QUANTIFYING AND USING UNCERTAINTY: REVIEW AND HIGHLIGHTS

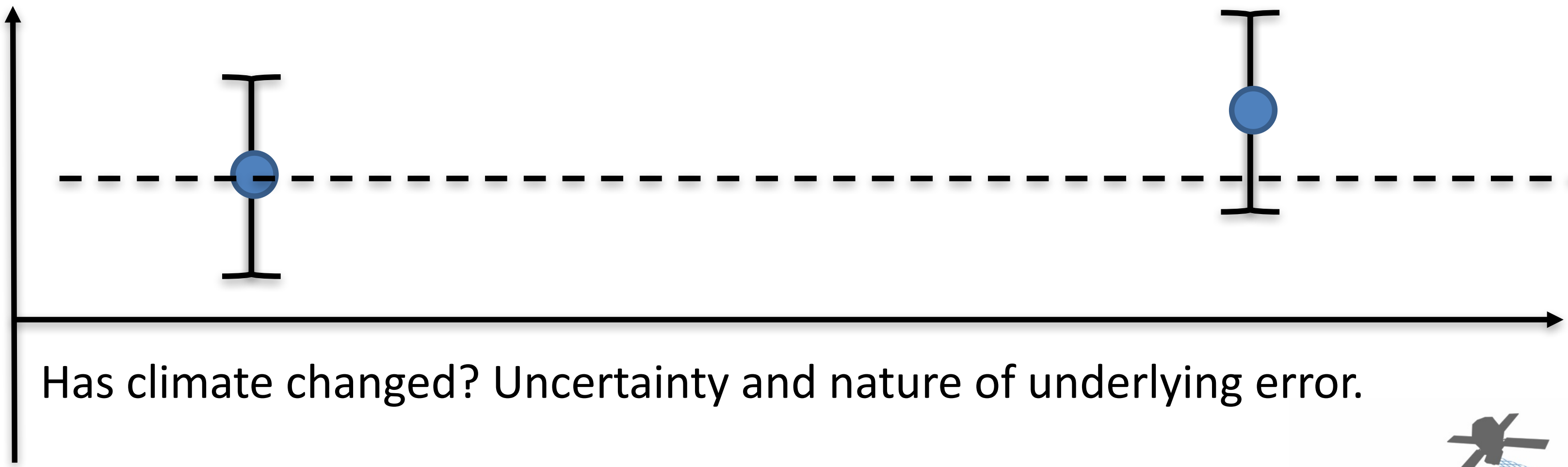
MERCHANT, GHENT, GOOD, HOYER, KENNEDY, LINDGREN, MORICE, SCHRIER

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

Uncertainty: meaning and importance

Uncertainty answers this question:

Given a value, within what spread of values is it reasonable to consider that the truth lay?

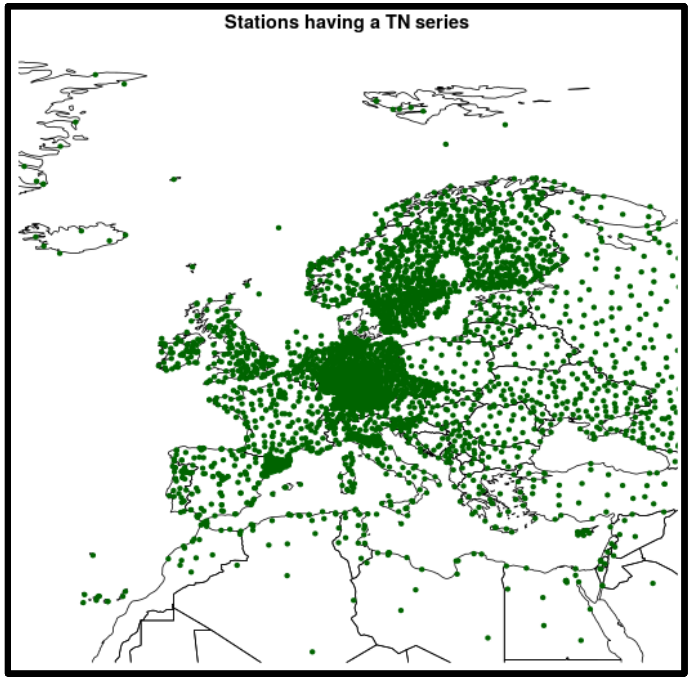


Has climate changed? Uncertainty and nature of underlying error.

STATION DATA

http://www.ecad.eu/

See also Antonello Squintu's poster!



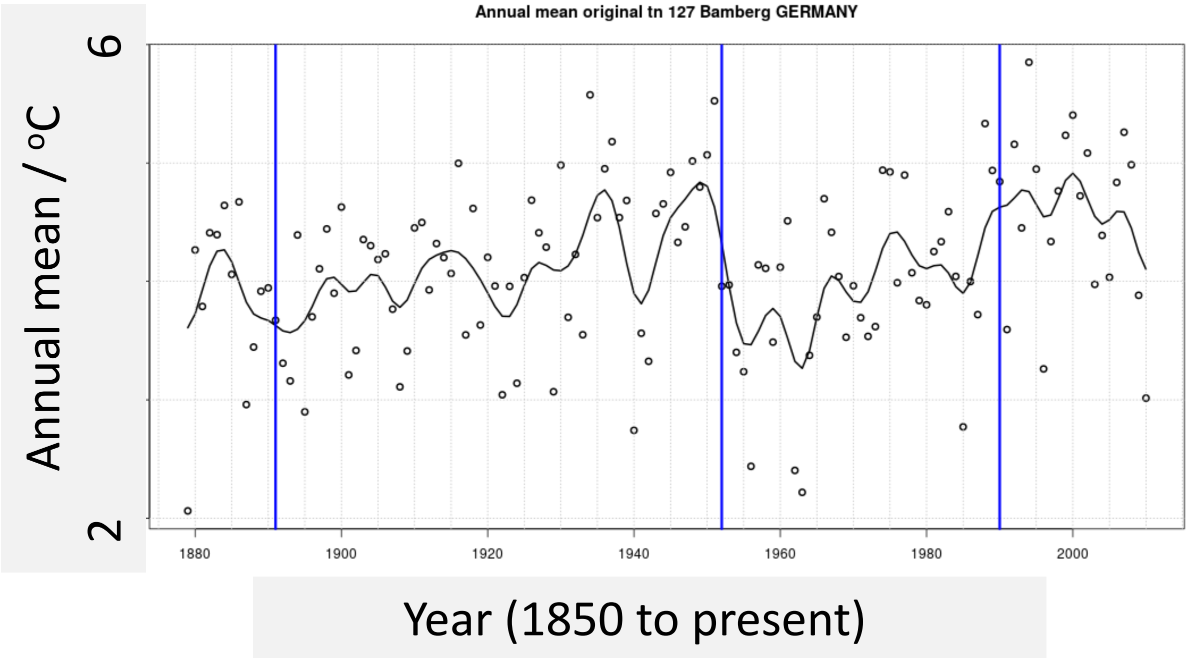
Many of these series have inhomogeneities

before

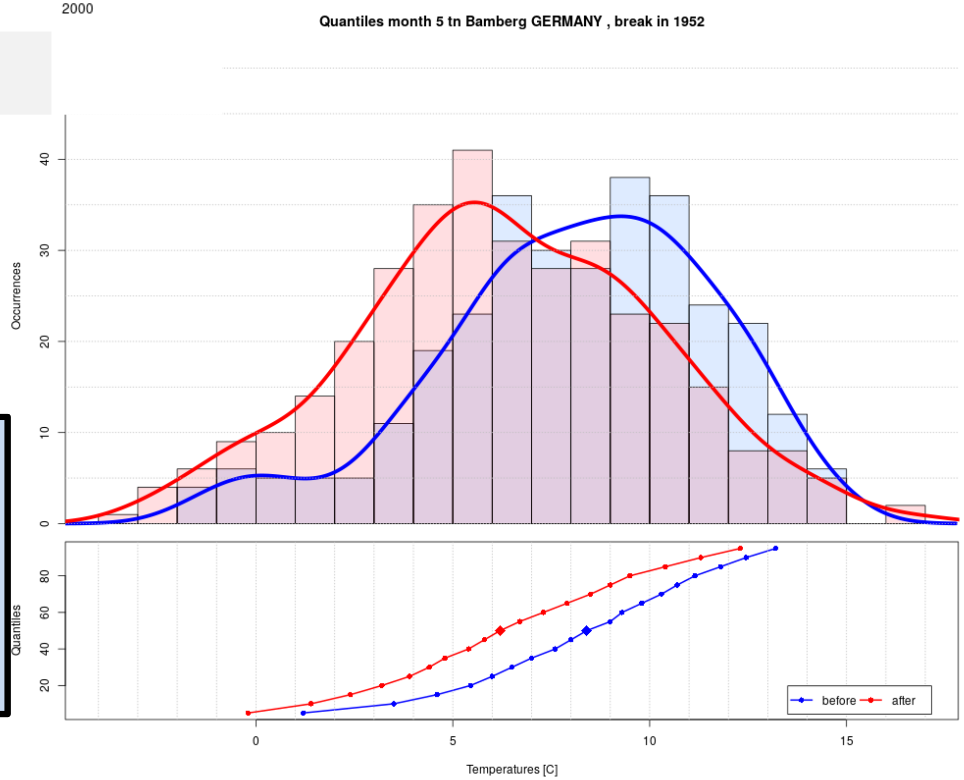
after

Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

BREAK DETECTION



T night,
station Bamberg

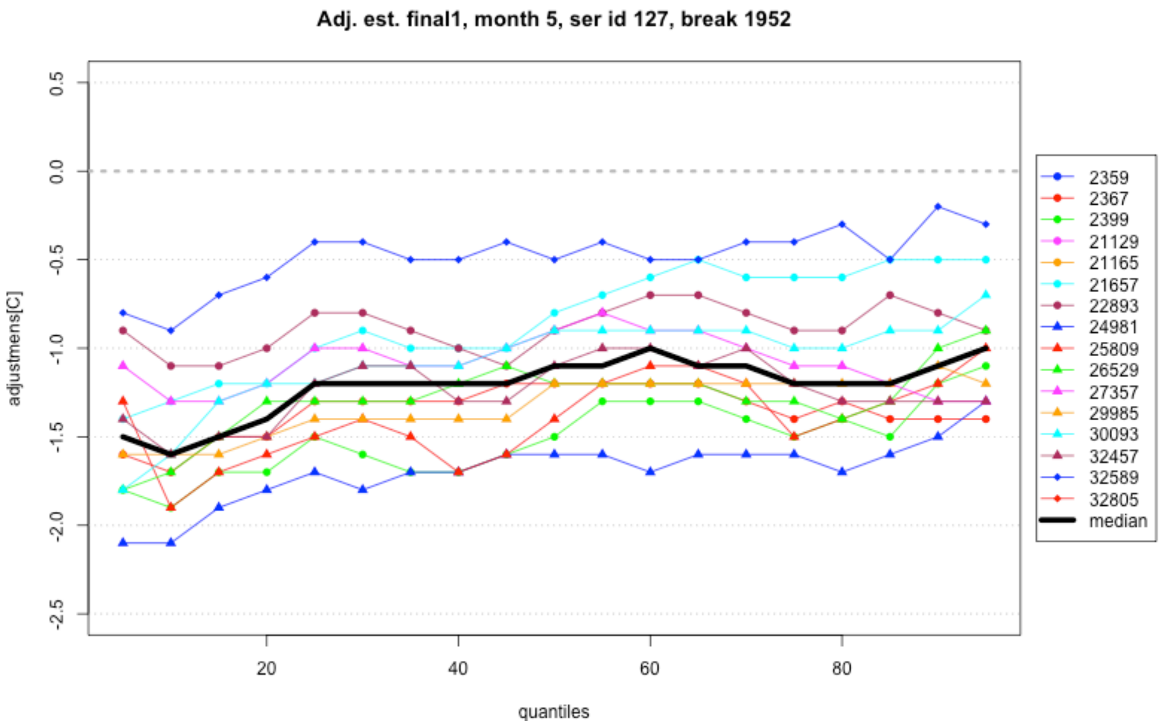


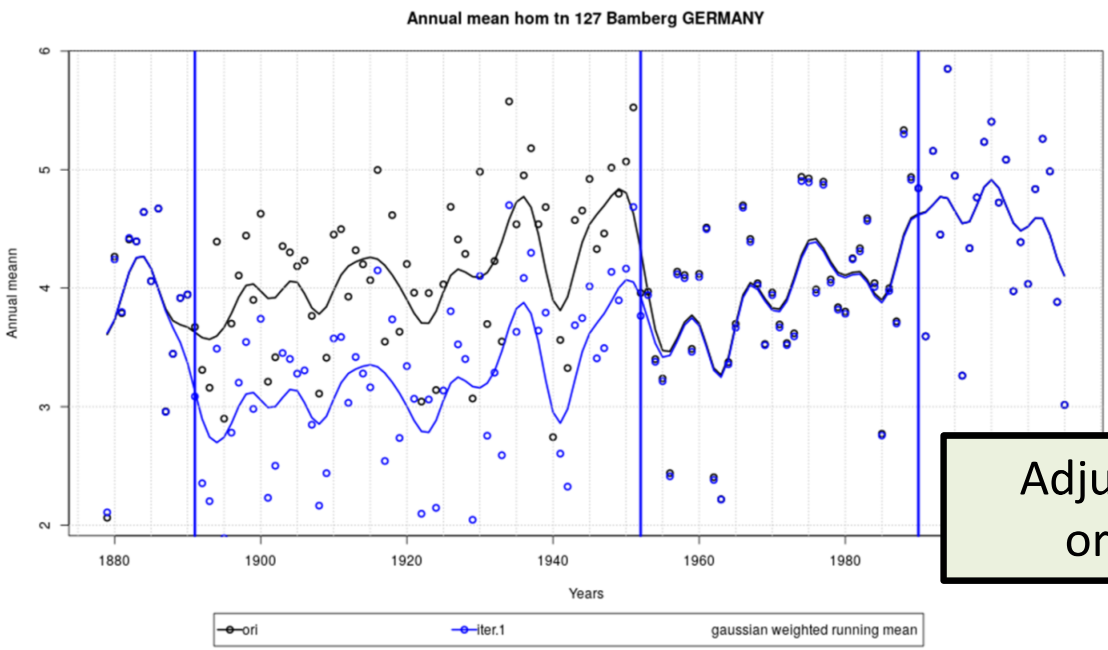
T night distributions:
10 years before (blue)
and after (red) the
break in 1952

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

THE ADJUSTMENTS

Each coloured line gives the adjustment against a different neighbouring station





Adjustment of the original series

EUSTACE

EUSTACE and Uncertainty

Integrated approach across project

- **Common understanding & models of uncertainty**

Outline:

Introductory comments (done that!)

Station Data (done that!)

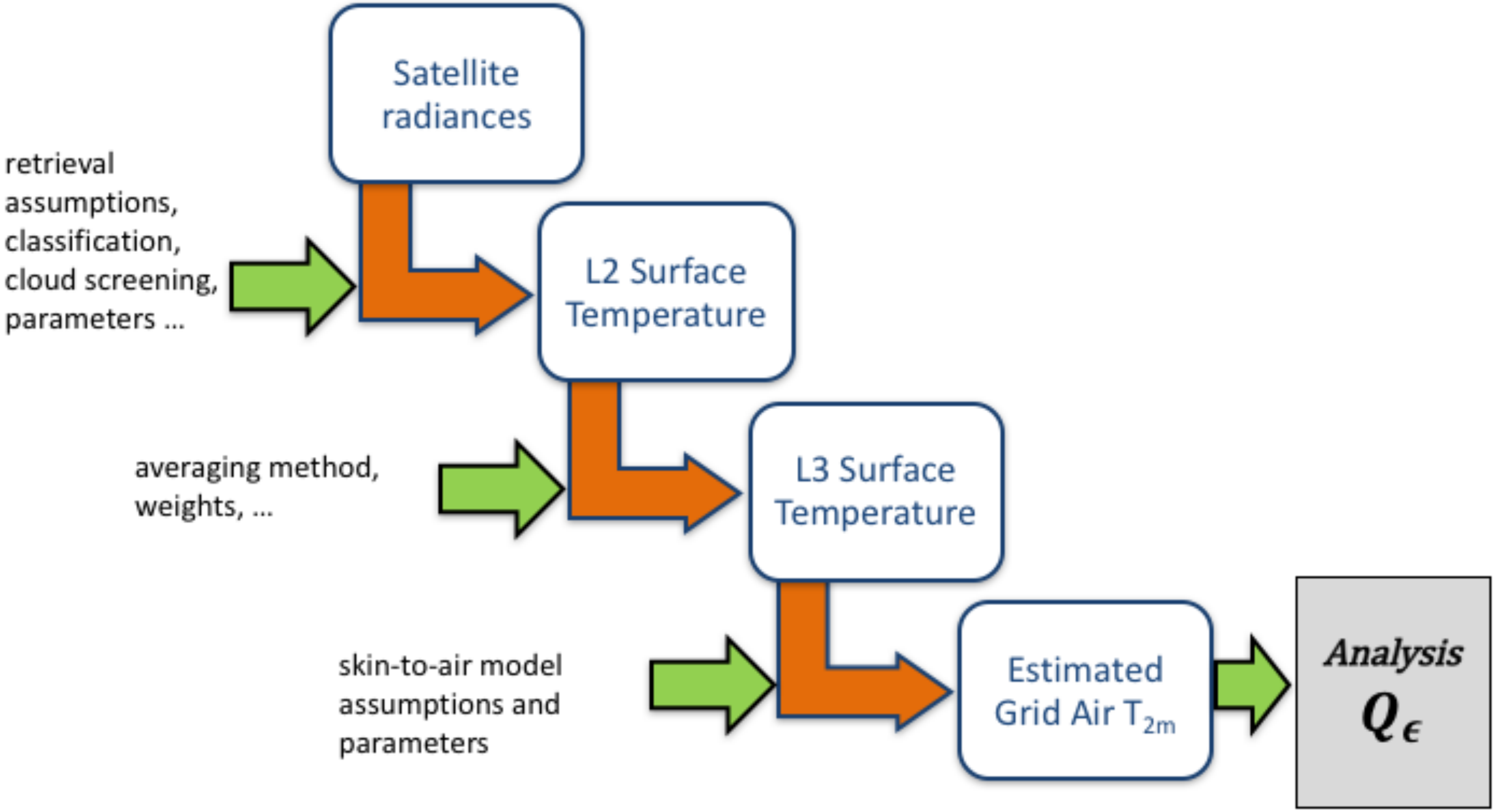
Satelltie ST retrieval and uncertainty validation

L2 to L3 uncertainty propagation

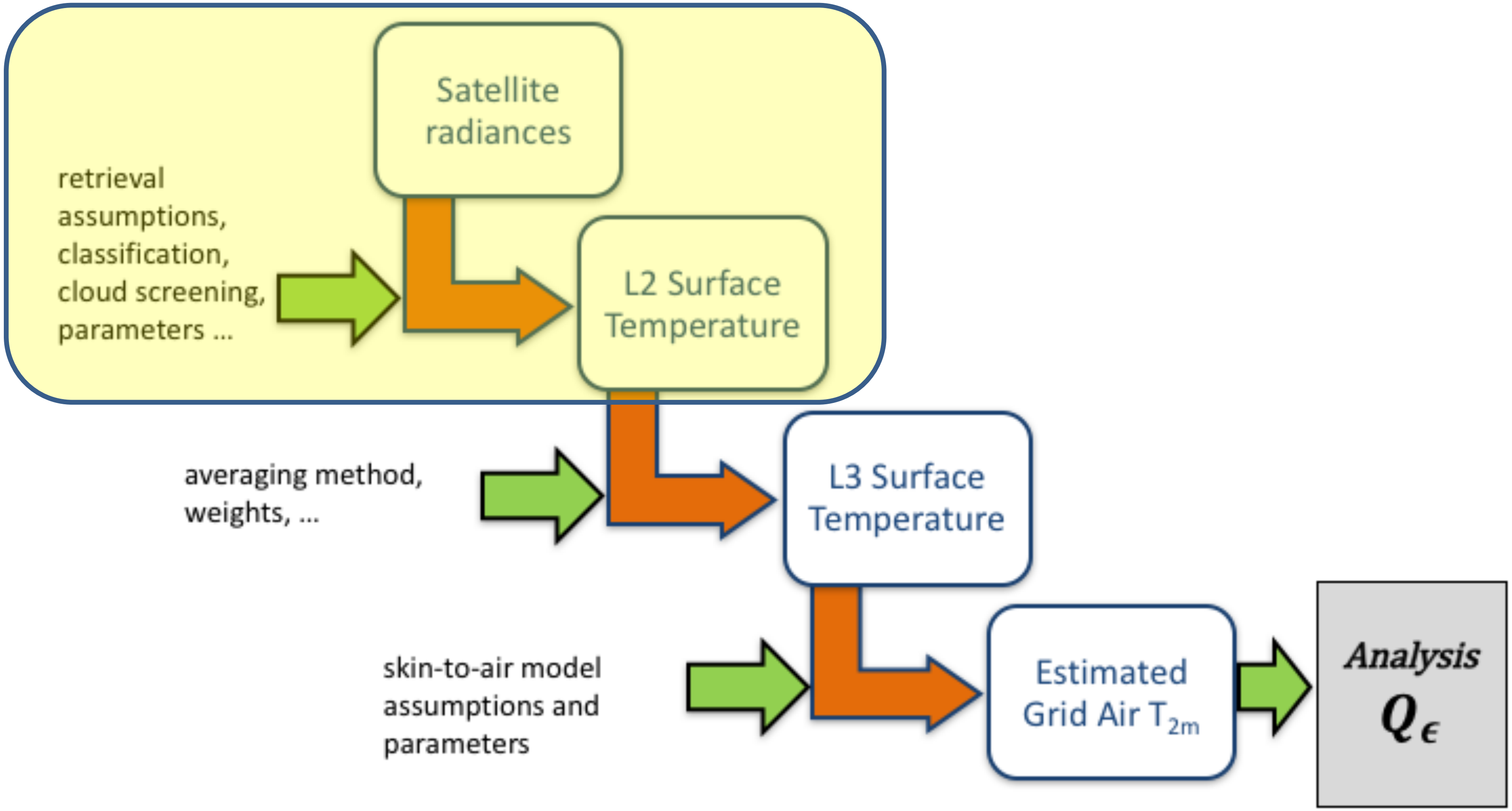
Satellite-to-T_{2m}

Usage of uncertainty information in SAT analysis

Sources and propagation of uncertainty



Sources and propagation of uncertainty



Satellite Surface Temperature Data

- Common three-component uncertainty model
  - random
  - locally correlated
  - systematic
- Validation of uncertainty
- Three-component model
  - Applies to all domains, land, ice, lake, sea
  - Applies across processing levels
  - Provides information propagated into analysis

EUSTACE

Land ST Uncertainty Components

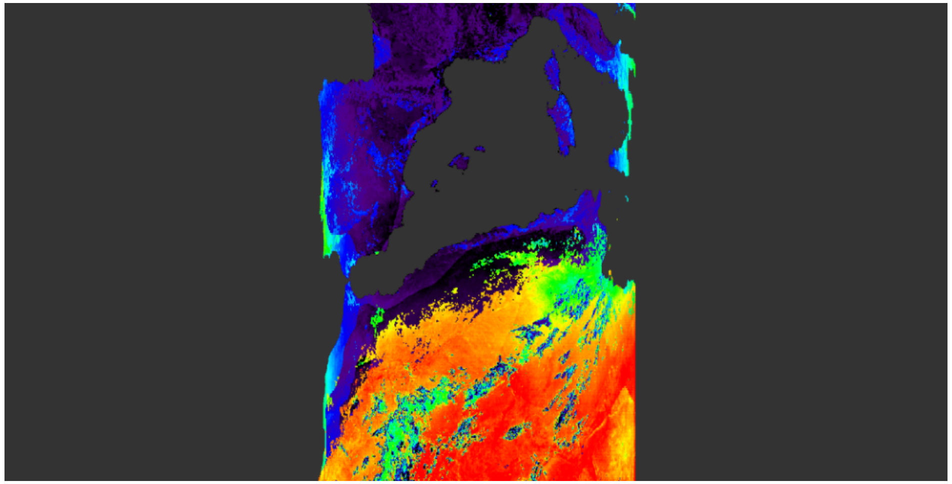
VARIABLE	METHOD	COMMENTS	
LST_UNC_RAN	**L2 Random 1 / Radiance noise Propagation**	$u_{ran,y}(x) = \sqrt{\sum_{c=1}^{n}\left(\frac{\partial R}{\partial y_c} u_{ran}(y_c)\right)^2}$	Random component of L1 channel uncertainties propagated through the retrieval
	L2 Random 2 / Emissivity noise Propagation	$u_{ran,\varepsilon}(x) = \sqrt{\sum_{c=1}^{n}\left(\frac{\partial R}{\partial \varepsilon_c} u_{ran}(\varepsilon_c)\right)^2}$	Estimate of the magnitude of pixel-to-pixel scale emissivity variability within areas based on land cover class
LST_UNC_LOC	**L2 Local 2 / Uncertainty from atmosphere/fit for regression-based retrieval**	$u_{loc,fit}(x) = \sqrt{Var(\hat{x} - x_{in})}$	Atmospheric fields correlated on timescales >1 day and length scales >100 km. For coefficient based retrieval methods the retrieval ambiguity is a contributor of residuals in the fit
	L2 Local 2 / Uncertainty from Emissivity	$u_{loc,\varepsilon}(x) = \sqrt{\sum_{c=1}^{n}\left(\frac{\partial R}{\partial \varepsilon_c} u_{loc}(\varepsilon_c)\right)^2}$	Across a particular land class area, there may be a mean difference between the assumed and true mean emissivity
LST_UNC_SYS	**L2 Systematic 1 / Reasoned estimate**	Assumed that known corrections have been applied by data producers and what remains is describable as an uncertainty in the bias of the satellite surface temperatures relative to other data sources of temperature (ie from validation)	

UNIVERSITY OF LEICESTER

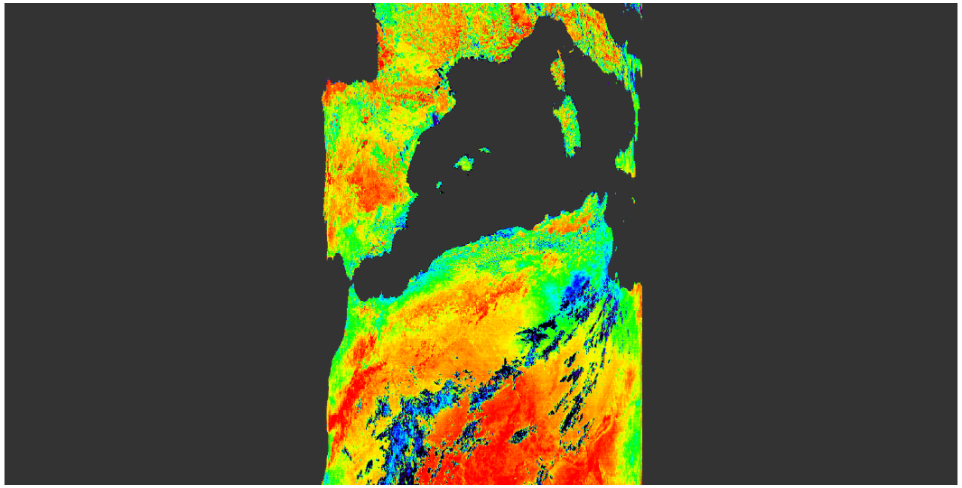
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

EUSTACE

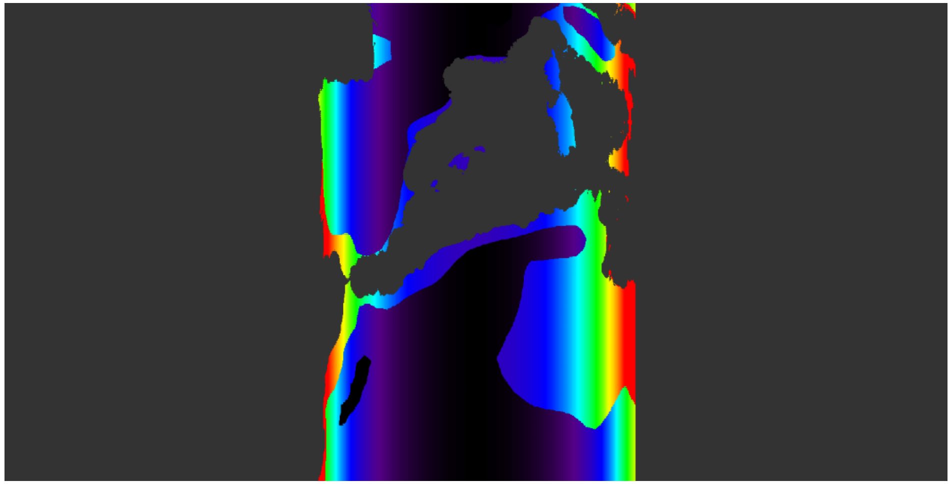
MODIS LST Uncertainties



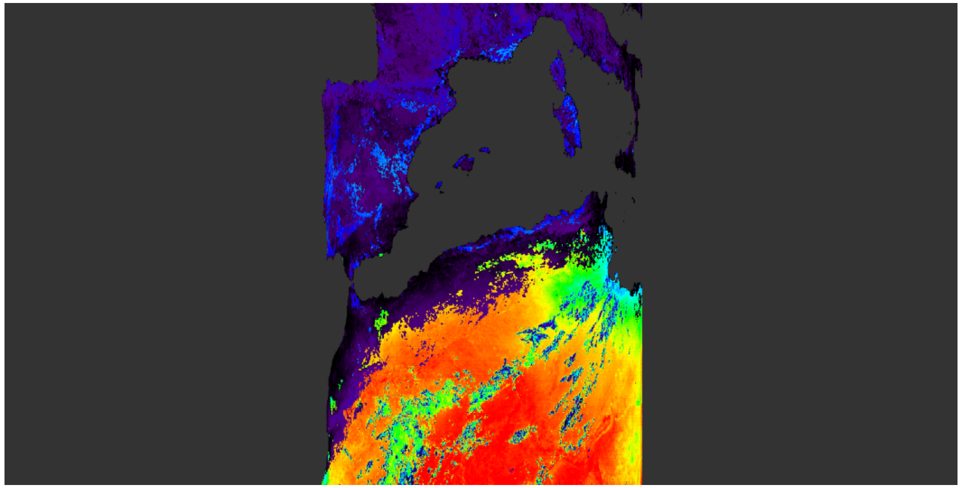
Total



Random



Locally correlated - atmosphere



Locally correlated - surface

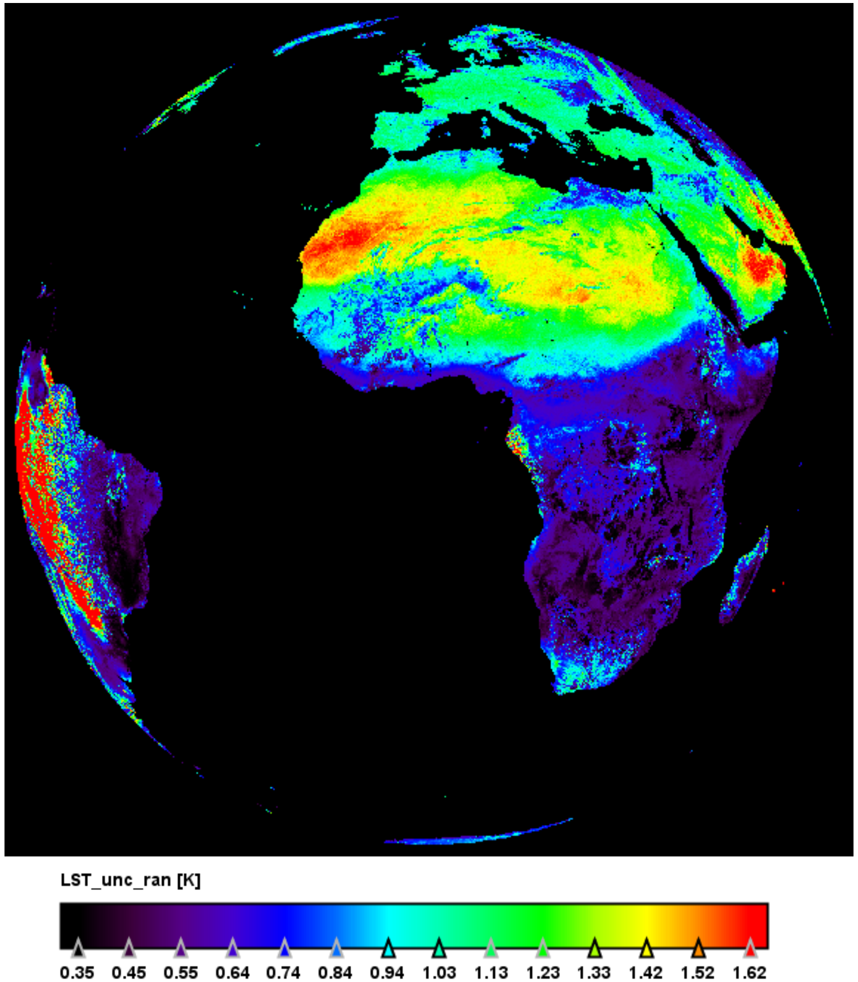
UNIVERSITY OF LEICESTER

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

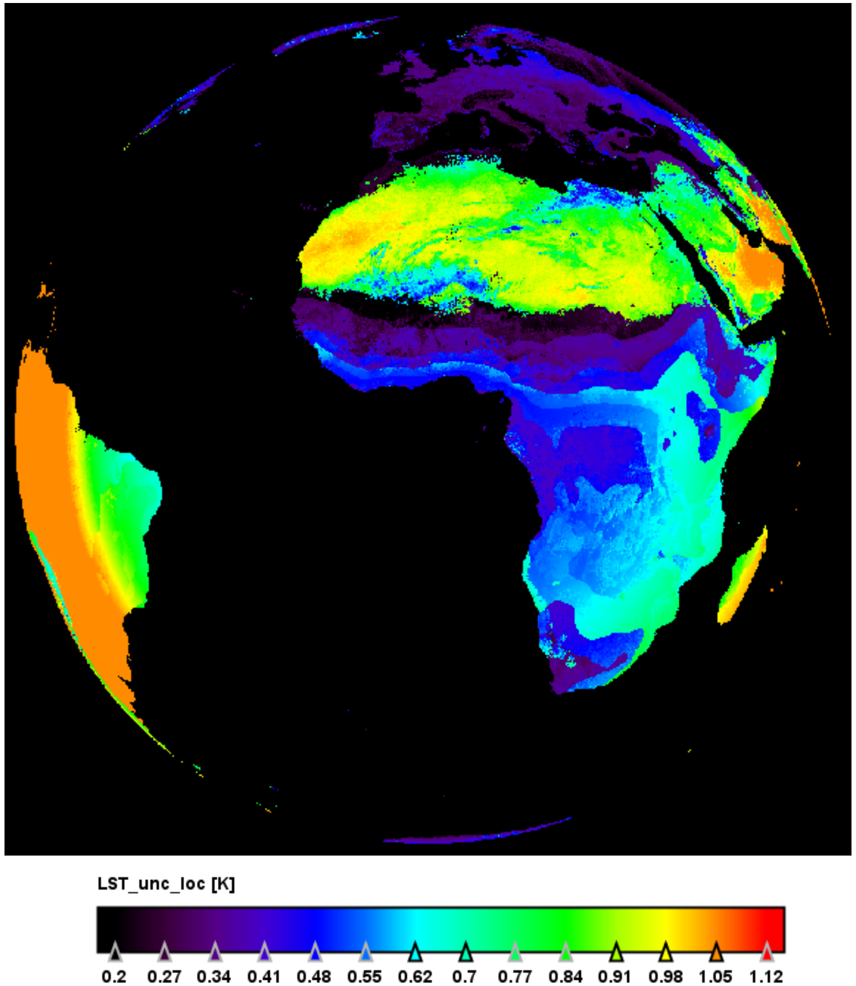
EUSTACE

SEVIRI LST Uncertainties

Random



Locally correlated

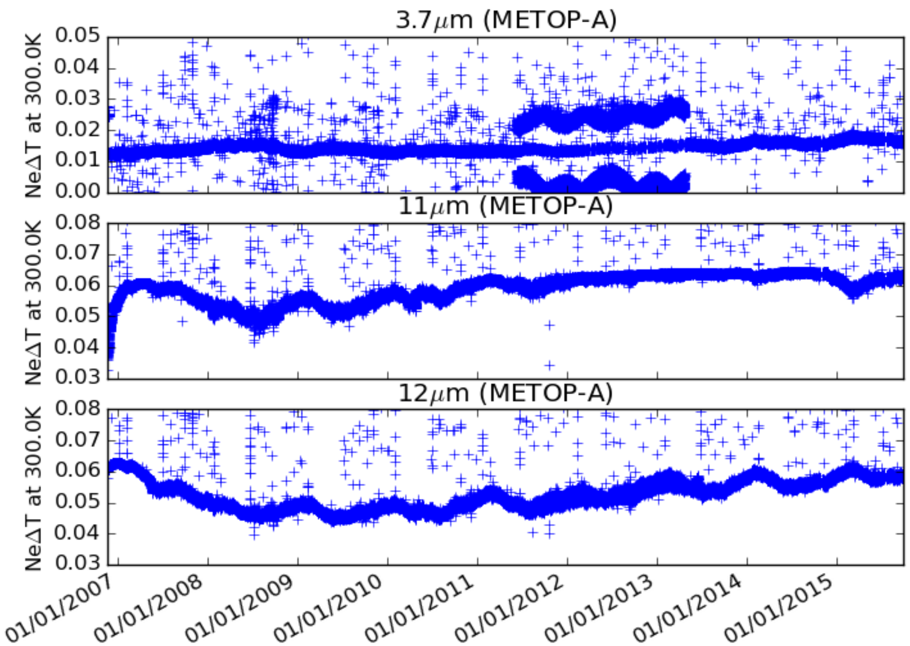


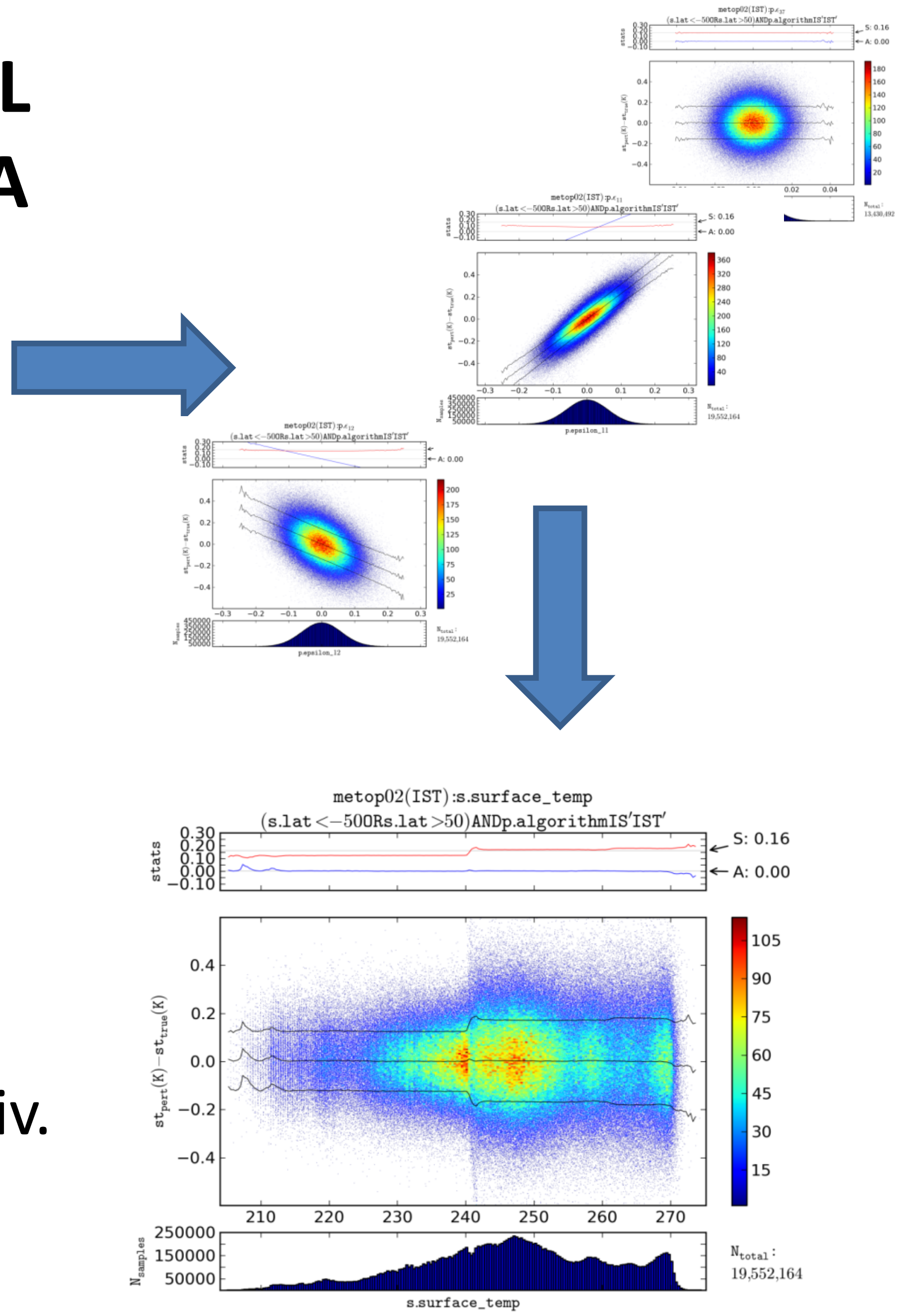
UNIVERSITY OF LEICESTER

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

EUSTACE

ICE ST UNCERTAINTY MODEL
NOISE IN MEOP-A- METOP-A





- NEdTs obtained from Jon Mittaz, Univ. Reading.
- Noise propagated through algorithm **by perturbation** – but same philosophy

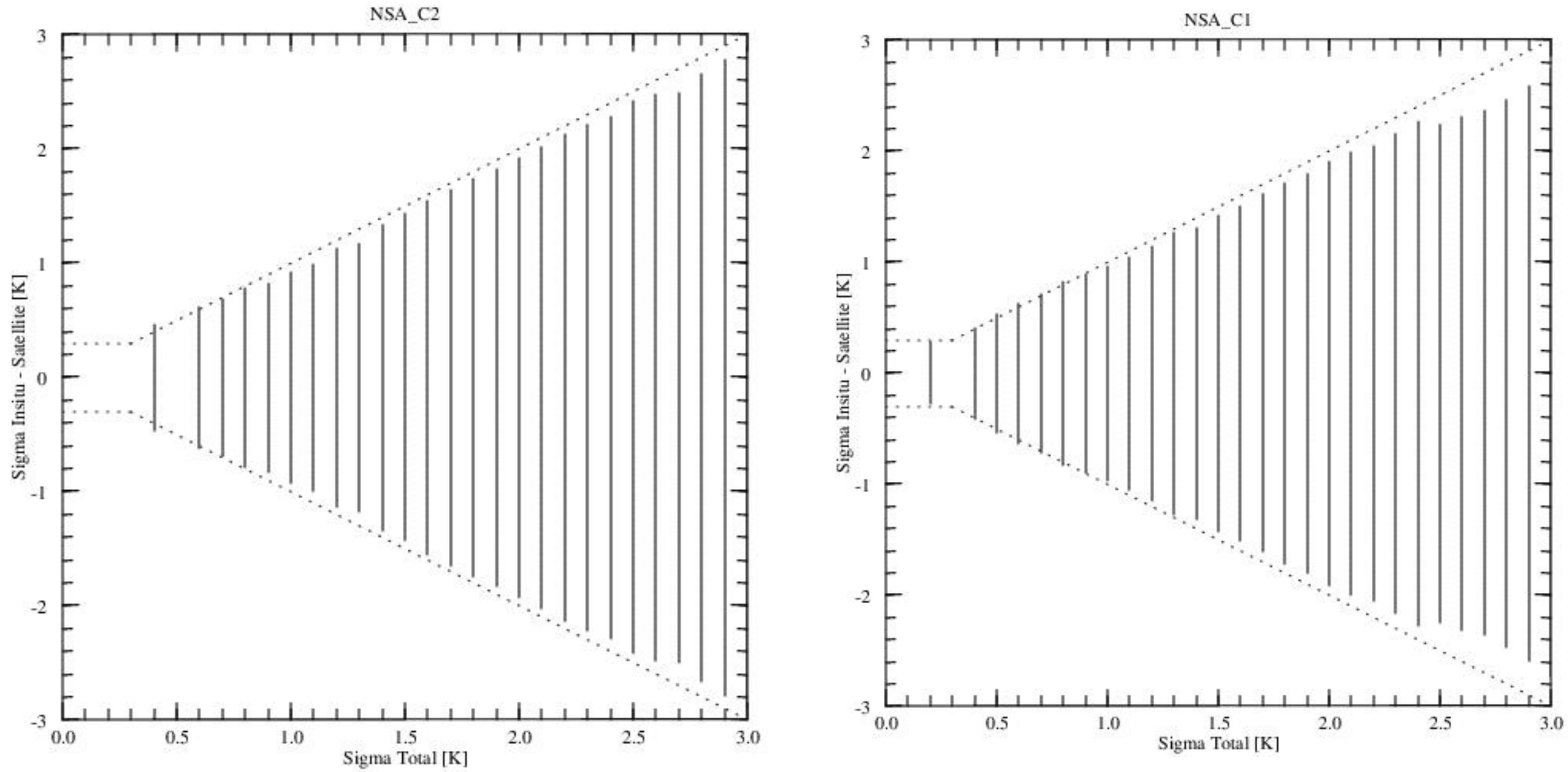
EUSTACE

VALIDATION OF SATELLITE UNCERTAINTIES

- Test the goodness-of-fit between the uncertainty from in situ validation ($\sigma_{sat\text{-}ground}$) and the total satellite product uncertainty for each associated matchup (σ_{total})
- $\sigma_{total} = \sqrt{\sigma_{sat}^2 + \sigma_{ground}^2 + \sigma_{space}^2 + \sigma_{time}^2 (+\sigma_{depth}^2)}$
- σ_{sat} is the total LST uncertainty for satellite pixel
- σ_{ground} is the uncertainty associated with the ground-based measurement
- σ_{space} is the uncertainty associated with matching a satellite and ground observation in a spatial context
- σ_{time} is the uncertainty associated with matching a satellite and ground observation in time
- σ_{depth} is the uncertainty due to the difference in depth of the measurements (SST only)

IST UNCERTAINTY VERIFICATION

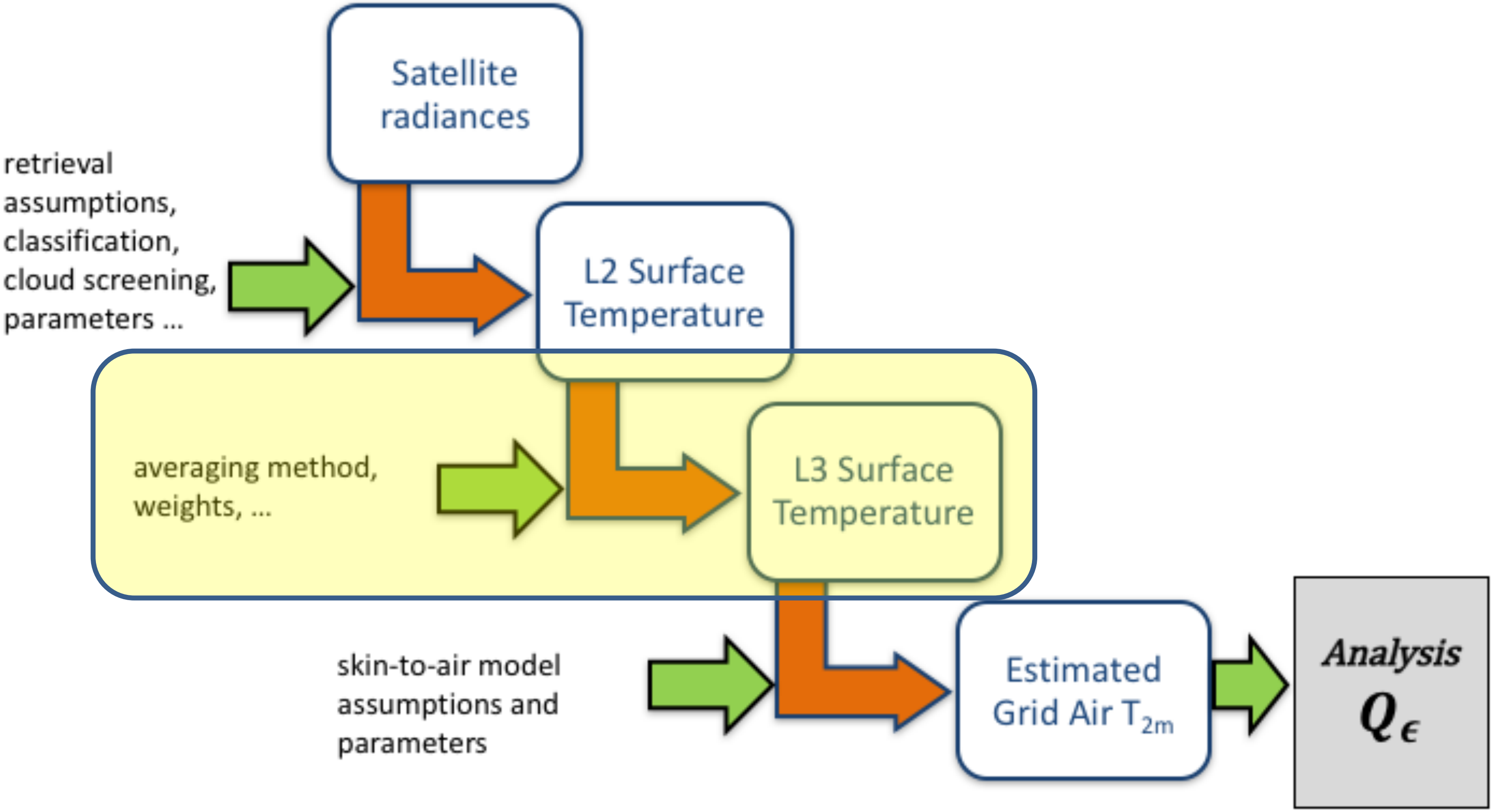
- Validation against independent radiometer observations from ARMS



AASTI IST uncertainty validation with respect to ARM in-situ data for 2008. Dashed lines show ideal uncertainty model accounting for uncertainties in the in situ data and geophysical uncertainties arising from spatial and temporal collocation. Solid black lines show one standard deviation of the retrieved minus in situ IST differences for each 0.1 K bin.

EUSTACE

Sources and propagation of uncertainty



EUSTACE

PROPAGATION OF L2 -> L3 UNCERTAINTIES

LST: Uncertainties propagated from 1km LST pixels (*upixel*) -> global 0.25 grid LST (*ucell*):

$$ucell_{random} = \sqrt{\frac{1}{n_{clear}}\left(\frac{\sum upixel_{random}}{n_{clear}}\right) + \frac{n_{cloud}VAR_{LST}}{n_{clear}+n_{cloud}-1}}$$

Uncertainty due to noise, etc

Sampling uncertainty (e.g. cloud)

RANDOM

$$ucell_{surface} = \sqrt{\frac{\sum upixel_{surface}}{n_{clear}}}$$

$$ucell_{atm} = \sqrt{\frac{\sum upixel_{atm}}{n_{clear}}}$$

Locally correlated uncertainties due to surface and atmosphere corrections

LOCAL

0.1 K

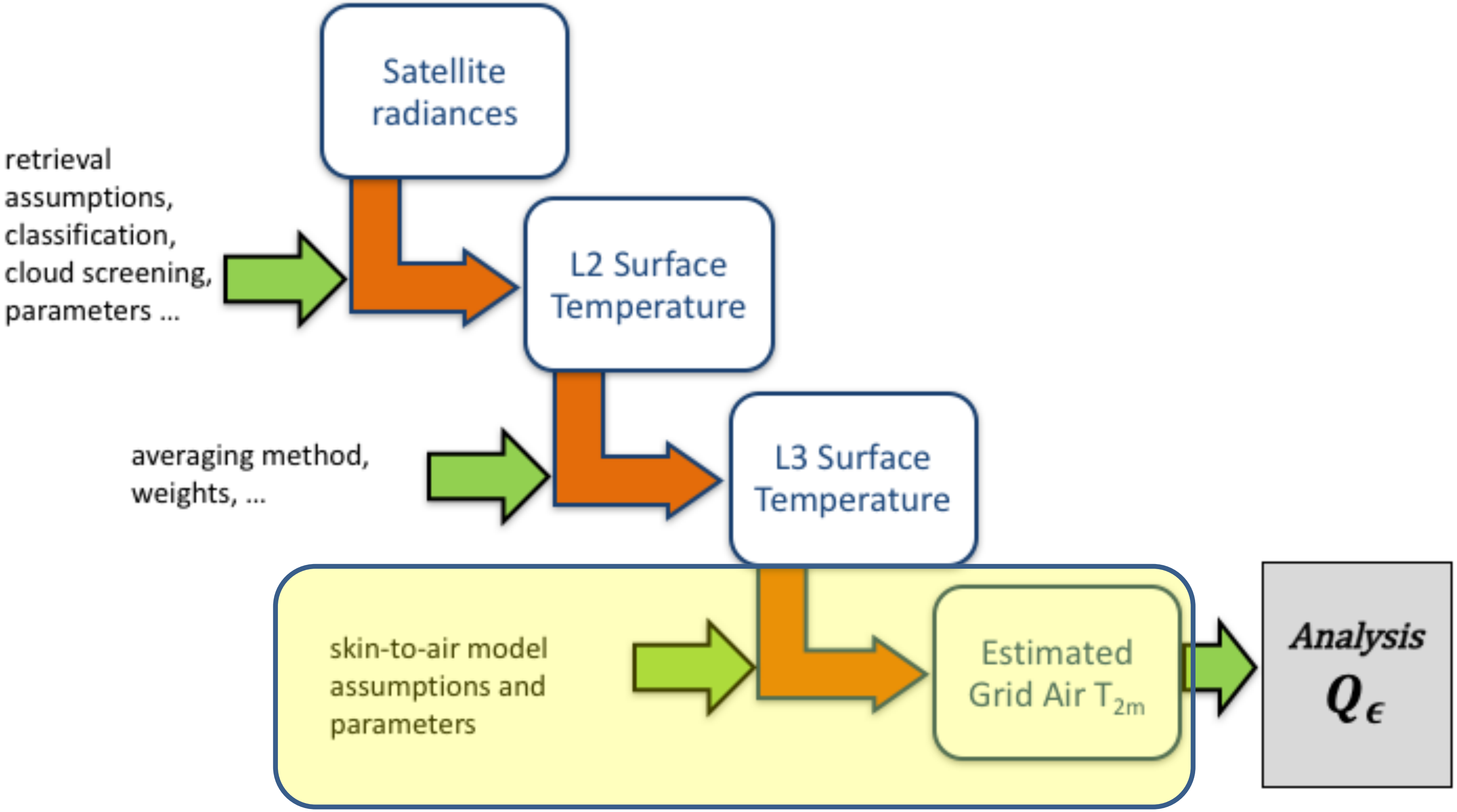
Global, systematic uncertainty

SYSTEMATIC

Met Office Hadley Centre

EUSTACE

Sources and propagation of uncertainty



SIMPLE AIR-SEA TEMPERATURE DIFFERENCE MODEL

THE THING DESIRED

Climatological Offset – Fourier components

$$\text{MAT} = \boxed{\text{SST} + \delta} + \boxed{\varepsilon}$$

Measured by ships only

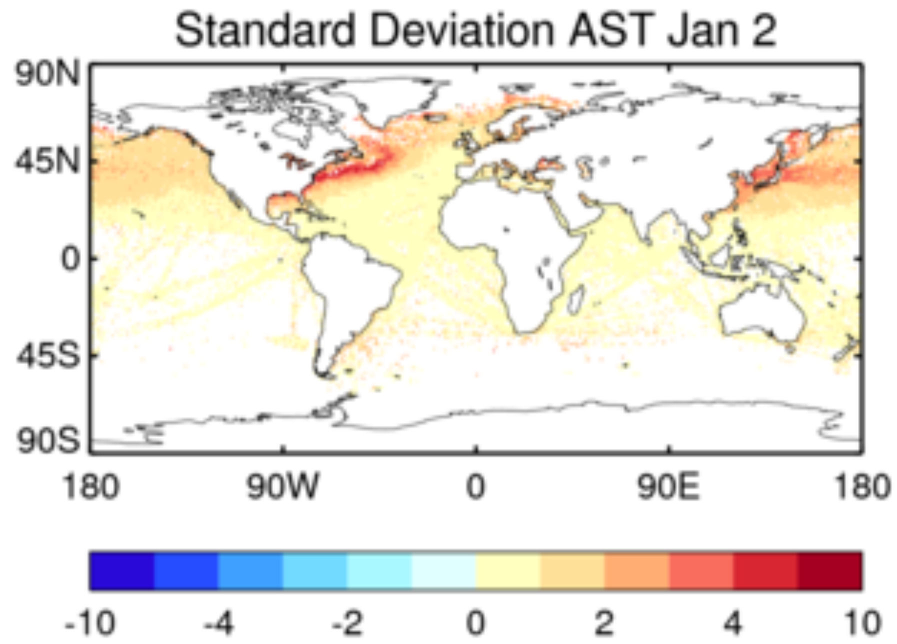
Measured by satellite/ship

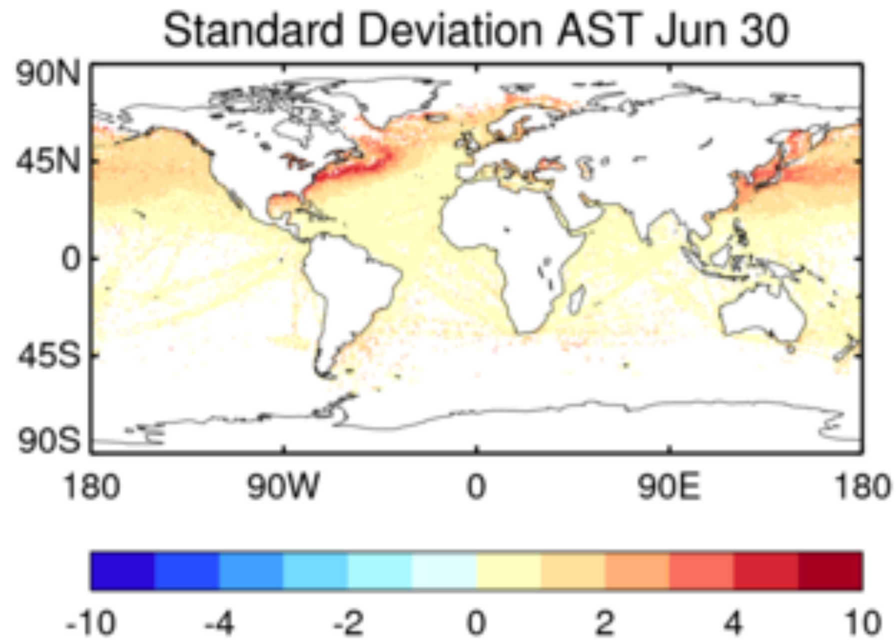
Temporally and spatially correlated variability

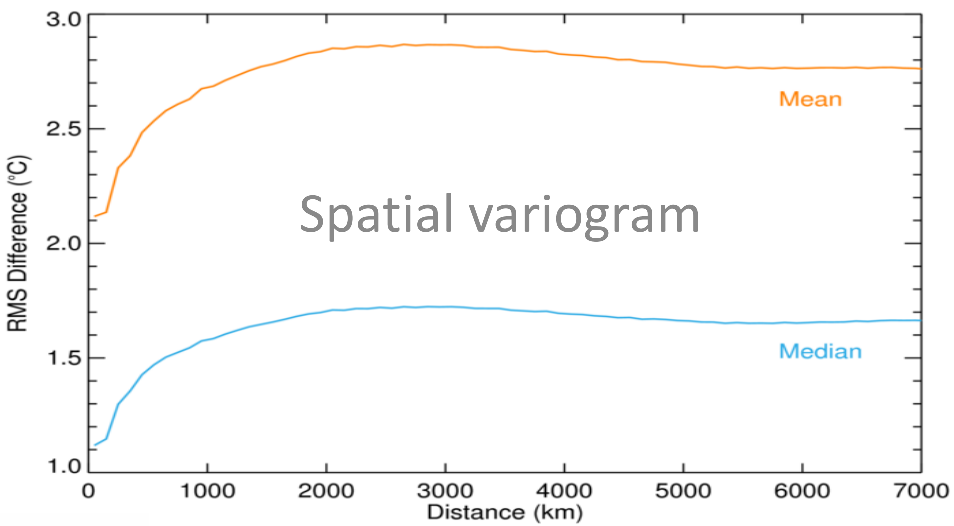
Data from ICOADS 2.5
1963-2000

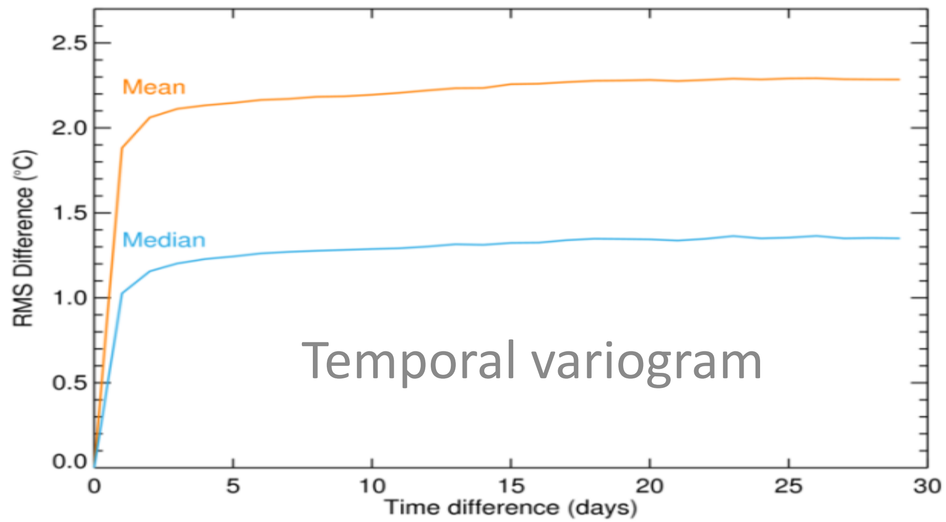
Met Office
Hadley Centre

EUSTACE









Met Office
Hadley Centre

EUSTACE

UNCERTAINTY IN SATELLITE LSAT

- Simple multiple linear regression model to estimate LSAT:

$$T_{max}=\alpha_0+\alpha_1.LST_{day}+\alpha_2.LST_{ngt}+\alpha_3.FVC+\alpha_4.SZA_{noon}+\alpha_5.Snow+\varepsilon_{Tmax}$$

$$T_{min}=\beta_0+\beta_1.LST_{day}+\beta_2.LST_{ngt}+\beta_3.FVC+\beta_4.SZA_{noon}+\beta_5.Snow+\varepsilon_{Tmin}$$

Vegetation fraction

Solar Zenith angle at noon

Error

Random

$$Tmax_{ucell\_random}=(\alpha_1{}^2.LST_{day\_ucell\_random}{}^2+\alpha_2 2.LST_{ngt\_ucell\_random}{}^2+\alpha_3{}^2.FVC_{ucell\_random}{}^2)^{½}$$

$$Tmin_{ucell\_random}=(\beta_1{}^2.LST_{day\_ucell\_random}{}^2+\beta_2{}^2.LST_{ngt\_ucell\_random}{}^2+\beta_3{}^2.FVC_{ucell\_random}{}^2)^{½}$$

Atmosphere

$$Tmax_{atm}=(\alpha_1{}^2.LST_{day\_ucell\_atm}{}^2+\alpha_2{}^2.LST_{ngt\_ucell\_atm}{}^2+\sigma Tmax^2)^{½}$$

$$Tmin_{atm}=(\beta_1{}^2.LST_{day\_ucell\_atm}{}^2+\beta_2{}^2.LST_{ngt\_ucell\_atm}{}^2+\sigma Tmin^2)^{½}$$

σTmax /σTmin = StDev of model residuals

Surface

$$Tmax_{surf}=(\alpha_1{}^2.LST_{day\_ucell\_surf}{}^2+\alpha_2{}^2.LST_{ngt\_ucell\_surf}{}^2+\alpha_3{}^2.FVC_{ucell\_local}{}^2)^{½}$$

$$Tmin_{surf}=(\beta_1{}^2.LST_{day\_ucell\_surf}{}^2+\beta_2{}^2.LST_{ngt\_ucell\_surf}{}^2+\beta_3{}^2.FVC_{ucell\_local}{}^2)^{½}$$

Met Office
Hadley Centre

EUSTACE

EXAMPLE UNCERTAINTY FIELDS
(1 JULY 2010)

Tmin$_{rando}$

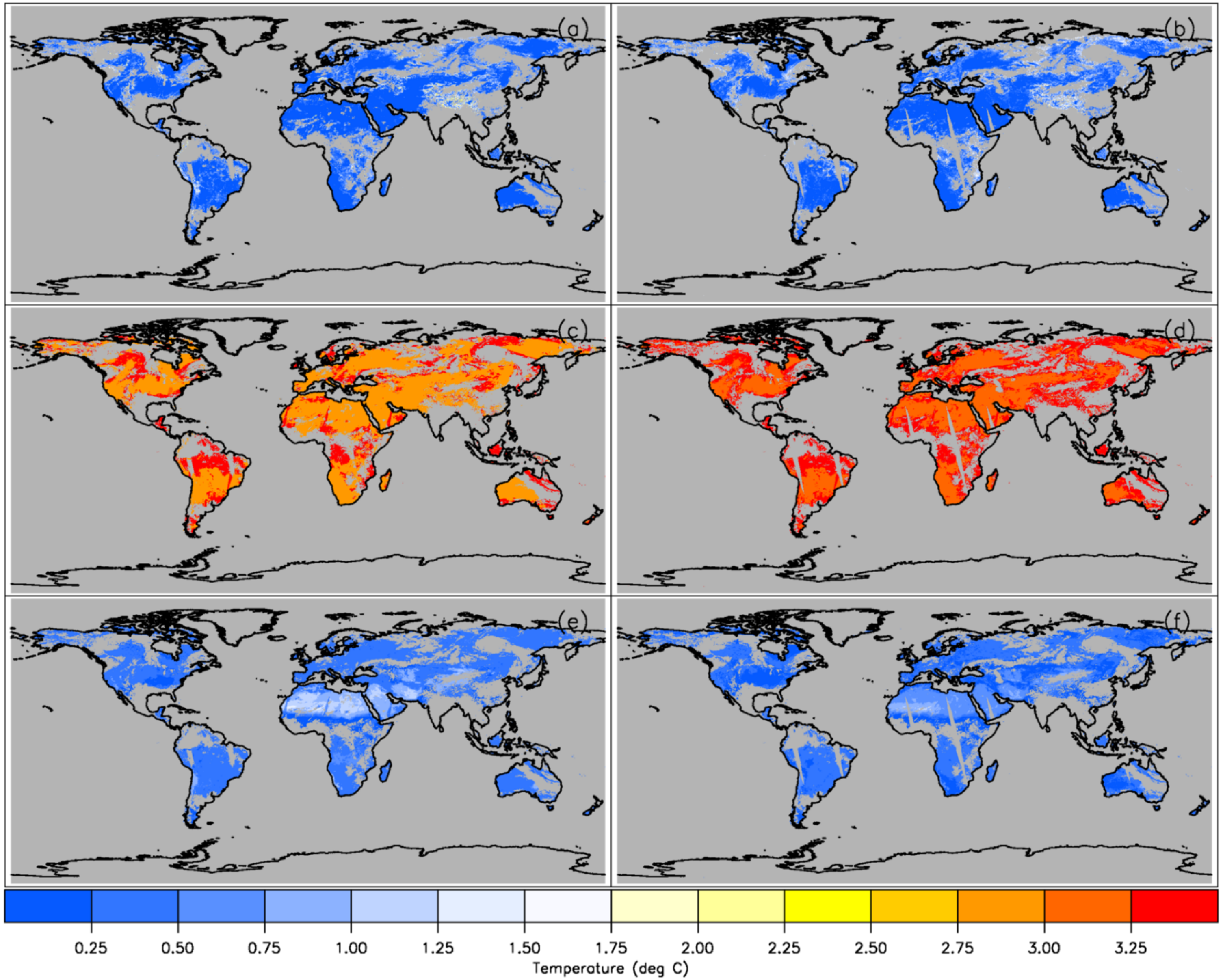
Tmax$_{random}$

Tmin$_{atm}$

Tmax$_{atm}$

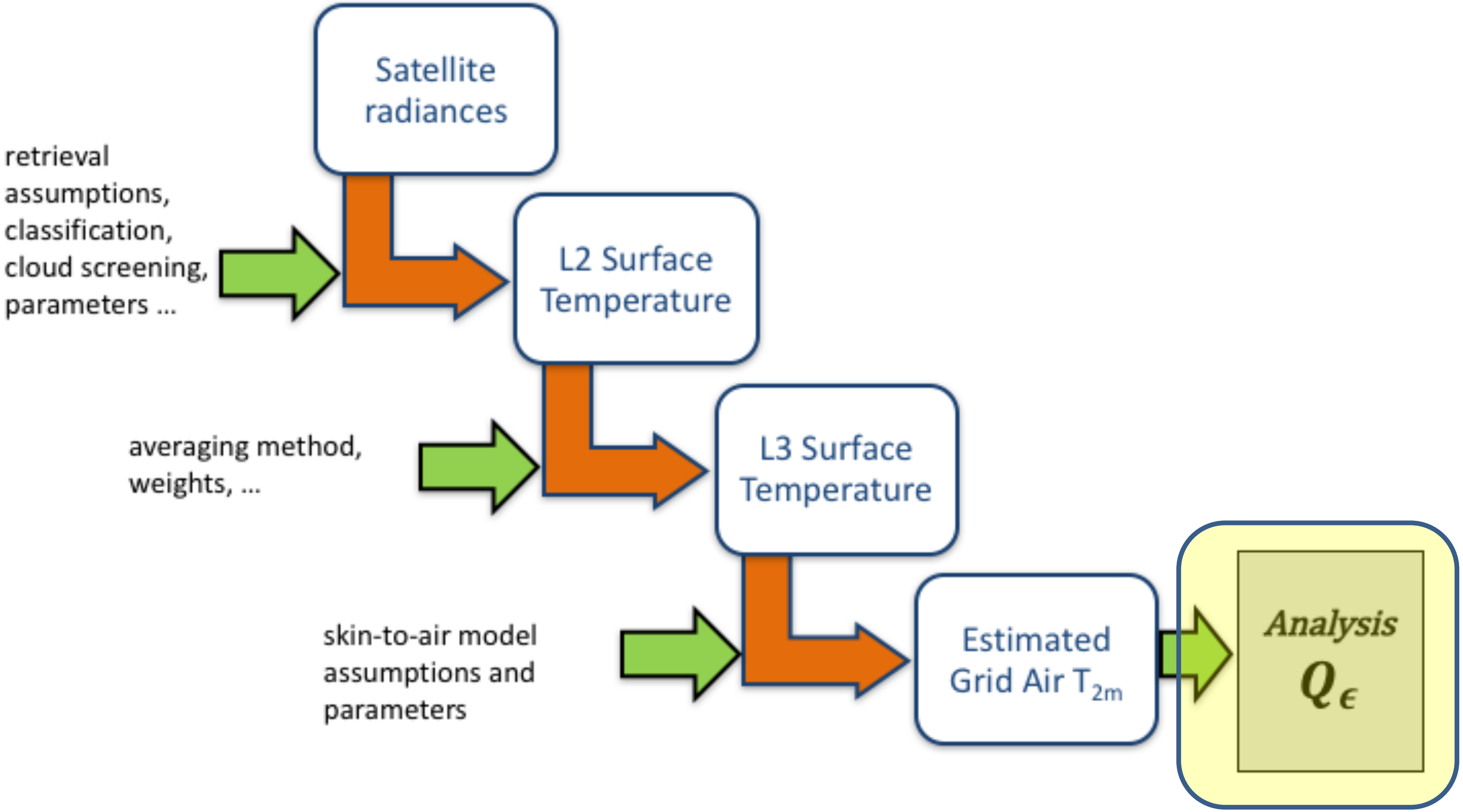
Tmin$_{surf}$

Tmax$_{surf}$



Met Office
Hadley Centre

Sources and propagation of uncertainty



EUSTACE

Advanced Standard Air Temperature Model

Temperature Process Decomposition

- Temperature variability is decomposed into model sub-components with defined structure in space/time:

$$\boldsymbol{T}(\boldsymbol{s},t) = \boldsymbol{T}^{\mathrm{clim}}(\boldsymbol{s},t) + \boldsymbol{T}^{\mathrm{large}}(\boldsymbol{s},t) + \boldsymbol{T}^{\mathrm{local}}(\boldsymbol{s},t)$$

$\boldsymbol{T}(\boldsymbol{s},t)$ = Temperature at space/time location $(\boldsymbol{s},t)$

$\boldsymbol{T}^{\mathrm{clim}}(\boldsymbol{s},t)$ = Climatological temperature

$\boldsymbol{T}^{\mathrm{large}}(\boldsymbol{s},t)$ = Large spatial/temporal scale component

$\boldsymbol{T}^{\mathrm{local}}(\boldsymbol{s},t)$ = Daily, short spatial scale component

Temperature Observation Model

Observation model

Daily mean air temperatures are decomposed into variability at different scales:

$$y^i = T(\boldsymbol{s}^i, t^i) + b^i + \epsilon^i$$

Where b^i is a sum of observational biases affecting observation i and ϵ^i are non-bias related observational errors.

y^i = An air temperature observation index by i

$T(\boldsymbol{s}^i, t^i)$ = Temperature at space/time location $(\boldsymbol{s}^i, t^i)$

b^i = Additive bias associated with observation i

ϵ^i = Error associated with observation i

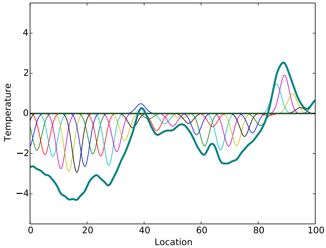
Met Office
Hadley Centre

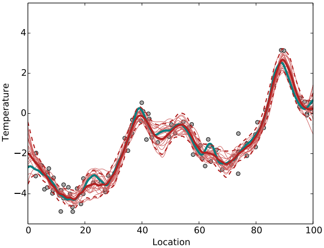
EUSTACE

Analysis method

Spatial interpolation based on the SPDE approach (Lindgren et al 2011):

- Temperatures are modelled as weighted sum of local functions.
- A Bayesian method, where variability/smoothness is controlled by a prior distribution for the weights.
- Compute the probability density function of the weights conditioned on the temperature observations.





Lindgren, F., H. Rue, J. Lindström, (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 4

Met Office Hadley Centre

EUSTACE

Analysis method

Estimation of temperatures and observation biases

Jointly estimate temperature model variables $\boldsymbol{u}$ and observation bias variables $\boldsymbol{\beta}$, with $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta})$ and observations:

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{J_u u} + \boldsymbol{J_\beta \beta} + \boldsymbol{\epsilon} \\
&= \boldsymbol{J_x x} + \boldsymbol{\epsilon}
\end{aligned}
$$

Apply Bayes' Rule to compute the analysis:

$$
\begin{aligned}
\text{Prior}: &\quad (\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu_x}, \boldsymbol{Q_x}^{-1}) \\
\text{Observations}: &\quad (\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{x}, \boldsymbol{Q_\epsilon}^{-1}) \\
\text{Posterior}: &\quad p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \propto p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) p(\boldsymbol{x} \mid \boldsymbol{\theta})
\end{aligned}
$$

Observational uncertainties are encoded into $\boldsymbol{Q_\epsilon}$. Structured errors have their structure encoded into $\boldsymbol{J_\beta}$ and their magnitude $\boldsymbol{\beta}$ is estimated.

Model Component Solutions

Linear Model

- Each of the three model components are constructed as a linear (or linearised) model, with a design matrix $\boldsymbol{J}$ and latent variables to be estimated $\boldsymbol{x}$:

$$\boldsymbol{y} = \boldsymbol{J}\boldsymbol{x} + \boldsymbol{\epsilon}$$

- Measurement error is additive Gaussian $p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_{\boldsymbol{\epsilon}}^{-1})$
- Model variables $\boldsymbol{x}$ have a Gaussian prior distribution $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{Q}_{\boldsymbol{x}}^{-1})$

Estimation

Compute the distribution of $\boldsymbol{x}$ conditioned on the observations:

$$p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}) = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}}, \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}}^{-1})$$

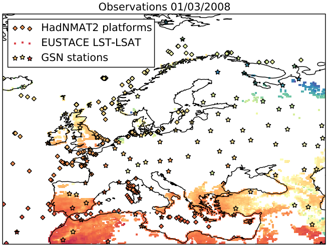
$$\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}}^{-1}\boldsymbol{J}^T\boldsymbol{Q}_{\boldsymbol{\epsilon}}(\boldsymbol{y} - \boldsymbol{J}\boldsymbol{\mu}_{\boldsymbol{x}})$$

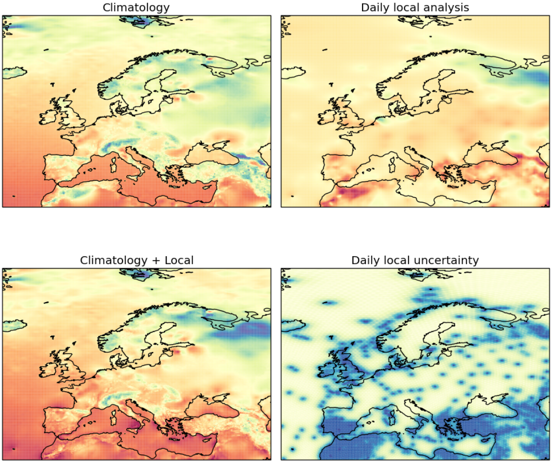
$$\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{Q}_{\boldsymbol{x}} + \boldsymbol{J}^T\boldsymbol{Q}_{\boldsymbol{\epsilon}}\boldsymbol{J}$$

Efficient solution depends on the sparse structure of $\boldsymbol{J}$, $\boldsymbol{Q}_{\boldsymbol{x}}$ and $\boldsymbol{Q}_{\boldsymbol{\epsilon}}$.

Demonstration Application

- Demonstrated on small region/subset of input data.
- Applied to in situ air temperature (HadNMAT2, GHCN Daily) and satellite LST derived air temperature.
- Climatology fitted using observations in 1961-1990.
- Placeholder uncertainty information.

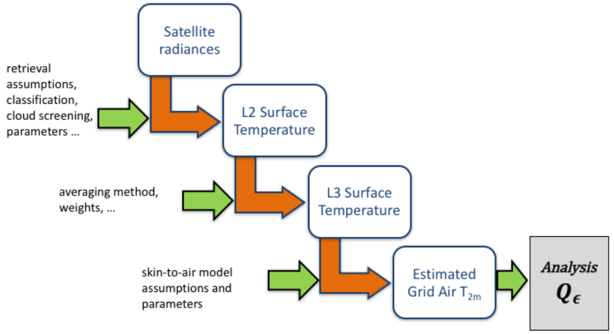




Met Office
Hadley Centre

EUSTACE

CONCLUSIONS



- EUSTACE attempting an integrated and coherent treatment of uncertainty at all levels of data
- Input data uncertainties have been estimated and validated
- Propagated and introduced uncertainty characterised as required at each step
  - L1→L2→L3→Analysis
- Coherency across observational epochs through consistent statistical treatment of uncertainty

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE