C, Trull TW, Wallace DWR, Weller RA. 2010. In situ sustained Eulerian observatories. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 1), Venice, Italy, 21–25 September 2009, Hall J, Harrison DE, Stammer D (eds). ESA Publication WPP-306, doi: 10.5270/OceanObs09.pp.27.

Legler DM. 1991. Errors of five-day mean surface wind and temperature conditions due to inadequate sampling. *J. Atmos. Ocean. Technol.* **8**: 705–712, doi: 10.1175/1520-0426(1991)008<0705:EOFDMS>2.0.CO;2.

Legler DM, Hill K. 2014. Tropical Pacific observing for the next decade. *Eos Trans. AGU* **95**(23): 196, doi: 10.1002/2014EO230006.

Lindau R. 2003. Errors of Atlantic air: sea fluxes derived from ship observations. *J. Clim.* **16**: 783–788.

McPhaden MJ, Busalacchi AJ, Cheny R, Donguy J-R, Gage KS, Halpern D, Ji M, Julian P, Meyers G, Mitchum GT, Niller PP, Picaut J, Reynolds RW, Smith N, Takeuchi K. 1998. The tropical ocean global atmosphere observing system: a decade of progress. *J. Geophys. Res. Oceans* **103**: 14169–14240, doi: 10.1029/97JC02906.

Meehl GA, Goddard L, Boer G, Burgman R, Branstator G, Cassou C, Corti S, Danabasoglu G, Doblas-Reyes F, Hawkins E, Karspeck A, Kimoto M, Kumar A, Matei D, Mignot J, Msadek R, Navarra A, Pohlmann H, Rienecker M, Rosati T, Schneider E, Smith D, Sutton R, Teng H, van Oldenborgh GJ, Vecchi G, Yeager S. 2014. Decadal climate prediction: an update from the trenches. *Bull. Am. Meteorol. Soc.* **95**: 243–267, doi: 10.1175/BAMS-D-12-00241.1.

Merchant CJ, Embury O, Rayner NA, Berry DI, Corlett GK, Lean K, Veal KL, Kent EC, Llewellyn-Jones DT, Remedios JJ, Saunders R. 2012. A twenty-year independent record of sea surface temperature for climate from Along Track Scanning Radiometers. *J. Geophys. Res.* **117**: C12013, doi: 10.1029/2012JC008400.

Merchant CJ, Embury O, Robert-Joens J, Fiedler E, Bulgin CE, Corlett GK, Good S, McLaren A, Rayner N, Morak-Bozzo S, Donlon C. 2014. Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geosci. Data J.* **1**(2): 179–191, doi: 10.1002/gdj3.20.

Met Office. 2002. *A new unified model, NWP Gazette*, June 2002. http://webarchive.nationalarchives.gov.uk/20060715135118; http://met-office.gov.uk/research/nwp/publications/nwp_gazette/jun02/um.html (accessed 30 March 2015).

Moat BI, Yelland MJ, Pascal RW, Molland AF. 2005. An overview of the airflow distortion at anemometer sites on ships. *Int. J. Climatol.* **25**: 997–1006, doi: 10.1002/joc.1177.

Morrissey ML, Greene JS. 2009. A theoretical framework for the sampling error variance for three-dimensional climate averages of ICOADS monthly ship data. *Theor. Appl. Climatol.* **96**: 235–248, doi: 10.1007/s00704-008-0027-3.

Prytherch J, Kent EC, Fangohr S, Berry DI. 2015. A comparison of SSM/I-derived global marine surface specific humidity datasets. *Int. J. Climatol.* **35**: 2359–2381, doi: 10.1002/joc.4150.

Reid PC, Valdés L. 2011. ICES status report on climate change in the North Atlantic. ICES Cooperative Research Report No. 310, 262 pp.

Reynolds RW, Zhang H, Smith TM, Gentemann CL, Wentz F. 2005. Impacts of in situ and additional satellite data on the accuracy of a sea surface temperature analysis for climate. *Int. J. Climatol.* **25**: 857–864, doi: 10.1002/joc.1168.

Rienecker MM, Suarez MJ, Gelaro R, Todling T, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs L, Kim G-K, Bloom S, Chen J, Collins D, Conaty A, da Silva A, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens R, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J. 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24**: 3624–3648, doi: 10.1175/JCLI-D-11-00015.1.

Roberts JB, Clayson CA, Robertson FR, Jackson DL. 2010. Predicting near-surface atmospheric variables from special sensor microwave/imager using neural networks with a first-guess approach. *J. Geophys. Res. Atmos.* **115**: D19113, doi: 10.1029/2009JD013099.

Roemmich D, Boehme L, Claustre H, Freeland H, Fukasawa M, Goni G, Gould WJ, Gruber N, Hood M, Kent EC, Lumpkin R, Smith S, Testor P. 2010. Integrating the ocean observing system: mobile platforms. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society* (Vol. 1), Venice, Italy, 21–25 September 2009, Hall J, Harrison DE, Stammer D (eds). ESA Publication WPP-306, doi: 10.5270/OceanObs09.pp.33.

Saha S, Moorthi S, Pan H-L, Wu X, Wang J, Nadiga S, Tripp P, Kistler R, Woollen J, Behringer D, Liu H, Stokes D, Grumbine R, Gayno G, Wang J, Hou Y-T, Chuang H-Y, Juang H-MH, Sela J, Iredell M, Treadon R, Kleist D, Delst PV, Keyser D, Derber J, EK M, Meng J, Wei H, Yang R, Lord S, Dool HVD, Kumar A, Wang W, Long C, Chelliah M, Xue Y, Huang B, Schemm J-K, Ebisuzaki W, Lin R, Xie P, Chen M, Zhou S, Higgins W, Zou C-Z, Liu Q, Chen Y, Han Y, Cucurull L, Reynolds RW, Rutledge G, Goldberg M. 2010. The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* **91**: 1015–1057, doi: 10.1175/2010BAMS3001.1.

Servain J, Busalacchi AJ, McPhaden MJ, Moura AD, Reverdin G, Vianna M, Zebiak SE. 1998. A pilot research moored array in the tropical Atlantic (PIRATA). *Bull. Am. Meteorol. Soc.* **79**: 2019–2031, doi: 10.1175/1520-0477(1998)079<2019:APRMAI>2.0.CO;2.

Weller RA, Taylor PK. 1993. *Surface Conditions and Air-Sea Fluxes, CCC0-JSC Ocean Observing System Development Panel*. Texas A&M University: College Station, TX, 131 pp.

Weller RA, Bradley EF, Edson JB, Fairall CW, Brooks I, Yelland MJ, Pascal RW. 2008. Sensors for physical fluxes at the sea surface: energy, heat, water, salt. *Ocean Sci.* **4**: 247–263, doi: 10.5194/os-4-247-2008.

WMO. 2000. Voluntary observing ships (VOS) climate subset project (VOSClim): project document. JCOMM Technical Report No. 5, WMO/TD No. 1010, WMO, Geneva, 18 pp.

Woodruff SD, Worley SJ, Lubker SJ, Ji Z, Freeman JE, Berry DI, Brohan P, Kent EC, Reynolds RW, Smith SR, Wilkinson C. 2011. ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**: 951–967, doi: 10.1002/joc.2103.