Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data

Sampling Uncertainty in Gridded Sea Surface Temperature Products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data

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

[a]*Department of Meteorology, University of Reading, Reading, UK*

Abstract

Sea surface temperature (SST) data are often provided as gridded products, typically at resolutions of order 0.05° from satellite observations to reduce data volume at the request of data users and facilitate comparison against other products or models. Sampling uncertainty is introduced in gridded products where the full surface area of the ocean within a grid cell cannot be fully observed because of cloud cover. In this paper we parameterise uncertainties in SST as a function of the percentage of clear-sky pixels available and the SST variability in that subsample. This parameterisation is developed from Advanced Along Track Scanning Radiometer (AATSR) data, but is applicable to all gridded L3U SST products at resolutions of 0.05-0.1°, irrespective of instrument and retrieval algorithm, provided that instrument noise propagated into the SST is accounted for. We also calculate the sampling uncertainty of ~0.04 K in Global Area Coverage (GAC) Advanced Very High Resolution Radiometer (AVHRR) products, using related methods.

Keywords:

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

---

1. Introduction

This paper addresses sampling uncertainty when deriving gridded sea surface temperature products from satellite infrared imagery data. Remotely sensed sea surface temperature data have uncertainties that should be quantified for scientific applications. Typically, uncertainties in satellite retrieval of sea surface temperature (SST) are quantified in a general sense via validation activities with reference to in-situ data (Donlon et al., 2007; GHRSST Science Team, 2010). In a companion paper, we present a method to estimate context-specific uncertainties using physics-based models of uncertainty arising from different sources of error, evaluated for each SST retrieval. The uncertainty estimates can be validated independently using in-situ data (Bulgin et al., 2016). This is one component of the uncertainty budget in a grid cell mean SST, which also includes random components from radiometric noise (hereafter referred to as noise), locally systematic components that arise in the SST retrieval step and uncertainty arising from unknown large scale systematic errors. A full discussion of these other components is provided in Bulgin et al. (2016). This paper focuses on the derivation of an empirical model of the uncertainty from spatially subsampling a grid cell for which an area-average SST is to be estimated.

In this paper we will carefully distinguish the terms 'error' and 'uncertainty', which are often used ambiguously. Error can be defined as the difference between an SST estimate (in this case from satellite data) and the true

SST (Kennedy, 2013; JCGM, 2008). In practice, the true SST is unknown and therefore we cannot know the measurement error. We can instead calculate the uncertainty, which is a measure of the dispersion of values that could reasonably be attributed to that measurement error, we use a 'standard uncertainty' -ie, quoted uncertainties represent an estimate of the error distribution standard deviation (JCGM, 2008). Although within this paper the terms 'error' and 'uncertainty' are used according to these definitions, usage differs in some cited references.

For many applications, SST data are not used or provided at the full resolution of the sensor but are averaged over defined areas to produce a gridded product. For large datasets with observations spanning many years, this approach can be necessary to reduce the volume of data for some users. Gridding in this way destroys more detailed information on the location of measurements, and so a gridded SST value is taken as an estimate of the average SST across the grid cell over some time period. Spatial sampling uncertainty is present in gridded products, since the full grid cell may not be observed (eg. because of partial cloud cover). If the gridded SST covers a window of time (rather than being a measurement at a stated time) there is also temporal sampling uncertainty, since the full time period may not be observed (eg. one or two passes available during a day from which to make a daily estimate). Temporal sampling issues are not discussed in this paper.

Sampling uncertainty has been widely considered in the construction of global or regional SST records from in situ records for evaluating temperature trends (Brohan et al., 2006; She et al., 2007; Rayner et al., 2006; Morrissey and Greene, 2009; Jones et al., 1997; Folland et al., 2001; Karl et al., 1999).

In this context, sampling uncertainties arise from the number of observations available in each grid cell and how well they represent the mean temperature within the grid cell in both space and time (Jones et al., 1997). Sampling uncertainty estimates consider the spatio-temporal correlation of measurements at different locations within the grid cell (Morrissey and Greene, 2009), the temporal variability in SST for each grid cell (Jones et al., 1997) and consistency in observation depth (She et al., 2007).

Here we use data from the Advanced Along Track Scanning Radiometer (AATSR) instrument to study sampling uncertainty in a gridded satellite SST product. We calculate sampling uncertainties in data gridded at two different spatial resolutions (0.05° and 0.1°) previously used in SST products (eg. Embury and Merchant (2012); Merchant et al. (2014)). We separate sampling uncertainty from other sources of uncertainty in SST so that it can be estimated as a distinct contribution to the total uncertainty estimate. We address only spatial sampling uncertainty because we aim to estimate total uncertainty in SST in a grid cell at the stated time of the satellite observations from a single overpass. We use the approach established in this paper to consider sampling uncertainty in data provided at lower spatial resolution than the native observations, for example in the case of Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) products.

The remainder of the paper proceeds as follows. In Section 2 we discuss the AATSR data and how they are used to synthesise sampling error distributions. In Section 3 we derive steps for calculating sampling uncertainty. In Section 4 we present our results using AATSR data and define a parameterisation for sampling uncertainty applicable over a range of spatial scales. In Section 5 we consider uncertainties arising from GAC sampling from the AVHRR instruments. We provide a discussion of the results in Section 6 and conclude the paper in Section 7.

2. Data and Methods

Level 3 uncollated (L3U) satellite data products (the subject of this paper) are defined as an average of the L2P data points of the highest quality level that fall within the L3 grid cell (GHRSST Science Team, 2010). The gridded SST product as defined by the Group for High Resolution Sea Surface Temperature (GHRSST) specification is therefore a simple average of the available observations as an estimate of the areal mean. Although other methods could be considered for generating areal means, such as Kriging, this is not the commonly accepted practice in this field.

When generating gridded SST products from infrared imagery, typically only a subsample of the potential SST observations are available, predominantly due to cloud obscuring the surface, but occasionally due to a failed retrieval or other problems with the observed data. If SST data points are available covering the whole grid cell, an average SST can be calculated over the grid cell. If a subset of points are available, the mean of these data may differ from the true mean across the grid cell and therefore an element of uncertainty is introduced into the mean of the available SSTs interpreted as a grid cell mean. In this study, we mainly use data extracts from the Advanced Along Track Scanning Radiometer (AATSR) over clear-sky regions in order to calculate the uncertainty introduced by estimating grid cell mean

SST from a subsample.

We extract 10 x 10 and 5 x 5 pixel samples globally which approximately correspond to the size of 0.1 x 0.1° and 0.05 x 0.05° grid cells across the tropics and mid-latitudes. AATSR has a pixel size of 1 km. At the equator, 5 km corresponds to 0.045 x 0.045° and at 60 degrees, this is 0.09 x 0.045°. For the 10 km samples, these are 0.09 x 0.09° at the equator and 0.18 x 0.09° at a latitude of 60 degrees. Samples are selected from all latitudes on the condition that all constituent pixels are classified as clear-sky using the Bayesian cloud detection scheme applied to ATSR data in the Sea Surface Temperature (SST) Climate Change Initiative (CCI) project (Merchant et al., 2014). The 5 x 5 pixel cells are embedded in the 10 x 10 pixel cells enabling a direct analysis of the impact of cell size on sampling uncertainties. Subsamples of different numbers of clear-sky pixels ('m') are selected from the full sample size ('n') using two methodologies to exclude pixels 1) randomly and 2) using cloud-mask structures transposed from other cloudy images. Random masks are compared with observed cloud-mask structures to determine whether sampling uncertainties can be calculated accurately using a more simple approximation. We calculate the sampling uncertainties for all values of $m => 1$ and $m <= n - 1$ for each cell size (5x5 or 10x10 pixel extracts).

The details of this approach are as follows. For each grid cell size and value of 'm', we generated 500 random masks and extracted 500 realistic cloud masks from other AATSR data screened using the SST CCI Bayesian cloud detection scheme (Merchant et al., 2014). As noted above, all of the extracted samples are fully clear-sky, so neither mask corresponds to the cloud conditions of any extract. However, the cloud masks obtained from other images have the spatial structures representative of cloud fields observed at the scales of the imagery. Clear-sky samples were extracted from global AATSR observations between 1st - 3rd January 2003 and sea surface temperatures were calculated using an optimal estimation retrieval Merchant et al. (2014). For each cell size we extracted 250,000 samples. We then applied each mask (500 for each value of 'm') to each of the 250,000 extracts.

Figure 1 shows the global distribution of the 250,000 extracted samples. These are classified according to the standard deviation of the SST over the 5 x 5 pixel cell to give an indication of the spatial distribution of sub-grid SST variability. Clear-sky samples are extracted from orbit data globally with the majority of extracts between 60° South and 60° North. These differences between the masked and unmasked SSTs will be used to characterise sampling uncertainty having accounted for the effect of SST noise on both the full sample and subsample mean SST.

3. Sampling Uncertainty Derivation

This section presents the method of estimating sampling uncertainty from these differences, accounting for the fact that the pixel SSTs are noisy. We have to account for SST noise to develop a model for sampling uncertainty that applies to sensors with different noise characteristics. Each mean SST (of both a full extract and a subsample) will have an element of uncertainty that ultimately derives from instrument noise in the observed brightness temperatures from which the SSTs are estimated. To obtain a more accurate sampling uncertainty we account for SST noise by the following method.

Considering first a single case, the mean SST across the full extract ($\widehat{SST}_n$) of 'n' pixels can be expressed as:

$$\widehat{SST}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where 'i' indexes the pixels for which 'x_i' is the (unknown) true SST. For a subsample of 'm' pixels, $\widehat{SST}_m$ is:

$$\widehat{SST}_m = \frac{1}{m}\sum_{j=1}^{m} x_j \tag{2}$$

where the subscript 'j' represents the observations found in the subsample 'm'. The subsampling error, 'E' is calculated by subtracting $\widehat{SST}_n$ from $\widehat{SST}_m$.

$$E = \frac{1}{m}\sum_{j=1}^{m} x_j - \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

Using the subscript 'h' to index only those observations that are not present in subsample 'm' (indexed using j) this equation can be rearranged to give:

$$E = \left(\frac{1}{m} - \frac{1}{n}\right)\sum_{j=1}^{m} x_j - \frac{1}{n}\sum_{h=1}^{n-m} x_h \tag{4}$$

This equation does not account for noise in the retrieved SST. In practice, we have only have an estimate ($\hat{E}$) of the true sampling error that is noisy because of SST noise in both $\widehat{SST}_n$ and $\widehat{SST}_m$. Each retrieved SST, $\hat{x}_i$, is $\hat{x}_i = x_i + e_i$, where e_i is the error in the SST due to noise. We don't know e_i explicitly, but we have an estimate of ϵ_i, which is the standard uncertainty in a single pixel SST retrieval due to noise. The uncertainty due to noise in the extract mean is:

$$\epsilon_n = \frac{1}{n}\sqrt{\sum_{i=1}^{n} \epsilon_i^2} \quad (5)$$

with a similar expression for noise in a subsample mean, ϵ_m. The SST noise can be propagated through the form of equation (4) (Ku, 1966) to give the uncertainty in $\hat{E}$. Noise is negligibly correlated between pixels and the covariance term is therefore ommitted.

$$\epsilon_E = \left[\left(\sum_{j=1}^{m} \left(\frac{\delta E}{\delta x_j} \right)^2 \epsilon_j^2 \right) + \left(\sum_{h=1}^{n-m} \left(\frac{\delta E}{\delta x_h} \right)^2 \epsilon_h^2 \right) \right]^{1/2} \quad (6)$$

$$= \left[\left(\frac{1}{m} - \frac{1}{n} \right)^2 \sum_{j=1}^{m} \epsilon_j^2 + \left(\frac{1}{n} \right)^2 \sum_{h=1}^{n-m} \epsilon_h^2 \right]^{1/2} \quad (7)$$

This uncertainty due to noise in $\hat{E}$ is then subtracted from $\hat{E}$ in variance space. Now, the sampling uncertainty (SU) we require is:

$$SU = [var(E)]^{\frac{1}{2}} \quad (8)$$

which we have to derive from realisations of $\hat{E}$ for different extracts. E and $\hat{E}$ are related by:

$$\hat{E} = E + e_E \quad (9)$$

where e_E is the error in the estimate of $\hat{E}$. Over multiple samples for a single given mask e_E is independent and uncorrelated with $\hat{E}$. Therefore, the variance in $\hat{E}$ is equal to the sum of the variance in E and e_E.

$$var(\hat{E}) = var(E + e_E) \tag{10}$$

$$= var(E) + var(e_E) \tag{11}$$

$$= var(E) + \frac{1}{K}\sum_{k=1}^{K} \epsilon_{E_k}^2 \tag{12}$$

Where the k index represents different extracts. The variance of $\hat{E}$ is estimated from the sample variance to give an unbiased estimate, as follows:

$$var(\hat{E}) = \frac{1}{K-1}\sum(\hat{E} - < \hat{E} >)^2 \tag{13}$$

Here, K is the total number of extracts and $< \hat{E} >$ the mean $\hat{E}$. Therefore, the sampling uncertainty can be estimated, accounting for noise,

$$SU = [var(\hat{E}) - var(e_E)]^{\frac{1}{2}} \tag{14}$$

Therefore substituting in equations (12) and (13):

$$SU = \left[\left(\frac{1}{K-1}\sum\left(\hat{E} - \frac{1}{K}\sum_{k=1}^{K}\hat{E}\right)^2\right) - \left(\frac{1}{K}\sum_{k=1}^{K}\epsilon_E^2\right)\right]^{1/2} \tag{15}$$

We apply this equation for calculating sampling uncertainty to the data as described in the following section.

4. Results

4.1. Sampling Uncertainty over Different Grid Sizes

We consider first sampling uncertainties over 5 x 5 pixel extracts corresponding to gridded SST products at a resolution of 0.05°. For each value of 'm' between $2 \leq m \leq 24$ (number of pixels available in the subsample) we apply each of our 500 masks to the 250,000 extracts, treating random and realistic cloud masks separately. For each of the masked samples we calculate the difference between the full sample and the subsample mean SST. The case where $m = 1$ is considered in a following section (4.3).

As demonstrated in Section 3, sampling uncertainty is dependent on the number of pixels 'm' available in the subsample. It is also likely that the magnitude of the sampling uncertainty will be dependent on the underlying SST variability within the grid cell. There may be a significant gradient in SST within a grid cell, for example in coastal regions, areas of upwelling or near SST fronts. We would expect subsampling to introduce higher uncertainties in the SST estimate in such locations than in grid cells where the SST is more homogeneous. Our analysis is based on clear-sky data extracts to which we have applied our various cloud masks, so we can calculate the SST variability over the full grid cell. However, when considering subsampled data in satellite imagery, the SSTs of the obscured pixels are unavailable. We therefore examine the sampling uncertainty dependence on SST variability by calculating the SST standard deviation across the 'm' pixels available in the subsample.

The SST standard deviation across the subsample, minus the uncertainty due to noise (subtracted in variance space) is calculated using Equation (15) for each of the masked extracts and binned in 0.1 K bands between 0-0.6 K giving six groups of data. SST noise is propagated into the sample and subsample SSTs from the pixel level uncorrelated uncertainties in the SST product. In each bin we have 500 sampling uncertainty curves from the application of 500 different masks, which are then combined to give a weighted mean (Figure 2). With such a large dataset, for extracts where 'm' is small, we find some cases for which the variance in the estimated SST noise is greater than the variance in the subsample SST just because of statistical fluctuations. To avoid negative variance, in these cases, we set the subsample SST variance to zero, as the SST variability across the grid cell is extremely low.

In Figure 2, panel (a) shows the results for the application of random masks and panel (b) the application of realistic cloud masks. We see that in both cases (random and realistic cloud masks) sampling uncertainty increases as the percentage of clear-sky pixels (those available in subsample 'm') decreases. The larger the SST standard deviation in subsample 'm', the larger the associated sampling uncertainty for any given percentage of clear-sky pixels. For the random masks, the sampling uncertainty increases approximately linearly with a decreasing percentage of clear-sky pixels until a value of 30-35 % where a more exponential increase is evident. Where realistic cloud masks are applied, the relationship between the percentage of clear-sky pixels and sampling uncertainty is more linear, with the gradient of the line increasing with increasing subsample SST standard deviation.

We can also plot the same sampling uncertainty data as a function of the SST standard deviation with the SST due to noise removed in the subsample for selected values of 'm', as shown in the bottom two panels of Figure 2, again for random (c) and realistic (d) cloud masks. These plots demonstrate that there is little difference between the application of random and cloud masks for '$m = 24$' where only a single pixel is masked, as would be expected. In the application of realistic cloud masks the gradient in sampling uncertainty as a function of increasing subsample SST standard deviation is steeper than where random masks are applied, with larger overall sampling uncertainties even in regions of low subsample SST variability (0.0-0.1 K).

The higher sampling uncertainties associated with the application of realistic cloud masks in comparison with random masks are not unexpected. Cloud fields tend to have coherent (non-random) spatial distributions, which vary with the type of cloud. Realistic cloud masks are more likely to mask adjacent pixels than random masks, which would increase the sampling error where sample SST variability is high across the grid cell. SST is spatially correlated between pixels at a resolution of 1 km. Therefore for a grid cell with a high SST standard deviation there is likely to be a strong gradient across the cell rather than a randomly distributed SST field. This coupled with realistic coherent cloud spatial distributions increases sampling uncertainty in comparison with using random masks. These results suggest that spatial sampling uncertainties cannot be well represented by masking pixels at random.

To assess the effect of the cell size on the sampling uncertainty we also consider 10 x 10 pixel extracts that approximately correspond to SST gridded products at 0.1° resolution. Figure 3 shows the equivalent plots to Figure 2 for the larger cell size. For the 10 x 10 pixel cell we see a wider sample of clear-sky percentages due to the increased number of pixels in the sample. The shape of the sampling uncertainty curve in relation to the percentage of clear-sky pixels is similar for both random and realistic cloud masks to the 5 x 5 pixel equivalent. There is a larger discrepancy in the absolute sampling uncertainties here, with much higher values in the application of realistic cloud masks. For random masks the the shift from a linear to more exponential curve in SU as a function of the percentage of clear-sky pixels occurs at $\sim$20% for this cell size due to the greater number of pixels in each extract.

For the larger sample size, we see higher maximum sampling uncertainties when applying realistic cloud masks due to smaller clear-sky percentages being represented by whole numbers of pixels. When pixels are masked randomly, the sampling uncertainty for a given percentage of clear-sky pixels and subsample SST deviation is lower when calculated over the 10 x 10 pixel cell than then 5 x 5 pixel cell. For any given percentage of clear-sky pixels, more pixels are available in the subsample 'm' over the 10 x 10 pixel cell than the 5 x 5 pixel cell. This increases the likelihood that the observations in the subsample will be distributed across the entire sampled cell for broken cloud or where the length scale of the cloud structure is of the order of the cell size.

4.2. Modelling Sampling Uncertainties

In practice, when generating gridded SST products we cannot calculate sampling error by comparing the sample and subsample means as we do not have SST available for pixels obscured by cloud. We need therefore to model sampling uncertainty as a function of the variables we do have available: the percentage of clear-sky pixels and the SST standard deviation across subsample 'm', accounting for noise. We consider each SST standard deviation band separately and plot sampling uncertainty with respect to the percentage of clear-sky pixels, as a function of the number of pixels in the full grid cell extract.

We can model the sampling uncertainty for the 5 x 5 and 10 x 10 pixel cells by fitting a cubic in the form $SU = ax^3 + bx^2 + cx + d$ to the data where x is the percentage of clear-sky pixels. Figure 4 shows the 5 x 5 and 10 x 10 pixel data and sampling uncertainty model. The coefficients for each subsample standard deviation for the 5 x 5 pixel grid cells are given in Table 1 and for the 10 x 10 pixel grid cells in Table 2. Figure 4 indicates that the cubic fit (shown in solid lines) is a close match to the data (dashed lines). The data are slightly noisier than the model (as would be expected) and this is more obvious in the 10x10 cell where more percentages of clear-sky pixels are represented. In Figure 4, we see that the SST variability is the dominant factor determining the shape of the modelled sampling uncertainty curve. As this increases, the gradient of the sampling uncertainty curve increases giving larger uncertainties particularly for lower percentages of clear-sky pixels. The effect of varying 'n' is important in the context of generating products, regularly gridded in latitude/longitude where the number of pixels falling within each grid cell may vary with latitude or instrument coverage or viewing geometry. The sampling uncertainty curves for the two cell sizes show little deviation from one another suggesting that the impact of small variations in pixel number between grid cells at these scales is likely to be negligible.

The modelled sampling uncertainties are calculated from data where the

Table 1: Cubic coefficients as a function of subsample SST standard deviation for a 5x5 pixel cell using realistic cloud masks.

SST Std. Dev.	a	b	c	d
0.0-0.1 K	-1.67e^{-7}	3.51e^{-5}	-2.82e^{-3}	9.65e^{-2}
0.1-0.2 K	-1.86e^{-7}	3.95e^{-5}	-3.63e^{-3}	0.15
0.2-0.3 K	-1.31e^{-7}	2.74e^{-5}	-3.37e^{-3}	0.2
0.3-0.4 K	-9.94e^{-8}	1.86e^{-5}	-3.12e^{-3}	0.23
0.4-0.5 K	-5.51e^{-8}	6.57e^{-6}	-2.53e^{-3}	0.25
0.5-0.6 K	-3.26e^{-8}	-1.59e^{-6}	-1.94e^{-3}	0.25

Table 2: Cubic coefficients as a function of subsample SST standard deviation for a 10x10 pixel cell using realistic cloud masks.

SST Std. Dev.	a	b	c	d
0.0-0.1 K	-2.80e^{-7}	5.44e^{-5}	-3.7e^{-3}	0.1
0.1-0.2 K	-3.36e^{-7}	6.56e^{-5}	-4.81e^{-3}	0.16
0.2-0.3 K	-2.80e^{-7}	5.48e^{-5}	-4.72e^{-3}	0.2
0.3-0.4 K	-2.49e^{-7}	4.72e^{-5}	-4.63e^{-3}	0.24
0.4-0.5 K	-2.47e^{-7}	4.36e^{-5}	-4.61e^{-3}	0.27
0.5-0.6 K	-2.22e^{-7}	3.63e^{-5}	-4.33e^{-3}	0.29

effect of noise has been removed from the subsample SST. The model is therefore applicable to SSTs generated from any instrument or retrieval on the same scale provided that the propagation of uncertainty due to noise within the SST calculation has been correctly accounted for. This has been verified using the nadir only ATSR Reprocessing for Climate (ARC) coefficient based SST estimate as a comparison (Embury and Merchant, 2012). The propagation of noise differs in an optimal estimation and coefficient-based retrieval. So although the data are from the same instrument (on different days) this is a good test of the robustness of the methodology. We find a maximum RMSE of 0.017 K and maximum mean percentage difference of 0.16% over both extract sizes between the sampling uncertainty model presented here and the equivalent model generated using ARC data. In cases where the variance in the noise exceeds the SST variance (for low numbers of clear sky pixels), the SST variance should be set to zero in order to use the model (using the same approach as that adopted in the generation of the model).

4.3. Calculating Sampling Uncertainties for a Subsample of 1

So far, the discussion on sampling uncertainty as a function of subsample size has excluded the case where the subsample size 'm' is equal to one. Under these conditions sampling uncertainty cannot be calculated as a function of the subsample SST standard deviation. We therefore calculate the sampling uncertainty for the case '$m = 1$' across all extracts (using a weighted mean from 500 masks applied to each of the 250,000 extracts and report the mean value in Table 3, treating random and realistic cloud masks separately. The sampling uncertainties calculated will be weighted towards lower full sample SST standard deviations as there are more extracts with lower SST variability. In regions of high SST gradients these values are therefore likely to be an underestimate, and in very homogeneous regions an over-estimate. The overall tendency towards full samples with lower SST variability is however typical of the global distribution of gridded SST values as samples are extracted across the globe (Figure 1). The sampling uncertainty where '$m = 1$' is larger over the 10 x 10 pixel grid cell (0.141 K) than the 5 x 5 pixel grid cell (0.103 K). Where the full sample SST variability is high, a single pixel is unlikely to represent well the mean SST across the grid cell, and the larger the grid cell, the less likely this is to be representative.

Table 3: Sampling uncertainties for the case where $m = 1$ for cell sizes of 5 x 5 and 10 x 10 pixels, using realistic cloud masks (125 x 10^6 samples).

Cell Size	Sampling Uncertainty
5 x 5 pixels	0.103
10 x 10 pixels	0.141

5. AVHRR GAC Type Subsampling

So far we have considered the case where the number of pixels in the subsample 'm' is governed purely by data availability, ie. only observations obscured by cloud are eliminated from the available subsample. In the case of Global Area Coverage (GAC) data from the Advanced Very High Resolution Radiometers (AVHRR) (Robel et al., 2014), there is a predefined sub-sampling in the transmitted data. Observations are made at 1.1 km resolution at nadir, but due to limitations to data transmission from the early AVHRR instruments and latterly for consistency in data records, the GAC product is provided at a nominal resolution of 4 km. This is achieved by subsampling four pixels along the first scan line and then skipping a pixel before subsampling the next four pixels. The next two scan lines are skipped before resuming the sampling pattern described for the first line. Each four-pixel subsample is then considered to be representative of a 15 pixel cell (5 pixels across track by 3 pixels along track) (Robel et al., 2014). The signal received for each GAC pixel is the average brightness temperature or reflectance over the four pixels from the 15 pixel cell. Cloud screening is carried out on this average, rather than the constituent pixels, before calculating SST. This introduces a further source of sampling-related uncertainty, the calculation of which is beyond the scope of this paper. Here we consider the uncertainty introduced by regularly subsampling four in every fifteen pixels, and interpreting the four pixel average as an estimate for the full 3 x 5 pixel area.

We use Full Resolution Area Coverage (FRAC) Metop-A data to calculate the sampling uncertainty in GAC products. We take data from 33 orbits spanning the Metop-A data record and different times of year. From these orbits we identify all of the 5 x 3 pixel clear-sky extracts using the operational EUMETSAT cloudmask, which gives good global coverage of scenes (Ackermann et al., 2007). We apply the OSI-SAF coefficient based SST retrieval algorithms to these clear-sky extracts considering day and night separately, determined using solar zenith angle thresholds of $< 80°$ and $> 100°$ respectively (Le Borgne et al., 2007). We follow the methodology outlined in Section 3 to calculate the sampling uncertainty by taking a subsample of the first four pixels in every extract, having accounted for uncertainties due to noise, subtracted in variance space. We discard extracts where the operational cloud detection has seemingly failed to identify cloudy pixels, giving extreme SST variations across the fifteen pixel cell. We set an upper limit on the SST variation across a given grid cell of 2 K with the threshold determined using model SST data (unaffected by clouds) at $\frac{1}{48}th°$ resolution from Estimating the Circulation and Climate of the Ocean (ECCO2) (Menemenlis et al., 2008). For the closest match to GAC sampling we extract samples of

3 x 2 pixels ($\sim$6 x 4 km). Over a global sample of $> 11 \times 10^6$ extracts we find a maximum SST gradient of 2.06 K across the full samples.

We specify AVHRR GAC sampling uncertainties under daytime and nighttime conditions, for three satellite viewing angle bands (1-1.1,1.1-1.5 and 1.5-3 in secant theta space), corresponding to approximately 0-25°, 25-50° and 50-70°. The results are shown in Table 4, in addition to the number of extracts included in each calculation. For the OSISAF NL retrieval algorithm we find that the sampling uncertainty is $\sim$0.04 K and for the OSISAF T37_1 algorithm $\sim$0.03 K at night. One possible reason for the difference, given that retrieval noise is accounted for, is the effect of cloud contamination which is not explicit in the uncertainty budget. The OSISAF NL algorithm shows slightly lower sampling uncertainties during the day than at night which may be due to diurnal warming reducing SST variability (Katsaros et al., 2005).

The OSISAF T37_1 nighttime algorithm uses the 3.7 μm channel in addition to the 11 and 12 μm channels, which is less sensitive to any cloud which may be present eg. at cloud edges etc. This may explain the reduced variance in subsample minus full sample SSTs, and the slightly lower sampling uncertainties when using this algorithm. For the GAC data, there is little dependence on atmospheric path length with sampling uncertainties decreasing by $\sim$0.002-0.006 K at the swath edge. This is due to greater overlap of pixels at this viewing geometry effectively reducing the unsampled area across the 15 pixel cell.

Table 4: AVHRR GAC sampling uncertainties as a function of viewing zenith angle. SSTs are calculated using OSISAF coefficient based retrievals, with the NL algorithm applied at solar zenith angles $< 80°$ and the T37_1 and NL algorithms applied at solar zenith angles $> 100°$.

Time	Algorithm	Viewing Angle	Sampling Uncertainty	Number of Obs
Day	OSISAF NL	0-25°	0.045	4278613
Day	OSISAF NL	25-50°	0.042	3396053
Day	OSISAF NL	50-70°	0.038	1841405
Night	OSISAF NL	0-25°	0.049	3639976
Night	OSISAF NL	25-50°	0.047	3015435
Night	OSISAF NL	50-70°	0.045	1668920
Night	OSISAF T37_1	0-25°	0.028	3639976
Night	OSISAF T37_1	25-50°	0.029	3015435
Night	OSISAF T37_1	50-70°	0.023	1668920

6. Discussion

Sampling uncertainties are yet to be routinely characterised in gridded SST products and the model presented here provides a method for calculating these uncertainties, applicable to all SST retrievals at the same scales as those studied here, where uncertainties due to noise have been removed. The impact of cell size is shown to be less important than the subsample SST variability in determining the sampling uncertainty and therefore these modelled uncertainties can be applied to grid cells at different latitudes and varying viewing geometries where the number of pixels falling within each grid cell can show local variation.

The results presented in Section 4.1 highlight significant differences in the sampling uncertainties calculated when applying randomly generated and observed cloud masks to the extracted samples. Sampling uncertainties calculated on the basis of random masking are an underestimate of the true uncertainty, a consequence of the spatial structure of both clouds and the underlying SST field. Observed masks more often eliminate clumps of pixels delineating a cloud feature rather than random pixels across a given cell. As the percentage of clear-sky pixels is reduced this increases the likelihood of masking a large coherent section of the image. In all but the most homogeneous cases of SST, the mean temperature of the remaining section is less likely to be representative of the whole cell than a random distribution of pixels scattered across the grid cell, due to the coherent structure of the underlying SST.

Sampling uncertainties are inherent in all gridded products generated from a subset of available data, eg. AVHRR GAC SSTs, Level 3 data. These data can be used for a variety of applications and both data users and providers should be aware of the uncertainties introduced by subsampling the higher resolution data.

7. Conclusions

In this paper we present a methodology for calculating sampling uncertainty in gridded SST products once the uncertainty due to noise in the observations has been removed. We model sampling uncertainty as a function of the percentage of clear-sky pixels within a given grid cell and the SST variability within those available pixels, considering cell sizes of 0.05° and 0.1°.

We establish that the dominant factor in determining sampling uncertainty is the subsample SST standard deviation and that latitudinal variations in the number of pixels falling within a given grid cell have a negligible effect. Our model is applicable to SST retrievals from any instrument on the same spatial scales, using any retrieval scheme providing that the propagation of instrument noise through the retrieval is correctly accounted for. We also consider the impact of routine subsampling of higher resolution data in the provision of GAC AVHRR products. We characterise sampling uncertainty as a function of atmospheric path length corresponding to viewing zenith angle, as information regarding the SST variability within the subsample is not provided within the GAC product. We find that sampling uncertainty is typically of the order of 0.04 K. We recommend the inclusion of sampling uncertainties in the uncertainty estimates provided with SST products, and demonstrate the validation of the ATSR uncertainty budget including this component in the companion paper (Bulgin et al., 2016).

8. Acknowledgements

The work undertaken in this paper was funded by the European Space Agency Sea Surface Temperature Climate Change Initiative project. We thank the ECCO2 project for making their model output available.

References

Ackermann, J., & Klaes, D., & McKernan, E., & Montagner, F., & Heinemann, T., & Schĺussel, P., & Schraidt, R., & Fiedler, L., & Lang, R., &

Bonsignori, R., & (2007). ATOVS/AVHRR Onboard Metop-A: Validation and Use of Products. EUMETSAT, Am Kavalleriesand 31, D-64295, Darmstadt, Germany.

Bulgin, C. E., & Embury, O., & Corlett, G., & Merchant, C. J. (2016). Independent uncertainty budget estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments. *Remote Sensing of Environment*.

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*, D12, 21pp.

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*, *88*, 8, 1197-1213.

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.

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, *120pp*.

Joint Committee for Guides in Metrology (2008). Evaluation of measurement data - Guide to the expression of uncertainty in measurement. www.bipm.org, *134pp*.

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

Karl, T. R., & Knight, R. W., & Christy, J. R. (1999). Global and Hemispheric Temperature Trends: Uncertainties Related to Inadequate Spatial Sampling. *Journal of Climate*, *7*, 1144–1163.

Katsaros, K. B, & Soloviev, A. V., & Weisberg, R. H., & Luther, M. E. (2005). Reduced Horizontal Sea Surface Temperature Gradients Under Conditions of Clear Skies and Weak Winds. *Boundary-Layer Meteorology*, *116*, *2*, 175-185.

Kennedy, J. J. (2013). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, *51*, *1*, 1-32.

Ku, H. H. (1966). Notes on the Use of Propagation of Error Formulas *Journal of Research of the National Bureau of Standards - C Engineering and Instrumentation*, *70C*, 263-273.

Le Borgne, P., & Legendre, G., & Marsouin, A. (2007). Operational SST Retrieval from METOP/AVHRR. *Proceedings of EUMETSAT Conference*, 10pp.

Menemenlis, D., & Campin, J., & Heimbach, P., & Hill, C, & Lee, T., & Nguyen, M, & Schodlok, M., & Zhang, H (2008). ECCO2: High resolution global ocean and sea ice data synthesis.. *Mercator Ocean Quarterly Newsletter*, *31*, 13-21.

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). *Journal of Atmospheric and Oceanic Technology*, *21*, *11*, 1734-1746.

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.

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.

Robel, J., & Graumann, A., & Kidwell, K., & Aleman, R., & Ruff, I., & Muckle, B., & Green, P., & Brown, S., & Throwe, J., & Planet, W., & Knoll, J., & Fineberg, D., & Winston, W., & Robinson, G., & Miller, T., & Goodrum, G., & Askew, J., & Sapper, J., & Patterson, S., & Harrod, E, & Brown, E., & Wark, D., & Ramsay, B., & Krimchansky, S., & Flynn, L., & Wrublewski, T., & Ross, D. & Mo, T., & Graumann, A., & Bowman, D., & Sullivan, J. & Peterski, N., & Hollinger, M., & Clouse, W., & Weinreb, M. & Rao, N., & McMillin, L., & Goldberg, H., & Kleespies, T. (2014). NOAA KLM User's Guide with NOAA-N, N Prime, and MetOp Supplements. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data and Information Service. http://www1.ncdc.noaa.gov/pub/data/satellite/publications/podguides/N-15%20thru%20N-19/pdf/0.0%20NOAA%20KLM%20Users%20Guide.pdf.

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.

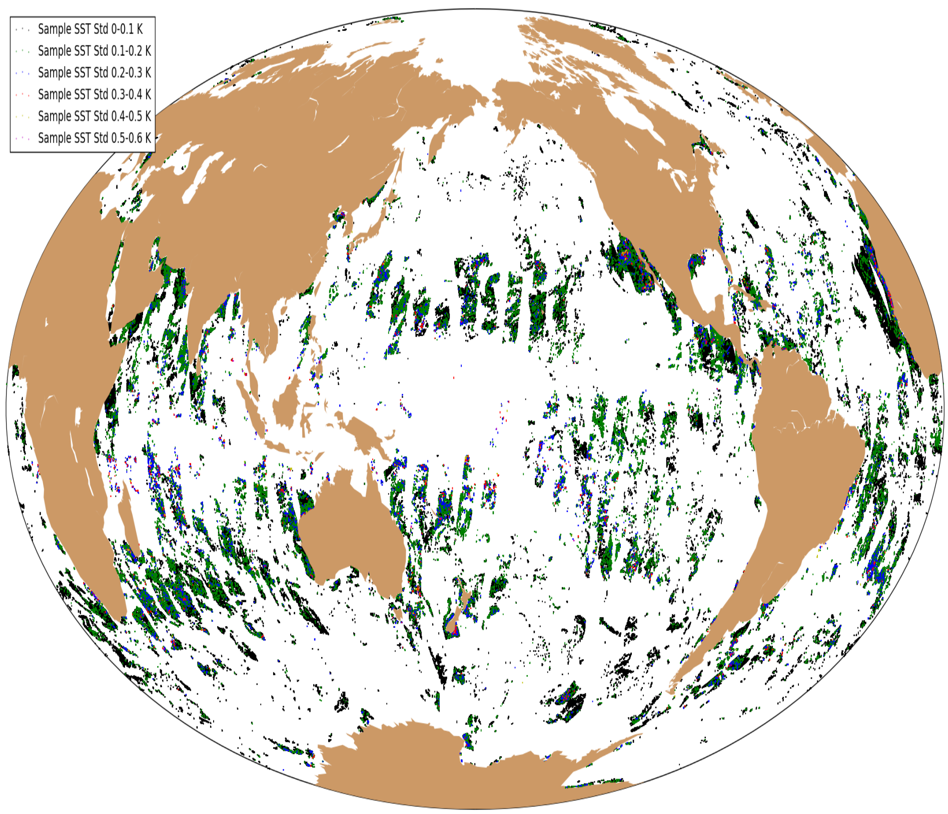

Figure 1: Global distribution of 10 x 10 pixel clear sky sea surface temperature samples extracted from AATSR data between 1st-3rd January 2003. Samples are colour coded according to the SST standard deviation across the sample cell.

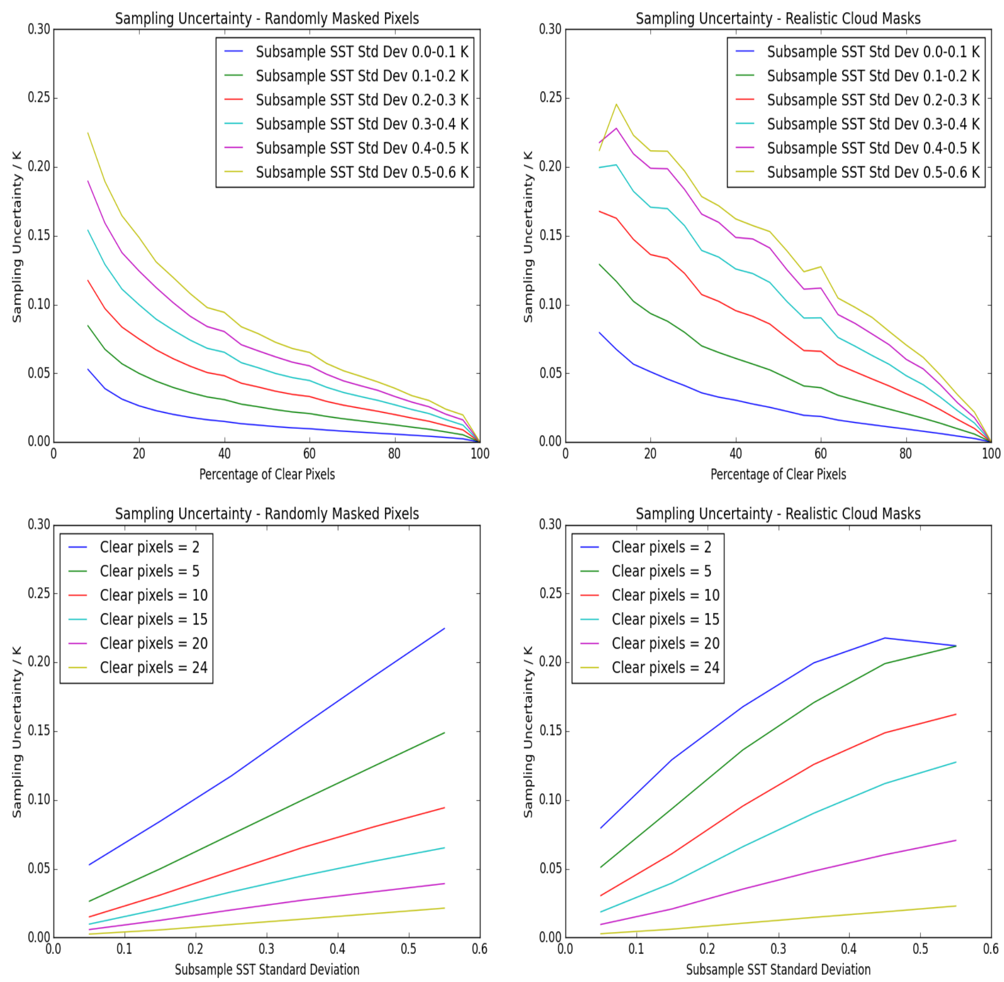

Figure 2: Top: Sampling uncertainties as a function of clear-sky pixel percentage over a 5 x 5 pixel cell with the application of randomly generated (left) and observed (right) cloud masks. Data are separated into six subsample SST standard deviation bands between 0-0.6 K. Bottom: Sampling uncertainty as a function of the subsample SST standard deviation with application of randomly generated (left) and observed (right) cloud masks. Results are presented for a number of clear-sky pixels.

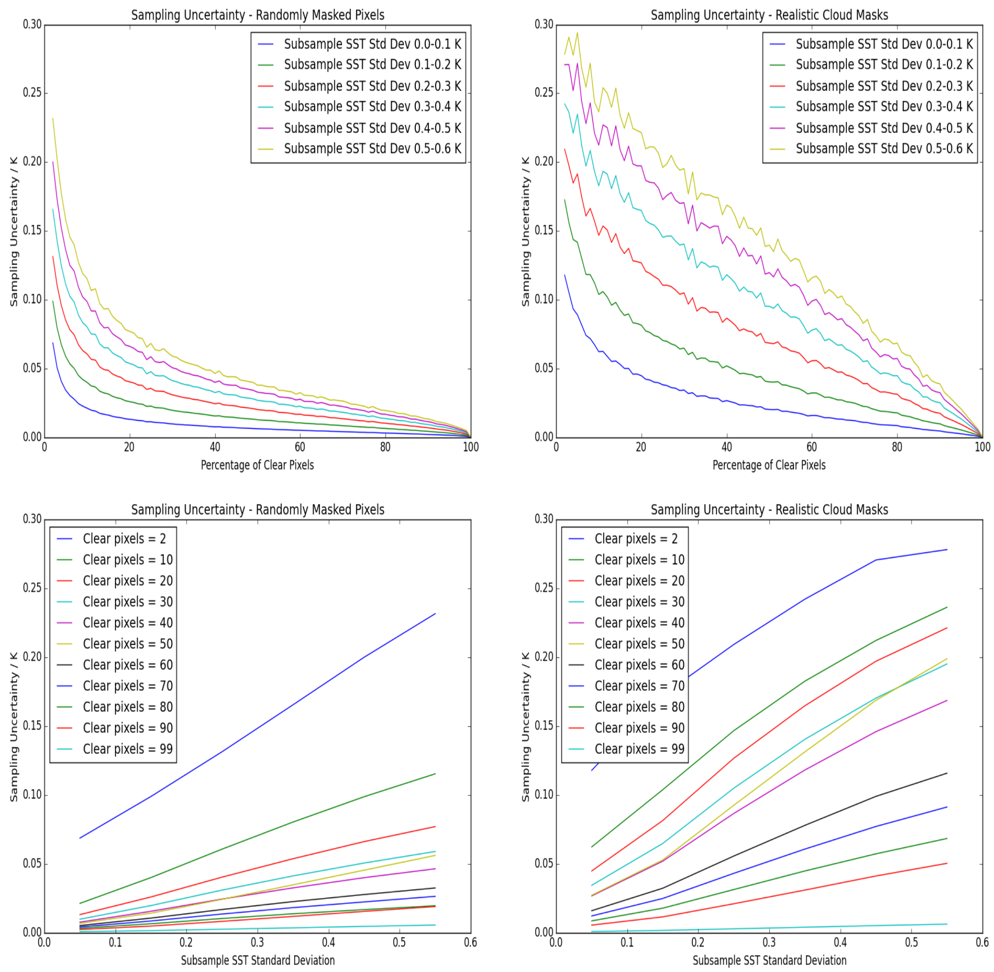

Figure 3: Top: Sampling uncertainties as a function of clear-sky pixel percentage over a 10 x 10 pixel cell with the application of randomly generated (left) and observed (right) cloud masks. Data are separated into six subsample SST standard deviation bands between 0-0.6 K. Bottom: Sampling uncertainty as a function of the subsample SST standard deviation with application of randomly generated (left) and observed (right) cloud masks. Results are presented for a number of clear-sky pixels.

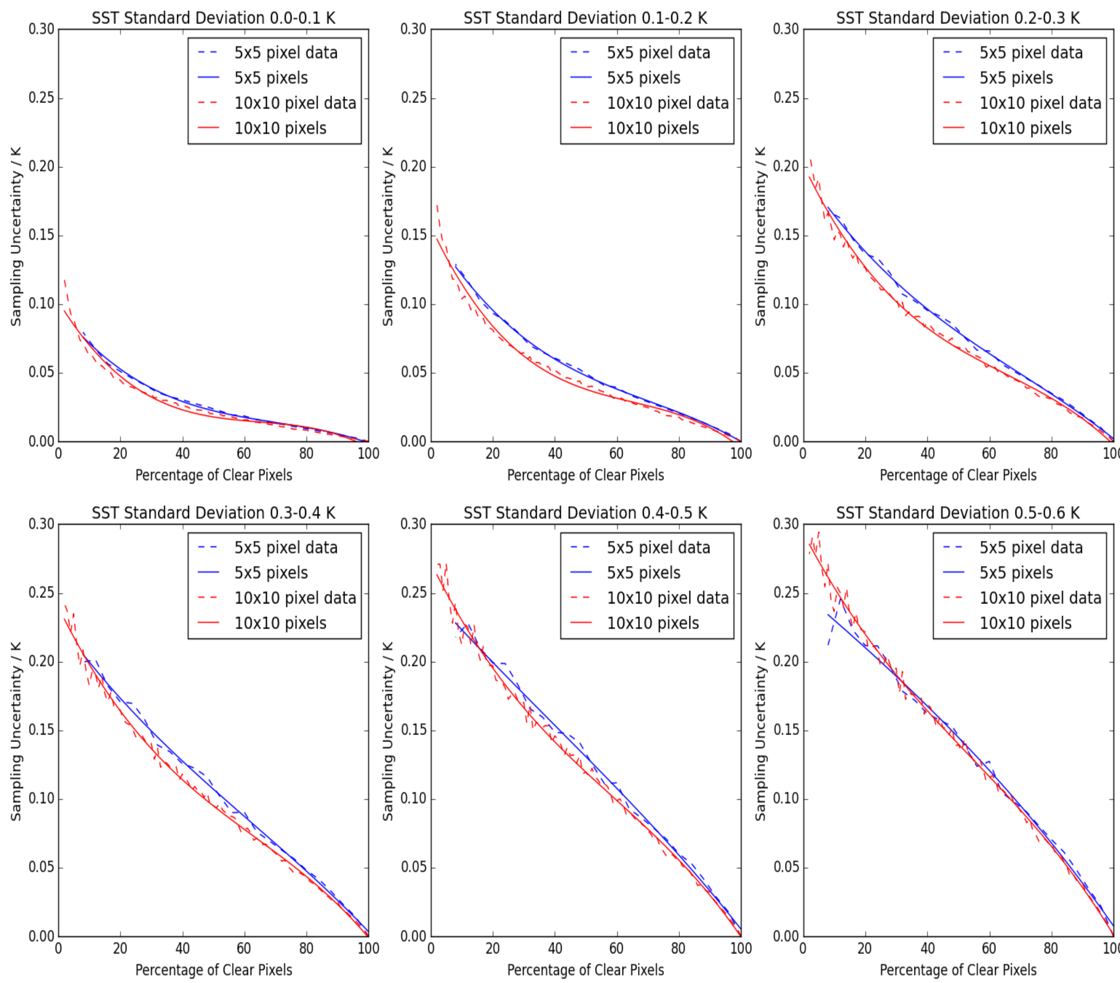

Figure 4: Modelled sampling uncertainties for 5 x 5 and 10 x 10 pixel cells over six subsample SST bands ranging between 0-0.6 K. Data for 25 and 100 pixel cells are overplotted in each panel.