Sea Surface Temperature User Workshop on Uncertainty

Nick A Rayner, Chris J. Merchant, Gary K. Corlett, Jon Mittaz, Claire Bulgin, Chris P. Atkinson, Simon A. Good and John J. Kennedy

Summary

The workshop was held at the Met Office, Exeter, UK, on the 18th – 20th November 2014. Forty participants attended, spanning a range of interests, including climate research, numerical weather prediction, physical ocean modelling, ecological analysis, Earth system model evaluation and producers of SST datasets for both the satellite and historical eras. In line with the aim of the workshop for two-way dialogue, the event was structured with scene-setting presentations, a poster session, practical exercises applying uncertainty information, and break-out group discussions leading to plenary feedback and debate.

The need for such a workshop was identified in general terms in the community white paper of the EarthTemp network (Merchant et al., 2013), and was organised within the auspices of the European Space Agency's Climate Change Initiative (Hollman et al., 2013) project on sea surface temperature (SST CCI). The SST CCI project is working to generate new climate data records for SST from satellite observations (Merchant et al., 2014) with a number of characteristics: independence from in situ observations, harmonisation across satellite sensors to maximise stability, the potential to link to historical in-situ based datasets and provision of realistic context-sensitive uncertainty estimates for every SST at all spatio-temporal scales. Ensuring that this latter objective is achieved in a way that is useful to SST users was the major motivation for the project to run this workshop.

There are significant parallels between the priorities of a developer of a climate data record (CDR) and the priorities of a metrologist (metrology being the science of measurement and its uncertainty): both want measurements that are stable over time, insensitive to particular sensors and measurement methods, and are of uniform calibration and quality worldwide. Links were identified between Earth observation and metrology in day 1 of the workshop, via presentation and discussion. Good practice in tracing the uncertainty budget through satellite processing levels from instrumental measurement to geophysical product can be based on metrological norms encapsulated in documents such as the Guide to the expression of Uncertainty in Measurement (GUM; BIPM, 2008). In communicating about uncertainty, careful adherence to standard vocabulary can reduce ambiguity and increase understanding. In particular, it is useful to preserve the distinction between error ('mistaken-ness') and uncertainty ('doubt'). Uncertainty is typically quantified as the standard deviation of an estimated error distribution. In the case of Earth observation, uncertainty cannot be estimated from the dispersion of replicate measurements, as in a laboratory. Uncertainty modelling for EO relies heavily on understanding the instruments and retrieval processes, supporting error propagation by simulation and/or analytic techniques (Merchant and Embury, 2014). The metrological discipline of creating an uncertainty budget that is traceable (complete and defensible at each link the in the chain) can be used as a precedent for establishing the rigour and credibility of CDRs from EO.

The effects leading to errors in SST measurements of all sorts were reviewed in the workshop, it being clear that errors from different effects have very different correlations in space and

time. Instrumental noise is usually modelled as independent random error between each SST measurement. Calibration drift over time, as an instrument ages, produces error that is highly correlated over global scales and years. EO datasets will usually also include errors on intermediate scales between these extremes, which may be termed locally systematic. In the case of SST, these effects are generally related to ambiguity arising from atmospheric variability, and therefore the errors correlate on synoptic scales. The current SST CCI approach is to specify components of uncertainty from random, locally systematic and systematic effects separately, and using these requires the user to engage with these concepts. In deriving uncertainty estimates for SSTs across the full range of scales needed by users, these different components of uncertainty need to be tracked and propagated separately, appropriate to their correlation structure.

Participants recommended that full characterisation and clear documentation of the error model was needed and either that these uncertainty components should be provided together with correlation information, or that their complex behaviour should be encapsulated in an ensemble as currently done by some providers of centennial-scale SST data sets. Since error covariance matrices can be large and difficult to use, it was recommended that these be parameterised to allow easy communication. It was therefore apparent in the workshop discussions that there is no simple answer to delivering uncertainty information to users, and that different users, for legitimate reasons, have different preferences amongst these options. The Climate Forecasting (CF) conventions require extension to accommodate more nuanced uncertainty information, in which effects and correlation structure can be specified. Participants recommended that use of uncertainty information would be facilitated by the provision of tools, for appropriate error propagation, ensemble selection or to create user-defined flags.

Even when provided with data producers' uncertainty estimates, users do not necessarily use these at face value. A discussion was held about what is required for SST users to trust uncertainty estimates attached to data as being realistic, and directly usable within their applications. Uncertainty validation and verification was welcomed, but more reference data is needed. The other two major influences were the scientific reputation of the data producers, established through the norms of peer reviewed publication, etc, and precedents where uncertainty information have been successfully exploited in applications. Formal mechanisms, such as publishing uncertainty traceability chains, were judged as less influential. However, this does not undermine the need for data producers to engage with uncertainty estimation in a rigorous, defensible manner, since this is part of building the necessary scientific credibility, as well as being good science practice.

Since precedent is persuasive to users, data producers can actively promote uptake of their products and the uncertainty information by engaging with trail-blazer users of their improved uncertainty information in products.

Recommendations that arose within the workshop will be synthesized and used to update the User Requirements Document maintained by the SST CCI project team (update due December 2015). The organisers thank all the attendees for their enthusiastic and constructive contributions.

1. Aims and Context of Workshop

Many different groups around the world develop observational data sets and analyses of sea surface temperature for varying applications, e.g. short-term weather and ocean forecasting, fisheries and ecosystem research, monthly-to-decadal forecasting, evaluation of climate model simulations, engineering and military use. In most cases, uncertainties in this information, arising from various aspects of the measurement or estimation processes involved, will have an impact on the way it is used and may impact on the conclusions drawn.

Hitherto, the use of provided uncertainty information has been minimal. This means a disconnect has existed between users and those sea surface temperature data providers who have attempted to quantify and communicate uncertainties in their products; either uncertainties are not being provided in a useful way, or users do not understand how to use them. This is true for other surface temperature data too, as identified in Merchant et al (2013).

Providing information on uncertainties in sea surface temperature (or any other) measurements in a way that is useful to users requires in-depth discussion with a range of different users. This is necessary in order to understand what those users are aiming to achieve and, therefore, how these uncertainties might affect them. Accordingly, the European Space Agency Climate Change Initiative sea surface temperature project identified this as a priority for their current programme of work and proposed a workshop to try to engender a common understanding of:

- where uncertainty in observational sea surface temperature products comes from;
- how to talk about it;
- how well the uncertainty information that is provided addresses users´ needs; and
- how to practically use such information.

It comprised a mixture of oral and poster presentations, activities and group discussions.

The aims of the meeting were to:

- Exchange information about uncertainties in sea surface temperature observations;
- Create new expert sea surface temperature users who, through publication of their work, can inspire others to take uncertainty information into account;
- Source requirements from sea surface temperature users on uncertainty information and other aspects of the Climate Data Records;
- Spread best practice through a follow-on meeting report or journal article.

2. Summary of sessions

2.1. SST CCI products

There are many sea surface temperature products, and the SST CCI project aims to make a unique contribution to climate observation by creating a dataset for the last few decades that has the following properties:

1. independence from SST measurements made *in situ*
2. of useful, quantified accuracy and sensitivity
3. with context-sensitive uncertainty estimates (at all spatio-temporal scales)
4. harmonised to provide useful stability
5. able to be linked to the longer historical record
6. generated by a robust, sustainable processing system

To achieve this SST CCI products are based on linking measurements from all sensors to the series of Along Track Scanning Radiometers, particularly ATSR-2 and AATSR. For these sensors, physics-based retrieval (i.e. estimation of SST from the radiances measured by the satellite instrument) can be defined using radiative transfer simulations of the change in radiance between the surface and the instruments, because these instruments are particularly well characterised.

Measurements from other sensors, particularly AVHRRs, are cross-calibrated (at both radiance and SST stages) to the ATSR results. This is useful since the AVHRRs extend back prior to the ATSRs in time and have better spatial sampling, but can be less accurate.

Many AVHRR sensors have flown on satellites whose orbits have decayed over time, resulting in a progressive change in the local time at which they observe any particular location. SST can have a large diurnal cycle under some conditions. To avoid aliasing of this diurnal cycle into the apparent changes seen in the long term record, thus compromising the long-term observational stability, all SST CCI instantaneous skin temperatures are adjusted via a model to a reference local time of day (1030 h or 2230h). To allow linkage to historical data, the same model also provides an adjustment to 20 cm depth, since these satellite instruments estimate the temperature of the first few microns below the ocean surface.

The issue of independence is driven by the desire to build confidence in satellite-only and in situ-only SST datasets by showing that these are compatible in their view of ocean surface temperatures, despite the need in both cases to account for observing system changes over time.

Swath (also known as Level 2), gridded (Level 3) and analysed (Level 4, gridded then in-filled using statistical techniques) versions of data products are provided by SST CCI, and an uncertainty estimate is provided with every SST data value.

To use SST CCI products, read the short data paper (Merchant et al., 2014), the Product User Guide (http://www.esa-sst-cci.org/sites/default/files/Documents/public/SST_CCI-PUG-UKMO-001_Issue-3-signed-accepted.pdf) and following links therein to data and further documentation.

2.2. Vocabulary and presentation of uncertainty

Vocabulary

We established a common vocabulary to be used when discussing uncertainty with the aid of a presentation drawing on experience from metrology.

After a brief historical introduction some of the central concepts were presented. In particular the core principle of a rigorous assessment of uncertainty and traceability achieved through formal documentation, audits, peer reviews and formal comparisons together with community defined references (preferably SI) was introduced. The concepts of uncertainty and traceability were further expanded on with the introduction of the GUM (Guide to the expression of Uncertainty in Measurement, BIPM 2008 - the foremost authority and guide to the expression and calculation of uncertainty) together with the key concepts of traceability as an unbroken chain linking all processes back to a reference using as complete as possible uncertainty budget calculation. It was pointed out that in the case of Earth Observation the situation may be more ill-defined than for standard metrological measurements due to either ill-conditioned problems (such as found in data assimilation) or just simple lack of knowledge of the complete system and that the GUM, for example, does not cover such cases. There are, however, several current projects that are beginning to look into such issues and will provide some guidance in the future. It was also pointed out that there are still benefits to going through a metrologically robust process since it forces a reassessment of all possible sources of uncertainty including documentary evidence which will reduce any hidden uncertainties due to any previous assumptions made.

The presentation then moved onto defining a consistent vocabulary for uncertainty and made the point that error is not the same thing as uncertainty and both terms have clearly defined meanings. In particular, uncertainty describes the spread of a probability distribution and can be parameterised by values such as the standard deviation. In other words uncertainty is the doubt you have on the value. Error, on the other hand, is a difference from truth which can derive from a range of different processes such as measurement imperfections. Further, when an error is known it can be corrected, though there will always be a residual error which will add to the final uncertainty.

Errors can arise from random and/or systematic effects. Random effects are different for every observation and cannot be corrected, even if the measurement is fully understood. Random effects can, however, have the same associated uncertainty (drawn from the same probability distribution). Note, it is incorrect to use the phrase “random uncertainties” – “uncertainty” describes the probability distribution. Strictly it should be “uncertainties associated with random effects”.

Systematic effects, on the other hand, are errors which can in principle be corrected if the cause of the error is fully understood. In reality, of course, there are many instances where the cause of an error is not fully known or understood giving rise to uncertainties associated with systematic effects. With many systematic effects there can also be an associated time and space scale where a given uncertainty is applicable. Examples include detector degradation where the error changes over time or where the calibration system is compromised such as the solar contamination seen in the Advanced Very High Resolution Radiometer (AVHRR) e.g. Cao et al. (2004). Spatial errors can also exist such as those caused by a geophysical retrieval where the effect of the atmosphere is not fully accounted for. Unfortunately metrology does not have as yet an exact terminology for such effects though the term ‘locally systematic’ has been proposed.

We discussed the different methods that are used to determine uncertainty. These can be separated into two cases known as Type A and Type B methods. Type A use statistical methods applied to a set of measurements either experimentally or numerically derived. Type B methods, on the other hand, use experience and knowledge of the underlying physical processes to derive estimations of the uncertainties. Both methods are considered valid in determining uncertainties.

Finally the question of how to propagate uncertainties was raised. The GUM contains the fundamental equation describing the propagation of uncertainties including the case where the uncertainties are correlated; this can be used directly in many instances. Of course where the underlying problem is ill-defined, as can be the case for data assimilation the GUM itself does not help, but for many Earth Observation problems it can be used easily. An example was given where the calibration of the AVHRR was studied. In this case, taking a numerical approach (a Type A method) combined with an analytical, equation based approach (Type B method) can begin to give real insights into the full uncertainty budget.

Further information on uncertainty propagation is given in Section 2.3.

In related discussions both informally and in group discussions other aspects of taking a metrological approach towards uncertainty were discussed. The use of a consistent vocabulary together with the rigour of a metrologically robust framework under which an uncertainty budget can be developed was considered to be an improvement and something that would help provide extra confidence in SST data and its associated uncertainties. However, it was also felt by many people who use SST data that the use of metrological techniques is not in itself sufficient to make people trust the final results, as non-experts would not be able to independently assess whether anything had been missed in the processing, and that in the end the reputation of the associated production team was perhaps more important. But the combination of a production team with an excellent reputation together with the use of metrological techniques would provide the highest confidence.

<u>Current methods of uncertainty communication in SST products</u>

Uncertainties in SST measurements and analyses are complicated and have different correlations in space and time. Different data providers use different means to communicate this information to users.

Uncertainties in SST CCI products are provided as three components, according to the correlation structure of the underlying errors. These components arise from random, locally systematic and large-scale systematic effects. A total uncertainty is also provided.

Other SST data providers, e.g. the Met Office Hadley Centre (Kennedy et al, 2011a and b), provide uncertainty information via an ensemble of interchangeable realisations of Level 3 and Level 4 products. Covariance matrices are also provided to describe residual uncertainties arising from biases in individual ships and under-sampling of grid boxes.

2.3. Origins and Propagation of Uncertainty

There are a number of potential sources of error in sea surface temperature (SST) measurements, which should be reflected in their associated uncertainties. Some sources of error are common to both satellite and in-situ observations such as calibration, geolocation, software issues, data corruption, instrument noise and degradation. Satellite observations can also be subject to retrieval errors on synoptic scales introduced by atmospheric conditions and errors from imperfect cloud detection. *In situ* data records are subject to human error in reading temperatures, inconsistency in the measurement depth, measurement conditions, logging and transcribing. The instrument scale can limit measurement accuracy, and buoys may wash ashore or become encased by barnacles.

Within the workshop we considered three primary types of uncertainty: originating from errors caused by random effects, errors from locally systematic effects and errors from large-scale systematic effects. These sources of uncertainty are uncorrelated with one another and are applicable to all levels of products. At higher product levels, uncertainties due to sampling effects were also considered. Details of the SST CCI approach are available in the SST CCI Uncertainty Characterisation Report (UCR; http://www.esa-sst-cci.org/sites/default/files/Documents/public/SST_cci%20UCR%20Issue%203%20%282013%2012%2004%29.pdf).

Corrections applied to *in situ* measurements to account for the impact on the climate record of changes in measurement method through time are specified using parameters which are themselves uncertain. This effect is considered large-scale systematic. These corrections are applied to measurements made using general types of measurement method, e.g. a canvas bucket, or engine room intake. Individual ships' bias characteristics differ from these large-scale corrections. There is therefore also a residual uncertainty in ship measurements from this locally-systematic effect.

The full equation for the propagation of uncertainties is shown in (1). The first term in the equation describes the addition of uncertainty terms in quadrature, with the differential describing the sensitivity of the uncertainty to the observation. The second term describes the uncertainty correlation. Where uncertainties are completely uncorrelated this term reduces to zero.

$$\sigma = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=n+1}^{n}\left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right)\operatorname{cov}\left(\sigma_i \sigma_j\right)} \qquad (1)$$

For satellite observations, sources of error in radiance data (Level 1, pre-transformation into the geophysical variable, such as SST) include radiometric noise (resulting in uncertainty from random effects), intermittently determined calibration parameters (locally systematic effects) and errors in emissivity or the spectral response function (large scale systematic effects). Propagation of uncertainties in Level 1 data from random effects through to Level 2 (swath) and Level 3 (gridded) data is illustrated in Figure 1. The top row shows a simulated error field for the 11 and 12 µm channel for the Advanced Along Track Scanning Radiometer (AATSR). The error field is simulated by randomly sampling a Gaussian distribution with a mean of zero and standard deviation of 0.1 K (determined by our knowledge of the noise as a function of the instrument's responsivity to temperature in both channels with reference to the blackbodies and instrument model).

In the second row, these errors are propagated into N2 SST retrievals (using both channels, but only the information from the nadir view) and D2 SST retrievals (using both channels and both nadir and forward views). The error propagation term is the sum of the errors in each channel multiplied by the corresponding channel coefficient from the SST retrieval. The D2 retrieval contains observations from four channels and therefore the propagated random error is larger than for the N2 retrieval. The third row illustrates how these errors propagate into Level 3 data over a 5x5 pixel grid box, with the fourth row showing the corresponding uncertainty field. The errors do not average down, but the uncertainties do reduce as a function of $1/\sqrt{n}$ because they arise from random effects. The uncertainties resulting from random effects are therefore largest where fewer observations are available in a given grid box (for example where cloud partially obscures the box).

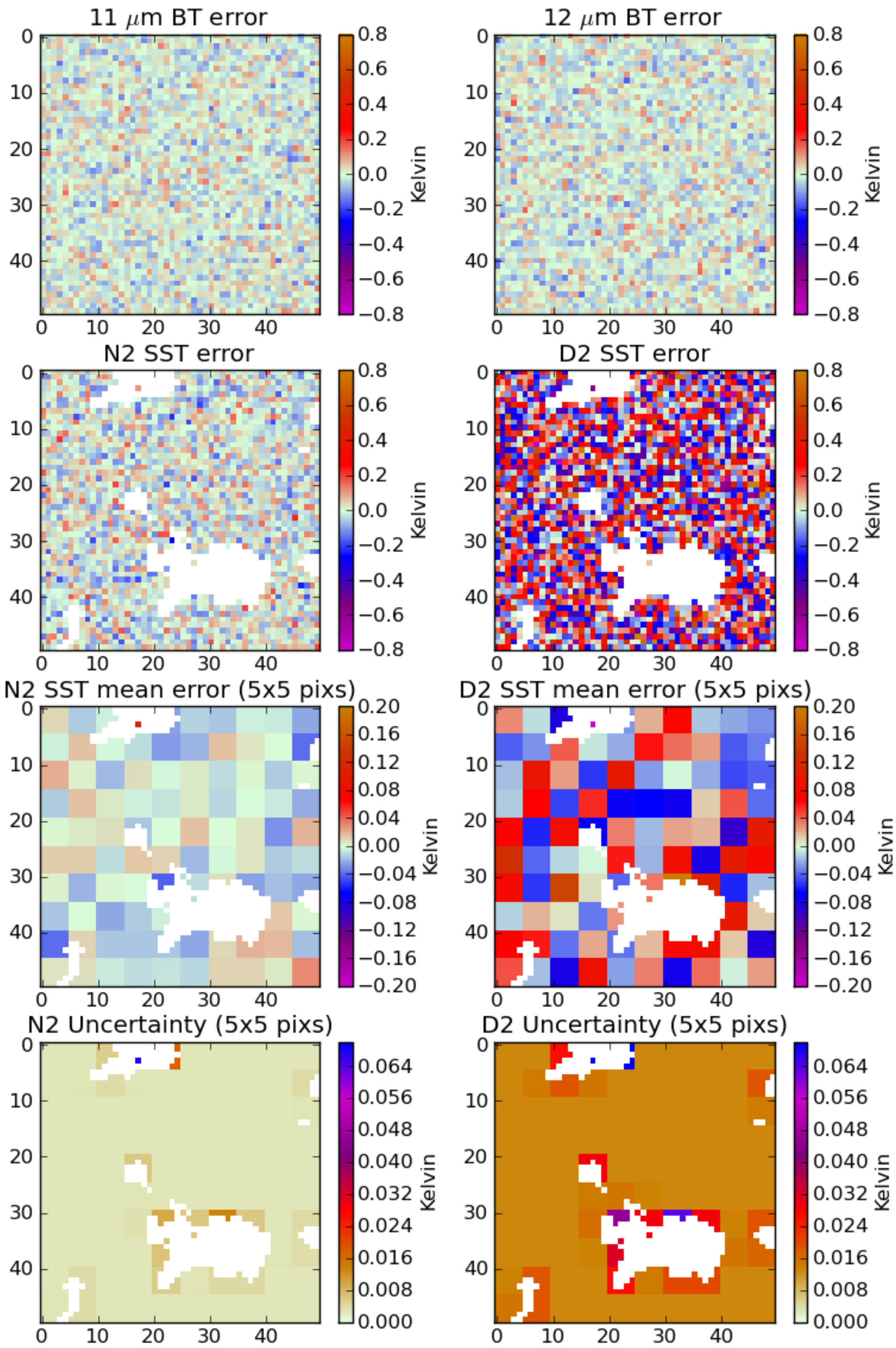


Figure 1: Propagation of error and uncertainty from random effects on brightness temperature (BT, top row) into Level 2 (second row) and Level 3 (third and fourth rows) SST data.

Locally systematic uncertainties in Level 2 data are caused by ambiguities in or limitations to the retrieval scheme. They can be estimated using simulation studies with a true SST and simulated brightness temperatures, e.g. under different atmospheric conditions. These simulated brightness temperatures can then be used to calculate SST using the retrieval algorithm. Comparisons between the true and retrieved SSTs enable characterisation of the algorithm uncertainty.

Production of Level 3 gridded data introduces a further source of uncertainty due to under-sampling. Where a gridded domain is not fully observed (in the case of satellite data, typically due to cloud cover), the mean SST in the observed pixels is likely to deviate from the mean SST that would be calculated if the domain were fully observed. This sampling uncertainty can be modelled for satellite data as a function of the variability in the SSTs observed, the percentage of clear sky pixels and the domain size.

Figure 2 shows the modelled sampling uncertainties. The SST variability in the observed pixels is the main determinant of the shape of this curve. Domain size only becomes important where the SST variability is high and a low percentage of clear-sky pixels is available. Where the percentage of clear-sky pixels is 100%, the sampling uncertainty naturally tends to 0.0 K as all possible observations are available.

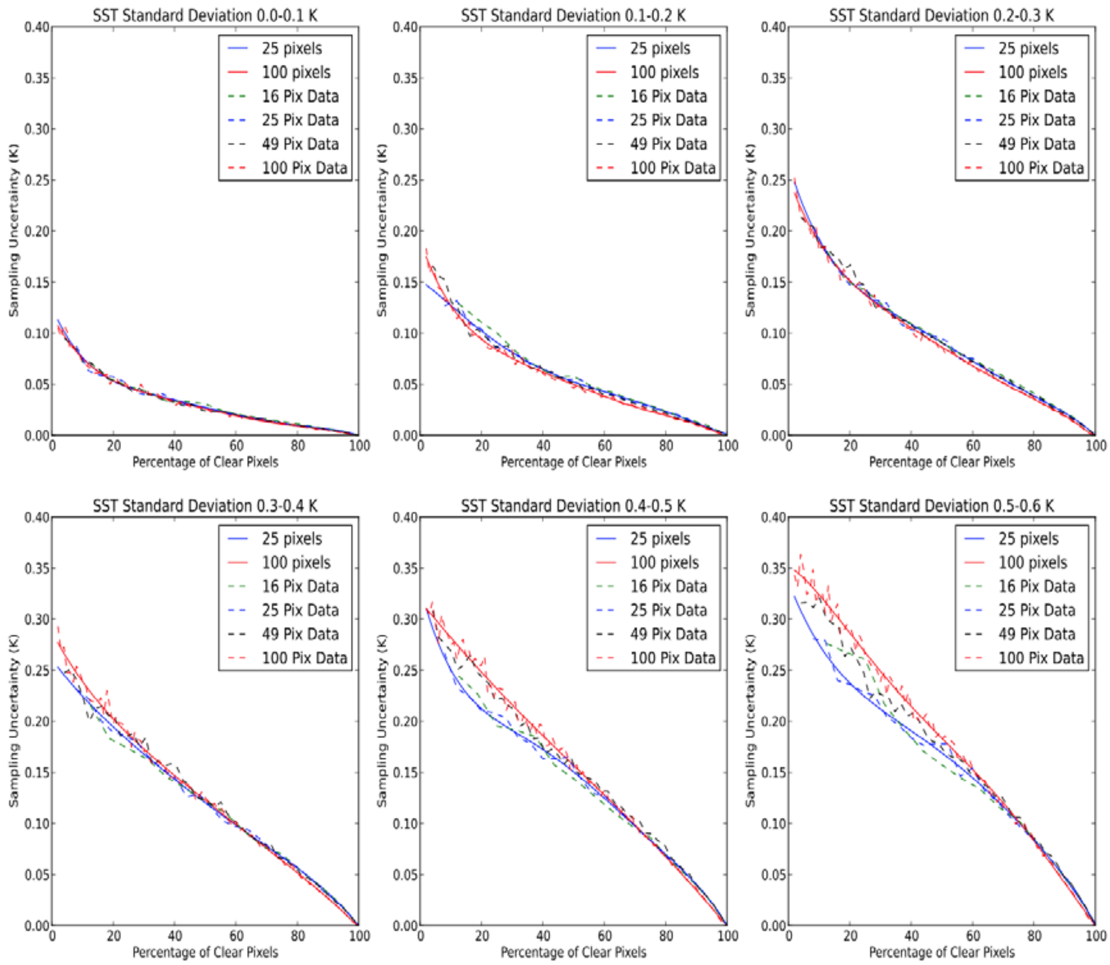


Figure 2: Modelled sampling uncertainties in Level 3 data at two different resolutions: 25 pixels (~0.05 degrees, blue) and 100 pixels (~0.1 degrees, red).

Level 3 data created for the Obs4MIPs archive are monthly, 1-degree resolution products, generated from Level 2 and higher resolution Level 3 datasets. In this case, the propagation of uncertainties is more complicated as the scales averaged over are longer than synoptic time scales and therefore the locally correlated uncertainties average down a certain amount, but not by as much as $1/\sqrt{n}$. The uncertainty propagation needs to take into account the number of constituent grid boxes and synoptic areas and therefore includes the spatial and temporal correlation between each pair of observations contributing to the SST derivation, with reference to the appropriate correlation length scales. However, in order for users of these fields to be able to appropriately propagate these differently correlated uncertainty components through their application, they need to be provided separately (this requires an evolution of the current Obs4MIPS format to accommodate this).

Uncertainties arising from large-scale correlated effects are applicable to each pixel in the SST retrieval over a very large area (e.g. a hemisphere or the globe) and do not average down as they are propagated through product levels.

SST CCI Level 4 analysis products are daily, global, spatially complete blended datasets using Level 2, Level 3 and EUMETSAT OSI-SAF ice concentrations as input. The uncertainty components provided in the Level 2 and Level 3 data are not propagated directly into the SST CCI analysis products, but the total uncertainty is used to weight the observations in the optimal interpolation assimilation. Analysis uncertainties in the Level 4 data are also calculated using optimal interpolation with a background error covariance weighted by the influence of the observations on the analysis.

In other SST analyses, e.g. HadISST.2.1.0.0, analysis uncertainty and its relationship to uncertainties in Level 3 data are propagated via an ensemble.

Practice at propagating uncertainty components was provided in the practical exercises (see Appendix D).

2.4. Validation of Uncertainty

A key development within the SST CCI project is the provision of enhanced uncertainty information for each pixel or cell in every SST CCI product. As discussed above, this enhanced uncertainty information includes estimates of uncertainty components that are uncorrelated between observations, locally correlated on synoptic spatio-temporal scales, and correlated on large scales. This facilitates a more realistic propagation of uncertainty from Level 2/Level 3 products to derivative products (e.g. Level 4 analyses) with coarser averaging. As the uncertainty information constitutes part of the product, it must be validated in its own right.

Validation here is defined as the assessment of the system outputs (i.e. the products) by independent means. So far the SST CCI team have only validated the total uncertainty, through comparisons to reference data from drifters, moorings, radiometers and floats. These measurements are independent from the SST CCI products. Further details can be found in the Product Validation and Intercomparison Report (PVIR; http://www.esa-sst-cci.org/sites/default/files/Documents/public/SST_CCI-PVIR-UoL-001-Issue_1-signed-accepted.pdf).

The approach adopted by the SST CCI team is based on establishing a validation uncertainty budget that considers the uncertainty as a result of errors arising from the comparison to the reference data. The main contributors are:

(1) the uncertainty given in the SST CCI product;
(2) the uncertainty on the reference measurement;
(3) a contribution from errors arising from comparing the reference measurement to the larger spatial average represented by the SST CCI product;
(4) a contribution from any differences in depth between the SST CCI product and the reference measurement; and
(5) a contribution from any time difference between the SST CCI data and the reference measurement.

A presentation demonstrated how term (3) is minimised by using a large set of match-ups and how terms (4) and (5) are minimised using a combined skin effect and diurnal variability model. Consequently, the uncertainty budget for a large dataset of match-ups to drifting buoys for example, can be represented as:

$$\sigma_{SST\_CCI-drifter} = \sqrt{\sigma^2_{SST\_CCI} + \sigma^2_{drifter}}$$

where σ_{SST_CCI} is the uncertainty on the SST CCI product and $\sigma_{drifter}$ is the uncertainty on the drifting buoys, which we estimate to be 0.2 K for the entire dataset. At low satellite uncertainties the standard deviation of the differences is dominated by the uncertainty in the reference data. As you move to higher satellite uncertainties the satellite uncertainty will then dominate, as the reference uncertainty becomes a less significant contribution to the total uncertainty. The other uncertainty contributions are effectively made negligible, as described earlier.

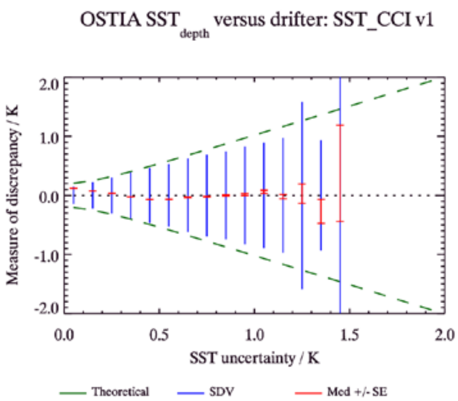


Figure 3: Plot of SST CCI Level 4 analysis product uncertainty against the robust standard deviation of the discrepancies between SST CCI analysis and drifting buoys. The green lines indicated the theoretical dispersion of uncertainties assuming an average drifter buoy measurement uncertainty of 0.2 K. The blue lines indicated the measured dispersion for each uncertainty level. The red lines indicate the standard error for each uncertainty level and also provide an indication of the number of match-ups.

Several results from the uncertainty validation were presented, and one example, for the SST CCI Level 4 analysis product is shown in Figure 3. The spread of uncertainties in Figure 3 is from ~0.05 K to 1.5 K. The agreement between the theoretical and measured RSD values is excellent across the full range of uncertainties. Some divergence is seen for uncertainties above 1.2 K but the increase in spread of the standard error (shown by the red lines) indicates a low number of match-ups at these levels.

The methodology only works where there are independent reference data to directly compare to. However, it is vital that we know that the uncertainties are realistic everywhere. The concept of uncertainty verification was then introduced, which can be used to inform the user where is has (and has not) been possible to verify that the product uncertainties are of the right order of magnitude. The term uncertainty verification and not uncertainty validation is used as the verification may not come from direct comparisons to reference data (validation) alone but may come from additional knowledge. For example, if we know that the uncertainties are realistic in one region from independent comparison then it is reasonable to assume that uncertainties in regions of similar measurement conditions (e.g. equivalent water vapour loading, view angle, aerosol loading, etc.) are also realistic. An example uncertainty verification map is shown in Figure 4.

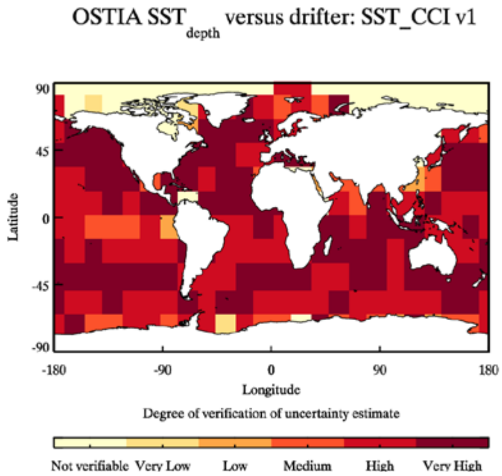


Figure 4: Verification maps for SST CCI analysis (OSTIA) SST_{depth} uncertainties assessed using drifting buoy SST_{depth}. This plot shows the degree to which the SST CCI product uncertainties can be verified using independent reference data. It should not be taken as an indication of SST CCI product data quality and is intended to help the user interpret their own results from using product uncertainties in their application.

The coverage of match-ups shown in Figure 4 is very good with very few unverifiable regions; however, the map covers the entire time period of the V1.0 SST CCI analysis dataset from 1991 to 2010. On average the uncertainties are of high quality compared to the reference dataset and in general regions of medium and low quality occur in areas that contain few drifting buoys. It is important to stress that regions of lower verifiability do not mean the product uncertainties are unrealistic in these areas, it simply means we cannot confirm them independently. Users are always advised to use the product uncertainties everywhere.

The issue of uncertainty validation was discussed during the breakout discussions, where participants were asked to say what would convince them that the SST CCI product uncertainties were realistic. The participants felt the current approach adopted by the SST CCI team is useful. The participants agreed that independent validation of product uncertainties is essential but emphasised that the results from the uncertainty validation, along with their derivation, must be published in peer-reviewed literature. The publication step is extremely important as this provides the confidence to the user that independent experts have reviewed the uncertainties. This is particularly true for a traceability chain for the derivation of the product uncertainties, where users felt they would not necessarily have the expertise to trust the uncertainties based on the chain alone – it needs the extra step of peer review.

Participants felt that the SST CCI project team could provide some examples where the uncertainties have been used to show a demonstrable improvement in an application (e.g. such as in a Level 4 analysis). It was noted that such an approach would, of course, be application dependent (e.g. a particular forecast model), but nevertheless by building up a series of such cases studies users can gain trust in the data. The participants stressed the need to validate product uncertainties at all scales e.g. global/regional/local/coastal, and include the sensitivity to assumptions of scale in any correlations between the sources of error considered when developing the uncertainty budget. Also, it would be useful to provide results from validating uncertainties in regions of known issues affecting satellite SST retrievals, e.g. stratocumulus clouds in the Southern Atlantic.

The participants felt the uncertainty verification maps were useful but noted that they would need to be used in combination with the validation results. Clear advice should be given, for example, on what to do where areas of high uncertainty cannot be validated. It was apparent from the discussion that the verification maps are confusing to some users and that the project team must think carefully how they are communicated in the future. For example, it must be made clear to users that the inability to validate a product uncertainty does not mean it is unrealistic, it is just that we cannot demonstrate its validity - users should always use the uncertainties in the products.

In general the SST CCI team has made a good start in convincing users that the product uncertainties are realistic. Several recommendations detailed in this section for additional information/steps that are needed by users will provide them with all the necessary material they need.

2.5. Current and Ideal Use of Uncertainty Information in Applications

Information on participants' current activities and needs for uncertainty information was gathered from poster and oral presentations and through small group discussions. Some participants already use information on sea surface temperature or other observational uncertainty in their applications:

- Ensemble twentieth century reanalyses are run at ECMWF using an ensemble of SST boundary conditions to allow exploration of the sensitivity of the reanalysis to uncertainty in SST. SST is a key driver of atmospheric reanalyses (*Shoji Hirahara*);
- The impact of observational uncertainty on the detection and attribution of human and natural influences on surface temperature change has been explored (*Andrew Schurer and Gareth Jones*);
- Ensembles of high resolution local analyses of SST in the Bay of Biscay region have been generated in order to explore air-sea interactions there (*Ganix Esnaola Aldanondo*);
- SST information has been assimilated into ocean analyses used to initialise seasonal forecasts and hindcasts. Here estimates of observational and representivity uncertainty are needed in order to appropriately weight each observation (*Drew Peterson*);
- Satellite retrievals and their uncertainties are used to characterise errors in measurements of SST made *in situ*, through direct comparison (*Dave Berry*);
- Assessments of the performance of climate model simulations, through the calculation of ~100 metrics summarising the simulation of different aspects of the climate system, utilises information on observational uncertainty in some cases in order to determine the significance of differences in those metrics as calculated from simulations and observations. This is currently done by comparison to a number of different observational data sets *(Alistair Sellar*);
- Ensembles of gridded fields of global surface temperature have been generated in order to determine the significance of observed changes (*Colin Morice*);
- The sensitivity of ensembles of weather forecasts to forcing by different SST analyses is being assessed (*Martin Lange*);
- Creation of complete, gap-free analyses of SST utilises information on the uncertainty in satellite retrievals of SST (*Andy Harris*);
- Reanalyses of both regional shelf seas over the last few decades and the global ocean over the last century assimilate information on SST and its uncertainties (*David Ford and Chunxue Yang*)

In some cases, participants identified what they *would like* to be able to do now, ignoring any technical challenges they might face:

- Compare model simulations with observations and see whether or not they agree within their uncertainties, taking error covariance into account;
- To be able to use correlated uncertainties properly in in-filled analyses. Analysis methods currently assume uncertainties are uncorrelated between observations;
- To use Level 2 and Level 3 along with their associated uncertainties in conjunction with *in situ* data and their uncertainties, e.g. for construction of infilled analyses;
- Be able to use the uncertainty information in SST data to look at the impact of this on meteorological parameters downstream in Numerical Weather Prediction (NWP);
- Use observational uncertainties in data assimilation to inform background error covariances;
- Examine the uncertainties at each step of a traceability chain for SST;
- Propagate SST uncertainty appropriately into regional averages, taking correlation of errors into account.

2.6. Requirements on Provision of Uncertainty Information

In order to achieve these goals, participants were asked how they would like to receive information on uncertainty in SST observations. They identified the following current requirements. (These have been translated into draft user requirements in Appendix C.):

Timeliness:
- Operationally-available daily SST analyses together with uncertainty estimates for each value (*for generation of ensembles for NWP*).

For other space and time scales:
- Information on uncertainties on specific/larger spatial scales, e.g. for ocean basin averages, and over longer temporal scales (*Hurricane forecasting and detection and attribution*);
- A tool (perhaps web-based) to grid to any spatial/temporal scale (e.g. model grids) with full uncertainty propagation (and ensemble generation) from native resolution. This would include the ability to extract information for specific regions and to extract SST information for different depths;
- Information to develop and use tools to calculate uncertainties at a range of time and space scales. If dataset uncertainty information is described in a standard way, uncertainties can be propagated appropriately automatically.
- Coastal SSTs with uncertainties under all conditions;

For information about the behaviour of uncertainties:
- Full characterisation of uncertainty for construction of in-filled analyses;
- Error covariance matrices (e.g. for data assimilation);
- Error distribution;
- Median estimate, plus uncertainty;
- Consistent treatment/presentation of uncertainty components across ECVs (e.g. for multi-variate data assimilation or model evaluation)

If the correlation structure were captured by an ensemble, users would need:

- A large, easy to use, fully-documented ensemble of sea surface temperature information, which samples the full error model (*detection and attribution and seasonal forecasting*);
- Different ensembles for different uses, e.g. in some applications, the tails of distributions may be more relevant;
- Perhaps, a tool which allows the user to generate their own ensemble.

For historical SST data:
- Uncertainty information for individual SST observations going back to 1800, possibly as an ensemble;
- Quantified uncertainties back as far as possible in the historical record;

For flags:
- Quality flags as a proxy for uncertainty, possibly a tool to create tailored flags for specific users/groups of users. Flags should be well and prominently documented;
- Yes/no type information for some users for whom ensembles may provide too much information;
- Indicator for the source of possible error (e.g. information on uncertainties from clouds to help to distinguish them from fronts);

For documentation:
- Clear documentation to mitigate against users using the data blindly;
- Information content of analysed values, e.g. time of last measurement or percentage coverage;
- Information on how uncertainty estimates were derived and what the contributing factors were;
- The ability to disentangle any retrieval bias from the systematic uncertainty term.

Relating to external bodies:
- Validated uncertainties. More high-quality reference data are needed to allow this;
- Updates to the CF conventions (standard name tables) to provide sufficient vocabulary to describe all uncertainty components adequately.

Some more general requirements were also identified:

- Sufficient information to make an informed choice about which data set to choose, including known limitations (e.g. analysis uncertainties which are assumed uncorrelated, but are not). This needs to be readily accessible, currently one has to drill down into the PUG to get the required information;
- Perhaps a PUG written by users;
- Link to information in other fora, e.g. NCAR climate data guide, GHRSST multi-product ensemble;
- Feedback mechanism, e.g. a forum or discussion group;
- Code repository;
- Provision of SST variability within a given grid cell and its associated uncertainty would be a useful metric

2.7. Reflections on practical experiences using uncertainty information

Workshop participants were encouraged to explore the uncertainty information provided in SST CCI products in practical exercises undertaken during the workshop. Practical exercises were undertaken in smaller groups of up to ten participants with facilitators to help them. Appendix D contains the specification of the exercises. Appendix E contains worked examples of the exercises in the form of ipython notebooks; these were not available to the participants during the workshop.

After the practical activities, the facilitators of the different groups reflected on how their groups of users had found the experience. They reported the following:

Regarding format:

- Some users had difficulties in reading the data (provision of the products as NetCDF4 is a concern for some users) – the project should provide reader tools and more assistance in reading the files;
- The time variable used in the SST CCI product format is fiddly to use (currently compatibility with GHRSST formats is required);
- When calculating 5-day averages, the provision of ATSR data in Level 3 format for every swath was cumbersome, so daily (or day/night) L3C (i.e. collating different swaths) could be useful.

Regarding documentation:

- Level 4 analysis uncertainties are too small when propagated into larger area averages, because no information on their correlation structure is provided. Should users treat these as correlated, or not?;
- Workshop organisers should provide model answers for later so people could see if they get the same answers (see Appendix E);
- An indication of number of observations contributing to each value in the Level 4 analysis would be useful;
- How do uncertainty values in Level 2, Level 3 translate to Level 4? (see Section 2.3);
- More documentation about uncertainties and the full equation for how to propagate them would be useful;
- The project should provide guidance on what we would expect their size to be to guide the user to determine whether or not they are realistic;
- For users not used to using large data sets, we need clear documented examples. E.g. How to take products from archive and use in applications.

Regarding tools:

- The project should provide functions for common types of data manipulation.

Other feedback:

- There was seen to be striping in some Level 4 analysis uncertainty fields, but the range of this is very small and not noticeable except when looking at small regions;
- Level 4 analysis uncertainties are named “analysis error” in the NetCDF files. This terminology is confusing;

These reflections have been translated into draft user requirements in Appendix C.

3. The future

3.1. Future landscape

In order to ensure that recommendations arising from the workshop were future-proofed in some way, participants were asked to think ahead to likely future improvements to their technical infrastructure and methods. They identified the following likely developments in:

Resolution and timeliness:

- Moving towards generating data as close to real time as possible. Data with 1-2 days delay is useful for e.g. seasonal forecasting and the continuation of reanalysis;
- Provision of more diurnal information;
- Perhaps, smaller grids (e.g. 1km) for some applications;

Diverse and better characterized observing system:

- Better information on the accuracy of SST measured by drifting buoys would be used:
    - Ideally point-by-point
    - Including metadata on type of buoy or other characteristics
    - From multiple sensors
- Incorporation of data from citizen measurements (e.g. dive computers) assuming uncertainties can be characterized;
- More diverse *in situ* observations (e.g. Argo, wave gliders) and consequent impact on homogeneity of derived data sets and analyses;
- Looking at cross-consistency between variables/features using uncertainties (eg. SST, ocean colour, sea ice concentration);

More computing power, resulting in the ability to:

- Better treat uncertainty when creating in-filled analyses;
- Deal with uncertainties analytically rather than (or as well as) via an ensemble;
- Send data more easily;
- Process more data:
    - Models will be at higher resolution. Data resolution should match model resolution;
    - Evaluate multiple sources of data;
    - Provide large ensembles;
    - Generate an ensemble on demand;
- Port uncertainty aggregation software from tools currently in Java to e.g. python and make modular to imbed in users' own processing.

Better in-filled analyses will lead to improved quality historical reconstructions.

Data assimilation:

- Coupled reanalysis will occur in next five years;
- Modern data allow improvement of e.g. assimilation methods, leading to improved historical analyses;
- Inclusion of observational error correlation in data assimilation.

NWP:

- Change to ensemble systems;
- Taking diurnal variability in SST into account;
- Ocean/atmosphere coupling;

- Distilling observations from new sensors into super-obs using uncertainty information.

Provision of information:
- Central data provision services and an increased capability for using new datasets and associated uncertainties;
- Annotated products with user comments as to their applicability;
- More interaction between users and data providers and consistent documentation across projects.

3.2. Exploration of Possible Future Developments in Provision of Uncertainty Information

Recurrent themes were identified from participants' statements of requirements. Consequently, we discussed three aspects in more detail, i.e. the provision of: ensembles; error covariance information in other forms and web-based tools.

<u>Provision of ensembles</u>

An ensemble is a potentially convenient way to encapsulate and convey to users information on the sometimes complicated correlation structure of errors. It allows users to explore the impact of these uncertainties on their application in a straightforward way; repeating their analysis multiple times using different ensemble members. A family of different ensembles might be generated, consistently, for different users. Sub-setting of the ensemble for specific regional interests would also need to be facilitated.

The number of ensemble members needed is application-dependent, but is likely to range between 10 and 1000. Producing a large number would allow users to choose a sub-set. However, users need to be guided to choose an unbiased set and not left to make random choices.

In such a set-up there would be:
- A data-producer preferred, central single realisation;
- Randomly-ordered ensemble members so they aren't grouped by, e.g. structural uncertainties in case users do randomly pick ensemble members;
- Perhaps, a sub-setting tool which could choose a sub-set from the larger ensemble.

It was discussed whether or not there should be a best estimate and an ensemble around it, or just the ensemble members. The difficulty is that what is "most likely" might depend on what you're looking at whether it be e.g. trends or variability, etc. It must be noted that the single "best estimate" realisation isn't an ensemble member, because it isn't random.

It is important to start the ensemble generation from Level 1 (radiance) data in order to include the structural uncertainty relating to how to address e.g. inter-satellite calibration. Ensembles would then need to be created on different Levels to suit different applications, i.e.: Level 2 (e.g. for assimilation); Level 3 (e.g. for detection and attribution); Level 4 (e.g. for driving models); and time series over regions from the Level 4 ensemble.

If the means were provided for a user to generate an ensemble, this should only provide the ability to change high-level features such as the output resolution, not to change the choice of the range of parameters used in the fundamental ensemble generation.

Users need: clearly stated underlying assumptions; the ensemble to be on 0.05 degrees latitude by longitude, daily and coarser; an operationally available ensemble (this is technically feasible) or a "best estimate" operationally and the ensemble later. In providing an updating ensemble, however, one needs to bear in mind that updates have to truly belong to the same ensemble; mere statistical compatibility is not sufficient.

Data providers need to be careful to describe the ensemble using correct, non-confusing terminology. For example, are the members equally likely or not?

Provision of covariance information

For some applications, there is a requirement for the provision of uncertainty and reliability information in a form other than an ensemble. For some analysis methods it is necessary (or at least useful) to have access to information about the covariance of errors. Explicitly forming error-covariance matrices can be prohibitively expensive, so it will likely be necessary to find some form of parameterisation that allows the majority of the information to be conveyed in a compact manner. For Level 4 analyses, potentially with a range of spatial scales, this could be a complex operation. In order to guide users through the complexities of using such compact covariance representations, good guidance and good examples would be needed.

Some users also need information regarding the shape of error distributions. In some cases this information will be used to answer standard questions: "are the data close enough to Gaussian for our application?", "Is the distribution symmetric?" However, for other applications, for example assessing extremes of temperature, information about the shape of the distributions will be important: fat-tailed distributions throw up outliers more often than expected. Providing a full pdf for each estimated SST would, as for providing covariance matrices, be expensive in terms of storage. A solution would be to parameterise the distributions using a limited number of functional forms and parameters. For those users not needing such detailed information, a selection of example PDFs could be provided to give users an idea of how the distributions might look under a variety of conditions and why they take the form they do.

With Level 4 analyses, a common concern is whether features that appear in the data are real, or an artefact of the interpolation and data fusion techniques. On the other side of this, is the concern that an absence of a feature, for example an SST front, does not necessarily mean that the feature was absent and might instead mean that the analysis was unable to resolve it due to a lack of data, limitations of the statistical methods or both. One way that users can assess this in current products is to use proxy uncertainty information provided with Level 4 data. For example, estimates of "observational influence", length-scale information (in methods that use adaptive length-scales, providing more detail where data density allows) or, more directly, whether a grid cell contained actual observations and how many there were. However, this kind of ancillary information is not easy to interpret without knowledge of how the analysis techniques work. Guidance on the strengths, limitations and characteristics of the methods is needed, as is guidance on the interpretation of features and their relationship with estimated uncertainty. The same problem also likely affects the interpretation of ensembles.

Provision of web-based tools

There are a number of common tasks that users may want to perform on the datasets generated by the SST CCI (or any other) project. These include sub-setting the data temporally, spatially or by the variables of interest, aggregating the data onto different grids and the corresponding propagation of uncertainty information, and applying flags to the data (for example using a criterion based on the uncertainty in the SSTs). For at least some of these tasks it makes sense to perform this processing before serving the data to users as the volume of sub-setted or aggregated data is lower than that of the full dataset. There is then less strain placed on internet connections and the data processing and storage capabilities of the end user. To achieve this, a web interface would need to be designed that allowed users to define what they would like done to the data. The web interface would need to be connected to a large data store and sufficient computing power to be able to complete the user-specified tasks in a timely manner. It was considered that this is all technically feasible using current technology and there are examples of similar systems already in use or planned.

The required timeliness of the system would depend on the task that was to be performed. For very quick tasks, the web interface should be able to complete the task with minimal delay and display a download link to allow users to download the data. For longer tasks there may be problems with web pages timing out and users may wish to close their web browser before the data were ready. In this case, the system should have the ability to send an email alert when the task is complete. Although it is debatable how long a wait is acceptable for a task to be completed, it was felt that one day would be reasonable but not a month. Certainly, the expectation would be that the system would complete any task faster than a user could themselves. In some cases it would be necessary for a user to be given priority over others. An example of this would be a user participating in a research cruise who cannot afford to wait for their data. It should therefore be made possible to apply for priority status for a particular activity to ensure that there are no delays in cases where there is urgency for data. Another potential issue is that long processing tasks may take up a large proportion of the processing capability of the system, preventing short tasks from completing in a timely manner; short tasks should therefore be given greater priority than longer tasks. The system should give an estimate of the amount of processing and final data quantity so that a user knows what to expect and can modify their choices if they wish.

A way to speed up access to processed data would be to cache frequently-downloaded data and point users to those if they request the same or similar processing. The cache would need to be monitored and any infrequently accessed data deleted. In addition, there would need to be an option to not expose data to public access if the nature of the activity was sensitive, for example for commercial reasons. The system could usefully output example code to read the output data in common programming languages and provide alerts to users if data they have downloaded have been updated. Other issues to consider include which data formats to support. For example, when on a research cruise with limited internet access, output in a very compressed format such as JPEG might be most appropriate. This case of limited internet access should also be considered when designing the web interface.

A useful addition to the web system would be a facility that allowed users to run their own code. This could be done by allowing users to have access to a command prompt in a secure virtual environment and run their own scripts on the data. Alternatively, users could upload functions that were executed by the system when processing each grid cell; these could, for example, be used to generate user defined statistics that are included in the output data files. In this case the user would not need to have any direct interaction with the data themselves nor any knowledge of the environment in which the code is running, which would help make the facility simple to use and secure.

Other functionality that could be considered is the ability to generate match ups between satellite and *in situ* data or to extract/interpolate along a ship track, the ability to upload a specification of an area over which data would be aggregated, and the ability to produce analyses on a user-specified grid and/or with variations to the standard settings of the analysis system. However, in this last case it was recognised that there are significant scientific hurdles to overcome to set up the system to do this and that the system may have to be restricted to running over a limited region because of the processing cost. To extend the usefulness of the tools, they could be designed to run on user's systems and data in addition to the SST CCI web system. The tools could also be designed in a modular way so that users could use individual modules in their own systems rather than just being supplied with an executable program. However, a difficulty is that the tools would need to be able to exploit the different technologies in use on the different systems to give the best performance.

In summary, it has been identified that a web application could provide a wide range of useful functionality to users. While there are plenty of challenges and ideas to consider, there is nothing that is unsolvable. It is also noted that a lot of the functionality that has been identified is not unique to SST data, so, while it would widen the technical challenges considerably, there is the potential to create a powerful system that also serves the users of other ECV data.

3.3. Summary and Next Steps

This workshop provided a two-way exchange of information between SST data providers and users. Common understanding was engendered through presentations about current SST products and their uncertainties, how to talk about and propagate uncertainties and how uncertainties might be validated. Approaches by others to the same problem were also discussed. Presentations, posters and group discussions allowed information to be gathered from users as to their needs for uncertainty information and documentation, both now and in the future.

A draft set of user requirements has been created as a result of this workshop (Appendix C). These will be analysed and incorporated, as appropriate, into the next revision of the SST CCI User Requirements Document, due in late 2015. Thereafter, these User Requirements will inform the development of future versions of the SST CCI and other SST products.

A presentation on the outcomes from this workshop will be made at the ESA Climate Change Initiative Integration meeting (Sweden, May 2015), in order to disseminate this information to teams working on the development of Climate Data Records of other Essential Climate Variables.

A short summary piece has been drafted for the EOS magazine of the American Geophysical Union. Should this be accepted for publication, this will enable the dissemination of these outcomes to a wider audience of data providers and users. Other organisations, such as the Group for High Resolution SST and Obs4MIPS will also be contacted directly.

The workshop organisers wish to thank all the participants most sincerely for giving up their time and for their enthusiastic participation; it is much appreciated.

References

Bureau International des Poids et Mesures (2008), Guide to the Expression of Uncertainty in Measurement (GUM), JCGM 100:2008. Available online at http://www.bipm.org/en/publications/guides/gum.html

Cao et al (2004)

R. Hollmann, C. J. Merchant, R. Saunders, C. Downy, M. Buchwitz, A. Cazenave, E. Chuvieco, P. Defourny, G. de Leeuw, R. Forsberg, T. Holzer-Popp, F. Paul, S. Sandven, S. Sathyendranath, M. van Roozendael, and W. Wagner, 2013: The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables. *Bull. Amer. Meteor. Soc.*, **94**, 1541–1552. doi: http://dx.doi.org/10.1175/BAMS-D-11-00254.1

Kennedy J.J., Rayner, N.A., Smith, R.O., Saunby, M. and Parker, D.E. (2011a). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors. J. Geophys. Res., 116, D14103, doi:10.1029/2010JD015218

Kennedy J.J., Rayner, N.A., Smith, R.O., Saunby, M. and Parker, D.E. (2011b). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation. J. Geophys. Res., 116, D14104, doi:10.1029/2010JD015220

Merchant, C.J. et al (2013), The surface temperatures of Earth: steps towards integrated understanding of variability and change, Geosci. Instrum. Method. Data Syst., 2, 305-321, 2013 doi:10.5194/gi-2-305-2013

Merchant, C. J. and O. Embury (2014), Simulation and inversion of satellite thermal measurements. 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. 489-526. ISBN 9780124170117 doi: 10.1016/B978-0-12-417011-7.00015-5

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S. and 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. doi: 10.1002/gdj3.20

Appendix A: Agenda

Tuesday 18th November

0845-0900 Registration

0900-0930 Welcome addresses:

- Welcome - *Pascal Lecomte, ESA Climate Office, Harwell, UK*
- Welcome – *Nick Rayner, Met Office Hadley Centre*
- SST CCI products and how they relate to other SST products – *Chris Merchant, Science Leader, ESA CCI SST project*

0930-0950 Activity: Where do the uncertainties come from? What can go wrong with SST measurements?
0950-1030 Presentation: Establishing a common vocabulary – *Jon Mittaz, NPL and University of Reading*
1030-1050 Presentation: How uncertainties are currently presented – *Chris Merchant and Nick Rayner*

1050-1130 Coffee

1130-1300 then 1400-1530 Activity: Exploring uncertainties (see handouts for your practical activity groupings)

1300-1400 Lunch

1530-1600 Coffee

1600-1620 Plenary discussion: Feedback from activity

1620-1700 Presentations: Example use cases of SST uncertainty information

- Use of SST ensemble to better understand uncertainty in an estimate of Transient Climate Response – *Andrew Schurer and Gabi Hegerl, University of Edinburgh*
- Applications of sea surface temperature in maritime defence operations - *Martin Veasey, Defence Applications, Met Office*

1700-1730 Presentation: Uncertainty validation – *Gary Corlett, University of Leicester*

1730-1830 Ice breaker reception in Street at the Met Office

Wednesday 19th November

0900-0930 Recap on what we discussed yesterday
0930-1000 Presentation: How do other people present uncertainty information? *Adam Povey, University of Oxford*
1000-1030 Presentation: Propagation of uncertainty information through levels of products – *Claire Bulgin, University of Reading and Emma Fiedler, Met Office*

1030-1100 Coffee

1100-1300 Poster session #1 and #2 (45 mins each session) and general discussion (30 mins): users' applications of SST observations, including statements of extra information needed and any technical challenges that exist to using SST uncertainty information. See handouts for your poster session allocation.

Feedback on current SST CCI User Requirements Document (see http://www.esa-sst-cci.org/?q=webfm_send/46) will also be gathered.

1300-1400 Lunch

1400-1700 Small group discussions to explore specific
questions, e.g.:

- Aims: What users would like to be able to do, ignoring technical challenges
- Proposals: How do you want to receive uncertainty information?

1520-1540 Coffee

- Future proofing: What are likely future improvements in users' technical infrastructure and methods?
- What would convince you that uncertainties provided were realistic?

Thursday 20th November

0930-1030 Recap and Plenary discussion: feedback from group discussions.

1030-1100 Coffee

1100-1215 Different small group discussions to then critically assess:

- Any problems with the ideas generated
- Technical challenges to using uncertainties
- How we could solve these problems

1215-1300 Plenary discussion: specific plans and feedback of morning's recommendations

Close

Appendix B: Participants

Name	Affiliation
Andrew Schurer	University of Edinburgh
Liz Kent	National Oceanography Centre
Andy Harris	University of Maryland
Grit Kirches	Brockmann Consult
Norman Fomferra	Brockmann Consult
David Berry	National Oceanography Centre
Ganix Esnaola Aldanondo	University of the Basque Country
Adam Povey	University of Oxford
Craig Donlon	European Space Agency
Martin Lange	Deutsche Wetterdienst
Jara Imbers	Risk Management Solutions
Finn Lindgren	University of Bath
Tracy Scanlon	National Physical Laboratory
Kate Collingridge	Cefas
Pascal Lecomte	European Space Agency
Wilfried Pokam	University of Yaounde
Chunxue Yang	CMCC
Shoji Hirahara	ECMWF
Alistair Sellar	Met Office
Malcolm Roberts	Met Office
Drew Peterson	Met Office
Richard Renshaw	Met Office
Christoforos Tsamalis	Met Office
Gareth Jones	Met Office
Emma Fiedler	Met Office
David Ford	Met Office
Colin Morice	Met Office
Martin Veasey	Met Office
John Eyre	Met Office
Pat Hyder	Met Office
Alison Dobbin	Risk Management Solutions
Caroline Poulsen	Rutherford Appleton Laboratories

The workshop was facilitated by:

Name	Affiliation
Nick Rayner	Met Office
Chris Merchant	University of Reading
Gary Corlett	University of Leicester
John Kennedy	Met Office
Jon Mittaz	NPL/University of Reading
Claire Bulgin	University of Reading
Simon Good	Met Office
Chris Atkinson	Met Office

Appendix C: Draft User Requirements

SST_CCI-UR-UWU-1	Information on uncertainties on specific/larger spatial scales, e.g. for ocean basin averages, and over longer temporal scales
SST_CCI-UR-UWU-2	A large, easy to use, fully-documented ensemble of sea surface temperature information, which samples the full error model and can be sub-selected according to need
SST_CCI-UR-UWU-3	A tool which allows the user to generate their own ensemble
SST_CCI-UR-UWU-4	A tool (perhaps web-based) to grid to any spatial/temporal scale (e.g. model grids) with full uncertainty propagation (and ensemble generation) from native resolution. This would include the ability to extract information for specific regions and to extract SST information for different depths
SST_CCI-UR-UWU-5	Quality flags as a proxy for uncertainty or a tool to create tailored flags for specific users/groups of users.
SST_CCI-UR-UWU-6	Full characterization and clear documentation of uncertainty, i.e. information on how uncertainty estimates were derived, what the contributing factors were, how to propagate them and what size to expect.
SST_CCI-UR-UWU-7	Quality flags should be well and prominently documented.
SST_CCI-UR-UWU-8	Indicator for the source of possible error (e.g. information on uncertainties from clouds to help to distinguish them from fronts)
SST_CCI-UR-UWU-9	Error covariance matrices
SST_CCI-UR-UWU-10	Error distribution
SST_CCI-UR-UWU-11	Median estimate, plus uncertainty
SST_CCI-UR-UWU-12	Validated uncertainties. (More high-quality reference data are needed.)
SST_CCI-UR-UWU-13	Quantified uncertainties back as far as possible in the historical record.
SST_CCI-UR-UWU-14	Coastal SSTs with uncertainties under all conditions.
SST_CCI-UR-UWU-15	Updates to the CF conventions (standard name tables) to provide sufficient vocabulary to describe all uncertainty components adequately.
SST_CCI-UR-UWU-16	Information content of analysed values, e.g. time of last measurement, number of contributing observations, or percentage coverage.

SST_CCI-UR-UWU-17	Operationally-available daily SST analyses together with uncertainty estimates for each value.
SST_CCI-UR-UWU-18	The ability to disentangle any retrieval bias from the systematic uncertainty term.
SST_CCI-UR-UWU-19	Consistent treatment/presentation of uncertainty components across ECVs.
SST_CCI-UR-UWU-20	Sufficient information, which is easily accessible, to make an informed choice about which data set to choose, including known limitations.
SST_CCI-UR-UWU-21	A PUG written by users.
SST_CCI-UR-UWU-22	Link to information in other fora, e.g. NCEP climate data guide, GHRSST multi-product ensemble.
SST_CCI-UR-UWU-23	Feedback mechanism, e.g. a forum or discussion group.
SST_CCI-UR-UWU-24	Code repository including, e.g. data readers and functions for common data manipulation tasks (in particular, how to use the time variable). Clear documentation of these.
SST_CCI-UR-UWU-25	Provision of SST variability within a given grid cell and its associated uncertainty.
SST_CCI-UR-UWU-26	Model answers to the practical activities.
SST_CCI-UR-UWU-27	Daily (or day/night) L3C products.
SST_CCI-UR-UWU-28	Rename the L4 analysis uncertainties to be more correct.
SST_CCI-UR-UWU-29	Clear documented examples of e.g. how to take products from the archive and use in applications.
SST_CCI-UR-UWU-30	Results from the uncertainty validation, along with their derivation, must be published in peer-reviewed literature.
SST_CCI-UR-UWU-31	Examples where the uncertainties have been used should be provided to show a demonstrable improvement in an application.
SST_CCI-UR-UWU-32	Validate product uncertainties at all scales e.g. global/regional/local/coastal.
SST_CCI-UR-UWU-33	Provide results from validating uncertainties in regions of known issues affecting satellite SST retrievals.
SST_CCI-UR-UWU-34	Clear advice should be given on interpretation of uncertainty verification maps, e.g. what to do where areas of high uncertainty cannot be validated.

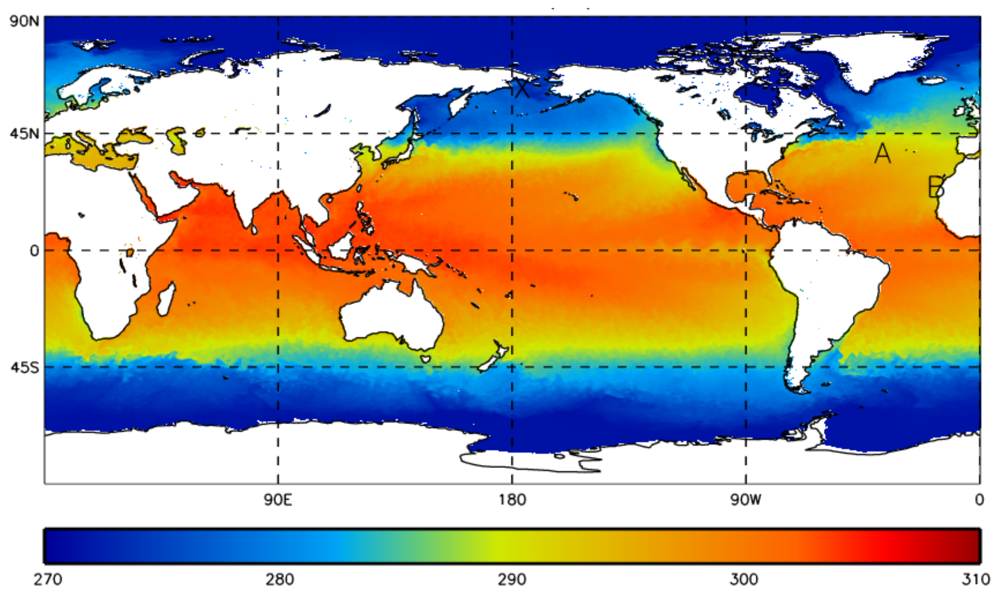
Appendix D: Practical activities

Activity 1: Compare the SST CCI L4 products and see if they agree within their uncertainties

Aim: To familiarise yourself with the uncertainties in ESA SST CCI L4 data and think about their limitations.

Objective: To compare time series of SST from the SST CCI L4 and SST CCI L4 demonstration products and assess whether they agree within their uncertainties.

Data provided: 1° extracts of daily SST CCI L4 fields (on a 0.05° grid) are provided for three different locations – (i) central North Atlantic subtropical gyre; (ii) offshore West Africa; (iii) central Bering Sea. See labels A, B and X in the figure below (the bottom left corner of a label is positioned in the centre of an extracted region). Data cover the period June-August 2007, which is the test period for which the demonstration SST CCI L4 product was produced. Separate data files are provided for each product and for each region (six files in total).



Background:

For each 1° region of interest you have been provided with a time series of data extracted from the SST CCI L4 product and from the SST CCI L4 demonstration product. The demonstration product is produced by the same analysis system as the SST CCI L4 product, but makes use of microwave (MW) SST retrievals in addition to infra red (IR) data. IR frequencies are attenuated by cloud and dust particles, reducing the frequency and accuracy of measurements in regions with high cloud and dust levels. MW measurements have a higher uncertainty, but are much less attenuated by clouds and dust. Observations show the offshore West Africa and Bering Sea regions extracted here to be locations of high probability atmospheric-dust and cloud-cover respectively.

Tasks:

To make the assessment, for each region you should:

i. Read in and unpack the analysed_sst and analysis_error variables for the two L4 products.
ii. Create a time series of SST and uncertainties for each 1° location.
iii. Assess whether the difference between the time series is significant given the uncertainties.

Questions:

- Do the time series differ within their uncertainties? How and why do they differ?

- How would covariance of the uncertainty between grid cells or between products affect your interpretation? Why might this covariance arise?

- Instead of averaging over a 1° area, try comparing time series from a single 0.05° grid cell. How does this affect your conclusions?

- An approximate correlation length scale for the L4 uncertainties might be ~50 km. This is a comparable scale to the 1° area being considered here. If we assume the L4 uncertainties entirely correlated in space for each product, how does this affect your conclusions?

Activity 2: Create and SST index and calculate uncertainties on it using L3U and L4 data and compare

Aim: To familiarise yourself with the uncertainties in ESA SST CCI L3U and L4 data and how they should be combined/compared.

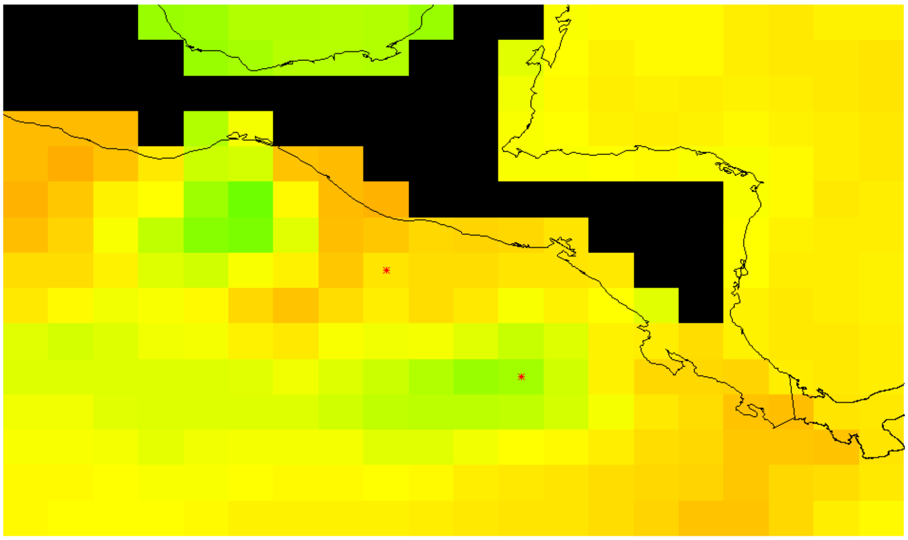
Objective: To create a simple SST index and calculate its uncertainty

Data provided: Re-gridded ATSR V1.1 L3U and analysis V1.0 L4 for 2006 to 2010 for a region covering 100W to 80W and 5N to 20N. The data have been re-gridded to a spatial resolution of 1 degree and a temporal resolution of 5 days.

Background:

You have been provided with re-gridded L3U and L4 data for a region covering 100W to 80W and 5N to 20N for the period 2006 through 2010 in 1 degree pentads. The chosen region is affected by the Papagayo wind that intermittently dramatically cools the SST. Your task is to create an index of the degree of Papagayo cooling by comparing the SST and its uncertainties at two different locations. The extracted L3U data contains an additional uncertainty term "coverage_uncertainty", which is an estimate of the uncertainty in each 1-degree cell due to the variable sampling as a result of clouds in the input infrared imagery.

An example L4 image for the chosen region is given below:



In addition to the re-gridded data we have also provided you with a pre-extracted time series from two locations, one in an area affected by the wind (88.5W/9.5N) and one in an area unaffected (88.5W/9.5N). The two locations are shown using the two dots on the previous image. For the L3U data, the total uncertainty has been added to the time series file along with its component terms.

Tasks:

To create the index the first step is to extract a time series of SST uncertainties. To do this you should:

iv. Read in the image data
v. Select your locations – one from an area affected by the wind and one from an area unaffected by the wind
vi. Extract a time series of SST and uncertainties at each location; for L3U calculate the total uncertainty for each point of the time series

Once you have extracted your time series the next step is to generate the index and its uncertainty. Finally, plot the index for both the L3U and L4 data and compare. You should:

i. Calculate the mean of the time series at each location and generate an anomaly time series for each location
ii. Calculate the index as the difference between the anomalies at each location
iii. Calculate the uncertainty on the index.
iv. The data is plotted to distinguish its positive (affected by the wind) and negative (unaffected) phases.

We strongly advise that you generate your own time series but you are free to use the pre-extracted time series should you wish to do so.

Questions:

- How did you calculate the uncertainty on the index?
- What are the main differences between the L3U and L4 results?
- What information is currently missing in either the L3U or L4 data sets that would have helped with this task?

Activity 3: Create an averaged time series (1 degree and 5-day) from SST CCI L3U data

Aim: To understand how different uncertainty terms should be propagated when averaging data in time and space.

Objective: To create a 5-day average SST time series at 0.2 metre depth for a 1 degree region using L3U data.

Data provided: AATSR v1.1 L3U data (on a 0.05° grid), extracted for the region 30-31°N and 46-45°W in the subtropical North Atlantic. Data are for August 2004. A separate file with data extracted for the whole of 2004 is also provided should participants prefer to calculate a longer time series.

Background:

You have been provided with a time series of L3U data extracted for a 1° region. The SST in each extracted 0.05° grid cell has several different uncertainty terms associated with it that arise due to different kinds of measurement error. These uncertainty terms separate a grid cell's total uncertainty into components which are correlated with other grid cells on different time/space scales. By providing these components separately, this allows us to propagate uncertainty appropriately when aggregating multiple grid cells. This is what we will do here.

The files provided for this exercise are in SST CCI L3U format. These provide both a skin SST (the upper few microns of the sea surface, which is what is measured by infra red instruments) and SST at 0.2 m depth. To get from skin SST to SST at 0.2 m we apply an adjustment to the skin SST measurement. This adjustment has an associated uncertainty. Because we are producing a time series of SST at 0.2 metre depth, we need to make sure we take this uncertainty into account in this activity.

Note that the data files provided for this exercise contain a large amount of missing data. This is because each time slice comprises data from a single AATSR orbit, thus in many cases a time slice contains no data where the AATSR swath did not pass over the region of interest. Data may also be missing due to cloud obscuring the AATSR view. We have purposely chosen an area which is well sampled by the AATSR, and so in this activity we will ignore sampling uncertainty (the uncertainty that arises due to incomplete sampling of an area in space and time). However, this is an additional source of error that a user would need to consider when working with the SST CCI data (note that when regridding SST CCI data using the SST CCI regridder tool, an estimate of sampling uncertainty is provided).

Tasks:

To create a time series of 1-degree and 5-day averages you should:

vii. Read in and unpack the time, sst_dtime, sea_surface_temperature_depth, quality_level, large_scale_correlated_uncertainty, synoptically_correlated_uncertainty, adjustment_uncertainty and uncorrelated_uncertainty variables.

viii. Produce a time variable (hint: add sst_dtime to time to get seconds after 1981-01-01 00:00:00, then subtract 744163200 to convert to seconds since 2004-08-01 00:00:00).

ix. For each 5-day averaging window, find SST obs whose times fall in the averaging window and whose value is not equal to -32768 and whose quality level is equal to 5.

x. Average the obs that meet the above criteria for each 5-day window.

xi. Calculate uncertainties for the average in each 5-day window (hint: this can be done following guidance in Section 7.4 of the SST CCI PUG, which we have provided for this activity. A mean spatial separation, d_{xy}, and mean temporal separation, d_t, will be required for the calculations. You can use d_{xy}=50 km and d_t=2 days).

[To simplify the exercise, you may prefer to calculate a single monthly average with uncertainties for August 2004. In this case average all SST obs in the file whose value is not equal to -32768 and whose quality level is equal to 5, and calculate the uncertainties using d_{xy}=50 km and d_t=6 days]

Questions:

- For 1-degree and 5-day averages, can we distinguish an SST signal at this location given the uncertainty?
- Which term(s) is the dominant source of uncertainty in the averages?
- How might this differ if we averaged over smaller or larger time/space scales?

Appendix E: Worked Examples of Practical Activities

Activity 1 - Compare the SST CCI L4 products and see if they agree within their uncertainties

Set up

Import the modules we need and set graphics to display on the page.

In [1]:

```
import matplotlib.pyplot as plt
from netCDF4 import Dataset
import numpy as np
%matplotlib inline
```

Define a function to read in the data we wish to use.

In [2]:

```
def read_data(region):
    '''Read extracts of the SST CCI level 4 product and the SST CCI demonstration level 4 product.
       Region can be '179_5W58_5N' or '19_5W20_5N' or '40_5W33_5N'.
       The SST and SST uncertainty data from both products are returned.
    '''
    datadir = '/project/sstcci/workshop'
```

```
    file1   = datadir + '/ESA_SST_CCI-uncertainty-workshop-activity-I-L4-' + region + '.nc'
    file2   = datadir + '/ESA_SST_CCI-uncertainty-workshop-activity-I-demoL4-' + region + '.nc'

    nc   = Dataset(file1)
    sst1 = nc.variables['analysed_sst'][:]
    unc1 = nc.variables['analysis_error'][:]
    nc.close()

    nc   = Dataset(file2)
    sst2 = nc.variables['analysed_sst'][:]
    unc2 = nc.variables['analysis_error'][:]
    nc.close()

    return sst1, unc1, sst2, unc2
```

A function to display the four fields read by the read_data function.

In [3]:

```
def plot_data(sst1, unc1, sst2, unc2, timepoint):
    '''Produces a plot of the SST and uncertainty data.
       sst1, unc1, sst2, unc2 are the data read by the read_data function.
       timepoint is the index to plot in the time dimension.
    '''
    fig = plt.figure(figsize=(10, 10))

    ax = fig.add_subplot(2, 2, 1)
    p = plt.imshow(sst1[timepoint, :, :])
    cb = plt.colorbar(p)
    cb.formatter.set_useOffset(False)
    cb.update_ticks()
    cb.set_label('K')
```

```
    ax.set_title('SST1')

    ax = fig.add_subplot(2, 2, 2)
    p = plt.imshow(unc1[timepoint, :, :])
    cb = plt.colorbar(p)
    cb.formatter.set_useOffset(False)
    cb.update_ticks()
    cb.set_label('K')
    plt.gca().set_title('UNC1')

    ax = fig.add_subplot(2, 2, 3)
    p = plt.imshow(sst2[timepoint, :, :])
    cb = plt.colorbar(p)
    cb.formatter.set_useOffset(False)
    cb.update_ticks()
    cb.set_label('K')
    plt.gca().set_title('SST1')

    ax = fig.add_subplot(2, 2, 4)
    p = plt.imshow(unc2[timepoint, :, :])
    cb = plt.colorbar(p)
    cb.formatter.set_useOffset(False)
    cb.update_ticks()
    cb.set_label('K')
    plt.gca().set_title('UNC1')

    plt.show()
```

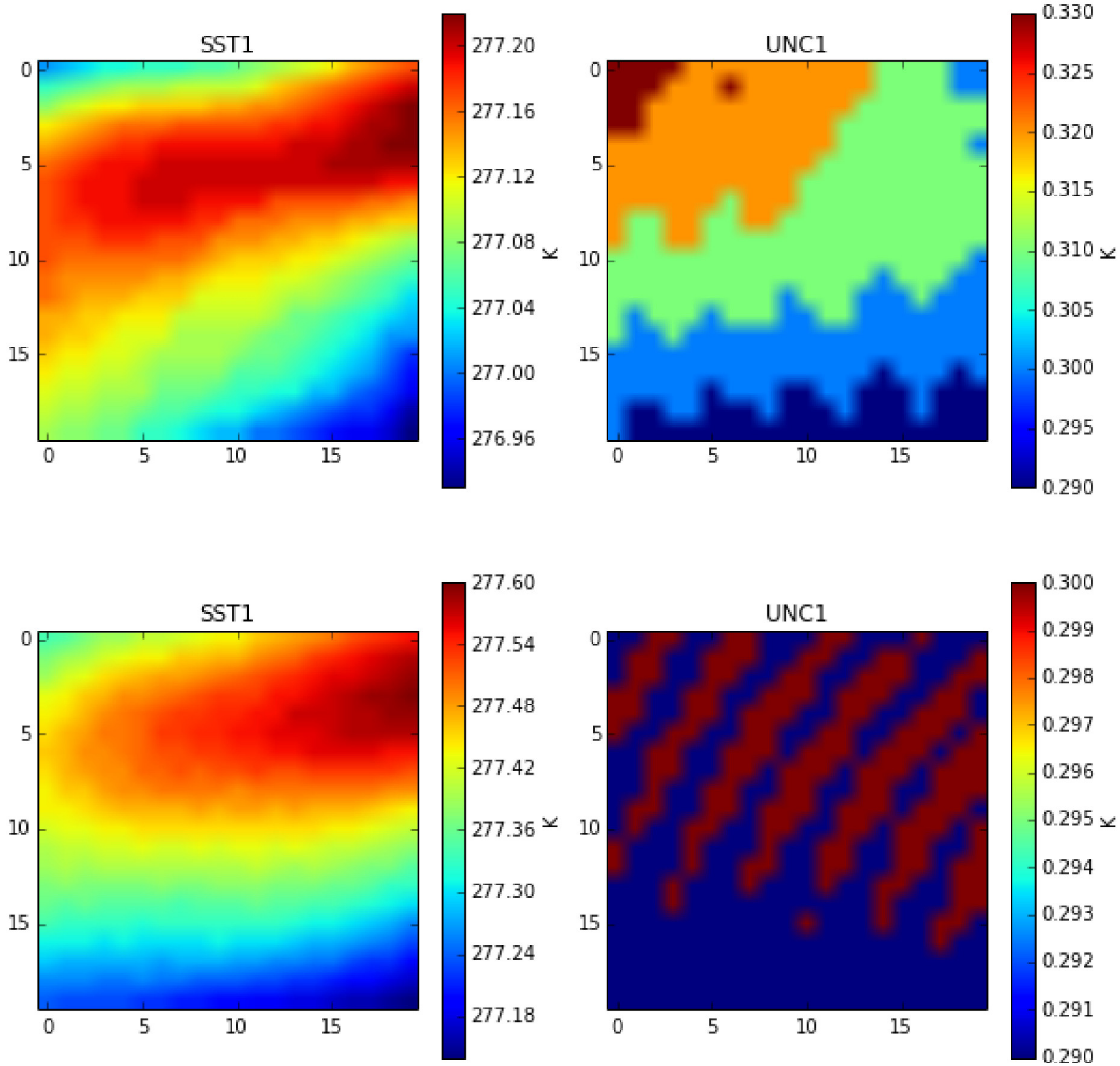
Results

Region: 179_5W58_5N

Read the data and plot.

In [4]:

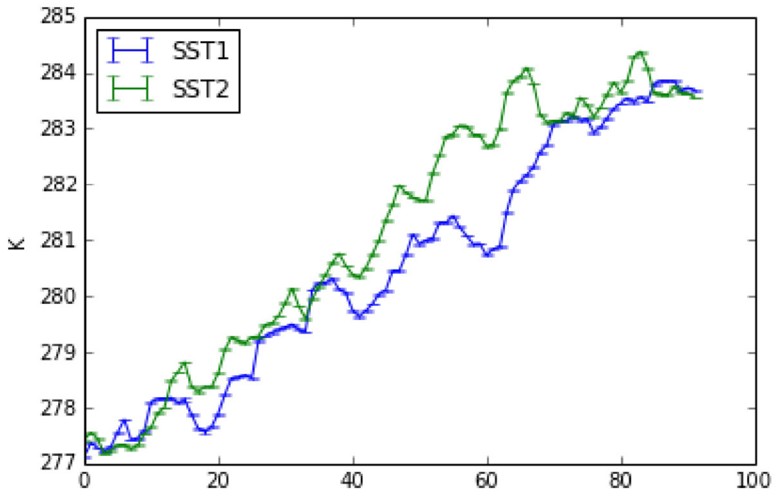
```
sst1, unc1, sst2, unc2 = read_data('179_5W58_5N')
plot_data(sst1, unc1, sst2, unc2, 0)
```


SST1
277.20
277.16
277.12
277.08
277.04
277.00
276.96
K
UNC1
0.330
0.325
0.320
0.315
0.310
0.305
0.300
0.295
0.290
K
SST1
277.60
277.54
277.48
277.42
277.36
277.30
277.24
277.18
K
UNC1
0.300
0.299
0.298
0.297
0.296
0.295
0.294
0.293
0.292
0.291
0.290
K

First look at averages of the images. For efficiency the data are first reshaped into 2-dimensional arrays with time in one dimension. The mean of the SSTs is found. Uncertainties are combined assuming that errors in the SSTs are uncorrelated between grid cells.

In [5]:

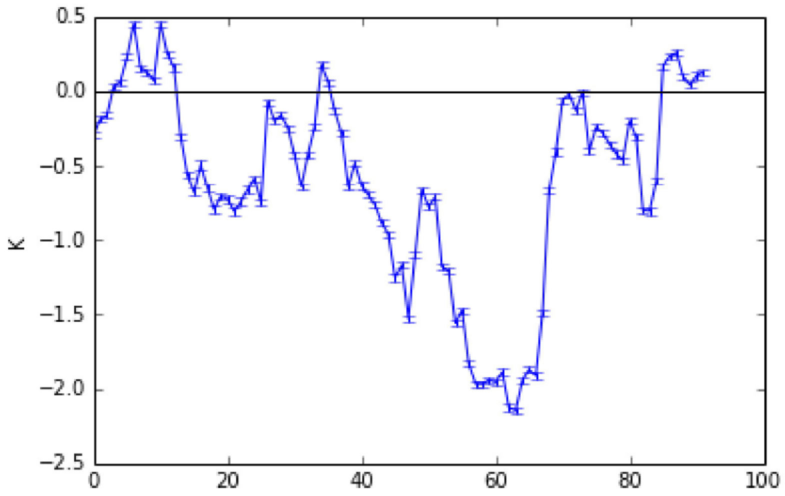
```
# Find the number of time points.
ntimes = sst1.shape[0]

# For computational efficiency we make the arrays 2D with time in one dimension, space in the other.
sst1rs = sst1.reshape(ntimes, -1)
unc1rs = unc1.reshape(ntimes, -1)
sst2rs = sst2.reshape(ntimes, -1)
unc2rs = unc2.reshape(ntimes, -1)
npoints = unc1rs.shape[1]

# Using the reshaped arrays we can generate the time series.
sst1ts = np.mean(sst1rs, axis=1)
unc1ts = np.sqrt(np.sum(unc1rs**2, axis=1)) / npoints
sst2ts = np.mean(sst2rs, axis=1)
unc2ts = np.sqrt(np.sum(unc2rs**2, axis=1)) / npoints

# Plot of the data.
plt.errorbar(range(ntimes), sst1ts, yerr=unc1ts, label='SST1')
plt.errorbar(range(ntimes), sst2ts, yerr=unc2ts, label='SST2')
plt.gca().set_ylabel('K')
plt.legend(loc=2)
plt.show()
```



We are interested in the difference between the two SST time series. The uncertainties are combined in quadrature i.e. we are assuming that errors in the two time series are uncorrelated with each other (this is perhaps ill-advised for this example, however).

In [6]:

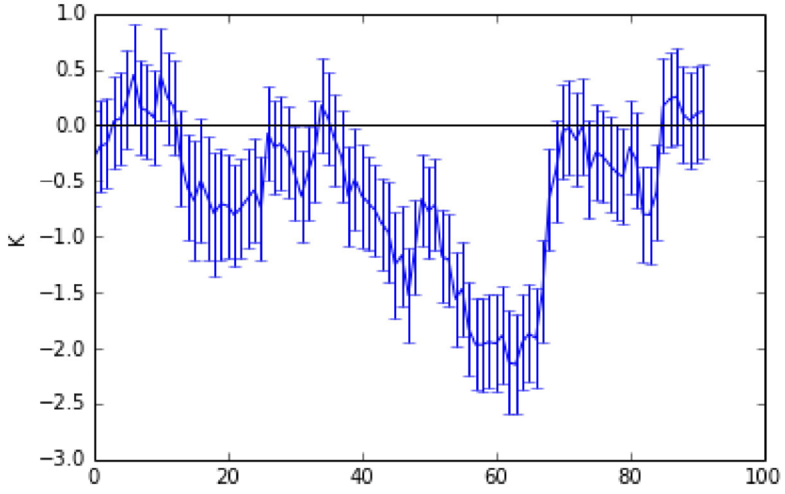
```
sstdiff  = sst1ts - sst2ts
unctotal = np.sqrt(unc1ts**2 + unc2ts**2)

plt.errorbar(range(ntimes), sstdiff, yerr=unctotal)
plt.axhline(color='k')
plt.gca().set_ylabel('K')
plt.show()
```



An alternative assumption is that errors are correlated between grid cells.

In [7]:

```
unc1ts = np.sum(unc1rs, axis=1) / npoints
unc2ts = np.sum(unc2rs, axis=1) / npoints

unctotal = np.sqrt(unc1ts**2 + unc2ts**2)

# Plot.
plt.errorbar(range(ntimes), sstdiff, yerr=unctotal)
plt.axhline(color='k')
plt.gca().set_ylabel('K')
plt.show()
```

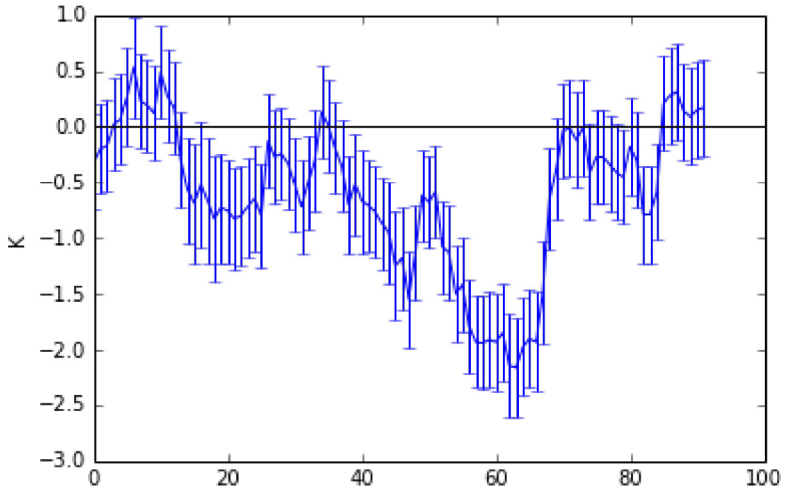


Single pixels can also be compared.

In [8]:

```
# Pixel coordinates.
x = 9
y = 9

# Extract from the data.
sst1ts = sst1[:, x, y]
unc1ts = unc1[:, x, y]
sst2ts = sst2[:, x, y]
unc2ts = unc2[:, x, y]

# Difference.
sstdiff  = sst1ts - sst2ts
```

```
unctotal = np.sqrt(unc1ts**2 + unc2ts**2)

# Plot.
plt.errorbar(range(ntimes), sstdiff, yerr=unctotal)
plt.axhline(color='k')
plt.gca().set_ylabel('K')
plt.show()
```



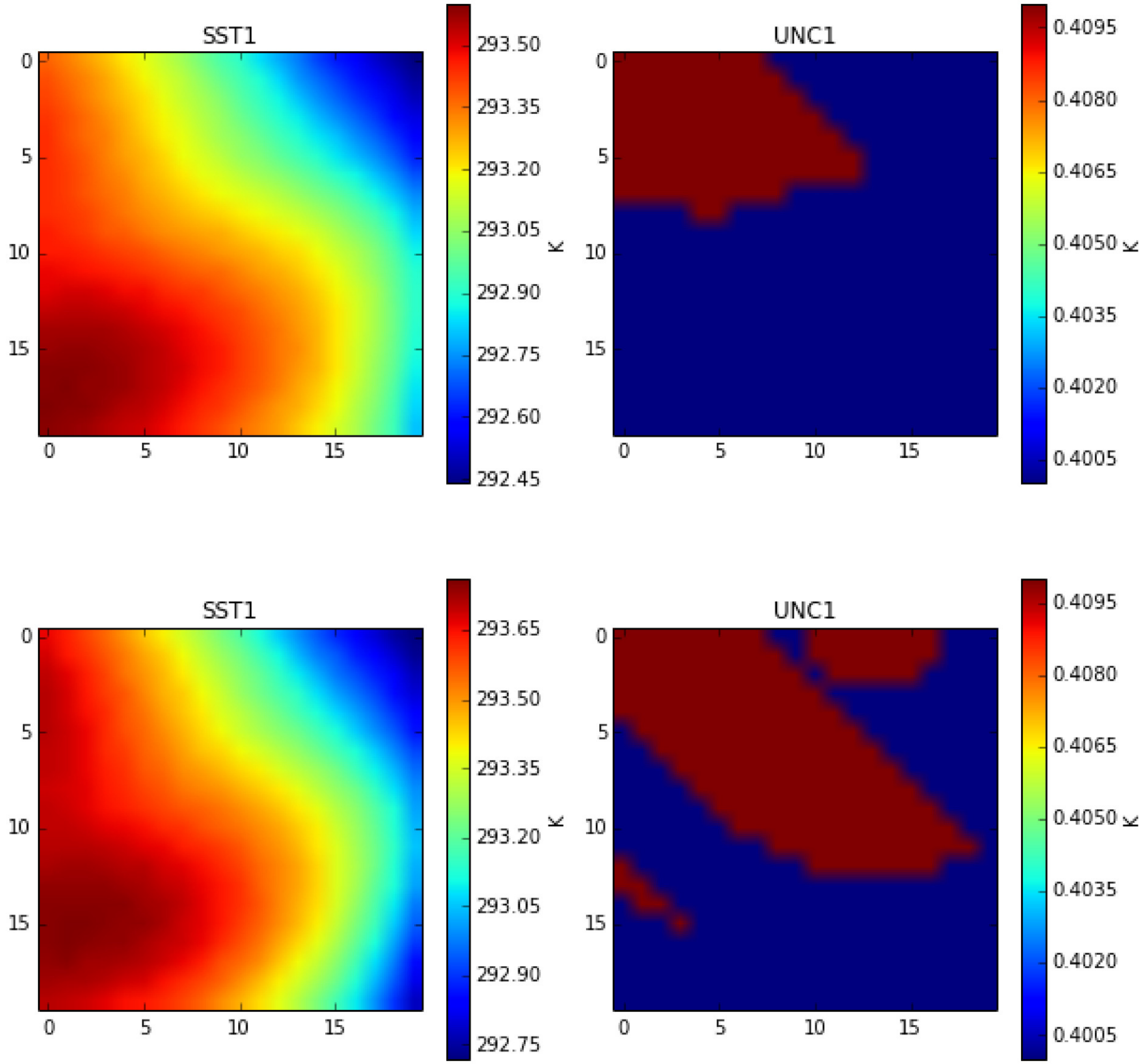
The result is similar to the correlated errors example.

Region: 19_5W20_5N

Results are shown for a single pixel from this region. Working follows the example above.

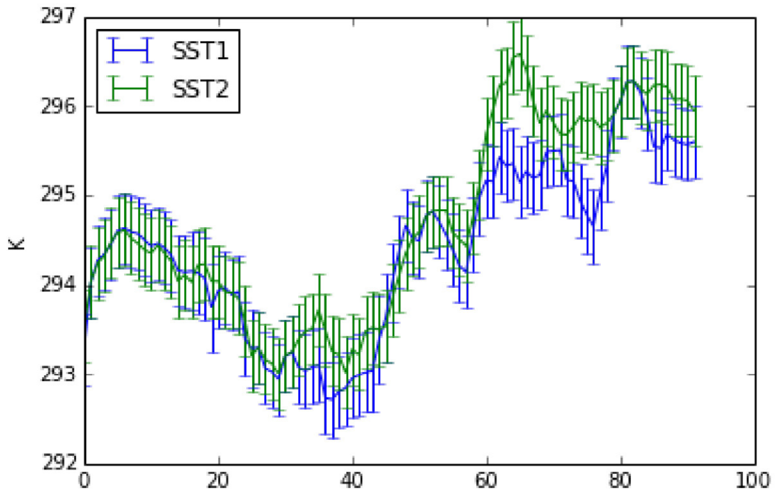
In [9]:

```
sst1, unc1, sst2, unc2 = read_data('19_5W20_5N')
plot_data(sst1, unc1, sst2, unc2, 0)
```


SST1
293.50
293.35
293.20
293.05
292.90
292.75
292.60
292.45
K
UNC1
0.4095
0.4080
0.4065
0.4050
0.4035
0.4020
0.4005
K
SST1
293.65
293.50
293.35
293.20
293.05
292.90
292.75
K
UNC1
0.4095
0.4080
0.4065
0.4050
0.4035
0.4020
0.4005
K

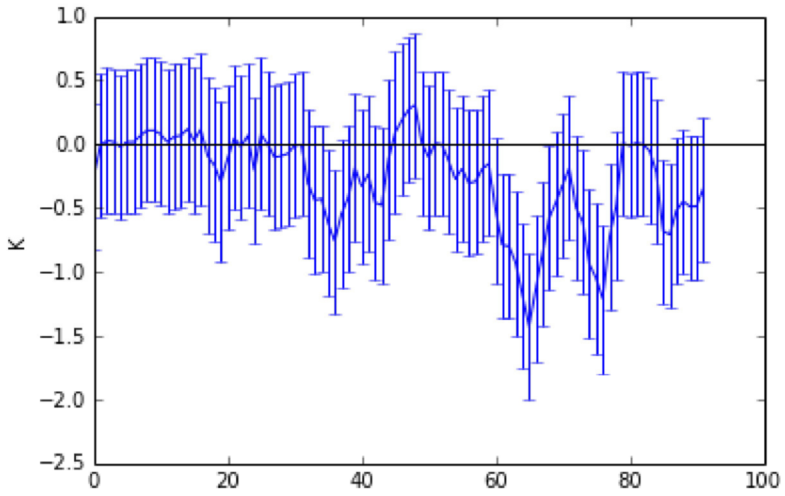
In [10]:

```
# Pixel coordinates.
x = 9
y = 9

# Extract from the data.
sst1ts = sst1[:, x, y]
unc1ts = unc1[:, x, y]
sst2ts = sst2[:, x, y]
unc2ts = unc2[:, x, y]

# Find the number of time points.
ntimes = sst1ts.shape[0]

# Plot of the data.
plt.errorbar(range(ntimes), sst1ts, yerr=unc1ts, label='SST1')
plt.errorbar(range(ntimes), sst2ts, yerr=unc2ts, label='SST2')
plt.gca().set_ylabel('K')
plt.legend(loc=2)
plt.show()
```



In [11]:

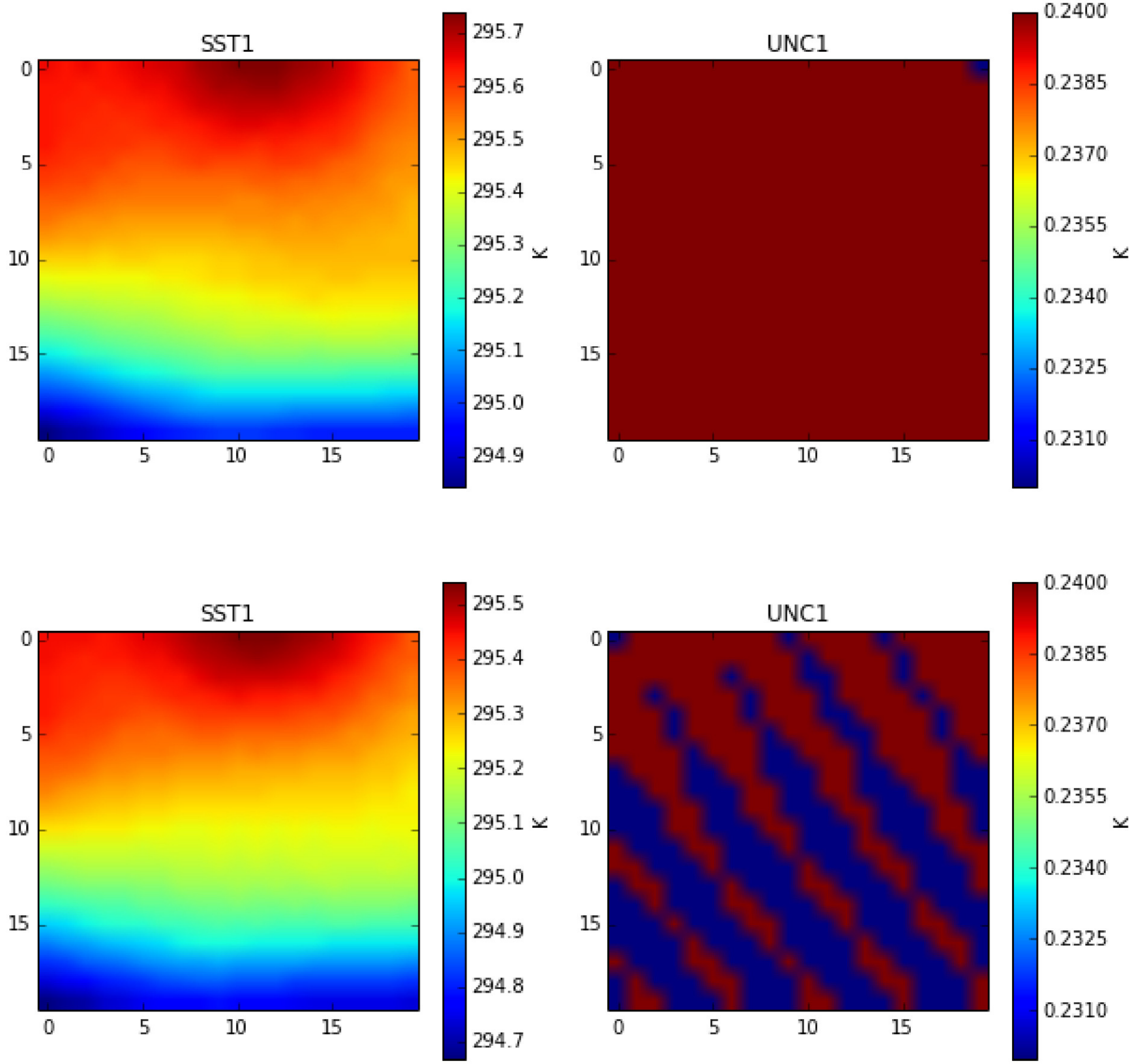
```
sstdiff  = sst1ts - sst2ts
unctotal = np.sqrt(unc1ts**2 + unc2ts**2)

plt.errorbar(range(ntimes), sstdiff, yerr=unctotal)
plt.axhline(color='k')
plt.gca().set_ylabel('K')
plt.show()
```



Region: 40_5W33_5N

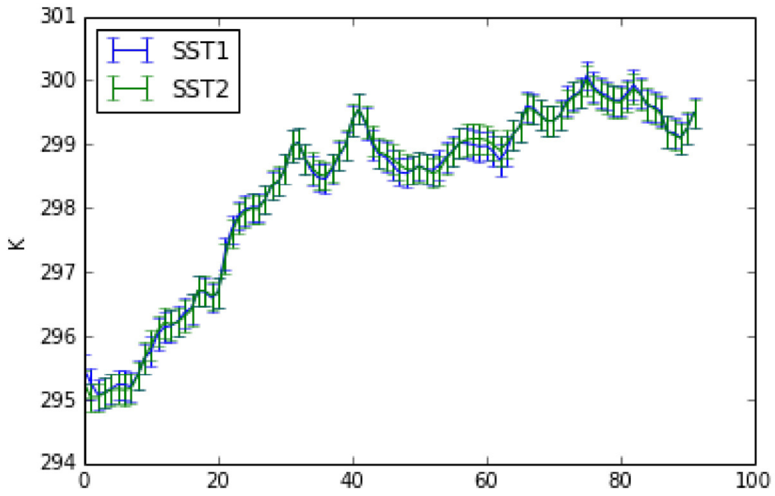
In [12]:

```
sst1, unc1, sst2, unc2 = read_data('40_5W33_5N')
plot_data(sst1, unc1, sst2, unc2, 0)
```


SST1
UNC1
SST1
UNC1
K
K
K
K
295.7
295.6
295.5
295.4
295.3
295.2
295.1
295.0
294.9
294.8
294.7
0.2400
0.2385
0.2370
0.2355
0.2340
0.2325
0.2310
0
5
10
15

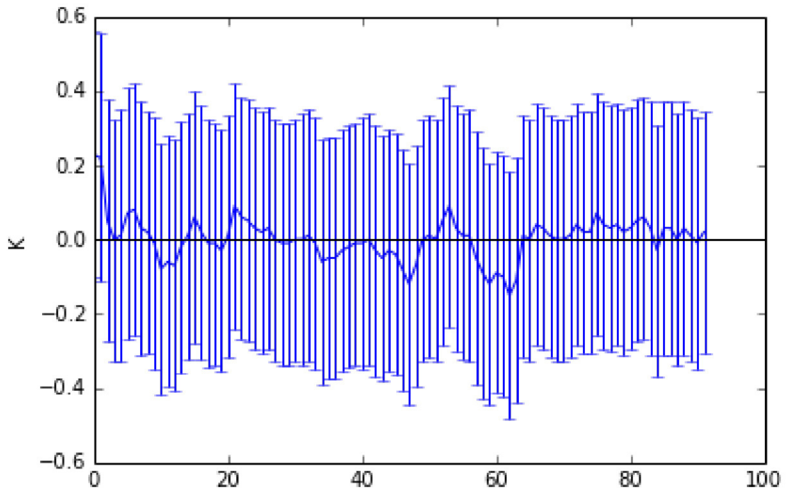
In [13]:

```
# Pixel coordinates.
x = 9
y = 9

# Extract from the data.
sst1ts = sst1[:, x, y]
unc1ts = unc1[:, x, y]
sst2ts = sst2[:, x, y]
unc2ts = unc2[:, x, y]

# Find the number of time points.
ntimes = sst1ts.shape[0]

# Plot of the data.
plt.errorbar(range(ntimes), sst1ts, yerr=unc1ts, label='SST1')
plt.errorbar(range(ntimes), sst2ts, yerr=unc2ts, label='SST2')
plt.gca().set_ylabel('K')
plt.legend(loc=2)
plt.show()
```



In [14]:

```
sstdiff  = sst1ts - sst2ts
unctotal = np.sqrt(unc1ts**2 + unc2ts**2)

plt.errorbar(range(ntimes), sstdiff, yerr=unctotal)
plt.axhline(color='k')
plt.gca().set_ylabel('K')
plt.show()
```



Answers to worksheet questions

- Do the time series differ within their uncertainties? How and why do they differ?

For 1° averages the time series clearly differ given their uncertainties. For the 1° regions centred on 179.5W-58.5N and 19.5W-20.5N we see that the SST CCI L4 product is frequently cooler than the demonstration L4 product. This is likely because the demonstration L4 product incorporates microwave measurements which are less attenuated by cloud and dust. A systematic difference between the two products is less apparent for the 1° region centred on 40.5W-33.5N where there is a lower probability of cloud or dust in the atmosphere.

- How would covariance of the error between grid cells or between products affect your interpretation? Why might this covariance arise?

When aggregating grid cells together, covariance of the error between grid cells will increase the uncertainty in the aggregate. This in turn increases the uncertainty of the difference between the two L4 products, which is calculated by combining the uncertainties of the aggregates in quadrature.

One way in which this covariance might arise is because the satellite SST observations ingested by the L4 products have a component of uncertainty arising from errors that are correlated over synoptic scales.

If covariance of the error exists between the products, this will reduce the uncertainty of the difference between the two L4 products. This is because, if we have two values A and B with uncertainties sig(A) and sig(B), then using the law of propagation of uncertainties, the uncertainty of A-B = SQRT [sig(A)^2 + sig(B)^2 – 2*cov(sig(A),sig(B))]; i.e. the covariance term (cov) reduces the uncertainty of the difference. This covariance might arise because the two L4 products being compared share some input observations and use the same analysis system.

These two un-quantified effects make it difficult to interpret the significance of the differences between the 1° averaged time series.

- Instead of averaging over a 1° area, try comparing time series from a single 0.05° grid cell. How does this affect your conclusions?

If we look at 0.05° extracts, we see that the time series themselves are similar to the time series of the 1° aggregates, but the magnitude of the uncertainties has increased. Despite this increase, for regions 179.5W-58.5N and 19.5W-20.5N, we still see that the SST CCI L4 product is sometimes cooler than the demonstration L4 product given the uncertainties. Because we are no longer aggregating grid cells, we do not need to worry about covariance of the error between grid cells. However, as discussed in the previous question, when looking at the difference between the two time series, the covariance of the error between the two products remains un-quantified which makes interpreting the significance of the differences difficult. However, because this covariance will act to reduce the uncertainty of the difference between the two products, we can consider the uncertainties for the 0.05° extracts to be at their upper limit. This means that the differences we see here are robust.

It is emphasised that were the uncertainties correctly quantified, the two products would agree given their uncertainties. The disagreement we see for regions 179.5W-58.5N and 19.5W-20.5N is likely occurring because the error model does not adequately capture the effects of cloud and dust on the SST analysis.

- An approximate correlation length scale for the L4 errors might be ~50 km. This is a comparable scale to the 1° area being considered here.
  If we assume the L4 errors entirely correlated in space for each product, how does this affect your conclusions?

In this case we get a similar answer as for the 0.05° extracts.

Actvity 2

In this activity we create an index for cooling of the SST by the Papagayo wind

First we setup the python libraries we need

In [11]:
```
from netCDF4 import Dataset
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.dates as dt
```

The first task is to read in the L3U and L4 extracts. There is one file provided for ATSR L3U data and one for analysis L4 data

Define location of data

In [12]:
```
datadir = '/Users/garycorlett/gkc1/SST_CCI/Phase-II/uncertainty_workshop/Data'
l3u_extract_file = datadir + '/ESA-SST_CCI-uncertainty-workshop-actvity-II-L3U-extracts.nc'
l4_extract_file = datadir + '/ESA-SST_CCI-uncertainty-workshop-actvity-II-L4-extracts.nc'
```

First read in L3U data (we are only going to read in the variables we need)

In [13]:
```
nc_l3u = Dataset(l3u_extract_file)
# Time
time_l3u = nc_l3u.variables['time'][:]
# SST - read 20 cm depth
sst_l3u = nc_l3u.variables['sst_depth_20'][:,:,:]
# Uncertainty due to incompelte spatial coverage of ATSR L3U data
cov_l3u = nc_l3u.variables['coverage_uncertainty'][:,:,:]
# Uncertainty due to uncorrelated (random) effects
unc_l3u = nc_l3u.variables['uncorrelated_uncertainty'][:,:,:]
# Uncertainty due to large scale correlated effects
cor_l3u = nc_l3u.variables['large_scale_correlated_uncertainty'][:,:,:]
# Uncertainty due to locally (synoptically) correlated effetcs
syn_l3u = nc_l3u.variables['synoptically_correlated_uncertainty'][:,:,:]
# Uncertainty from errors in diurnal adjusment model
adj_l3u = nc_l3u.variables['adjustment_uncertainty'][:,:,:]
nc_l3u.close()
```

Now read in the L4 data

In [14]:
```
nc_l4 = Dataset(l4_extract_file)
# Time
time_l4 = nc_l4.variables['time'][:]
# SST - 20 cm depth
sst_l4 = nc_l4.variables['analysed_sst'][:,:,:]
# Total uncertainty - incorrectly named as following GHRSST guidelines
tot_l4 = nc_l4.variables['analysis_error'][:,:,:]
nc_l4.close()
```

Now extract time series at chosen location. We are going to select locations at 12.5N/91.5W and 9.5N/88.5W. As we know the extracts are from 5N to 20N in latitude and 100W to 80W in longitude at 1-degree resolution we can simply select the locations by selecting their indices as [10,11] and [7,8], respectively. Note: Data are stored as [time, lat, lon] so the first index is for latitude and the second as longitude.

In [16]:

```
# Extract data at chosen locations
l3u_ssts_1 = sst_l3u[:,7,8]
l3u_ssts_2 = sst_l3u[:,10,11]

l4_ssts_1 = sst_l4[:,7,8]
l4_ssts_2 = sst_l4[:,10,11]

# For L3U we need to calculate the total uncertainty from the provided component elements
l3u_tot_1 = np.sqrt(cov_l3u[:,7,8]**2 + unc_l3u[:,7,8]**2 + cor_l3u[:,7,8]**2 + syn_l3u[:,7,8]**
2 + adj_l3u[:,7,8]**2)
l3u_tot_2 = np.sqrt(cov_l3u[:,10,11]**2 + unc_l3u[:,10,11]**2 + cor_l3u[:,10,11]**2 + syn_l3u[:,
10,11]**2 + adj_l3u[:,10,11]**2)

l4_tot_1 = tot_l4[:,7,8]
l4_tot_2 = tot_l4[:,10,11]
```

One we extracted out time series we can now create an index by differencing the SST anomaly at each location. To calculate the anomaly we subtract the mean SST at each location. We also calculate the uncertainty on the index as the uncertainty on the difference between two SSTs so we simply propogate the uncerainties in quadrature.

In [17]:

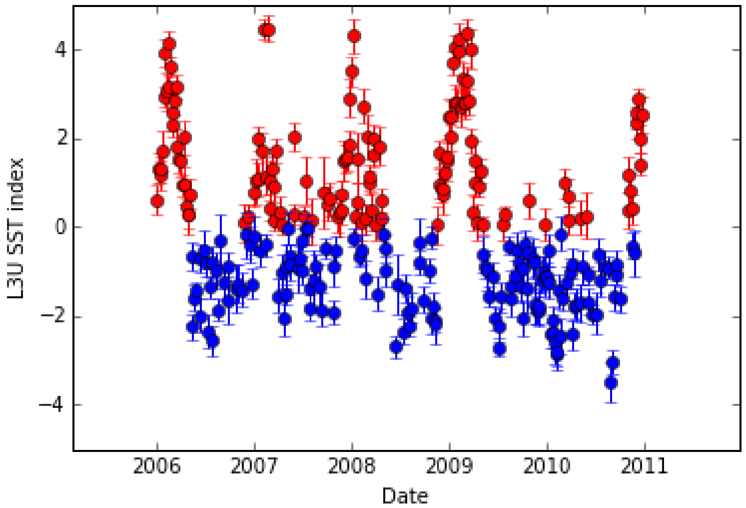
```
# First calculate means
mean1_l3u = np.mean(l3u_ssts_1)
mean2_l3u = np.mean(l3u_ssts_2)

mean1_l4 = np.mean(l4_ssts_1)
mean2_l4 = np.mean(l4_ssts_2)

# Now calculate index
diff_l3u = (l3u_ssts_1 - mean1_l3u) - (l3u_ssts_2 - mean2_l3u)

diff_l4 = (l4_ssts_1 - mean1_l4) - (l4_ssts_2 - mean2_l4)

# Finally, calculate uncertainty on index
tot_unc_l3u = np.sqrt(l3u_tot_1**2 + l3u_tot_2**2)

tot_unc_l4 = np.sqrt(l4_tot_1**2 + l4_tot_2**2)
```

Now we can plot the index - for simplcity we plot the negtive and positive phases of the index indifferent colours. We represent the uncertainty at each location as an error bar (for ease of plotting).

In [18]:
```
# Convert times as seconds from 00:00:00 01/01/1981 to Matplotlib format
epoch = dt.date2num(datetime.datetime( 1981, 1, 1))
l3u_dates = (time_l3u / 86400.0) + epoch
l4_dates = (time_l4 / 86400.0) + epoch
```

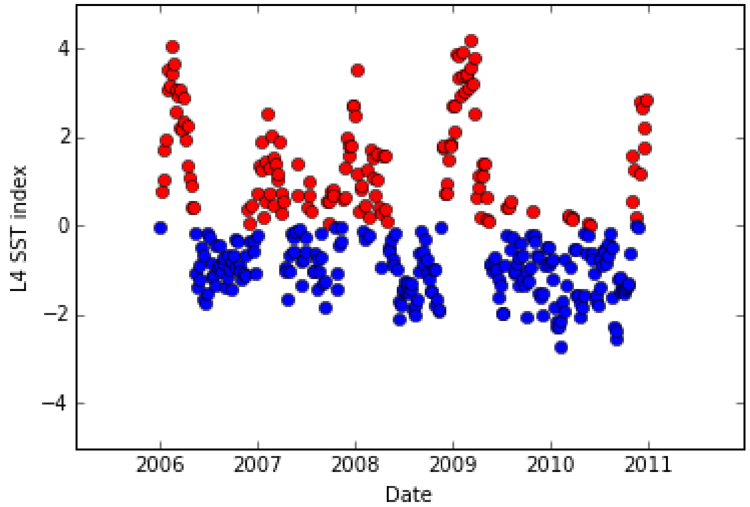
First, the L3U results:

In [19]:
```
# Create a mask for values greater than zero and values less than or equal to  zero
ind_pos = np.where(diff_l3u > 0.0)
ind_neg = np.where(diff_l3u <= 0.0)
# Plot using error bars to show magnitude of uncertainty, postive phase in red, negative phase i
n blue
fig = plt.figure()
plt.ylim(-5,5)
plt.xlabel('Date')
plt.ylabel('L3U SST index')
ax = fig.add_subplot(1, 1, 1)
ax.errorbar(l3u_dates[ind_pos], diff_l3u[ind_pos], yerr=tot_unc_l3u[ind_pos], color='red', fmt =
 'o')
ax.errorbar(l3u_dates[ind_neg], diff_l3u[ind_neg], yerr=tot_unc_l3u[ind_neg], color='blue', fmt
= 'o')
ax.xaxis_date()
plt.show()
```



Next, the L4 results:

In [20]:
```
# Create a mask for values greater than zero and values less than or equal to  zero
ind_pos = np.where(diff_l4 > 0.0)
ind_neg = np.where(diff_l4 <= 0.0)
# Plot using error bars to show magnitude of uncertainty, postive phase in red, negative phase i
n blue
fig = plt.figure()
plt.ylim(-5,5)
plt.xlabel('Date')
plt.ylabel('L4 SST index')
ax = fig.add_subplot(1, 1, 1)
ax.errorbar(l4_dates[ind_pos], diff_l4[ind_pos], yerr=tot_unc_l4[ind_pos], color='red', fmt = 'o
')
ax.errorbar(l4_dates[ind_neg], diff_l4[ind_neg], yerr=tot_unc_l4[ind_neg], color='blue', fmt = '
o')
ax.xaxis_date()
plt.show()
```



Answers to questions

Question 1: How did you calculate the uncertainty on the index?

To calculate the uncertianty on the index you add the uncertainties from each location in quadrature (you are taking the difference between SSTs as the index).

Question 2: What are the main differences between the L3U and L4 results?

The indexes are actually fairly consistent between the L3U and L4 results, including the non-observation of a positive phase at the start of 2010. The most noticeable difference is the order of magnitude of the uncertainties.

Question 3: What information is currently missing in either the L3U or L4 data sets that would have helped with this task?

The difference between the uncertainties is due to the fact that the L4 uncertainty (or analysis_error as it is named) is wrongly assumed to be random and reduces as part of the re-gridding to 1 degree. The L4 requires more detail on correlation lengths as given in the L3U products.

Activity 3 - Create an averaged time series (1 degree and 5-day) from SST CCI L3U data

Set up

Import the modules we need and set the graphics to appear inline.

In [1]:

```
import matplotlib.pyplot as plt
from netCDF4 import Dataset
import numpy as np
%matplotlib inline
```

Define a data reader.

In [2]:

```
def read_data(all2004=False):
    if all2004:
        filename = '/project/sstcci/workshop/ESA_SST_CCI-uncertainty-workshop-activity-III-L3u-45_5W30_5N-200
4.nc'
        reftime  = 725760000
    else:
```

```
        filename = '/project/sstcci/workshop/ESA_SST_CCI-uncertainty-workshop-activity-III-L3u-45_5W30_5N-Aug2
004.nc'
        reftime  = 744163200

    data = {} # Dictionary to hold the data.

    nc = Dataset(filename)

    # Combine the time and sst_dtime variables to find the time of each SST relative to a reference time.
    # The reference time corresponds to the beginning of the time period being examined in the units used in t
he netCDF file.
    time         = nc.variables['time'][:]
    dtime        = nc.variables['sst_dtime'][:]
    data['time'] = time[:, np.newaxis, np.newaxis] + dtime[:, :, :] - reftime

    # Other variables are extracted and stored in the dictionary as they are.
    data['sst']       = nc.variables['sea_surface_temperature_depth'][:]
    data['unc_lsc']   = nc.variables['large_scale_correlated_uncertainty'][:]
    data['unc_sc']    = nc.variables['synoptically_correlated_uncertainty'][:]
    data['unc_u']     = nc.variables['uncorrelated_uncertainty'][:]
    data['unc_a']     = nc.variables['adjustment_uncertainty'][:]

    # Ensure that all the variables are masked consistently and that only quality level 5 SSTs are used.
    quality = nc.variables['quality_level'][:]
    mask = (data['sst'].mask | data['unc_lsc'].mask | data['unc_sc'].mask |
            data['unc_u'].mask | data['unc_a'].mask | quality != 5)
    data['sst'].mask     = mask
    data['unc_lsc'].mask = mask
    data['unc_sc'].mask  = mask
    data['unc_u'].mask   = mask
    data['unc_a'].mask   = mask

    nc.close()

    return data
```

Analysis - August 2004

Read in the data.

In [3]:

```
data = read_data()
```

The data are to be divided into five day windows. Each window of data is analysed to find the average SST and its uncertainty.

In [4]:

```
windowsize = 5.0 * 24.0 * 60.0 * 60.0 # 5 days in seconds.
nwindows  = int(np.ceil(np.max(data['time']) / windowsize))
print('Number of windows is %i' % nwindows)
```

```
Number of windows is 6
```

Create arrays to hold the aggregated results.

In [5]:

```
asst     = np.ma.zeros(nwindows)
aunc_lsc = np.ma.zeros(nwindows)
aunc_sc  = np.ma.zeros(nwindows)
aunc_u   = np.ma.zeros(nwindows)
aunc_a   = np.ma.zeros(nwindows)
asst.mask     = True
aunc_lsc.mask = True
aunc_sc.mask  = True
aunc_u.mask   = True
aunc_a.mask   = True
```

Loop through the windows aggregating the SSTs and uncertainties that fall into each.

In [6]:

```
mean_dist = 50.0
mean_time = 2.0

for iwindow in range(nwindows):
    timemin  = iwindow * windowsize
    timemax  = (iwindow + 1) * windowsize
    inwindow = (data['time'] >= timemin) & (data['time'] < timemax)

    ninwindow     = np.count_nonzero(data['sst'].mask[inwindow] == False)

    if ninwindow == 0: continue

    n_synop_areas = ninwindow / (1.0 + np.exp(-0.5 * (mean_dist / 100.0 + mean_time))*(ninwindow - 1))

    asst[iwindow]     = np.sum(data['sst'][inwindow]) / ninwindow # I.e. the mean.
    aunc_lsc[iwindow] = np.sum(data['unc_lsc'][inwindow]) / ninwindow
    aunc_sc[iwindow]  = np.sqrt(np.sum(data['unc_sc'][inwindow]**2) / (ninwindow * n_synop_areas))
```

```
    aunc_u[iwindow]    = np.sqrt(np.sum(data['unc_u'][inwindow]**2)) / ninwindow
    aunc_a[iwindow]    = np.sqrt(np.sum(data['unc_a'][inwindow]**2) / (ninwindow * n_synop_areas))
```

Total uncertainty is obtained by combining the uncertainty components in quadrature.

In [7]:

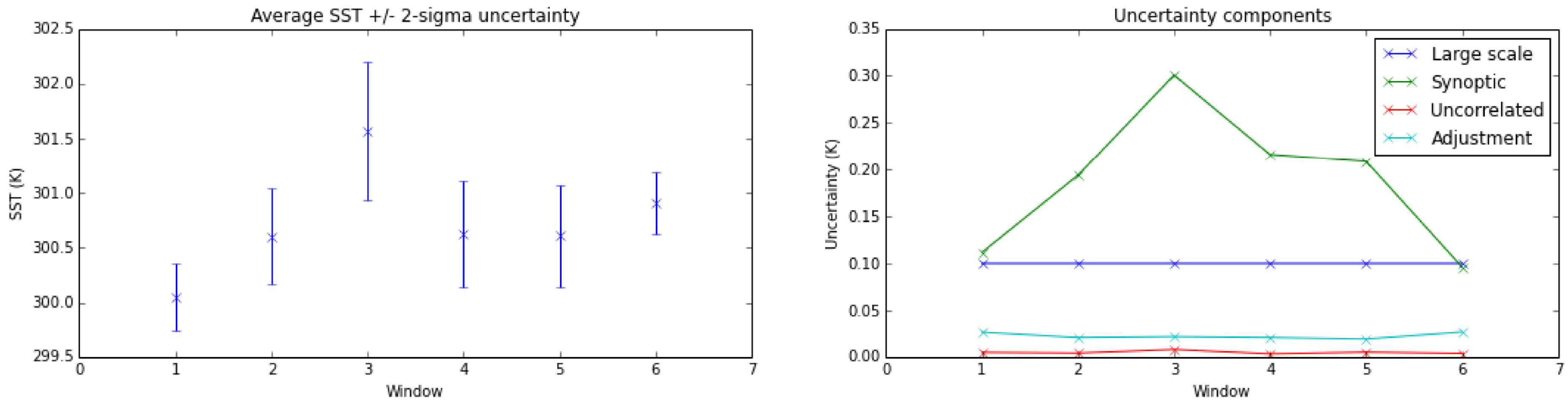
```
aunc_total = np.sqrt(aunc_lsc**2 + aunc_sc**2 + aunc_u**2 + aunc_a**2)
```

The results of the calculations are plotted below.

In [8]:

```
xaxis = np.arange(nwindows) + 1

fig = plt.figure(figsize=(18, 4))

# Plot of SST and total uncertainty.
fig.add_subplot(1, 2, 1)
use = asst.mask == False # Avoids a warning when errorbar is used below.
plt.errorbar(xaxis[use], asst[use], marker='x', ls='None', yerr=2*aunc_total[use])
plt.gca().set_xlim(0, nwindows + 1)
plt.gca().set_xlabel('Window')
plt.gca().set_ylabel('SST (K)')
plt.gca().set_title('Average SST +/- 2-sigma uncertainty')

# Plot of uncertainty components.
fig.add_subplot(1, 2, 2)
plt.plot(xaxis, aunc_lsc, 'x-', label='Large scale')
plt.plot(xaxis, aunc_sc, 'x-', label='Synoptic')
```

```
plt.plot(xaxis, aunc_u, 'x-', label='Uncorrelated')
plt.plot(xaxis, aunc_a, 'x-', label='Adjustment')
plt.gca().set_xlim(0, nwindows + 1)
plt.gca().set_xlabel('Window')
plt.gca().set_ylabel('Uncertainty (K)')
plt.gca().set_title('Uncertainty components')
plt.legend()

plt.show()
```



ANSWERS TO WORKSHEET QUESTIONS: 1. If we plot the August 2004 data we can distinguish some variablility at 2-sigma uncertainty between 5-day windows. If we plot the 2004 data we clearly see the seasonal cycle. We should remember that sampling uncertainty has not been included here. 2. For 1-degree and 5-day averages the synoptically correlated uncertainty is the dominant contributor to the total uncertainty. The large scale component can also make a significant contribution to the total uncertainty and for some averaging windows can be comparable in magnitude to the synoptically correlated component. 3. If we averaged over larger time and space scales the effective number of synoptic areas would increase, the synoptically correlated component of uncertainty would reduce in magnitude, and the large scale correlated component of uncertainty would become dominant. If we averaged over smaller space and time scales the synoptically correlated components of uncertainty would increase as the effective number of synoptic areas tends towards one. The uncorrelated component of uncertainty would also increase because its variance is reduced by the square of the number of observations being averaged. In the data provided, the magnitude of the uncorrelated component of uncertainty and the adjustment uncertainty are generally smaller than the synoptically correlated component, and so at small scales the synoptically correlated component of uncertainty is likely to dominate.

Analysis - 2004

The processing done for August 2004 is now repeated for the whole of 2004.

In [9]:

```
data = read_data(all2004=True)

windowsize = 5.0 * 24.0 * 60.0 * 60.0 # 5 days in seconds.
nwindows  = int(np.ceil(np.max(data['time']) / windowsize))
print('Number of windows is %i' % nwindows)

asst     = np.ma.zeros(nwindows)
aunc_lsc = np.ma.zeros(nwindows)
aunc_sc  = np.ma.zeros(nwindows)
aunc_u   = np.ma.zeros(nwindows)
aunc_a   = np.ma.zeros(nwindows)
asst.mask     = True
aunc_lsc.mask = True
aunc_sc.mask  = True
aunc_u.mask   = True
aunc_a.mask   = True
```

```
for iwindow in range(nwindows):
    timemin  = iwindow * windowsize
    timemax  = (iwindow + 1) * windowsize
    inwindow = (data['time'] >= timemin) & (data['time'] < timemax)

    ninwindow     = np.count_nonzero(data['sst'].mask[inwindow] == False)

    if ninwindow == 0: continue

    n_synop_areas = ninwindow / (1.0 + np.exp(-0.5 * (mean_dist / 100.0 + mean_time))*(ninwindow - 1))

    asst[iwindow]     = np.sum(data['sst'][inwindow]) / ninwindow # I.e. the mean.
    aunc_lsc[iwindow] = np.sum(data['unc_lsc'][inwindow]) / ninwindow
    aunc_sc[iwindow]  = np.sqrt(np.sum(data['unc_sc'][inwindow]**2) / (ninwindow * n_synop_areas))
    aunc_u[iwindow]   = np.sqrt(np.sum(data['unc_u'][inwindow]**2)) / ninwindow
    aunc_a[iwindow]   = np.sqrt(np.sum(data['unc_a'][inwindow]**2) / (ninwindow * n_synop_areas))

aunc_total = np.sqrt(aunc_lsc**2 + aunc_sc**2 + aunc_u**2 + aunc_a**2)

xaxis = np.arange(nwindows) + 1

fig = plt.figure(figsize=(18, 4))

# Plot of SST and total uncertainty.
fig.add_subplot(1, 2, 1)
use = asst.mask == False# Avoids a warning when errorbar is used below.
plt.errorbar(xaxis[use], asst[use], marker='x', ls='None', yerr=2*aunc_total[use])
plt.gca().set_xlim(0, nwindows + 1)
plt.gca().set_xlabel('Window')
plt.gca().set_ylabel('SST (K)')
plt.gca().set_title('Average SST +/- 2-sigma uncertainty')

# Plot of uncertainty components.
fig.add_subplot(1, 2, 2)
plt.plot(xaxis, aunc_lsc, 'x-', label='Large scale')
```

```
plt.plot(xaxis, aunc_sc, 'x-', label='Synoptic')
plt.plot(xaxis, aunc_u, 'x-', label='Uncorrelated')
plt.plot(xaxis, aunc_a, 'x-', label='Adjustment')
plt.gca().set_xlim(0, nwindows + 1)
plt.gca().set_xlabel('Window')
plt.gca().set_ylabel('Uncertainty (K)')
plt.gca().set_title('Uncertainty components')
plt.legend()

plt.show()
```

```
Number of windows is 74
```