Horizon 2020

H2020-EO-2014 New ideas for Earth-relevant Space Applications

EUSTACE

(Grant Agreement 640171)

EU Surface Temperature for All Corners of Earth

Deliverable D1.2

Common approach to providing uncertainty estimates across all surfaces

Deliverable Title	***Common approach to providing uncertainty estimates across all surfaces***	
Brief Description	*This report lays out the principles upon which uncertainty estimates will be provided with the satellite surface temperature datasets that will be input to the EUSTACE analysis.*	
WP number	1	
Lead Beneficiary	*Chris Merchant, University of Reading*	
Contributors	*Darren Ghent, University of Leicester* *John Kennedy, Met Office* *Elizabeth Good, UK Met Office,* *Jacob Hoyer, DMI*	
Creation Date	21-Aug-15	
Version Number	0.2	
Version Date	5-Sep-15	
Deliverable Due Date	31-Aug-15	
Actual Delivery Date	5-Sep-15	
Nature of the Deliverable	R	*R – Report*
		DEM – Demonstrator, Pilot, Prototype
		DEC – Dissemination, Exploitation or Communication
		O – Other
Dissemination Level/ Audience	PU	*PU – Public*
		CO - Confidential, only for members of the consortium, including the Commission services

Version	Date	Modified by	Comments
0.1	21/08/2015		Draft for review
0.2	03/09/2015		Agreed version for delivery

Table of Contents

1 Executive Summary

This report lays out the principles upon which consistent uncertainty estimates will be provided with the satellite-based surface temperature datasets that will be input to the EUSTACE analysis.

Each provider of satellite-derived data will provide level 2 (along track, ungridded; skin temperature only) and daily grid average ("level 3") surface temperature data of a named type (sea, land, ice, air; daily mean, min or max). The grid will be 0.25° latitude-by-longitude and the data provided will be the mean of the observed data in the grid cell, to be interpreted as an estimate of the full-area mean across the grid cell. For each cell temperature, three components of uncertainty will be provided, representing the uncertainty from effects whose errors have distinct correlation properties: random (no correlation of error component between cells); locally systematic (correlation of error component between "nearby" cells); and [large-scale] systematic (correlation of error component between "distant" cells.

The distinction between "nearby" and "distant" links to how the error component will be treated in the EUSTACE analysis procedure. Locally correlated errors will be modelled via spatio-temporal correlation length scales that influence the weight of an observation the air temperature analysis in the vicinity of its time-space location. Systematic errors will be accounted for by allowing a bias to be determined within the analysis procedure between different sources of data, whose magnitude is conditioned by the uncertainty attributed to systematic effects.

Detailed explanations and proposed means of estimating the different terms are given in this report, for each domain, based on extensive discussions with the different groups providing data.

Attaching uncertainty estimates of this nature to satellite-derived surface temperatures is relatively recent for sea surface temperature and is a necessary innovation for the other domains. Inevitably, some assumptions and approximations are adopted as data producers learn how to provide this improved uncertainty characterisation, and these are listed in the report in order to guide future developments, should the EUSTACE analysis turn out to be highly sensitive to particular assumptions or approximations made.

2 Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOA, Section B1.1):

No.	Objective	Yes	No
1	Intensively develop the hitherto immature use of Earth Observation estimates of Earth's surface **skin** temperature to enable new Climate Data Records of the surface **air** temperature Essential Climate Variable (ECV) to be created, for all locations over all surfaces of Earth (i.e. land, ocean, ice and lakes), for every day since 1850. EUSTACE will achieve this by: combining information estimated from multiple satellites with surface air temperature measurements made *in situ* and **creating complete analyses** of surface air temperature, through the application of novel statistical in-filling methods.	X	
2	Integrate these new daily surface air temperature Climate Data Records into a range of applications in Earth System Science and Climate Services and research, amongst others. EUSTACE will achieve this via the active and continuous engagement of trail-blazer users, and the provision of products through already-existing user community data portals and service mechanisms, in standard formats.		X
3	Undertake and report detailed research into the relationships between surface skin temperature estimated from Earth Observation satellite measurements and surface air temperature observed *in situ* by conventional measurements, over all surfaces of the Earth, including the polar regions. This is likely to provide information useful for refining coupling in Earth system models.	X	
4	Create a sustainable, automated system at an appropriate level of maturity for the potential production of the products beyond the lifetime of the project. To enable this, EUSTACE will also identify Earth Observation and conventional data streams that could be used to update the surface air temperature Climate Data Records in the future, including those from Sentinel missions.		X
5	Extensively validate the new surface air temperature Climate Data Records against independent, surface-based reference data, sourced by the project for this purpose.		X
6	Develop and report new, consistent, validated estimates of uncertainty both in already-existing Earth Observation surface skin temperature estimates and in the new surface air temperature Climate Data Records, at all locations and times across the Earth's surface.	X	

7	Develop links with related activities within Europe and beyond to help to ensure the execution of a joined-up work programme, the Copernicus Services and to enable the provision of requirements for the future surface skin temperature and surface air temperature observing system.	X	
8	Other – not directly linked to one of the above objectives		

3 What is uncertainty and from where does it arise?

The terms 'error' and 'uncertainty' are often unhelpfully conflated. Careful usage (following international standards; VIM, 2012) brings clarity to thinking about uncertainty information. Terms with precise definitions that need careful usage appear in italic in the next paragraph.

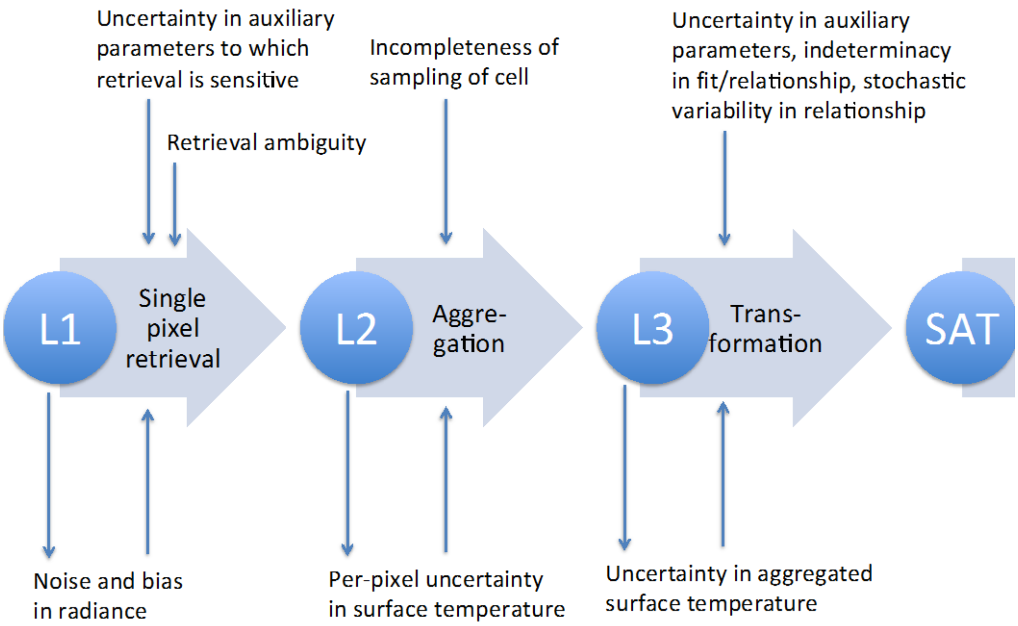
A *measured value* results from *measurement* of a target quantity, called the *measurand*. It is only an estimate of the measurand, because various effects introduce *errors* into the process of measurement. These errors are unknown, although their distributional properties may be able to be characterized (as discussed later). *Uncertainty* information characterizes the distribution of values it is reasonable to attribute to the measurand, given the measured value and our characterization of effects causing error.

In short, the error is the 'wrongness' of the measured value (and is unknown). The uncertainty describes the 'doubt' we have about the measurand's value, given the measured value and our understanding of effects causing errors.

Note that these technical definitions correspond well to the plain meaning of the words 'error' and 'uncertainty' as used by non-scientists. As well as improving communication between scientists, careful usage will help scientists communicate beyond their community.

Uncertainty arises from and propagates through every physical and data transformation involved in creating a satellite dataset. Propagation of uncertainty to a quantity y that is derived by some transformation from the inputs x is described by the law of propagation of uncertainty (GUM, 1995), applicable where a first-order (linear) approximation is valid, or can be assessed by means of simulation of the transformation (such as a Monte Carlo approach).

In the context of EUSTACE, the following cascade of uncertainty applies in moving from the satellite radiance product ("level 1" or "L1" inputs to EUSTACE) to a provided estimate of surface air temperature (SAT):



(Note: in some cases, transformation may precede aggregation in sequence of processing, but the same sources and of uncertainty and principles of propagation apply.)

Several properties of the uncertainty arising from a given effect need to be quantified, estimated, or described:

- The dispersion of the errors. This will be quantified in EUSTACE as the second moment of the estimated error probability density function (the standard deviation of the estimated error distribution). In the standard parlance, this is the *standard uncertainty* associated with these errors. As is common, from this point on, “standard” will be implicit when discussing a quantified uncertainty.
- The shape of the error distribution. Even descriptive statements based on judgements are useful here. In particular, is the distribution of errors thought to depart from Gaussian in that
    - there is significant skewness?
    - there is a significant outlier rate?
- The correlation of the errors between different measured values. This particularly affects how uncertainty is propagated in the aggregation step. The correlation of errors between gridded cell values is required information for the analysis of SAT. This will be expressed as the characteristic length of an exponential decrease in correlation coefficient with separation, one length for each of time and space. There may not be full information to characterise the length scales rigorously, in which case reasoned estimates will be assumed.

4 Model for uncertainty

Uncertainty will be modelled as three components:

- “random” -- meaning errors that are both random and independent between all data
- “locally systematic” – meaning errors that are highly correlated across short separations in time and distance; statisticians may refer to this case as “structured random”
- “[large-scale] systematic” – meaning errors that have a structure that is persistent in time and space; this includes but is not limited to “biases”

Thus, the components are distinguished by their error correlation length scales. (Hereafter, for convenience, the “large-scale” modifier will be dropped before “systematic”.) In truth, the division between locally systematic and systematic cases is somewhat artificial, and how best to decompose effects with a range of types of correlation into these components is a matter of judgement for data providers.

This approach is:

- a necessary minimum, since locally systematic effects are significant, and preclude use of a simple random/systematic model
- an approximation, in that there are several effects that have a systematic aspect, and all of these are required to be partitioned into either the locally systematic or systematic component
- a significant advance on what has generally been done for satellite datasets hitherto
- based on reasonable experiences with this approach in two European Space Agency projects (“SST CCI” and “GlobTemperature”)

This three-component model applies to all satellite processing levels (L1, L2, L3 and SAT).

The following sections describe the principles for estimating each component for each processing level. For each component, one or more methods of estimation will be defined, with comments on applicability and any limitations.

5 Level 1 uncertainty estimation

Level 1 products used in EUSTACE comprise calibrated radiances with geolocation. (We will often, as here, use ‘radiance’ generically to cover any measure of radiation, whether expressed as brightness temperature, reflectance or radiance.) Since the target resolution for the gridded data to be used in EUSTACE is 0.25°, we may assume that consideration of geolocation uncertainty has negligible impact (although where an estimate for this effect exists, it can be used). The uncertainty estimation focuses on uncertainty in radiance.

Ideally, L1 products would provide per pixel uncertainty information, but this is not generally the case. The H2020 project FIDUCEO (Fidelity and Uncertainty in Climate data records from Earth Observations, www.fiduceo.eu) will address this deficiency for some relevant datasets by mid-2017, but for EUSTACE, such information will need to be approximated or parameterised. (On this and other areas where FIDUCEO will advance understanding of uncertainty in satellite datasets, the EUSTACE project will take opportunities to give inputs on requirements and review FIDUCEO datasets and tools in its capacity as a project interested in climate data record uncertainties.)

Random: Detector noise, amplifier noise and digitisation cause errors that are well represented as being independent between image pixels. For brightness temperatures (BTs), this is usually expressed as noise equivalent differential temperature (NEDT) in temperature units, which is, in fact, standard uncertainty in BT. Since noise from these sources may be relatively constant in radiance and BT is a non-linear function of radiance, NEDT is usually expected to depend on the scene temperature. In some L1 products, the standard deviation of BTs when viewing calibration targets may be available per scan line, which can then be used to estimate NEDT.

Locally systematic: Effects such as stray light in the sensor may affect the calibration on limited space-time areas within an L1 product; at present there is no basis on which to estimate such effects, and they are neglected. Uncertainty in on-board calibration parameters arises because calibration targets are viewed for a finite time giving a statistical uncertainty from sensor noise. The error from this source is correlated across all the scans that use a particular on-board calibration cycle. This effect is by design small for a good sensor design, and is neglected. (FIDUCEO will support revisiting these approximations in future.)

Systematic: Each channel at L1 will have its specific systematic error, which propagates into retrieved surface temperature. Generally, efforts are made to correct for systematic effects in both radiance and retrieval jointly, at the stage of developing a retrieval algorithm rather than at L1 (although attempts at radiance bias correction are becoming more common). L1 contributions to systematic effects in retrieved surface temperature are difficult to decompose, so it is proposed that the overall uncertainty from systematic effects is addressed at level 2. (Again, FIDUCEO will develop a more rigorous approach in future.)

The following methods for L1 uncertainty from random effects may be used for EUSTACE:

METHODS FOR LEVEL 1 UNCERTAINTY		
Method ID / title	**Principle**	**Comments / limitations**
L1 Random 1 / Constant noise assumption	Use a constant value per channel for NEDT as a standard uncertainty estimate. Based on literature or engineering specifications for noise.	Does not reflect expected reduction in noise for higher scene temperature, nor temporal changes in noise from instrument. Some literature/specification based estimates for noise may be conservative (over-estimated).
L1 Random 2 / Constant noise-BT relationships	Use a fixed relationship per channel for NEDT as a function of BT as a standard uncertainty estimate. Based on literature or engineering specifications for noise and/or	Reflects expected reduction in noise for higher scene temperature. Some literature/specification based estimates for noise may be conservative (over-estimated), and the NEDT may not in reality behave

	deduction from instrument radiance-BT relationship.	in the simple way expected. No account of temporal changes in noise from instrument.
L1 Random 3 / Inference from calibration target data	Interpolate/extrapolate from the observed standard deviation of measured values when viewing a calibration target of known temperature.	Necessary data not available for every instrument. Does account for temporal change in sensor performance. If thermal gradients exist across targets (adding to spread of measured values), may be over estimate.
L1 Local 1 / Neglect	Assume no locally correlated errors sources at L1.	Should be adequate to assume this for well-designed sensors, but there are potential exceptions.
L1 Systematic 1 / Address at L2	Large-scale systematic errors will lead to "biases" in retrieved surface temperature. Wrap up both effects in method L2 Systematic 1.	While propagating L1 systematic uncertainty to L2 surface temperature is possible in principle, there is limited knowledge of this in practice.

6 Level 2 Uncertainty estimation

From L1 products, skin surface temperature is estimated over sea, ice and land at the pixel level. The retrieval R generally depends on the radiances y, an assumption about the surface emissivity ε per channel, and (in some cases) other auxiliary information including numerical weather prediction fields, β. Thus, $\hat{x} = R(y,\varepsilon,\beta)$, where the hat indicates the retrieval estimate of the true target measurand. Uncertainty in each of these contributes to the total uncertainty in the retrieved surface temperature. Additionally, surface temperature retrieval methods generally can never fully resolve the ambiguity in the forward relationship between surface temperature and observed radiances. It is generally possible to find a plausible perturbation in atmospheric conditions whose impact on radiance is the same as the effect of a perturbation in temperature: all retrievals therefore deal with some ambiguity that is dependent on the atmospheric conditions, and this is an additional source of uncertainty at L2.

Random: Let the random component of L1 uncertainty in the channel c be $u_{ran}(y_c)$. The effect of this combined across all n channels used for retrieval needs to be propagated through R to give a contribution to the estimate of uncertainty from random effects $u_{ran}(x)$ in the retrieved surface temperature. We assume that radiance noise is sufficiently Gaussian and small that the law of propagation of uncertainty is adequate for this propagation, which means

(1) $$u_{ran,y}(x) = \sqrt{\sum_{c=1}^{n}\left(\frac{\partial R}{\partial y_c}u_{ran}(y_c)\right)^2}$$

Emissivity is an auxiliary input to all estimates of thermodynamic temperature from BTs, whether explicit or implicit. In the case of SST, this is considered a negligible source of error, since emissivity variability is relatively small, and the modelling of emissivity is relatively well understood as a function of salinity, wave state, temperature and view angle. For other cases, there is a potentially significant random error component caused by the pixel-to-pixel variations in emissivity not captured in emissivity auxiliary information, because it is related to variability on the ground that is not captured in emissivity atlases/models. The associated uncertainty can be estimated as

(2)
$$u_{ran,\varepsilon}(x) = \sqrt{\sum_{c=1}^{n} \left(\frac{\partial R}{\partial \varepsilon_c} u_{ran}(\varepsilon_c) \right)^2}$$

where, evidently, some estimate of the uncertainty in emissivity per channel is required, as well as evaluation of the sensitivity of the retrieval to emissivity.

Since atmospheric fields are smooth on pixel to pixel scales, it is assumed that there is no random error associated with errors in $\boldsymbol{\beta}$: any errors will be locally correlated. Thus, the total uncertainty from random effects is

(3)
$$u_{ran}(x) = \sqrt{u_{ran,y}(x)^2 + u_{ran,\varepsilon}(x)^2}$$

Locally systematic: Atmospheric fields are correlated on synoptic timescales. The correlation scales vary with weather and climatic regime: we can expect the scales to be >1 day and >100 km, but this may require more quantification. It is assumed that errors in estimates of these fields from NWP are correlated on the same scales as the fields themselves; this assumption is plausible and pragmatic, rather than evidence-based at present. Therefore, if there is sensitivity in the retrieval to an explicit set of NWP parameters, $\boldsymbol{\beta}$, then the NWP-caused errors are locally correlated, and the associated uncertainty is in principle calculable as

(4)
$$u_{loc,\beta}(x) = \sqrt{\mathbf{C}_\beta \mathbf{U}_\beta \mathbf{C}_\beta^{\mathrm{T}}}$$

$$\mathbf{C}_\beta = \left(\frac{\partial R}{\partial \beta_1}, \ldots \right)$$

$$\mathbf{U}_\beta = \begin{bmatrix} u(\beta_1)^2 & r_{1,2} u(\beta_1) u(\beta_2) & \cdots \\ r_{2,1} u(\beta_2) u(\beta_1) & \ddots & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where $r_{r,c}$ is the correlation coefficient of the errors in two parameters, and $r_{r,c} = r_{c,r}$ so that $\mathbf{U}_{\beta}$ is symmetric.

This more complex form arises because correlations between errors in atmospheric parameters cannot be neglected. This component of uncertainty is calculated as a by-product in optimal estimation as part of the retrieval error covariance matrix. Referring the reader to Rodgers (2000), the relevant equation showing this is (4.43), and the equivalent term is in standard retrieval notation: $\left(\mathbf{I}_{N} - \mathbf{GK}\right)\mathbf{S}_{a}\left(\mathbf{I}_{N} - \mathbf{GK}\right)^{\mathrm{T}}$.

Coefficient based methods do not explicitly evaluate eq. (4), but the retrieval ambiguity is an equivalent form of uncertainty in this case, and is a contributor of residuals in the fit. For radiative-transfer based retrieval coefficients, simulated-retrieved and simulation-input surface temperatures can be compared. The standard deviation of this input and output difference is an estimate of the magnitude of this locally correlated form of uncertainty (Merchant and Embury, 2014). The calculation of the uncertainty can be done on stratified data to parameterise the variations in magnitude of this form of uncertainty (Embury et al., 2012). For each range, $\gamma_{i} < \gamma \leq \gamma_{i+1}$, where γ is a variable on which the residuals are stratified, the uncertainty is estimated as

(5) $$u^{i}_{loc,fit}\left(x\right) = \sqrt{Var\left(\hat{x} - x_{in}\right)}$$

where x_{in} is the surface temperature used for simulation and $\hat{x}$ is the retrieved estimate if the simulated radiances are fed to the retrieval equation. The uncertainty can then be fitted as a function of γ (which has the advantage of giving a smooth variation in estimated uncertainty) or read from a table for the appropriate value of i.

If coefficients are based on empirical regression, the standard deviation of the residuals to the fit arises from this effect plus the measurement noise in both the satellite radiances and the reference measurements used to fit the regression. If estimates of the variance contributed by these terms are available, these can be subtracted from the total variance of residuals to estimate the locally correlated component:

(6) $$u_{loc,fit}(x) = \sqrt{Var\left(\hat{x} - x_{ref}\right) - u_{ran}\left(x\right)^{2} - u_{ref}^{\ 2}}$$

where x_{ref} represents empirical reference data to which the retrieval is fit, and u_{ref} is the uncertainty of the reference data. Again, this can be done on a stratified basis.

LST retrieval assumes an emissivity which may be driven by auxiliary land classification information and/or and observed vegetation index. Across a particular land class area, there may be a mean difference between the assumed and true mean emissivity. This is thus a locally correlated effect on the scales of emissivity variability. The form of the propagation to L2 uncertainty is identical to eq. (2), but for locally correlated error components.

(7) $$u_{loc,\varepsilon}(x) = \sqrt{\sum_{c=1}^{n}\left(\frac{\partial R}{\partial \varepsilon_c} u_{loc}(\varepsilon_c)\right)^2}$$

The spatio-temporal scales of the emissivity-driven and atmospherically-driven locally correlated errors are likely to differ. The length scale for the emissivity errors should be estimated from the linear extents of the land class or features used to define emissivity variability. The time scale for the emissivity errors should be estimated from the time interval at which emissivity is updated (e.g., monthly climatological emissivity values for a particular location).

Two locally correlated components and their length scales (which may be variable) can be provided in the case where atmospheric and emissivity effects have markedly different scales. Alternatively, if it is decided to stick to a strictly three-component model, a plausible approach is to provide a weighted mean of the estimated scales (which are in any case likely to be approximate), weighted according to their respective magnitudes of uncertainty.

Systematic: Systematic errors are not merely biases, since errors that depend systematically on variable factors can have zero mean, yet still be systematically correctable in principle (if we had adequate knowledge of the effect). It is assumed here that known corrections have been applied by data producers, either at L1 or in the retrieval process to L2.

In the simplest case, the correction is a global bias correction (which may be zero conceptually if no correction was found necessary). In that case, the systematic uncertainty is the uncertainty in that correction – a single number for all observations.

When the surface air temperature analysis is done, offsets between different sources of information will be modelled and estimated. The uncertainty from systematic effects at L2 (and L3) will be used as information to weight data sources in deducing relative biases in the analysis step.

(It is likely that systematic effects lead to patterns of systematic error (bias) in the data (spatially, temporally and/or as a function of context) that have not been characterised, often because there is insufficient validation data to reveal the patterns. In some cases, the patterns could in principle be predicted. For example, an (unknown) constant calibration error in radiances for a particular sensor channel will produce a pattern of error that varies between observations according to the sensitivity of the retrieval process to that channel. The sign and overall magnitude of the systematic error is unknown, but the pattern is predictable. It is possible in principle that such patterns could be used in the analysis step (and the coefficient of the pattern estimated). This is noted for future research, since this capability is not presently developed by the data providers.)

METHODS FOR LEVEL 2 UNCERTAINTY		
Method ID / title	**Principle**	**Comments / limitations**
L2 Random 1 / Radiance noise propagation	Propagate the results of method L1 Random 1, 2 or 3 through the retrieval process, either analytically (eq. (1)) or	Required for all cases.

	numerically.	
L2 Random 2 / Emissivity noise propagation	Estimate the magnitude of pixel-to-pixel scale emissivity variability within areas that, based on land cover class, are treated as having a common emissivity. Then use eq. (2) to estimate L2 impact. Combine this with results of L2 Random 1.	Within-land-class variability of emissivity may be difficult to estimate. A possibility is to use samples from the ASTER 1km Global Emissivity Database (Hulley et al., 2012) for this characterisation.
L2 Random 3 / Parameterised emissivity noise	Emissivity may be parameterised in terms of NDVI or similar. The random effects in the input parameters then need to be propagated through the emissivity parameterisation to establish the random component of emissivity uncertainty.	An alternative to L2 Random 2.
L2 Local 1 / Propagate uncertainty from non-emissivity parameters in retrieval	Use eq. (4).	Requires an error covariance matrix for atmospheric/auxiliary parameters.
L2 Local 2 / Uncertainty from atmosphere/fit for regression-based retrieval	Use eq. (5) or (6) as appropriate.	Alternative to L2 Local 1.
L2 Local 2 / Uncertainty from emissivity	Estimate locally correlated error from emissivity assumptions then use eq. (7).	**Either**: combine with result of L2 Local 1 or 2 as appropriate, e.g.: $u_{loc}(x)=\sqrt{u_{loc,fit}^{2}+u_{loc,\varepsilon}^{2}}$ if scales are compatible or not estimated. **Or**: report separately as a further locally correlated error with its own scale (which needs to be stated).
L2 Systematic 1 / Reasoned estimate	Knowledge of the satellite engineering specifications and/or validation performance may allow a reasoned estimate of the likely magnitude of residual bias.	Essentially expert judgement.
L2 Systematic 2 / Validation-based	Where a statistically significant mean difference between satellite and (adequately representative) validation data has been found, the mean should be applied as a correction and the uncertainty in the mean (the standard error) provided as systematic	Preferable to L2 Systematic 1, provided validation data are considered to be adequately representative.

	uncertainty.	
L2 Systematic 3 / From overlaps between sensors	Where one or more overlaps between sensor pairs are available, the mean differences between matched surface temperatures are indicative of the uncertainty from systematic effects.	Preferable to L2 Systematic 2 when possible, since geographical range is likely to be more representative. Element of judgement may still be present in how the overlap relative biases are interpreted.

7 Level 3 uncertainty estimation

The uncertainty at L3 is found by propagation from the L2 components of uncertainty. The 0.25° daily grid cell will be represented by the mean of the retrievals, x_i, available for L2 pixels:

(8) $$x_{L3} = \frac{1}{N}\sum_{i=1}^{N}\hat{x}_i$$

The uncertainty from random errors in the available pixels is reduced by averaging the uncertainty in contributing pixels in quadrature:

(9) $$u_{ran}\left(x_{L3}\right) = \sqrt{\left(\frac{1}{N^2}\sum_{i=1}^{N}u_{ran}\left(x_i\right)^2\right) + u_{samp}{}^2}$$

and u_{samp} is the uncertainty from random errors introduced when the grid cell is not fully sampled in space and time by valid satellite surface temperatures. This assumes that the locations that are sampled vary in such a way that the differences between the available sample mean and the unknown true cell mean are random.

A model is required for u_{samp}. If the standard deviation of x across the full grid cell is σ, then an obvious parameterisation is

(10) $$u_{samp} = \left(\frac{N_{tot} - N}{N_{tot} - 1}\right)^{\alpha}\sigma$$

where N_{tot} is the total number of pixels that would fully sample the cell and α is a parameter to be determined. For all α, this has the properties that u_{samp} is zero when the grid cell is fully observed ($N = N_{tot}$) and that when only 1 pixel is observed, the sampling uncertainty equals the standard deviation σ; both these limits are as required. However, unless the cell is fully observed, σ is not known, and must be estimated from the standard deviation of the available pixels, σ_N. σ_N is a poor (imprecise) estimate of σ for small N, and for $N=1$ is undefined. An alternative means of developing u_{samp} has been proposed for sea surface temperature by Bulgin et al. (submitted), which uses realistic sub-sampling of fully observed cells across many cases to find statistics of sampling error. The statistics are binned by N/N_{tot} and σ_N, and a fit for u_{samp} is determined in terms of these parameters. It was found that this approach worked well for cell sizes of 0.1° and 0.05° for sea surface temperature. The application of the method to other domains and a cell size of 0.25° needs to be assessed.

It is assumed that the spatio-temporal correlation scales of the locally correlated effects are much greater than ~25 km and about 1 day, and therefore this component of uncertainty is not reduced by averaging. This is a conservative approximation – i.e., any error from this approximation means this component of uncertainty is somewhat over-estimated. This approximation is usually defensible for the atmospheric/fitting term, since synoptic scales of the atmosphere are >100 km and >1 day, but may prove less realistic for emissivity effects (to be investigated). Hence, they propagate thus:

(11) $$u_{loc}(x_{L3}) \cong \frac{1}{N}\sum_{i=1}^{N} u_{loc}(x_i)$$

The systematic errors are common to all the contributing pixels by definition, and thus:

(12) $$u_{sys}(x_{L3}) = u_{sys}(x)$$

METHODS FOR LEVEL 3 UNCERTAINTY		
Method ID / title	**Principle**	**Comments / limitations**
L3 Random 1	Propagate level 2 random effects using (9), and add sampling uncertainty estimate using, for example, eq. (10)	Sampling uncertainty models need to be assessed and may be improved relative to proposed equation.
L3 Locally correlated 1	Use average across cell from level 2, i.e., eq. (11).	
L3 Systematic 1	Obtain from systematic term at level 2 directly.	

8 Uncertainty in inferred SAT

The transformation from satellite-derived surface skin temperature to surface air temperature is

(13) $$T_{type} = M(x, \chi)$$

where *type* may be daily maximum, daily minimum, daily mean or daily average[1], and χ comprises the parameters of the transformation model *M*. The transformation may be applied either to L2 or L3 surface temperatures by different groups. Where applied prior to L3 gridding, the equations of section 7 apply to T_{type} after this transformation when the surface air temperatures obtained are gridded. Here, the presentation assumes the transformation is applied to L3 data.

The propagation of uncertainty from x_{L3} to T_{type} on the same spatial grid is straightforward:

(14)
$$u_{comp,x}\left(T_{type}\right)=\left|\frac{\partial M}{\partial x_{L3}}u_{comp}\left(x_{L3}\right)\right|$$

for each of $comp=ran,loc,sys$. Depending on the nature of the transformation and the correlation properties of the errors in the elements of χ, the uncertainty

(15)
$$u_{comp,\chi}\left(T_{type}\right)=\sqrt{\sum_{i=1}\left(\frac{\partial M}{\partial \chi_i}u_{comp}\left(\chi_i\right)\right)^2}$$

may have components that are random, locally systematic and/or large-scale systematic. For any given component:

(16)
$$u_{comp}\left(T_{type}\right)=\sqrt{u_{comp,x}\left(T_{type}\right)^2+u_{comp,\chi}\left(T_{type}\right)^2}$$

which yields the uncertainty estimates required for the surface air temperature analysis, with separated random, locally correlated and systematic terms.

9 Checklist of information to be provided to the analysis

- Type of temperature(s) provided
- Statement of the geometry of the temperature provided relative to its associated location information. (By default for satellite-derived air temperatures, cells are expected to be 0.25° by 0.25° cells, with location 0°N, 0°E being a cell corner. The reported location by default is the centre latitude and longitude of the cell.)

[1] The daily mean is the simple average of the daily maximum and minimum, following traditional station air temperature conventions in some countries. The daily average is an estimate of the true 24-hour average, which in turn will be used in the analysis step as an estimate of the daily mean, allowing for the fact that the daily mean and daily average may differ.

- Name and brief description of known error effects (source of uncertainty) that have been quantified and are included in the provided uncertainty estimates. Identify to which uncertainty component(s) each effect contributes.

- Name and brief description of known (or suspected) error effects whose associated uncertainty has been unable to be quantified (and are not included in the provided uncertainty estimates).

- Description of model to transform satellite surface temperature to air temperature estimate, including a description of the uncertainty model for the transformation.

- The data. For a given observation at a stated location and time, this includes:

    - the air temperature estimate
    - the uncertainty from independent, random effects
    - the uncertainty from locally correlated effects[2], with a correlation length scale in both time and space
    - the uncertainty from systematic effects

- The error covariance matrix of the transformation model parameters, χ, and the set of sensitivities: $\frac{\partial M}{\partial x_{L3}}$ and $\frac{\partial M}{\partial \chi}$

10 References

Bulgin C., O. Embury and C. J. Merchant (submitted), Sampling Uncertainty in Global Area Coverage (GAC) and Gridded Sea Surface Temperature Products.

Embury, O., C.J. Merchant and M.J. Filipiak (2012), A reprocessing for climate of sea surface temperature from the Along-Track Scanning Radiometers: basis in radiative transfer, Remote Sensing of the Environment, doi: 10.1016/j.rse.2010.10.016

GUM (1995) *Evaluation of measurement data – Guide to the expression of uncertainty in measurement* JCGM 100:2008

Hulley, G. C., T. Hughes, and S. J. Hook (2012), Quantifying Uncertainties in Land Surface Temperature (LST) and Emissivity Retrievals from ASTER and MODIS Thermal Infrared Data, J. Geophys. Res. Lett, 117, D23113, doi:10.1029/2012JD018506. - See more at: https://lpdaac.usgs.gov/dataset_discovery/community/community_products_table/ag1km#sthash.arHLNlWP.dpuf

[2] Note comment earlier in main text that, where error correlation lengths scales from different locally correlated effects greatly diverge, it may be necessary to provide >1 separate estimates of the locally corrected terms. Where reasonable, however, combine all effects into one locally correlated term.

Merchant, C. J. and Embury, O. (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

Rodgers, C.D. (2000), Inverse methods for atmospheric sounding: theory and practice. World Scientific

VIM (2012), *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms* (VIM 3rd edition) JCGM 200:2012 (JCGM 200:2008 with minor corrections)