Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments

Article

Accepted Version

To link to this article DOI: http://dx.doi.org/10.1016/j.rse.2016.02.022

Publisher: Elsevier

Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments

C. E. Bulgin[a], O. Embury[a], G. Corlett[b], C. J. Merchant[a]

[a] *Department of Meteorology, University of Reading, Reading, UK*
[b] *University of Leicester, Department of Physics & Astronomy, University of Leicester, UK*

Abstract

We establish a methodology for calculating uncertainties in sea surface temperature estimates from coefficient based satellite retrievals. The uncertainty estimates are derived independently of in-situ data. This enables validation of both the retrieved SSTs and their uncertainty estimate using in-situ data records. The total uncertainty budget is comprised of a number of components, arising from uncorrelated (eg. noise), locally systematic (eg. atmospheric), large scale systematic and sampling effects (for gridded products). The importance of distinguishing these components arises in propagating uncertainty across spatio-temporal scales. We apply the method to SST data retrieved from the Advanced Along Track Scanning Radiometer (AATSR) and validate the results for two different SST retrieval algorithms, both at a per pixel level and for gridded data. We find good agreement between our estimated uncertainties and validation data. This approach to calculating uncertainties in SST retrievals has a wider application to data from other instruments and retrieval of other geophysical variables.

Keywords:

Sea Surface Temperature, Uncertainty Budget, Remote Sensing, Climate Change Initiative

1. Introduction

Uncertainty is inherent in all geophysical measurements and must be appropriately characterised for their scientific application. Data providers have a responsibility to communicate the levels of uncertainties associated with their products and inform data users of the correct methodology for using uncertainty information provided. Within the Sea Surface Temperature Climate Change Initiative (SST CCI) project (Hollmann et al., 2013; Merchant et al., 2014) we aim to provide an uncertainty budget for every SST value provided in products (skin temperature, SST at 0.2 m depth and spatially averaged SST). We aim to derive uncertainty estimates independently of SST validation datasets, allowing validation of both the SST values and their associated uncertainty.

The terms 'error' and 'uncertainty' are sometimes used interchangeably, but have distinct standard definitions that will be adhered to throughout this paper. Error is the difference between a measured value and the true value of the measurand (JCGM, 2008; Kennedy, 2014). In practice we know neither the true value nor therefore the error for a particular measurement. However the distribution of the errors can often be estimated and this distribution characterises the uncertainty in the measured value. Formally, uncertainty is a parameter characterising the dispersion of values that could reasonably be attributed to the measured value (JCGM, 2008). To quantify uncertainty in this paper we quote one standard deviation of the error distribution.

It is common to provide generic uncertainty estimates for remotely sensed SST derived via comparison with in-situ datasets during validation activites. The standards of the Group for High Resolution Sea Surface Temperature (GHRSST) specify the provision in all datasets of single sensor error statistics (SSES). For pragmatic reasons, SSES are defined to comprise the mean difference and standard deviation of remotely sensed SST matched to a 'reference' dataset (GHRSST Science Team, 2010). Drifting buoy SSTs are often used as the 'reference'. Mean and standard deviation validation statistics are often provided as globally invariant dataset specific values (May et al., 1997; Reynolds et al., 2002; Casey and Cornillon, 1998). An additional field indicating the retrieval quality level can be specified at pixel resolution providing information on the likelihood of cloud contamination, noise lamplification at extreme satellite zenith angles or input data quality (Donlon et al., 2007; Kilpatrick et al., 2001). An extension of this approach is the MODerate Resolution Infrared Spectrometer (MODIS) algorithm, which provides validation-based uncertainty information stratified by season, latitude, surface temperature, satellite zenith angle, a selected brightness temperature difference, SST quality level and day/night (Castro et al., 2010).

Sources of uncertainty in remotely sensed SST are intrinsic to the retrieval process and the data utilised. Uncertainties vary from pixel to pixel due to local changes in instrument noise, satellite viewing geometry and atmospheric conditions. We present here a method of estimating SST retrieval uncertainty that accounts for these factors at the pixel level. There are a number of sources of uncertainty in SST measurement and the need to differentiate the effects of random, and systematic errors has been previously noted (Reynolds et al., 2002; Casey and Cornillon, 1998; Merchant et al., 2012; Kennedy, 2014). Gridding of products introduces sampling uncertainties and a number of studies have considered these when constructing global or regional SST datasets from in-situ observations (She et al., 2007; Folland et al., 2001; Rayner et al., 2006; Morrissey and Greene, 2009; Jones et al., 1997; Brohan et al., 2006).

In this paper, we consider uncorrelated and locally systematic effects contributing to SST uncertainty. The random or uncorrelated effects arise from noise in the satellite brightness temperature, which propagates into the retrieved SST. Locally systematic effects cause errors that are correlated on synoptic scales of atmospheric variability and are related to the retrieval method itself interacting with changes in atmospheric properties (Minnett, 1991; Barton , 1998; Le Borgne et al., 2011; Minnett and Corlett, 2012; Embury and Merchant, 2012; Merchant et al., 2012). We also discuss uncertainties from large scale systematic effects (spatially coherent on larger scales than synoptic features). In a companion paper (Bulgin et al., 2016) we derive a method for calculating sampling uncertainty in gridded products due to incomplete sampling of observations in each grid cell, primarily as a result of cloud cover. In this paper, we use reuslts from Bulgin et al. (2016), and, for completeness, show how sampling uncertainty combines with other components of uncertainty in gridded products.

The remainder of the paper is structured as follows. Section 2 describes the theory behind the calculation of uncertainties, their propagation and how this is applied to different levels of SST data (orbit data and gridded products). Section 3 describes how an initial uncertainty budget is constructed from errors originating from random, locally correlated and sampling effects. In Section 4 we present a validation of our uncertainty budget and in Section 5 provide a discussion of the results. We conclude the paper in Section 6.

2. Uncertainty Calculation and Propagation

We construct an uncertainty budget for SST measurements in CCI products comprised of uncertainty components arising from random, locally systematic, large-scale systematic and sampling effects. The full equation for the propagation of uncertainty in a variable y, $(u(y))$, given that y is related to input quantities x_i via $y = f(x_1, .., .., x_n)$, is defined as equation (1) in the Guide to the Expression of Uncertainty in Measurement (GUM) (JCGM, 2008).

$$u^2 = \sum_{i}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 u_i^2(x_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right) u(x_i, x_j) \tag{1}$$

Uncertainty is expressed with respect to (y) in the GUM, and we reproduce this notation throughout the paper. However, in Earth Observation, we conventionally relate a retrieval estimate $\hat{x}$ to observations y ie. $\hat{x} = f(y)$ which is the reverse convention. The first term in equation (1) describes the propagation of uncertainties from uncorrelated errors. These can be added in quadrature with the differential term $(\partial f/\partial x_i)$ defining the sensitivity of the total uncertainty to each uncertainty component. The second term describes the propagation of uncertainty terms arising from correlated errors. This term sums the uncertainty components from correlated errors for each pair of input variables (x_i and x_j) found as the product of the sensitivity for both x_i and x_j and the covariance between them, $u(x_i, x_j)$. The factor of '2' is included, as for each pair, each is equally correlated with the other.

Equation (1) can also be written in the form of equation (2) where the uncertainty is expressed as the sum over all pairs of input variables and the covariance term is expressed as the product of the standard uncertainty in x_i, written u_i, in x_j, written u_j, and of the correlation of errors in x_i and x_j, written r_{ij}.

$$u^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u_i u_j r_{ij} \tag{2}$$

Equation (2) applies fairly generically to any transformation $y = f(x_i, ...x_n)$ for which the sensitivity parameters $(\partial f / \partial x_i)$ are adequately constant over the range $x_i - u_i$ to $x_j + u_j$; it is a first order approximation. Because we will use the results later, we illustrate the use of equation (2) for calculating the uncertainty in the mean SST from a number of observations. If $f = \sum_{i=1}^{n} x_i / n$, where each x_i is a contributing SST value, then the sensitivity parameter is $\partial f / \partial x_i = 1/n$ giving:

$$u^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} u_i u_j r_{ij} \tag{3}$$

We can consider three limiting cases. First assume errors are uncorrelated between pixels. We can then put $r_{ij} = \delta_{ij}$, where $\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$. In this case, the uncertainty in the mean is scaled by the familiar '$\frac{1}{\sqrt{n}}$' reduction in uncertainty, because

$$u^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} u_i u_j \delta_{ij} \tag{4}$$

$$= \frac{1}{n^2} \sum_{i}^{n} u_i^2 \tag{5}$$

Second, consider the case $r_{ij} = 1$, which means errors fully correlate between contributing SSTs. Equation (3) becomes

$$u^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{i=j}^{n} u_i u_j \tag{6}$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^{n} u_i \right)^2 \tag{7}$$

implying $u = \frac{1}{n} \sum_{i=1}^{n} u_i$ ie. the uncertainty is the average uncertainty of the contributing SSTs.

Third, consider the case $r_{ij} = \delta_{ij} + (1 - \delta_{ij})r$ - all SSTs have the same error correlation with other SSTs. Substituting into equation (3) gives

$$u^2 = \frac{1}{n^2} \sum_{i}^{n} \sum_{j}^{n} u_i u_j [\delta_{ij} + (1 - \delta_{ij})r] \tag{8}$$

$$= \frac{1}{n^2} \sum_{i}^{n} \sum_{j}^{n} u_i u_j [r + (1 - r)\delta_{ij}] \tag{9}$$

$$= \frac{r}{n^2} \left(\sum_{i=1}^{n} u_i \right)^2 + \frac{(1-r)}{n^2} \left(\sum_{i=1}^{n} u_i^2 \right) \tag{10}$$

This form yields the previous results as special cases ($r = 0$ and $r = 1$). Constant r_{ij} for $i \neq j$ is in practice unlikely to be exact for a real situation, but may be a useful approximation in some cases, avoiding the need to estimate r_{ij} for every contributing pair.

3. Uncertainty Budget Components

3.1. Uncorrelated Effects

Random errors in SST estimation from satellite data arise from noise in the satellite observations. The signal recorded by a typical radiometer is a voltage measured across a detector, digitised and recorded as counts. In the operational calibration, a linear radiance is calculated in the form $radiance = gain \times counts + offset$ where the gain and count parameters are calculated during instrument calibration (Smith et al., 2012). A non-linearity adjustment is then applied to the longwave channels (Smith et al., 2012) for which the associated uncertainty has not been calculated. In this analysis we simply take the detector noise in the measured counts and propagate this into our geophysical retrieval. In a coefficient based retrieval, SST is calculated from a pre-defined linear or nearly linear (Anding and Kauth (1970); Deschamps and Phulpin (1980); Kilpatrick et al. (2001); May et al. (1997); McMillan and Crosby (1984), and further references within Merchant (2013)) combination of the observed brightness temperatures. Brightness temperature uncertainty is characterised using channel-specific noise equivalent differential temperature (NEdT). This uncertainty is then propagated into the SST retrieval uncertainty.

We illustrate the propagation of errors from random effects using data from the polar orbiting Advanced Along Track Scanning Radiometer (AATSR) aboard the Envisat satellite. Envisat was in a sun synchronous orbit with an equator overpass time of 10.00 am. AATSR made observations in seven spectral bands covering the visible and infrared spectrum at two viewing geometries: nadir ($0 - 22°$) and forward ($52 - 55°$). SST can be derived using the nadir infrared channels only, or using both the nadir and forward views. We consider here the propagation of uncertainties through two different retrievals: 'N2' using the 11 and 12 μm channels in the nadir view only and 'D2' using the 11 and 12 μm channels in both views. The formula used here for estimating coefficient based SSTs from satellite data is:

$$\hat{x}_{SST} = a_0 + \sum_k a_k y_k \tag{11}$$

Where y_k refers to each channel used in the retrieval, a_0 is an offset and a_k are channel specific coefficients. Note that here $\hat{x} = f(y)$, in contrast to usage in Section 2 (as previously noted). These coeffcients vary with the context in which the observation is made, according to the viewing geometry and total column water vapour (TCWV), but are predefined. The error (difference between the measured value and true value) for a given SST can be defined as:

$$e_{SST} = \sum_k a_k e_{y_k} \tag{12}$$

This is a linear combination of the errors in the brightness temperatures in each chanel (denoted by 'k') multiplied by the coefficient used in the retrieval. In practice, we do not know the true SST value nor therefore the error on each individual measurement, but we can simulate a 'typical' error field from our knowledge of the NEdT in each channel. We illustrate this in panels 1 and 2 of Figure 1 which show simulated error fields for the nadir view of the 11 and 12 μm channel at pixel resolution (1 km at nadir for AATSR). These are constructed using a Gaussian random number generator selecting values from a distribution with 0.0 °C mean and 0.05 °C standard

Table 1: Coefficients for each channel used to calculate SST in the 'N2' and 'D2' retrievals in Figure 1.

Retrieval	Channel	Sec(Sat Zenith Angle)	a_1 coefficient
N2	11 μm	1.0	2.04314
N2	12 μm	1.0	-1.02542
D2	11 μm	1.0	4.65371
D2	11 μm	1.76	-1.65009
D2	12 μm	1.0	-3.27043
D2	12 μm	1.76	1.27186

deviation representing NEdT estimates for the two channels (Embury and Merchant, 2012). Errors vary in magnitude from pixel to pixel and can be either positive or negative in sign.

Panels 3 and 4 of Figure 1 show the propagation of these simulated error fields in a N2 and D2 retrieval. For the purpose of this illustration we assume a fixed view angle and TCWV (23 kg m^{-2}) across the image giving coefficients (a_k) dependent only on channel, as shown in Table 1. Under normal retrieval conditions these would vary slightly on a per-pixel basis. The coefficients are specified to five decimal places (Merchant and LeBorgne, 2004). Further discussion of error inherent in the retrieval process is provided in Section 3.2. As indicated in equation (12) the uncorrelated errors in a given retrieval are the sum of the errors in each channel, and therefore the total errors are smaller in the N2 retrieval than the D2 retrieval (which uses four channels with generally heavier weights).

Many users require gridded Level 3 products generated from full resolution data. When generating gridded products, the average SST can be calculated using the arithmetic mean:

$$\hat{x}_{GriddedSST} = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_{SST(i)} \tag{13}$$

Where n is the number of observations (i) in the grid cell. The alternative would be to calculate a weighted mean based on the per-pixel uncertainties, but we choose the arithmetic mean as it gives equal weight to all measurements across the grid cell and therefore represents a mean across the geophysical variability within the grid cell. Panels 5 and 6 show the arithmetic mean of the errors over a 5 x 5 pixel grid cell, approximately representing the creation of 0.05° Level 3 products. The range in the mean error is naturally smaller in the gridded product, but remains larger for the D2 retrieval than the N2 retrieval.

In practice, when retrieving SST from satellite observations we don't explicitly know the error in either the brightness temperatues or SST. We need, however, to estimate the uncertainty in the SST retrieval. Given estimates of NEdT, this is an example of standard uncertainty propagation. 'Standard uncertainty' is the standard deviation of errors in each channel brightness temperature, estimated to be of the order of 0.05 K for both the 11 and 12 μm channels of AATSR (Embury and Merchant, 2012). The propagation of uncorrelated uncertainty components is shown in equation (5) where uncertainties are added in quadrature. Applying this to equation (11), in the first instance to give the per pixel uncertainty, and differentiating with respect to each channel (y_k) used in the retrieval gives:

$$u_i = \sqrt{\sum_k a_k^2 u_{y_k}^2} \tag{14}$$

For a gridded product using the arithmetic mean, the uncertainty in the mean of the contributing pixels is

$$u_{GriddedSST} = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum u_i^2}{n}} \tag{15}$$

For fixed coefficients and a constant error in the brightness temperatures (0.05 K) as in Figure 1, there is an invariant uncertainty value for each retrieval algorithm (N2 and D2) at the pixel level. When creating a real SST product, NEdT varies as a function of both channel and brightness temperature. For N2 retrievals in the example provided, this invariant uncertainty value is 0.11 K and for D2 retrievals 0.25 K. Uncertainties in gridded averages reduce by $\frac{1}{\sqrt{n}}$ giving uncertainty estimates of 0.02 K and 0.05 K for N2 and D2 retrievals over fully observed grid cells. In practice, many grid cells in Level 3 products are not fully observed due to cloud cover. This reduces the number (n) of observations available and increases the uncertainties from random effects. This is illustrated in panels 7 and 8 of Figure 1 for N2 and D2 retreivals. A cloud mask has been superimposed on the simulated data at the per-pixel level and uncertainties propagated into the 5x5 pixel product. Observing only part of a given grid cell additionally introduces sampling uncertainty, discussed briefly in Section 3.4 and more fully in the companion paper (Bulgin et al., 2016).

3.2. Locally Systematic Effects

Uncertainties from locally systematic effects arise from ambiguities in or limitations of the SST retrieval algorithm. Coefficient based retrievals for the ATSR instruments in Phase 2 of the SST CCI will use coefficients from the ATSR Reprocessing for Climate (ARC) project. These are calculated based on radiative transfer simulations which cover a comprehensive range of surface and atmospheric conditions (Embury and Merchant, 2012; Embury et al, 2012). Locally systematic effects therefore vary on synoptic scales consistent with changes in atmospheric conditions.

We can characterise the uncertainties arising from locally systematic effects in the retrieval scheme using simulation studies. To do this, we take a 'true' SST field from Numerical Weather Prediction (NWP) data and simulate the associated brightness temperatures globally as would be observed by the AATSR instrument using the RTTOV radiative transfer model. We can then use these simulated brightness temperatures as input into our retrieval scheme, comparing our retrieved SST with the 'true' SST eg. (Merchant et al., 2009). For any given scene, we can plot the retrieval error field using this methodology as shown in Figure 2. The contour lines in the top panels show atmospheric pressure and in the bottom two panels TCWV with the spatial distribution of the error field consistent with synoptic scales of pressure in hPa and total column wate vapour (TCWV) in kgm^2 variability. However, features in the SST error field are not simply linked to TCWV distributions, since we see that a single contour line can run through regions of both positive and negative errors. The ARC retrieval coefficients are banded by TCWV and the observed errors are not simply a bias that can be removed from the retrieval. Uncertainty arising from these error effects is characterised in the retrieval as a function of TCWV consistent with the coefficient banding. Panels in the left and right in Figure 2 show the SST retrieval error fields for different days, which vary in time as well as space on synoptic scales.

Within the retrieval scheme, uncertainties are calculated as the standard deviation of the error distributions from the simulated data, taking the differences between the 'true' and retrieved SSTs. This is the fitting error of the regression when the coefficients are applied to the simulated data used to generate the coefficients. Figure 3 shows the uncertainties as a function of TCWV for retrievals using different channel combinations at different viewing geometries. For the N2 retrieval using two channels (11 and 12 μm) the uncertainties increase as a function of TCWV, flattening at higher TCWV's above 45 kg m^{-2}. With the addition of information from multiple viewing angles (0-22° and 52-55°) locally systematic uncertainties are significantly reduced to $\sim$ 0.1 K or lower.

Figure 3 also shows the uncertainty from uncorrelated effects as a function of TCWV for different channel combinations. Comparing single-view retrieval uncertainties with dual-view uncertainties, the dual-view capability reduces the systematic uncertainty at the expense of the increased retrieval noise. Uncertainties from uncorrelated effects are dependent on both the NEdT for a given channel combination and the coefficients. For the N2 and D2 retrievals large weights are assigned to the 11 and 12 μm channels which magnifies the uncorrelated uncertainty. ARC coefficients are tuned to assume NEdTs of 0.01 K (smaller than actual values) as they are designed to produce

SST products at 0.1° resolution. This has the effect of reducing locally systematic uncertainties at the cost of increased uncorrelated uncertainties as these decrease as a function of $1/\sqrt{n}$ when calculating the gridded product.

Many SST retrievals also use information from the 3.7 μm channel at night. The consequence of adding this third channel to the retrieval (results not shown) reduces uncertainty from locally systematic effects to ~ 0.1 K or lower, with larger uncertainties for drier atmospheres. As TCWV increases, the 11 and 12 μm channels become less sensitive to the surface whilst the 3.7 μm channel remains relatively transparent. SSTs in regions of high TCWV, close to the equator also show less variability which may improve the fit of the retrieval to the training data. For the uncertainties due to uncorrelated effects, including the 3.7 μm channel in the retrieval results in smaller weights for the 11 and 12 μm channels reducing the noise amplification.

3.3. Large Scale Systematic Effects

Other effects can cause SST errors that are correlated on larger scales. For brevity, the uncertainty associated with unknown errors correlated on large scales is hereafter referred to as "systematic uncertainty". (It is taken for granted that any 'known' or 'estimated' systematic errors have been addressed i.e., that any general bias has been quantified and subtracted from data. The systematic uncertainty therefore quantifies the degree of doubt in the measurements associated with what might be termed 'residual biases'.)

All satellite sensors are calibrated prior to launch to a pre-defined standard. The required accuracy for SST measurements from space for climate applications is 0.1 K (Ohring et al, 2005). In some cases the SST algorithm itself is capable of adjusting for some of the systematic errors in calibration, for example an SST retrieval algorithm that fits regression coefficients to buoys directly will correct for some of the calibration biases as part of the fitting process. This process will also introduce an additional source of uncertainty from unknown errors in the buoy measurement. The buoy data are point measurements at depth whereas the satellite observations are area measurements of skin temperature. If the sensor is poorly characterised this additional uncertainty term can be smaller than the systematic calibration bias. Thermal channels on some sensors seem in practice to have a BT calibration accuracy of 0.1 K, judging by the SST accuracy achievable using radiative transfer-based coefficients.

The sensor having been calibrated to a certain level, there remain smaller errors, within the specified calibration accuracy, that are unknown. These may vary systematically with scene temperature, general instrument temperature, the thermal state of the on-board calibration target, the temperature of the detectors, the illumination of the sensor on the space-craft by the Sun, and potentially with many other factors. Sometimes, these effects are sufficiently evident in flight that they can be diagnosed and corrected for (Cao et al., 2005; Yu et al., 2012; Wang and Cao, 2008; Mittaz and Harris, 2011; Mittaz et al., 2013). There may be a gradual evolution of such systematic calibration effects over time, as the sensor ages, and/or as the platform orbit drifts, changing the illumination and thermal cycling of the sensor.

Where satellite datasets are reprocessed, there may be some effort to harmonise the BTs across different sensors. "To harmonise" here means to reconcile the calibration of the observed BTs given the known differences between the sensors; it does not mean that the BTs would be the same for two sensors viewing the same scene; it does mean that the differences would be traceable to known instrumental differences, such as different spectral response functions. The adjustments made to BTs in the light of harmonisation have their own associated uncertainty, and this also is likely to be systematic as defined here. Overall, harmonisation is intended to reduce systematic effects, particularly relative errors between sensors.

It is possible in principle to estimate the systematic uncertainty associated with calibration. There are two possible approaches. The first is to exploit the pre-flight calibration information where an analysis of the potential calibration errors has been made. Where such information is available in sufficient detail in the public domain, it can form the basis of an uncertainty budget. The second is to exploit near-coincident observations in space between different sensors. Having accounted for instrumental characteristics, differences in matched observations can be used to adjust a less-well-calibrated sensor to a better-calibrated sensor. These adjustments have a quantifiable statistical uncertainty, which then provides an estimate of the magnitude of the post-correction systematic uncertainty eg. (Goldberg, 2007).

In general, however, calibration uncertainty is not well quantified and propagation of such information into the systematic uncertainty in SST has not been undertaken, to our knowledge. Arguably, for SSTs generated operationally for use in numerical weather prediction and real-time oceanography, it has not been necessary. However, in the context of developing reprocessed SST datasets for climate applications, it is an area that needs to be developed. Climate data records require justified uncertainty estimates, particularly estimates of their multi-decadal stability, which implies a detailed engagement with understanding and propagating uncertainty from systematic effects throughout the record (Minnett and Corlett, 2012). A metrology (science of measurement) of Earth Observation needs to be developed, to bring relevant metrological principles for developing traceable chains of uncertainty to bear in the context of historic satellite missions.

3.4. Sampling Uncertainties

Many users of SST data require gridded products with SST specified as a mean value across the space and time represented by the grid cell. Often grid cells are not fully observed, typically in infrared measurements due to cloud cover, but also in the case of corrupted data or problems with the retrieval process. Data may also be removed from the subsample by conservative cloud detection schemes which can mask clear-sky pixels. The mean SST across the observed pixels may differ from the mean SST across all pixels in the grid cell introducing sampling uncertainty.

We cannot explicitly calculate the difference between the SST across the full grid cell and the SST in the available subsample within the retrieval as we do not know the SST of the unsampled pixels. We can however model the sampling uncertainty associated with this process using fully clear-sky data extracts, and we do this as a function of the percentage of the total number of pixels available in the subsample and the standard deviation of the SST in the available pixels.

The full details of the derivation of the sampling uncertainty model are provided in the companion paper (Bulgin et al., 2016). Here we provide only a brief overview, for completeness of the discussions in this paper. In Bulgin et al. (2016) we parameterise sampling uncertainty using a cubic function in the form $(ap^3 + bp^2 + cp + d)$ where a, b, c and d are coefficients dependent on the SST standard deviation in the subsample, and p is the percentage of clear-sky pixels within a given grid cell. This model is therefore applicable to any retrieval scheme with data at the same spatial scale provided that the noise contribution to the SST standard deviation has been subtracted.

3.5. Other effects contributing to uncertainty

The propagation of the effects of radiometric noise and the analysis of locally systematic uncertainty discussed has assumed the context of normal clear-sky conditions for each SST retrieval. This neglects the fraction of retrievals that will in practice be made under unusual conditions. These are principally retrievals made for pixels whose classification as clear-sky-over-seawater is doubtful, but which have nonetheless passed the cloud screening process. At present, we have no method for estimating this in the uncertainty budget.

The first case to consider is 'residual' unscreened cloud contamination. Clouds escape detection if they are sufficiently small and low (warm) or sufficiently optically thin (e.g., some cirrus). In these cases they can nonetheless affect BTs at the level of several tenths of kelvin. The corresponding impact on SST depends on how different the cloud impacts on BTs are from the impact of increased water vapour in the atmosphere (which the retrieval algorithms are adapted to deal with). The probability of such cases is considered to be greater around the edges of areas correctly identified as cloudy. Note that both the distribution of BT modification by cloud-contamination in pixels falsely considered to be clear sky, as well as the frequency of failure to detect are dependent on the cloud screening system. One could envisage that simulation of a representative range of cloudy situations be carried out to generate such information, but to our knowledge, this has not been done. Given these pieces of information, assessment of the contribution to SST uncertainty could be undertaken by error propagation methods similar to those described earlier. At present, however, the contribution of this effect to SST uncertainty is not estimated.

The second case to consider is atmospheric aerosol of a form and optical depth outside the range of circumstances for which the retrieval coefficients are designed. Again, to the degree that the aerosol affects the BTs differently to water vapour [e.g., Merchant et al. (2006)], SST errors may be induced of unknown size. While aerosol events trigger cloud detection if the optical depths are sufficiently great, there is a regime where SST retrievals can be affected, the effect in most cases being to make the retrieved SST too cold. Again, the contribution of this effect to SST uncertainty is not estimated.

The third case relates to sea ice being present within the pixel for which SST is retrieved. If the ice is not too cold and is relatively dark (circumstances that often go together in the formation of new ice), the ice may not be detected. Similar considerations apply as to missed residual cloud or aerosol, and this contribution to uncertainty again is not presently estimated.

There are a number of further effects contributing to SST uncertainty that are neglected in the SST CCI uncertainty model. These include differences in the instantaneous field of view for channels of different wavelength, and local to regional variations in trace gas concentrations.

4. Validation of the Uncertainty Budget

Having constructed an initial uncertainty budget for remotely sensed SSTs independently of in-situ data, we can now use these in-situ data to validate our uncertainties (as well as the retrieved SST). In Section 3, we characterised two quantifiable components of uncertainty relating to SSTs calculated from satellite data at a pixel level (a random component due to noise in the data and a locally systematic component arising from uncertainties varying on a synoptic scale within the retrieval) from which we construct our initial uncertainty budget. We validate this budget using data from the AATSR instrument spanning four years (2006 - 2009 inclusive) considering both the N2 and D2 retrievals. The data used in the validation are taken from the SST CCI multi-sensor match-up system (MMS) (Corlett et al., 2014) where drifting buoy and satellite observations are matched globally under clear-sky conditions (Corlett et al., 2014).

Matches are filtered to include only the closest in-situ match in time to the satellite observation and to check the quality of the in-situ data. Matches can have a maximum time difference of 4 hours and maximum spatial separation of 10 km. Bad quality in-situ data are removed based on the following criteria 1) absolute difference between NWP and in-situ SST greater than 10 K, 2) standard deviation of the in-situ SST history greater than 5 K and 3) standard deviation of the in-situ latitude history greater than 10 degrees. Validation of satellite data using in-situ data necessitates a comparison between a point measurement and the satellite footprint. There are uncertainties in this process arising from comparing two different types of observation and geolocation errors in both the satellite and in-situ data. The filtering is therefore necessary to minimise both spatial and temporal separation of the satellite and in-situ observations (Minnett, 1986; Donlon et al., 2002; Corlett et al., 2006).

For each match up, the uncertainties in the retrieved SST are calculated as follows. The noise in a given observation is a function of both the channels and associated brightness temperature, and is calculated by monitoring in-orbit blackbody temperature signals (Smith et al., 2012). For AATSR, the NEdT is fairly consitent throughout the lifetime of the mission. These NEdT values are are used to calculate the uncertainty due to uncorrelated effects at L2 using the methodology presented in Section 3. The uncertainty from locally systematic effects is quantified as a function of the TCWV consistent with the banding of the retrieval coefficients. In both cases the uncertainties are then propagated into the gridded product for validation of data in L3 format. For the gridded products, a sampling uncertainty is also calculated due to the presence of cloud preventing observation of all pixels within a given grid cell (Bulgin et al., 2016). This is an additonal uncertainty due to uncorrelated effects that is introduced in the gridding process. At both the per pixel and gridded scales the uncertainty components are added in quadrature to give a total uncertainty.

The validation data for the N2 and D2 pixel level retrievals are shown in the top two panels of Figure 4. Here we plot the standard deviation of the SST difference (retrieval minus drifting buoy) against the SST retrieval uncertainty which we have calculated independently represented by the thin black lines in Figure 4. The dashed lines indicate the uncertainty model we would expect to see based on retrieved SST minus drifting buoy differences.

There is a lower limit on this model of +/− 0.15 K which represents the uncertainty in the drifiting buoy measurements. We chose the time period of 2006-2009 inclusive for our validation as the drifting buoy uncertainty has been stable at around 0.15 K over this period (Lean and Saunders , 2012). The blue line on the plots indicate the median difference between the retrieved and in-situ SST across all match-ups in each uncertainty bin (width 0.02 K). The standard error in this value is represented by the error bars. Red lines at the end of the black bars indicate the statistical uncertainty in the calculated standard deviation and are visible primarily for bins where the number of contributing cases is small.

For the N2 pixel level data we find that our uncertainty estimates closely match the expected uncertainty model below a total uncertainty of 0.25 K. Above this threshold, our estimated retrieval uncertainties are too high: a better fit would be obtained if the bins shifted to lower estimated uncertainty values. For the D2 retrieval, we see that our uncertainties calculated within the retrieval process show excellent agreement with the expected uncertainty model. At a per-pixel level the dominant terms in the uncertainty budget come from the uncorrelated and locally systematic effects, assuming that a good cloud detection algorithm is used. Therefore the validation indicates that our estimate of these components is well constrained.

We also consider the validation of uncertainties for gridded N2 and D2 retrievals across a 5x5 pixel domain approximately corresponding to 0.05°. In this case we also include the sampling uncertainty component in our initial uncertainty budget (Bulgin et al., 2016). The results for this validation are shown in the bottom two panels of Figure 4. When considering gridded data we find a larger range of estimated uncertainty than for the per pixel data. This is because SST varies across the gridded domain, and for cells that are not well sampled, the uncertainty on the mean SST increases. For the N2 gridded data we see a similar pattern to the N2 per pixel data with uncertainties being slightly overestimated. For the D2 gridded retrieval the overall uncertainties are smaller, but we underestimate the total uncertainty.

5. Discussion

Overall, we see that our independent uncertainty estimates show good agreement with validation data using in-situ drifting buoy measurements. The best agreement is for the D2 retrieval at a per-pixel level. For the N2 retrievals we see a similar over-estimation of uncertainties above 0.2-0.25 K in both the pixel level and gridded products. The uncertainty budget constructed is based on the errors that we currently have the capability to estimate and propagate through the retrieval. Some of the sources of error discussed in the earlier sections such as residual unscreened cloud contamination, failure to detect clear-sky pixels and aerosol are not yet included. These may be larger across a gridded domain if they affect multiple pixels.

In this validation, the estimation of large scale systematic uncertainties has also been excluded, but in the SST CCI Version 1 products this is set to an invariant value of 0.1 K per pixel as a best estimate of the magnitude of this component, and then added in quadrature to the uncertainty budget (Merchant et al., 2014).

Although at present the uncertainty budget can not be fully constrained due to the limitations described in the Section 3, we are able to characterise well the components resulting from random, locally systematic and sampling effects across a range of retrievals for the ATSR instruments as evidenced by the good validation statistics. On the relatively short spatial and temporal scales (pixel to gridded averages at 0.1° and instantaneous measurements) the uncertainties from uncorrelated and locall systematic effects are the dominant terms in the uncertainty budget. The contributions from the 'missing' components are therefore relatively small under these SST retrieval conditions. Empirical systematic effects (biases) are within the estimated uncertainties and these uncertainties can sucessfully distinguish more and less certain SSTs. The approach outlined in this paper has a wider application to coefficient based SST retrievals using other algorithms and data from other instruments. If the data provider or user knows the NEdT distribution for each channel used in the retrieval they can propagate this through the algorithm to obtain the uncertainty due to uncorrelated effects in the retrieved SST. Data providers can use simulation studies to characterise the locally systematic uncertainty in their retrieval scheme, and the sampling model is applicable to any SST retrieval on the same spatial scales as discussed in this paper provided that the uncertainty due to noise is removed first. Provision of uncertainty information as part of the retrieval process then enables validation of these uncertainty estimates, as well as the SST, using in-situ data.

Figure 5 maps mean uncertainty estimates for 2010. The uncertainty maps show the square root of the mean of the error variance across all days with observations. Where more than one observation is available for a given day, the smallest error variance has been used. The uncertainty from uncorrelated effects (a) contains the noise and sampling uncertainty components and when added to the uncertainty due to lcally systematic effects (b) in quadrature, produces the total uncertainty map (c). Total uncertainties typically range between 0.1-0.25 K globally, with the highest values predominantly in equatorial regions and some northern hemisphere high latitudes. The uncertainty due to uncorrelated effects is the larger contributor to this signal, and in these regions scattered or patchy cloud cover increases sampling uncertainties. Figure 5 (d) also shows the ratio of the retrieved SST variability to the uncertainty, calculated by dividing the standard deviation of the SST in an given location over the whole of 2010 by the total uncertainty. The highest ratios are seen in mid-latitude regions where SSTs show greater seasonal variation.

6. Conclusions

In this paper we present a framework for the provision of uncertainty estimates in coefficient based SST retrieval from satellite data, based on propagation of noise, simulation of noise-free retrieval errors, and empirical characterisation of sampling effects. The uncertainty estimates can be validated in their own right, in addition to validating the retrieved SST. We provide a detailed discussion of different sources of uncertainty in the SST retrieval and how to propagate these through the retrieval process. We derive three uncertainty components here and in the companion paper; uncertainties due to uncorrelated, locally systematic and sampling effects. We apply our derivation to AATSR data within the context of the SST CCI project and find that our uncertainties validate well against in-situ data for both per pixel and gridded products, and for two different retrieval algorithms.

7. Acknowledgements

The work undertaken in this paper was funded by the European Space Agency Sea Surface Temperature Climate Change Initiative project.

References

Anding, D. & Kauth, R. (1970). Estimation of sea surface temperature from space. *Remote Sensing of Environment*, *1*, 217-220.

Barton, I. J. (1998). Improved techniques for the derivation of sea surface temperatures from ATSR data. *Journal of Geophysical Research*, *103*, 8139-8152.

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

Bulgin, C. E., & Embury, O., & Merchant, C. J. (2016). Sampling Uncertainty in Global Area Coverage (GAC) and Gridded Sea Surface Temperature Products. *Remote Sensing of Environment*.

Coa, C., & Sullivan, J. & Maturi, E. & Sapper, J. (2004). The effect of orbit drift on the calibration of the 3.7 μm channel of the AVHRR onboard NOAA-14 and its impact on night-time sea surface temperature retrievals. *International Journal of Remote Sensing*, *25*, 975-986.

Casey, K. S., & Cornillon, P. (1998). A Comparison of Satellite and In Situ-Based Sea Surface Temperature Climatologies. *Journal of Climate*, *12*, 1848-1863.

Castro, S., & Cornillon, P., & Gentemann, C., & Jessup, A., & Hacker, P., & Kaplan, A., & Lindstrom, E.. & Maturi, E., & Minnett, P., & Reynolds, R. (2010). Sea Surface Temperature Error Budget: White Paper. Interim Sea Surface Temperature Science Team (ISSTST), pages 44.

Corlett, G. K., & Barton, I. J., & Donlon, C. J., & Edwards, M. C., & Good, S. A., & Horrocks, L. A., & Llewellyn-Jones, D. T., & Merchant, C. J., & Minnett, P. J., & Nightingale, T. J., & Noyes, E. J., & O'Carroll, A. G., & Remedios, J. J., & Robinson, I. S., & Saunders, R. W., & Watts, J. G. (2006). The accuracy of SST retrievals from AATSR: An initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets.. *Advances in Space Research*, *37*, 764-769.

Corlett, G., & Atkinson, C., & Rayner, N., & Good, S., & Fiedler, E., & McLaren, A., & Hoeyer, J., & Bulgin, C. E. (2014). Product Validation and Intercomparison Report (PVIR) SST_CCI-PVIR-UoL-001, pp. 14.

Corlett, G. K., & Merchant, C. J. & Minnett, P. J. & Donlon, C. J. (2014). Assessment of Long-Term Satellite Derived Sea Surface Temperature Records. In: Zibordi, G., Donlon, C. J. and Parr, A. C. (eds.) Optical radiometry for ocean climate measurements. Experimental methods in the physical sciences 47(47).. Academic Press, pp. 639-674. ISBN 9780124170117 doi: 10.1016/B978-0-12-417011-7.00015-5

Deschamps, P. Y., & Phulpin, T. (1980). Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 um. *Boundary Layer Meteorology*, *18*, 131-143.

Donlon, C. J., & Minnett, P. J., & Gentemann, C., & Nightingale, T. J., & Barton, I. J., & Ward, B., & Murray, M. J. (2002). Toward Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research. *Journal of Climate*, *15*, 353-369.

Donlon, C., & Robinson, I., & Casey, K. S., & Vazquez-Cuevo, J., & Armstrong, E., & Arino, O., & Gentemann, C., & May, D., & LeBorgne, P., & Piollè, J., & Barton, I., & Beggs, H., & Poulter, D. J. S., & Merchant, C. J., & Bingham, A., & Heinz, S., & Harris, A., & Wick, G., & Emery, B, & Minnett, P., & Evans, R., & Llewellyn-Jones, D., & Mutlow, C., & Reynolds, R. W., & Kawamura, H., & Rayner, N. (2007). The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project. *Bulletin of the American Meteorological Society*, 1197-1213.

Embury, O., & Merchant, C. J., & Corlett, G. K (2012). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: initial validation, accounting for skin and diurnal variability. *Remote Sensing of Environment*, *116*, 62-78.

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

Embury, O., & Merchant, C. J. & Fillipiak, M. J. (2012). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: basis in radiative transfer. *Remote Sensing of Environment*, *116*, 32-46.

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

GHRSST Science Team (2010). The Recommended GHRSST Data Specification (GDS) Revision 2.0 Technical Specifications. Available from the GHRSST International Project Office, http://www.ghrsst.org, pp. 120

Goldberg, M (2007) Global space-based inter-calibration system (GSICS) *Proc. SPIE 6684, Atmospheric and Environmental Remote Sensing Data Processing and Utilization III: Readiness for GEOSS, 668402, http:dx.doi.org10.111712.735246*.

Hollman, R., & Merchant, C. J., & Saunders, R., & Downy, C., & Buchwitz, M., & Cazenave, A., & Chuvieco, E., & Defourny, P., & de Leeuw, G., & Forsberg, R., & Holzer-Popp, T. & Paul, F. & Sandven, S. & Sathyendranath, S. & van Roozendael, M. & Wagner, W. (2013). The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables. *Bulletin of the American Meteorological Society*, *94*, 1541-1552.

Joint Committee for Guides in Metrology (2008). Evaluation of measurement data - Guide to the expression of uncertainty in measurement. Bureau International des Poids et Mesures, available from www.bipm.org/en/publications/guides/, pp. 130.

Jones, P. D., & Osborn, T. J., & Briffa, K. R. (1997). Estimating Sampling Errors in Large-Scale Temperature Averages. *Journal of Climate*, *10*, 2548–2568.

Kennedy, J. J. (2014). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, *52*, 1-32.

Kilpatrick, K. A., & Podestá, G. P., & Evans, R. (2001). Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, *106*, 9179-9197.

Lean, K., & Saunders, R. W. (2012). Validation of the ATSR Reprocessing for Climate (ARC) Dataset Using Data from Drifting Buoys and a Three-Way Error Analysis. *Journal of Climate*,*26*, *1*, 4758-4772.

Le Borgne, P., & Roquet, H., & Merchant, C. J. (2011). Estimation of sea surface temperature from the Spinning Enhanced Visible and Infra Red Imager, improved using numerical weather prediction. *Remote Sensing of Environment*,*115*, *1*, 55-65.

May, D. A., & Parmeter, M. M., & Olszewski, D. S., & McKenzie, B. D. (1997). Operational Processing of Satellite Sea Surface Temperature Retrievals at the Naval Oceanographic Office. *Bulletin of the American Meteorological Society*, 397-407.

Merchant, C. J. (2013). Thermal remote sensing of sea surface temperature In: Kuenzer, C. and Dech, S. (eds.) Thermal Infrared Remote Sensing: Sensors, Methods, Applications. Remote Sensing and Digital Image Processing, 17. Springer Netherlands, Dordrect, pp.287-313. ISBN 9789400766389.

Merchant, C. J., & Embury, O., & Rayner, N. A., & Berry, D. I., & Corlett, G. K., & Lean, K., & Veal, K. L., & Kent, E. C., & Llewellyn-Jones, D. T., & Remedios, J. J., & Saunders, R. (2012). A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. *Journal of Geophysical Reserach*, *117*, C12013, pp.18.

Merchant, C. J., & Embury, O., & Le Borgne, P., & Bellec, B. (2006). Saharan dust in night-time thermal imagery: detection and reduction of related biases in retrieved sea surface temperature. *Remote Sensing of Environment*, *104*, 15-30.

Merchant, C. J., & LeBorgne, P. (2004). Retrieval of sea surface temperature from space based on modeling of infrared radiative transfer: capabiities and limitations. *Journal of Atmospheric and Oceanic Technology*, *21*, *11*, 1734-1746.

Merchant, C. J., & Harris, A. R., & Roquet, H., & Le Borgne, P. (2009). Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer. *Geophysical Research Letters*, *36*, *17*, 1-5 .

Merchant, C. J., & Embury O. & Roberts-Jones, J. & Fiedler, E. & Bulgin,

C. E. & Corlett, G. K. & Good, S. & McLaren, A. & Rayner, N. & Donlon, C. (2014). Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geoscience Data Journal*, *1*, *2*, 179-191.

McMillan, L. M., & Crosby, D. S. (1984). Theory and Validation of the Multiple Window Sea Surface Temperature Technique. *Journal of Geophysical Researchl*, *89*, *C3*, 3655-3661.

Minnett, P. J. (1991). Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *Journal of Geophysical Researchl*, *96*, *C10*, 18475-18489.

Minnett, P. J. (1986). A numerical study of the effects of anomalous North Atlantic atmospheric conditions on the infrared measurement of sea surface temperature from space. *Journal of Geophysical Researchl*, *91*, *C7*, 2156-2202.

Minnett, P. J., & Corlett, G. K. (2012). A pathway to generating Climate Data Records of sea-surface temperature from satellite measurements. *Deep Sea Research Part II: Tropical Studies in Oceanography*, *77-80*, 44-51.

Mittaz, J., & Harris, A. (2011). A Physical Method for the Calibration of the AVHRR/3 Thermal IR Channels. Part II: An In-Orbit Comparison of the AVHRR Longwave Thermal IR Channels on board MetOp-A with IASI. *Journal of Atmospheric Oceanic Technology*, *28*, 1072-1087.

Mittaz, J., & Bali, M. & Harris, A. (2013). The Calibration of Broad Band IR Sensors: Time variable biases and other issues. *EUMETSAT Meterological Satellite Conference, Vienna*, *Available online at http:www.eumetsat.intwebsitewcmidcidcplg?IdcService=GET_FILE &dDocName=PDF_CONF_P_S8_11_MITTAZ_V&RevisionSelectionMethod=L, pp. 8*.

Morrissey, M. L., & Greene, J. S. (2009). A theoretical framework for the sampling error variance for three-dimensional climate averages of ICOADS monthly ship data. *Theoretical Applied Climatology*, *96*, 235-248.

Ohring, G., & Wielicki, B., & Spencer, R., & Emery, B., & Datla, R. (2005). Satellite Instrument Calibration for Measuring Global Climate Change. *Bulletin of the American Meteroological Society*, 1303-1313.

Rayner, N. A., & Brohan, P., & Parker, D. E., & Folland, C. K., & Kennedy, J. J., & Vanicek, M., & Ansell, T. J., & Tett, S. F. B. (2006). Improved Analyses of Changes and Uncertainties in Sea Surface Temperature Measured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset. *Journal of Climate*, *19*, 446–469.

Reynolds, R. W., & Rayner, N. A., & Smith, T. M., & Stokes, D. C., & Wang, W (2002). An Improved In Situ and Satellite SST Analysis for Climate. *Journal of Climate*, *15*, 1609-1625.

She, J., & Hoyer, J. L.,& Larsen, J. (2007). Assessment of sea surface temperature observational networks in the Baltic Sea and North Sea. *Journal of Marine Systems*, *65*, 314-335.

Smith, D., & Mutlow, C.,& Delderfield, J. , & Watkins, B.,& Mason, G. (2012). ATSR infrared radiometeric calibration and in-orbit performance. *Remote Sensing of Environment*, *116*, 4-16.

Wang, L., & Cao, C. (2008). On-Orbit Calibration Assessment of AVHRR Longwave Channels on MetOp-A Using IASI. *IEEE Transactions on Geosciences and Remote Sensing*, *46*, 4005-4013.

Yu, F., & Wu, X., & Rama Varma Raja, M. K., & Wang, L., & Goldberg, M. (2012). Diurnal and Scan Angle Variations in the Calibration of GOES Imager Infrared Channels. *IEEE Transactions on Geosciences and Remote Sensing*, *51*, 671-683.

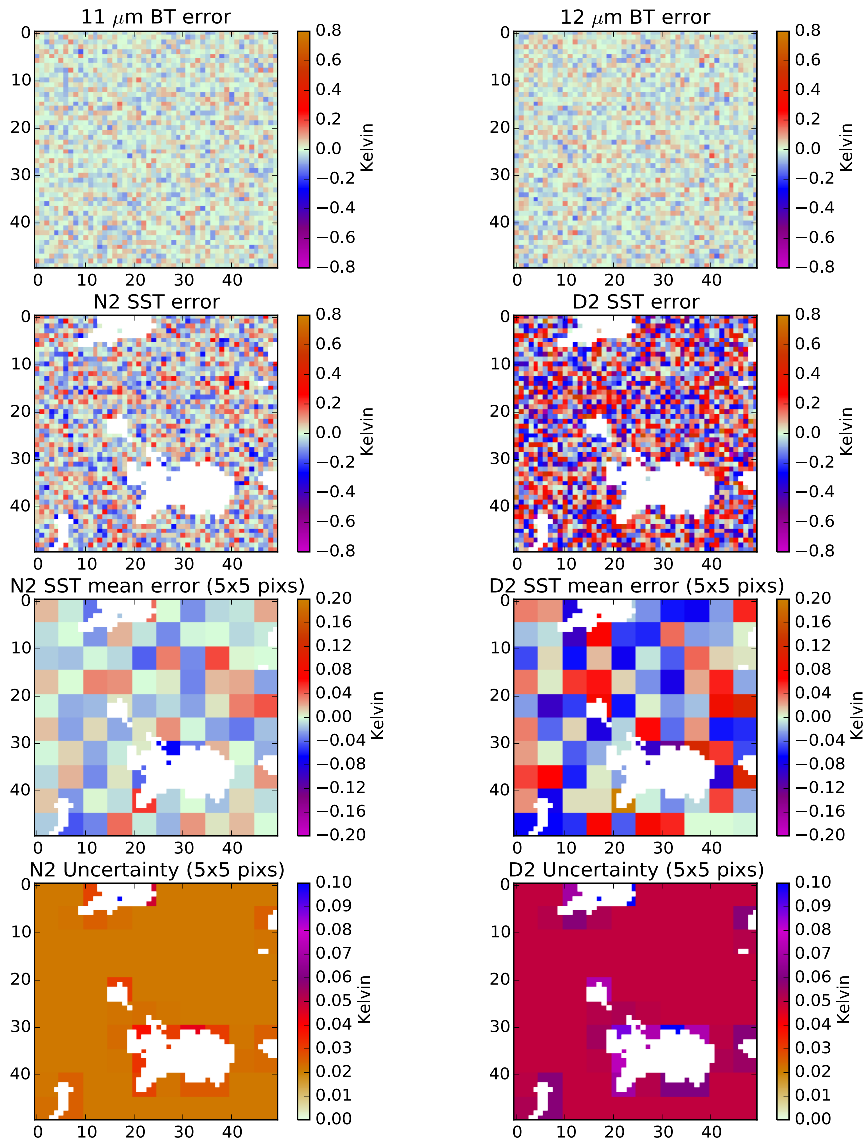

Figure 1: Uncorrelated random errors and uncertainties in brightness temperature observations and SST retrieval. Panels a) and b) show simulated errors in the 11 and 12 μm channels. Panels c) and d) show these errors propagated into SST retrievals for N2 and D2 retrievals. Panels e) and f) show the mean error at a 5x5 pixel resolution with a cloud mask superimposed on the data. Panels g) and h) show the associated uncertainty fields at a 5x5 pixel resolution.

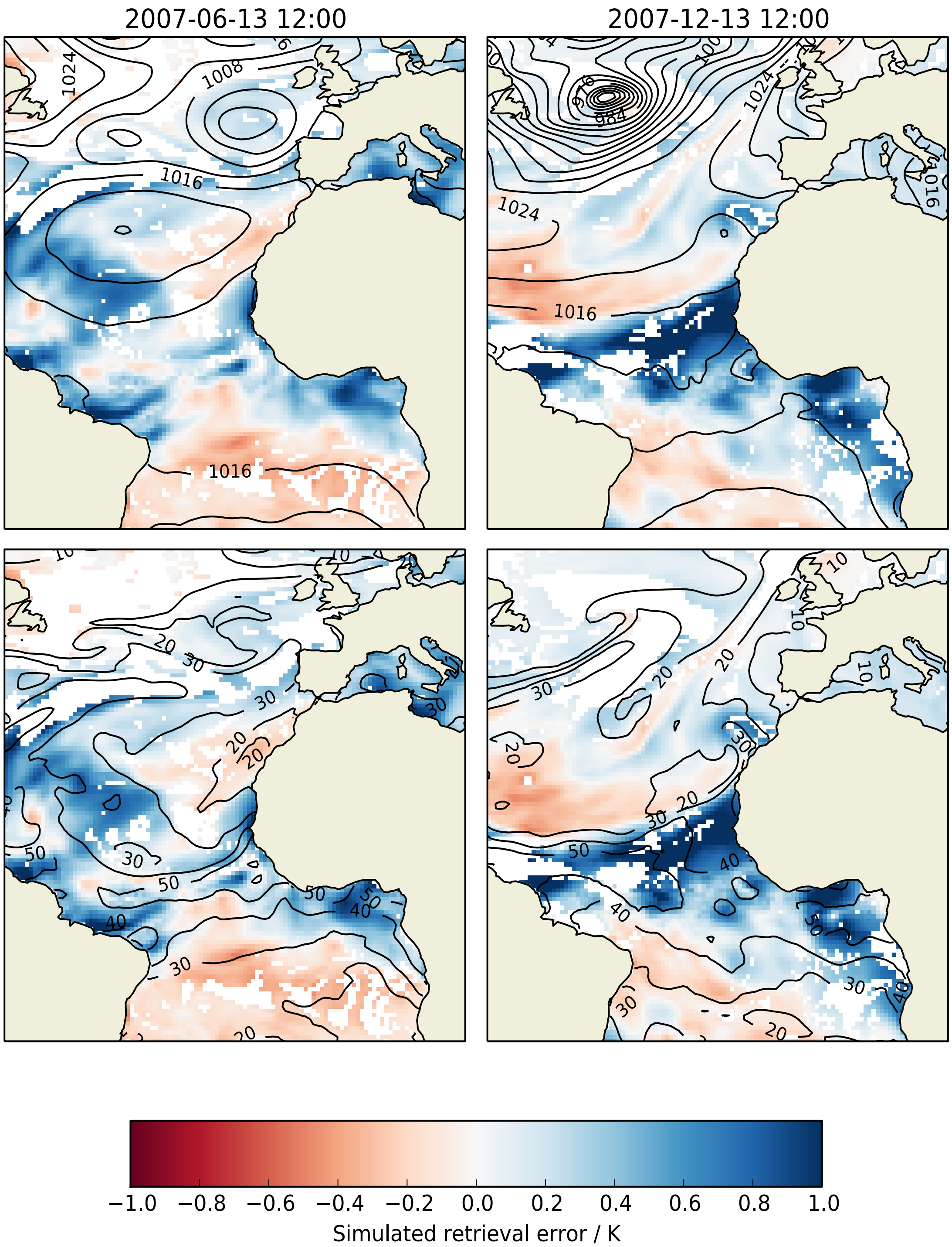

Figure 2: AATSR retrieval errors for two different days from simulation studies (left and right). Plots show the difference between the 'true' and retrieved SST field. Plots in the upper panels show pressure contours hPa, and plots in the lower panels TCWV contours kg m^{-2}.

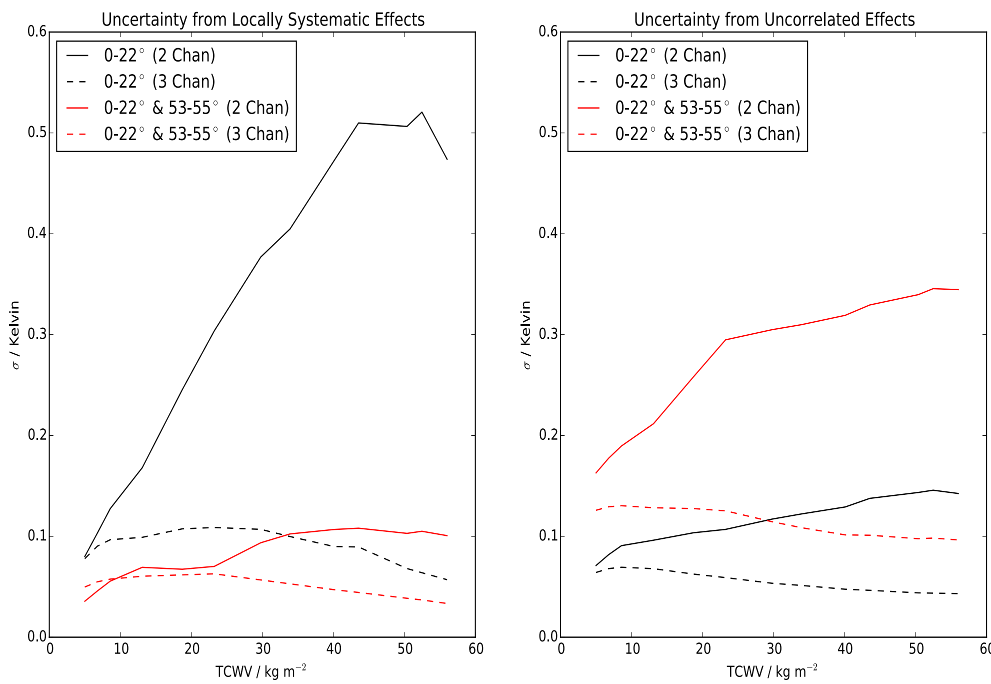

Figure 3: Uncertainties from a) locally systematic and b) uncorrelated effects as a function of total column water vapour for different channel combinations.

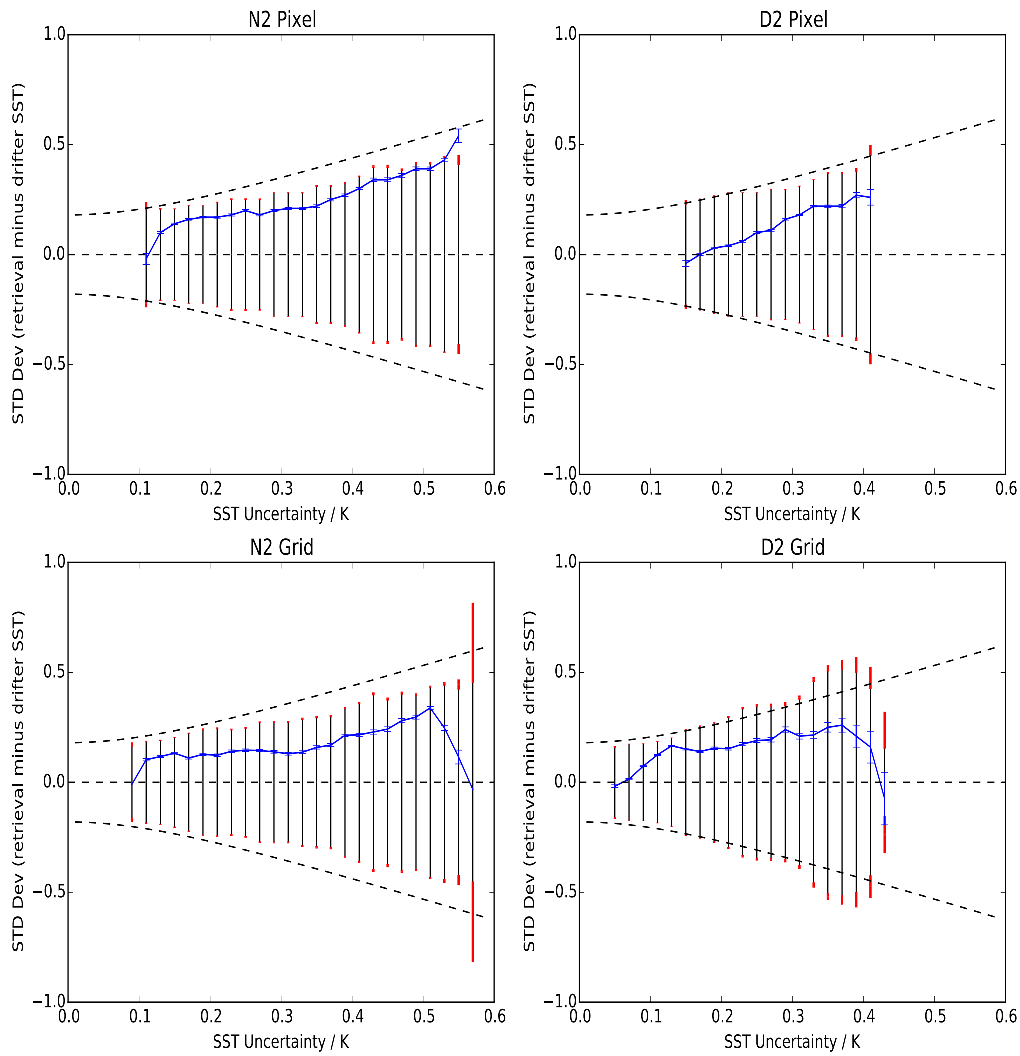

Figure 4: SST uncertainty validation against drifting buoy in-situ data. Top panels show pixel level uncertainties for N2 and D2 retrievals. Bottom panels show grid cell uncertainties (5x5 pixels approximately corresponding to a resolution of 0.05°) for N2 and D2 retrievals. Dashed lines show ideal uncertainty model accounting for uncertainties in the buoy data and geophysical uncertainties arising from a skin to depth comparison and colocation. Solid black lines show one standard deviation of the retrieved minus buoy SST differences, and blue lines the median satellite minus buoy SST difference. Error bars show the standard error in these differences. Uncertainties in the retrieval uncertainty are indicated by red bars at the base and top of the solid black lines.

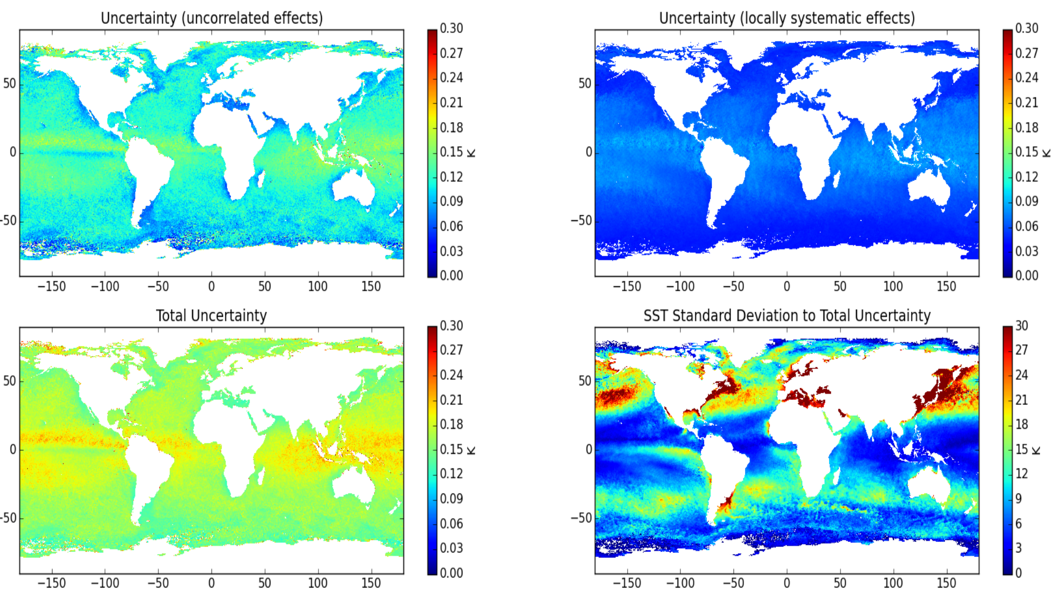

Figure 5: Annual means in SST retrieval uncertainties calculated from AATSR L3C data in 2010. Mean uncertainties are derived by adding all uncertainty observations in a given cell in quadrature, dividing by the number of observations and taking the square root. a) Shows uncertainty due to uncorrelated effects (noise and sampling uncertainty), b) shows noise due to locally systematic effects and c) total uncertainty. d) Shows the ratio of the SST standard deviation over 2010 to the total uncertainty.