Colocation Day 3: Uncertainty Panel Discussion

Claire Bulgin, Chris Merchant, Celine Lamarche and Harjinder Sembhi

25/03/2026

Seeding Questions

- On what aspects of uncertainty quantification can the CCI projects work together?
- Are there aspects of working with uncertainty data that should be universal across ECVs e.g. uncertainties on trend estimates? Stability analysis?
- Is there scope to have a unified framework for dealing with sensor resolution in CDRs?
    - Do we need something different for classification algorithms where resolution changes the answer?
    - To what degree should we compensate for temporal changes in sensor resolution in the uncertainty budget?
- Can we calculate metrological uncertainties when using machine learning algorithms?
    - What is a useful uncertainty estimate for machine learning applications e.g. downscaling?
- How do we use uncertainty information to inform observation-model comparisons?
- How well do data users understand uncertainties? How well do we understand the uncertainty information that data users need? How do we make stronger links between data providers and data users such that we close the gap? Does this need to be application specific?

What is uncertainty and why is it important?

Uncertainty is the degree to which a measurement is in doubt, and is fundamental to using the data appropriately, particularly in a decision-making context.

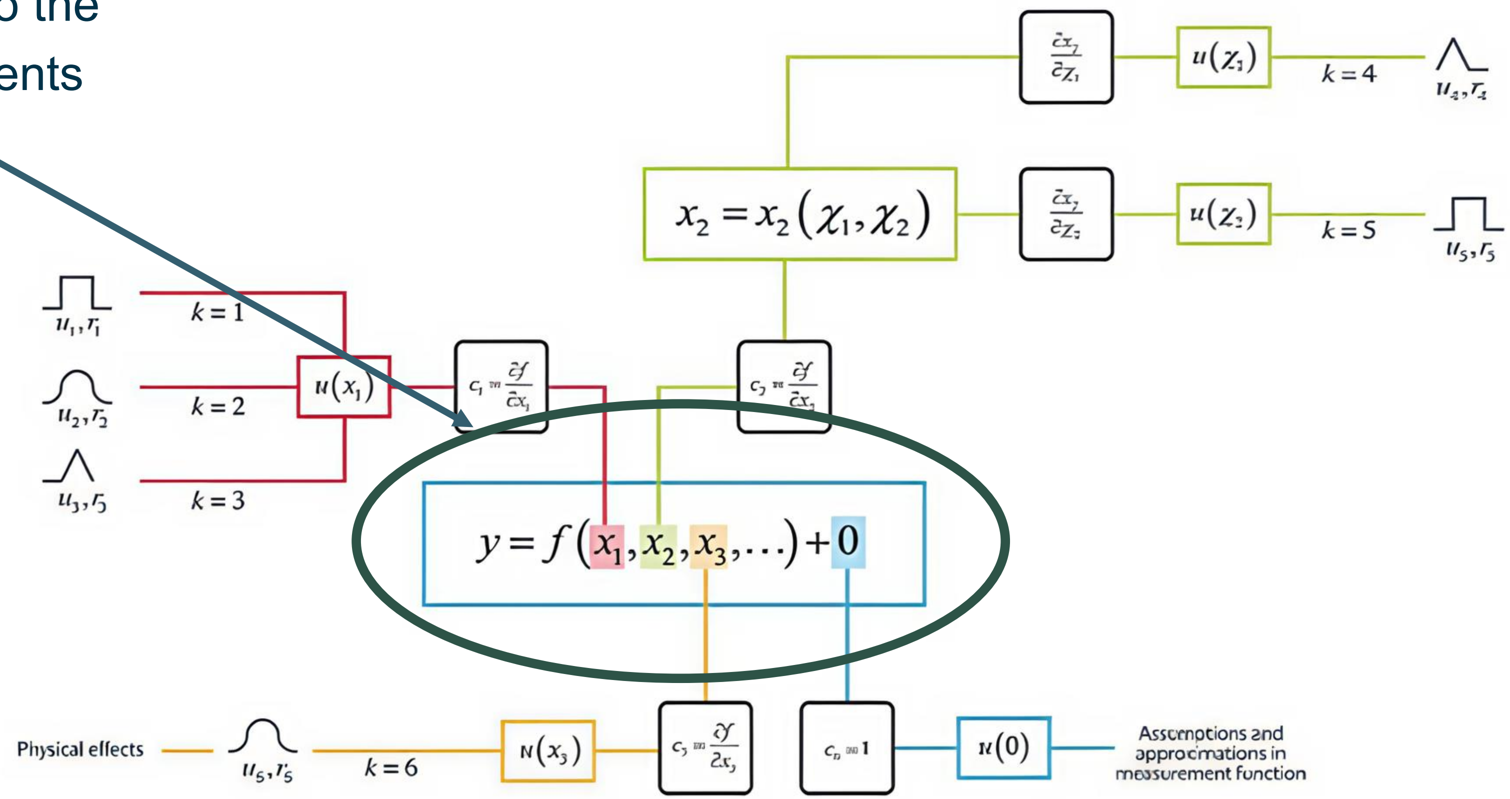
….uncertainty should not be confused with error (the degree to which a measurement differs from the truth). If we knew the error, we would correct for it and give the true measurement. Typically, we don't, so we estimate the uncertainty from the spread of the error distribution that can be attributed to a particular source of error.

An uncertainty budget can be constructed by combining the uncertainty estimates from different sources of error.

Further propagation of that uncertainty into derived products should follow metrological principals and is dependent on the correlation length scale of component of the uncertainty budget.

A metrological approach to defining uncertainty

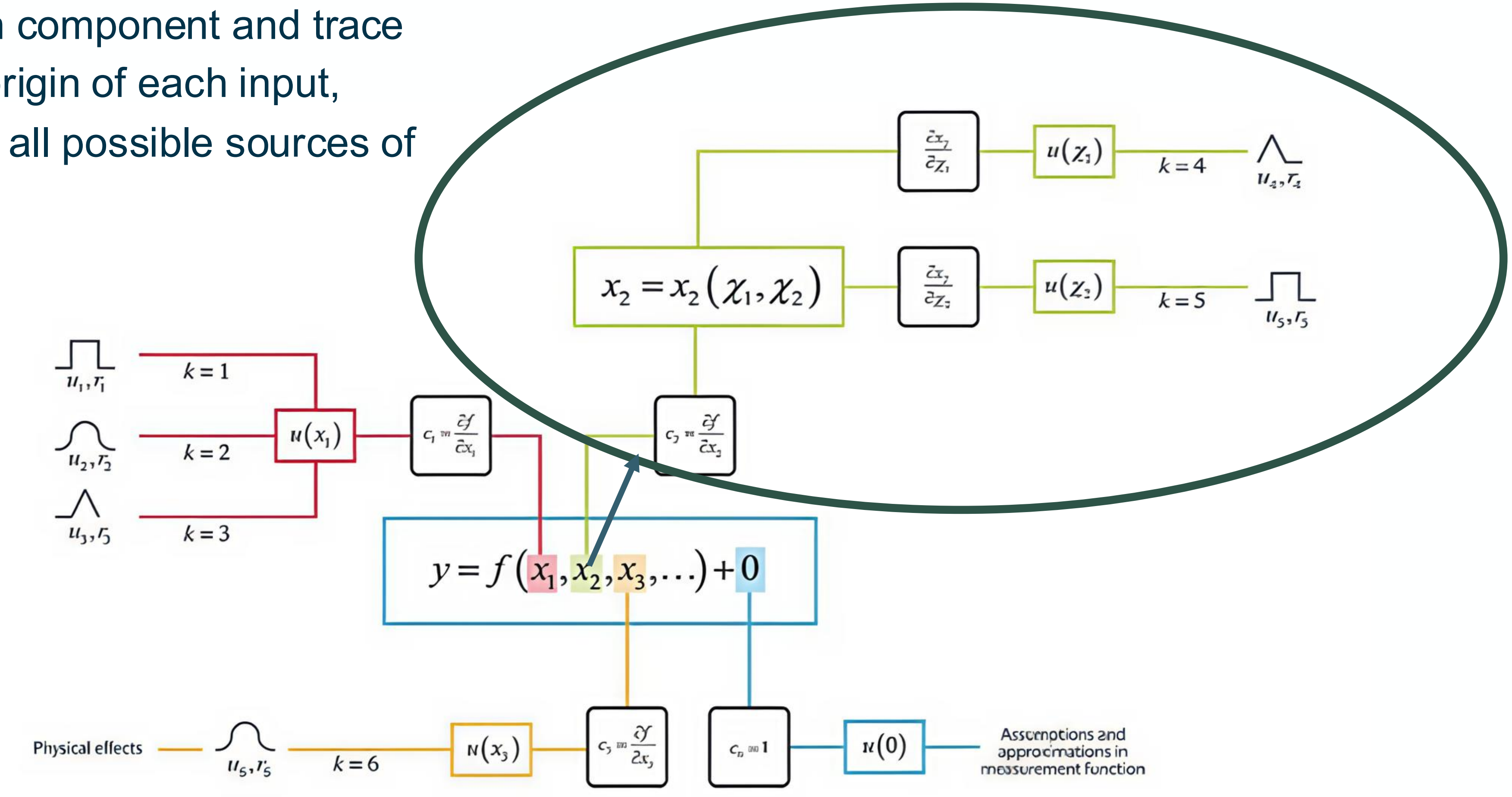
Define your measurement equation, splitting it into the different input components



Mittaz et al. (2019), Applying principles of metrology to historical Earth observations from satellites. Metrologia, 56. 032002.

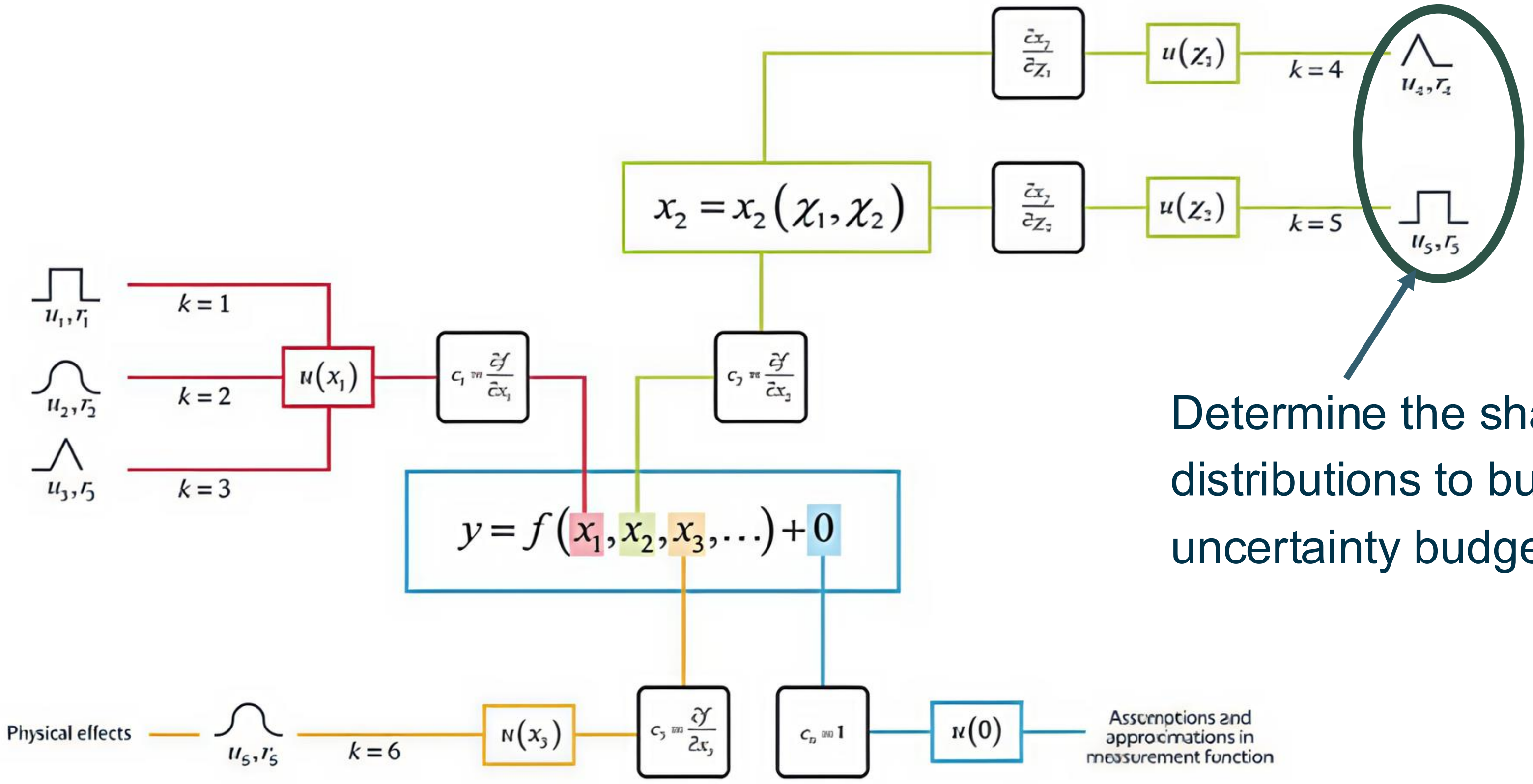
A metrological approach to defining uncertainty

Take each component and trace back the origin of each input, identifying all possible sources of error



Mittaz et al. (2019), Applying principles of metrology to historical Earth observations from satellites. Metrologia, 56. 032002.

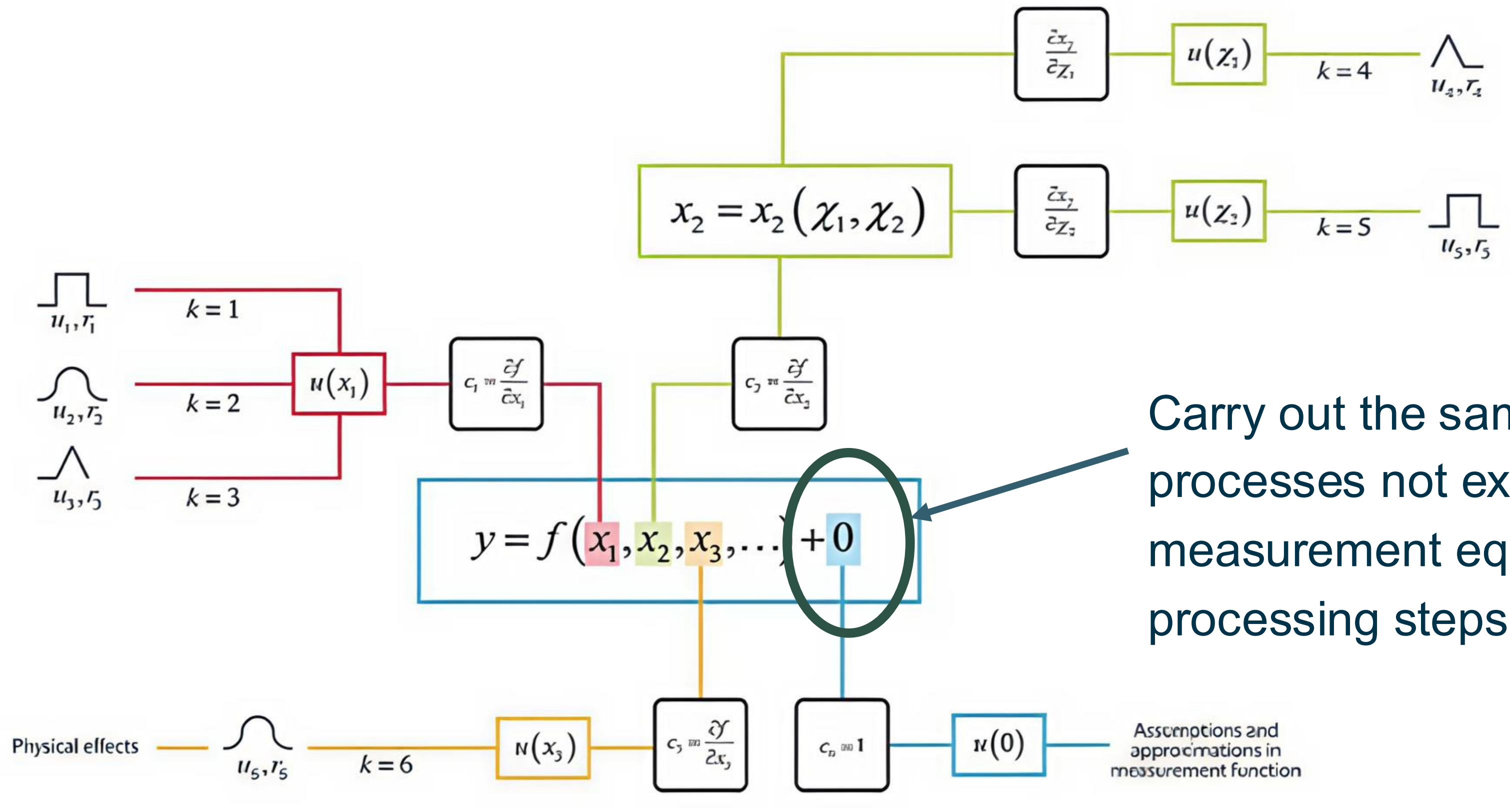
A metrological approach to defining uncertainty



Determine the shape of error distributions to build the uncertainty budget

Mittaz et al. (2019), Applying principles of metrology to historical Earth observations from satellites. Metrologia, 56. 032002.

A metrological approach to defining uncertainty



Carry out the same analysis for processes not explicit in the measurement equation e.g. pre-processing steps, classification.

Mittaz et al. (2019), Applying principles of metrology to historical Earth observations from satellites. Metrologia, 56. 032002.

What might require more consideration?

- Sometimes we have error sources that we can identify, but no means to quantify the resulting uncertainty.
- Some error sources only exist some of the time e.g. a wrong classification. If the classification is correct, then the classification error is zero, but if the classification is incorrect, this error can be large. How do we quantify an appropriate uncertainty, therefore? If we apply a large classification uncertainty to all data, this may render good data unnecessarily unusable in many applications. Classification errors also have implications for upscaling.
- What about processes that change over time? Satellite data tends to improve in quality over successive missions e.g. more channels, higher resolution. We tend to use the "best available data" at any given time but does this affect temporal stability of climate data records?
- Derived quantities may introduce new sources of uncertainty – how can we reliably calculate the uncertainty budgets for these .e.g. trends?

What might require more consideration?

- Sometimes we have error sources that we can identify, but no means to quantify the resulting uncertainty.
- Some error sources only exist some of the time e.g. a wrong classification. If the classification is correct, then the classification error is zero, but if the classification is incorrect, this error can be large. How do we quantify an appropriate uncertainty, therefore? If we apply a large classification uncertainty to all data, this may render good data unnecessarily unusable in many applications. Classification errors also have implications for upscaling.
- What about processes that change over time? Satellite data tends to improve in quality over successive missions e.g. more channels, higher resolution. We tend to use the ”best available data” at any given time but does this affect temporal stability of climate data records?
- Derived quantities may introduce new sources of uncertainty – how can we reliably calculate the uncertainty budgets for these e.g. trends?

What might require more consideration?

- Sometimes we have error sources that we can identify, but no means to quantify the resulting uncertainty.
- Some error sources only exist some of the time e.g. a wrong classification. If the classification is correct, then the classification error is zero, but if the classification is incorrect, this error can be large. How do we quantify an appropriate uncertainty, therefore? If we apply a large classification uncertainty to all data, this may render good data unnecessarily unusable in many applications. Classification errors also have implications for upscaling.
- What about processes that change over time? Satellite data tends to improve in quality over successive missions e.g. more channels, higher resolution. We tend to use the ”best available data” at any given time but does this affect temporal stability of climate data records?
- Derived quantities may introduce new sources of uncertainty – how can we reliably calculate the uncertainty budgets for these e.g. trends?

What might require more consideration?

- Sometimes we have error sources that we can identify, but no means to quantify the resulting uncertainty.
- Some error sources only exist some of the time e.g. a wrong classification. If the classification is correct, then the classification error is zero, but if the classification is incorrect, this error can be large. How do we quantify an appropriate uncertainty, therefore? If we apply a large classification uncertainty to all data, this may render good data unnecessarily unusable in many applications. Classification errors also have implications for upscaling.
- What about processes that change over time? Satellite data tends to improve in quality over successive missions e.g. more channels, higher resolution. We tend to use the ”best available data” at any given time but does this affect temporal stability of climate data records?
- Derived quantities may introduce new sources of uncertainty – how can we reliably calculate the uncertainty budgets for these e.g. trends?

Cloud masking stability affecting LST stability

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL
land surface temperature cci

- This figure illustrates how cloud mask instability can cause apparent trends in LST over time that are an artefact of the pre-processing.
- Comparisons are made at three ceilometer sites between the "true LST" – where the ceilometer sees clear-sky and clear-sky conditions as defined by three different cloud masking algorithms.

Table 7
LST stability in kelvin per decade arising from cloud detection instability, for the UoR, UoL and Oper algorithms at the NY, NSA and SGP ceilometer locations.

	NY	NSA	SGP
UoR	0.4	−0.73	0.21
UoL	0.38	0.18	0.5
Oper	0.36	0.01	0.29

- GCOS stability target is 0.3 K/decade at the threshold level.

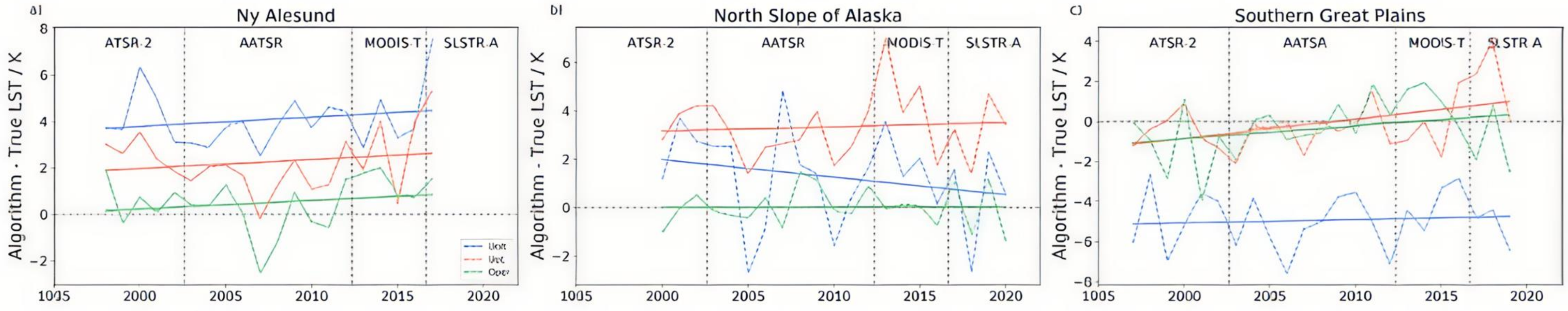


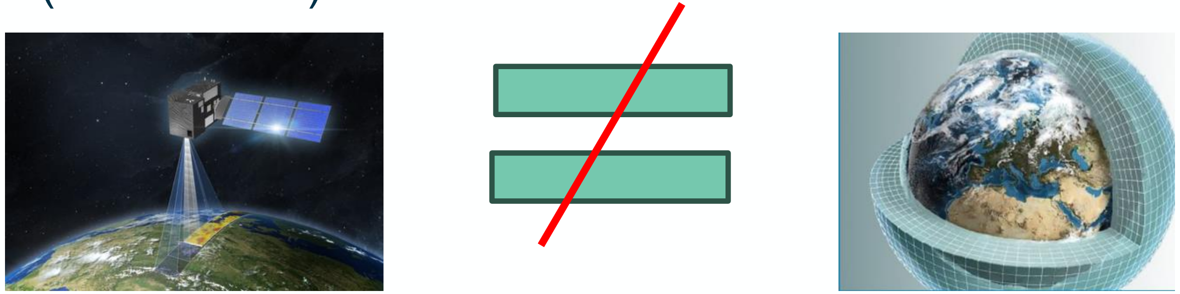
Fig. 9. LST stability for the satellite-ceilometer matches at NY, NSA and SGP. Timeseries show the 'algorithm-specific' LST anomalies minus the 'true' LST anomalies. Stability is determined using a linear fit to the resulting difference.

Bulgin et al (2024). Stability of cloud detection methods for Land Surface Temperature (LST) Climate Data Records (CDRs). Remote Sensing of Environment. 315. 114440.

Comparing observations and models

How do we use uncertainty information to inform observation-model comparisons?

- Firstly, are we comparing the same thing from model and observations? e.g. spatially complete fields for a given timestep (satellite observations rarely come in this format!)? Derived diagnostics vs retrieved variables.
- Observations and models both come with uncertainties, but these are rarely expressed in the same way (per pixel uncertainties vs ensembles). Should we generate more ensemble-based observational datasets? e.g. HadISST?
- The comparison process introduces additional uncertainty when you compare two datasets that represent different things. We can build globally-complete datasets from observations, but we need to think about how these are constructed relative to model outputs.
- Can we use machine learning to better map between the sampling space of observations and models for more representative comparison? (ESO4Clima).



https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Carbon_dioxide_monitoring_satellite_given_the_shakes

https://www.dkrz.de/en/projects-and-partners/projects/focus/new-climate-model

Communicating uncertainties with data users

How do we best communicate uncertainties with data users?

- Uncertainties are critical to making informed decisions where data form part of a decision-making process.
- Users can be diverse, ranging from scientists through to those without a strong scientific/mathematical background. The latter often just want to know "which data should I use"?
- Some progress has been made in demonstrating the benefit of uncertainties to scientific users, but most users are not confident in propagating uncertainties into derived products, beyond what data producers provide.
- Small advances have been made in making uncertainties more accessible to users e.g. adding LST uncertainties to ESMValTool for the purpose of re-gridding data. Integrating uncertainties across multiple ECVs into tools such as ESMValTool with the ability to propagate them into numbers of derived products still requires some significant effort.
- In some cases, user requirements may change how we frame uncertainty estimates e.g. probability of detection requirement for MEDUSA.
- How can we strengthen links between data providers and data users to ensure that uncertainty information is being used appropriately and aiding decision making?

Surveys in Geophysics Special Issue

Remote Sensing in Climatology – Essential Climate Variables and their Uncertainties

Special Issue published following an ISSI workshop in Dec 2024.

https://link.springer.com/collections/ghddcijcdb

Topics covered:

- Lost in translation – the need for common vocabularies in Earth Observation
- Metrological framework for uncertainty in satellite and in-situ observations
- Practical introduction to utilising uncertainty information
- Making sense of uncertainties – asking the right question
- Importance of scale in uncertainties and how we communicate this to users
- Stability specifications for CDRs and applicability to trend uncertainty estimation
- Trend uncertainty estimation
- Uncertainty quantification from variables derived from deep learning
- Error covariance structure from atmospheric correction in optical sensors
- Challenges and limitations of validating satellite datasets with independent measurements

Seeding Questions

- On what aspects of uncertainty quantification can the CCI projects work together?
- Are there aspects of working with uncertainty data that should be universal across ECVs e.g. uncertainties on trend estimates? Stability analysis?
- Is there scope to have a unified framework for dealing with sensor resolution in CDRs?
    - Do we need something different for classification algorithms where resolution changes the answer?
    - To what degree should we compensate for temporal changes in sensor resolution in the uncertainty budget?
- Can we calculate metrological uncertainties when using machine learning algorithms?
    - What is a useful uncertainty estimate for machine learning applications e.g. downscaling?
- How do we use uncertainty information to inform observation-model comparisons?
- How well do data users understand uncertainties? How well do we understand the uncertainty information that data users need? How do we make stronger links between data providers and data users such that we close the gap? Does this need to be application specific?

Surveys in Geophysics
https://doi.org/10.1007/s10712-025-09922-7

Check for updates

A Unified Framework for Trend Uncertainty Assessment in Climate Data Records: Demonstration on Global Mean Sea Level

Kevin Gobron[1,2] · Roland Hohensinn[3,4] · Xavier Loizeau[5] · Claire E. Bulgin[6,7] · Christopher J. Merchant[6,7] · Emma R. Woolliams[5] · Maurice G. Cox[5] · Wouter Dorigo[8] · Thomas Howard[5,9] · Mary Langsdale[10] · Adam C. Povey[11] · Michaël Ablain[12] · Janusz Bogusz[13] · Alexander Gruber[8] · Anna Klos[13] · Jonathan Mittaz[6]

Received: 12 May 2025 / Accepted: 18 December 2025

Abstract
Trends of essential climate variables are often estimated from climate data records to quantify changes in the Earth system. An understanding of the uncertainty in a trend is essential for accurately determining the significance of a trend and attributing its causes. Despite this importance, trend-uncertainty estimates rarely account for all known sources of uncertainty. Common approaches neglect measurement-system instability or neglect the

Surveys in Geophysics
https://doi.org/10.1007/s10712-026-09941-y

Check for updates

Stability Specifications for Climate Data Records: Their Meaning and Application in Evaluating Geophysical Trend Uncertainty

Christopher J. Merchant[1] · Emma R. Woolliams[2] · Wouter Dorigo[3] · Claire E. Bulgin[1] · Kevin Gobron[4,5] · Roland Hohensinn[6,7] · Xavier Loizeau[2] · Connor P. J. Tynan[2]

Received: 18 September 2025 / Accepted: 4 February 2026

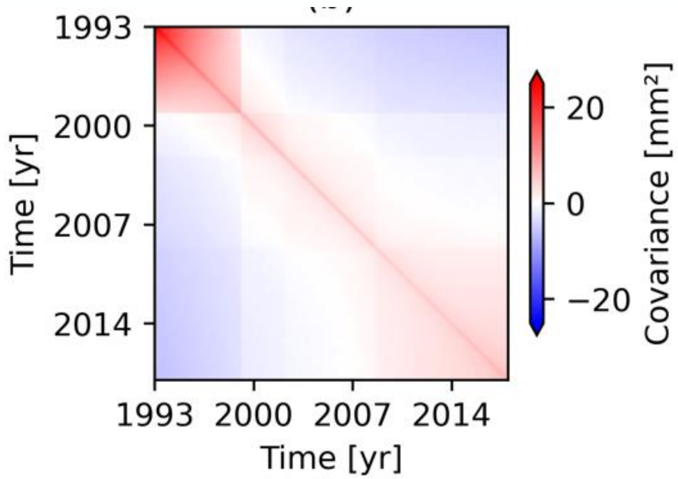
Abstract
When quantifying changes over time in the natural environment, the stability of the observations used should be considered. Stability conceptually refers to how accurately true geophysical changes and trends are reflected in observational data. We argue the need for a better approach to defining and quantifying stability consistently across climate data records. We propose that the appropriate stability metric is the stability uncertainty for specified spatial and temporal scales. We formally define stability uncertainty by analogy with metrological measurement uncertainty. Informally, stability uncertainty informs data analysts about the plausible magnitude of a non-geophysical contribution to trend values arising solely from the observing system. Neglecting the stability uncertainty leads to overconfident assessment of the significance of geophysical trends inferred from observations.

Full trend uncertainty needs metrology and empirical timeseries modelling

Necessary to avoid underestimation of trend uncertainty

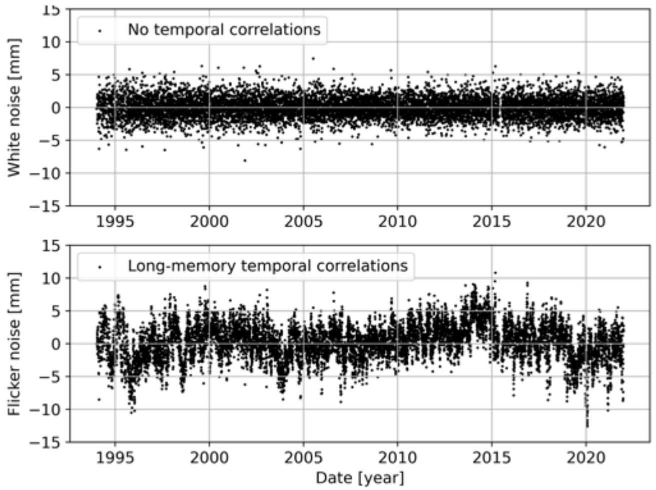
Complete covariance description:

Observation error covariance



AND

Empirical timeseries analysis of residuals



Measured values : geophysical trend+ trend in errors + unfitted errors (noise) + unfitted geophysics

$\underbrace{\text{trend in errors + unfitted errors (noise)}}$: **Trend uncertainty accounted for by the metrology of the observations**

$\underbrace{\text{unfitted geophysics}}$: **Trend uncertainty accounted for by modelling a stochastic process**

Recommendations to GCOS on stability

(1) Quantify the stability concept as ”stability uncertainty”

(2) … to be quantified over a stated temporal scale

(3) … to be quantified over a stated spatial scale

(4) … with a stated coverage factor (e.g., “stability standard uncertainty”)

(5) Foster efforts by data producers to evaluate and communicate the stability uncertainty

(6) Foster understanding by users that the stability uncertainty should inform their interpretation of trends in datasets, avoiding overconfidence in trend estimates and budget closures

Quantifying methane emission source rates & uncertainties

International agreements → rapid reduction, control of fugitive emissions

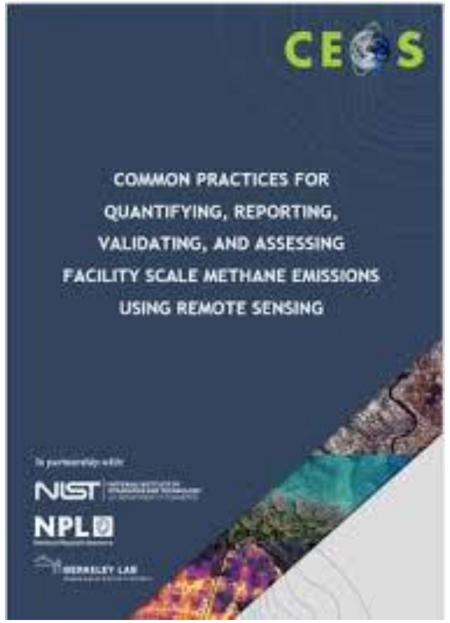
Community-accepted methodology for quantifying emissions → **measurements need to be robust, reproducible, and actionable.**

→ Feed into reporting, driven by policy and regulation

→ Need transparency, traceability, independence, and evidenced Quality Assurance (QA) → ensure fit-for-purpose data

Global Methane Pledge

Users want to take action.
1) Leak Detection And Repair (LDAR), 2) Unknown emissions identification
3) Infrastructure optimisation, 4) Monitoring of national emission inventories, 5) Enforcement and inspections, 6) Third-party verifications….



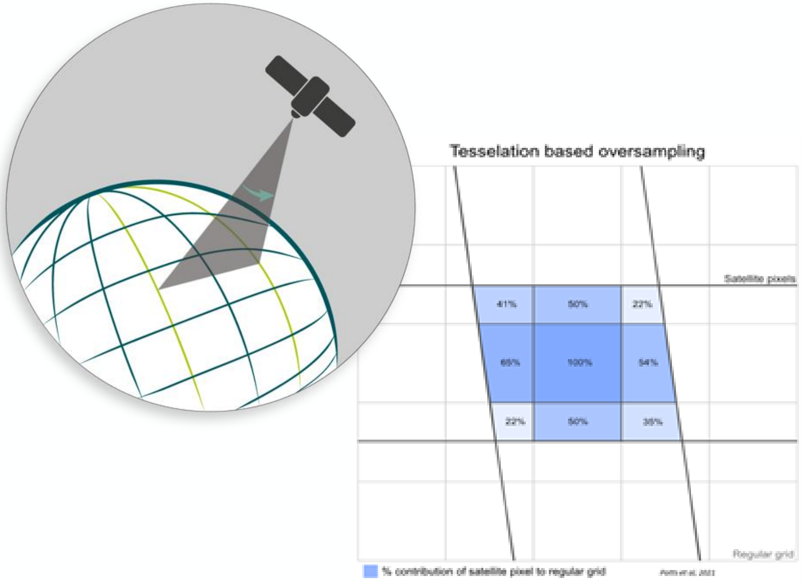
- Oil, Gas & Coal facilities
- Waste Manageme nt
- Climate Finance
- National Accounting
- NGOs & Think Tanks

How well do data users understand uncertainties?

Meteorological

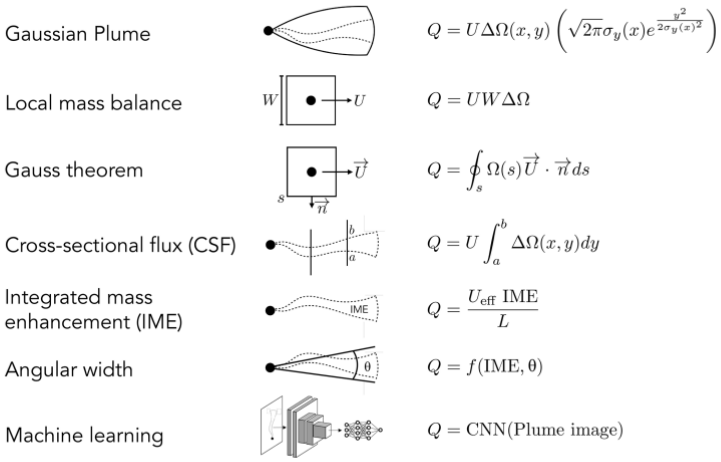
- Wind Speed Accuracy
- Wind Speed Variability -Turbulence
- Vertical Wind Profile - assume ground-level release but what about plume rise?

Instrumental & Observational



- Background methane
- Surface albedo
- Instrument SNR changes
- Instrument detection threshold
- Pixel resolution & interpolation methods

Modeling & Algorithmic

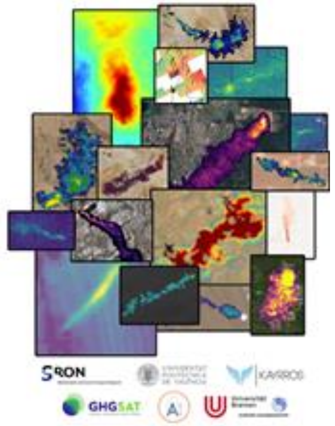


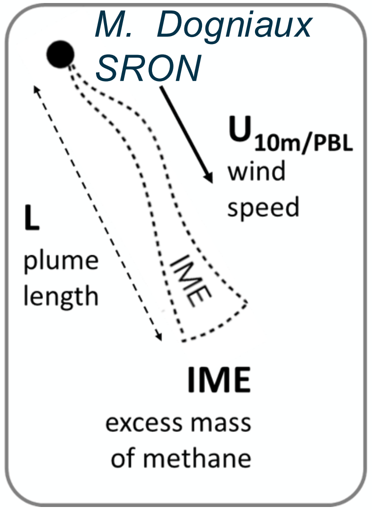
- XCH4 retrieval uncertainties
- Emission derivation methods
- Source location – downwind location
- Steady-state assumption using single snapshots

How well do data users understand uncertainties?

Community is using multiple methods with different sensitivities to build a picture of emissions & uncertainties

MEDUSA

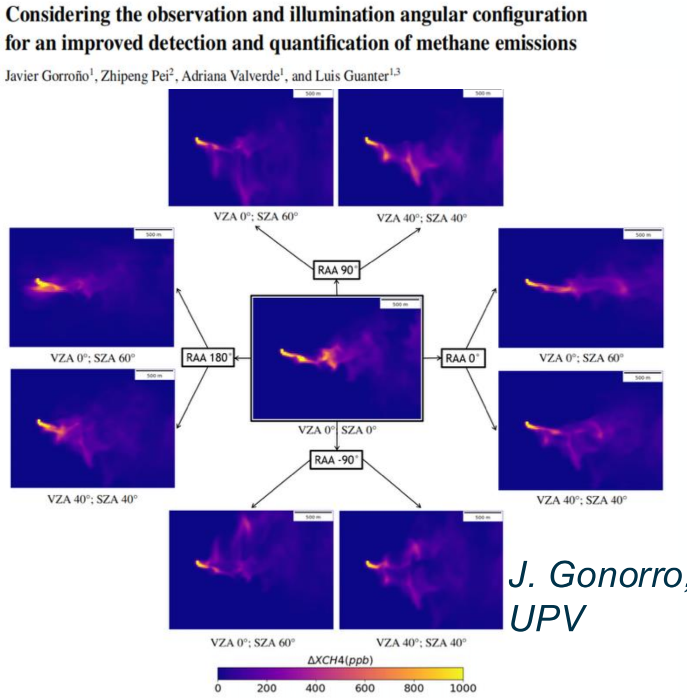




Large-scale methane plume uncertainty exploration

|Algorithm intercomparisons|
| Uncertainty decision trees|
| Probability of detections|

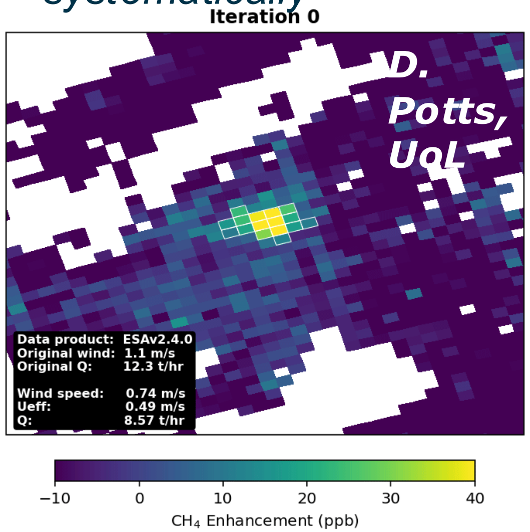
Parallax, scene angular mismatch



J. Gonorro, UPV

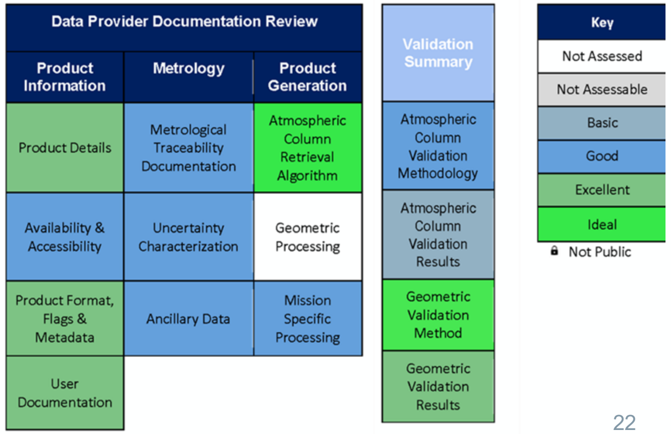
METRICS

Testing common practises systematically



How do the choices we make impact source rate & uncertainty?

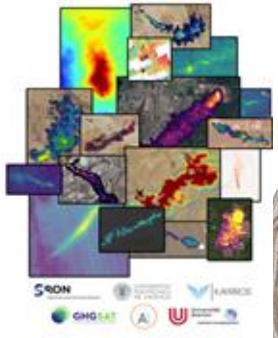
Robust Quality Assessment - To what extent are data are "fit for purpose"?

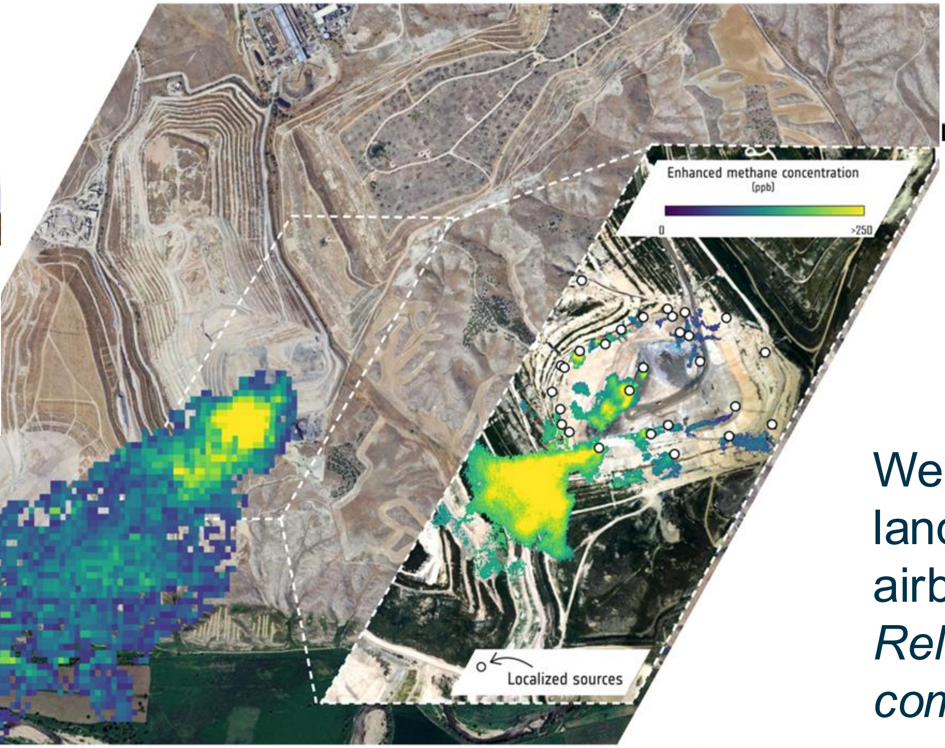


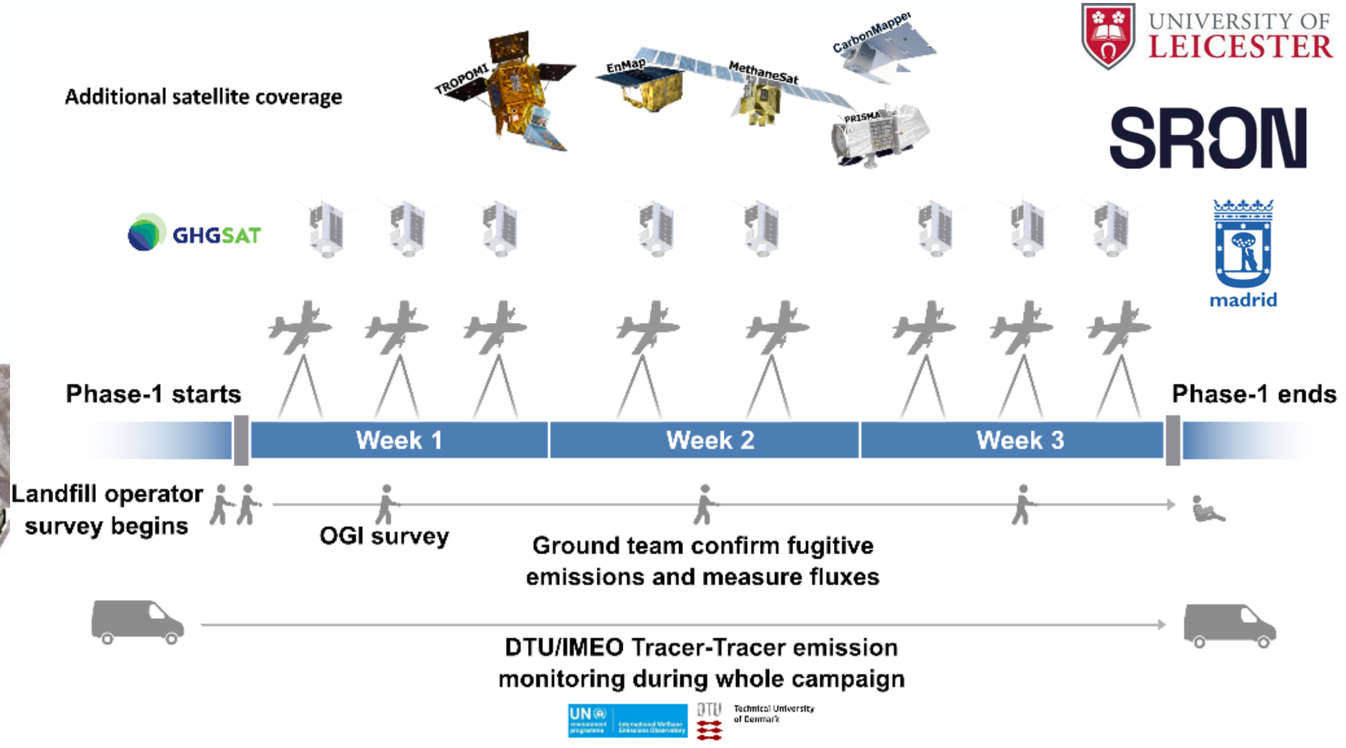
Real-world validation

Validating methane fluxes is challenging!
Single blind controlled releases (Sherwin et al. 2023) found quantification error across different satellites: ~ 55 % of mean estimates fell within ±50 % of the true value.

MEDUSA







We are working on real-world validation on a complex active landfill in Madrid - Comparing satellite-derived emissions with airborne and ground-based validation
Relies on strong links between data providers & facilities, satellite community and ground-based monitoring community

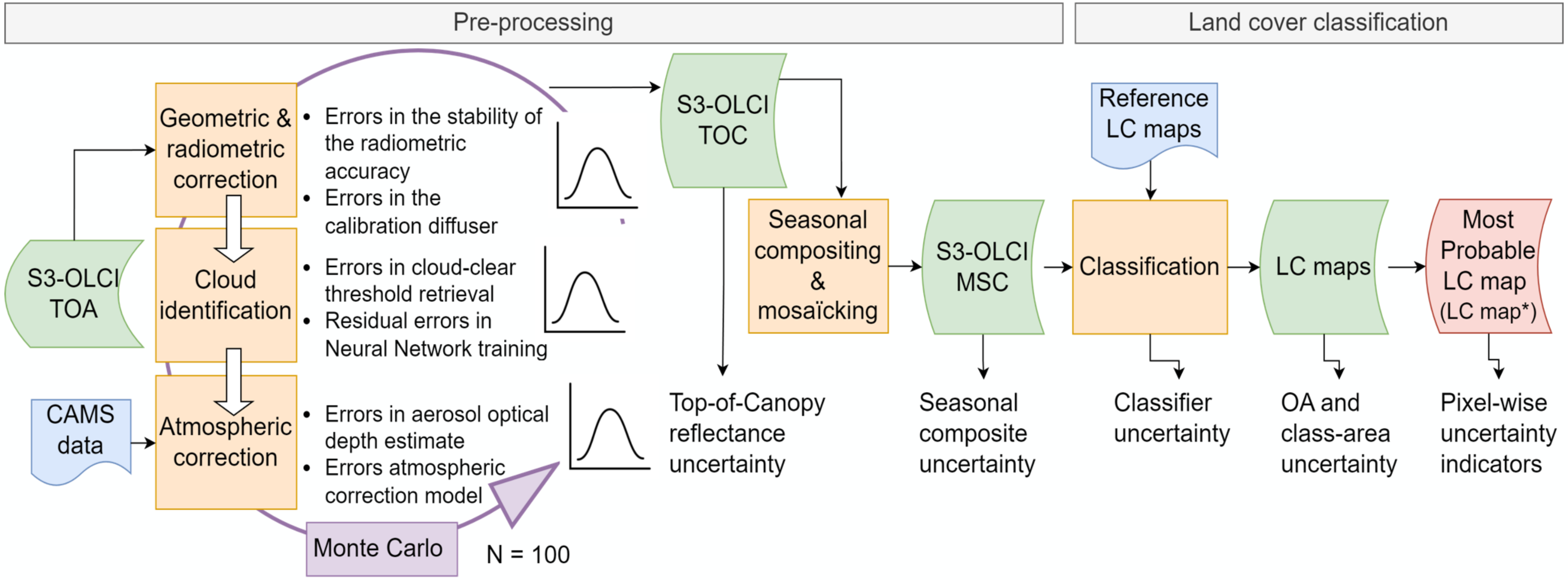
Colocation Day 3: Uncertainty Panel Discussion

CCI MRLand Cover experience for categorical products

Celine Lamarche, Ralf Quast, Carsten Brockman, Pierre Defourny et al.

25/03/2026

End-to-end propagation of SR and cloud detection uncertainty to Land Cover classification output using Monte Carlo variants

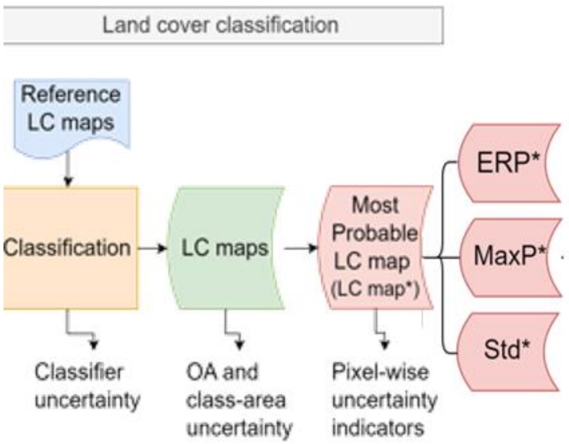


Uncertainty of the entire processing chain
(30 RF x 100 LC variants)

Site	OA (mean ± std)
EU	80.38 ± 0.33
AF	86.10 ± 0.31
SA	81.29 ± 0.39

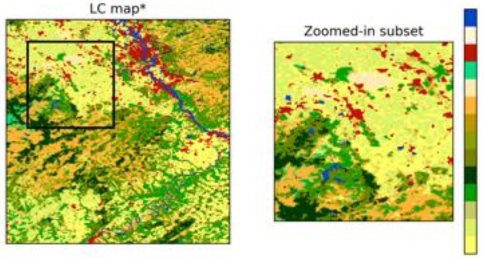
1. *Assessment of end-to-end uncertainty propagation on LC map accuracy*
2. *Evaluation of mean compositing impact on SR uncertainty reduction*
3. *Assessment of classifier uncertainty contribution*
4. *Identification of appropriate pixel-level uncertainty indicators for LC maps*

MaxP and ERP – most sensitive uncertainty indicator for Land Cover classes

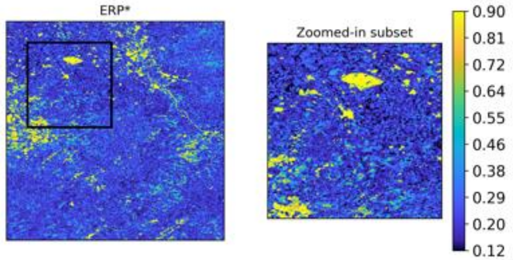


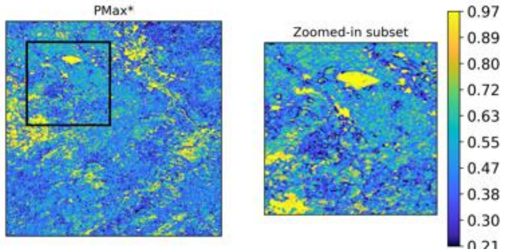
European site

Most probable LC map



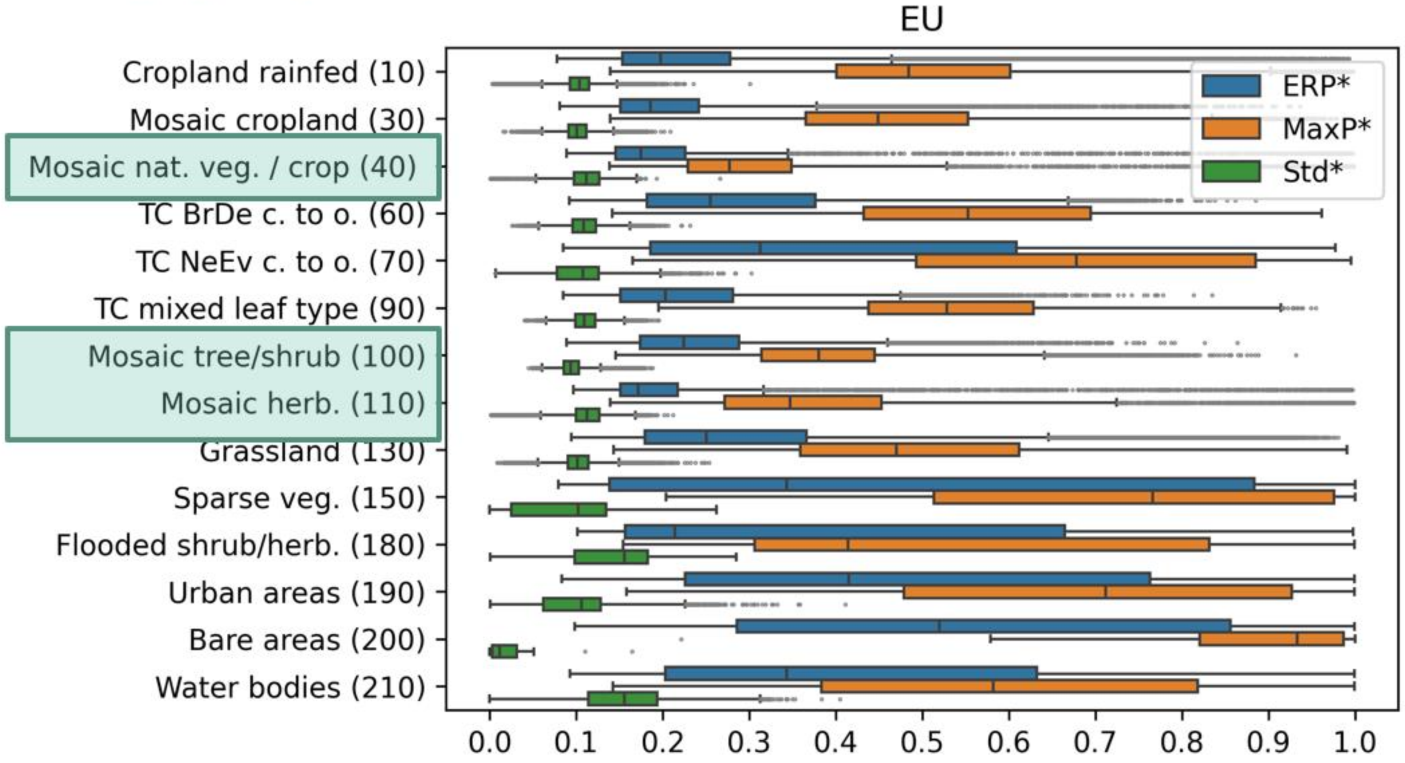
LC uncertainty indicators

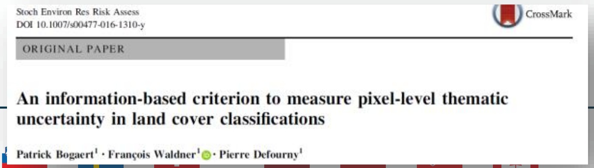




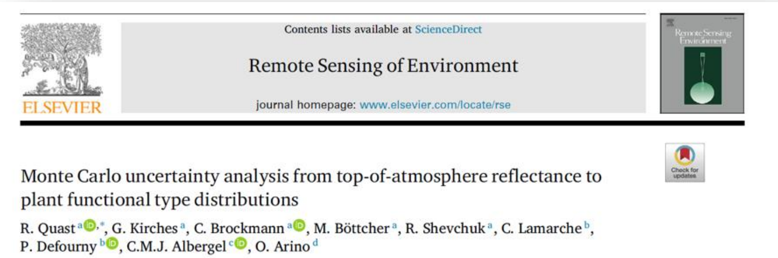
Uncertainty indicator relevance assessment

MaxP* and Equivalent Reference Probability (ERP) highlighting Mosaic classes as most uncertain !




Stoch Environ Res Risk Assess
DOI 10.1007/s00477-016-1310-y
CrossMark
ORIGINAL PAPER
An information-based criterion to measure pixel-level thematic uncertainty in land cover classifications
Patrick Bogaert[1] · François Waldner[1] · Pierre Defourny[1]

Monte Carlo uncertainty from S3 OLCI TOA reflectance to Plant Functional Type (PFT) distribution


Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

ELSEVIER

Monte Carlo uncertainty analysis from top-of-atmosphere reflectance to plant functional type distributions

R. Quast[a,*], G. Kirches[a], C. Brockmann[a], M. Böttcher[a], R. Shevchuk[a], C. Lamarche[b], P. Defourny[b], C.M.J. Albergel[c], O. Arino[d]

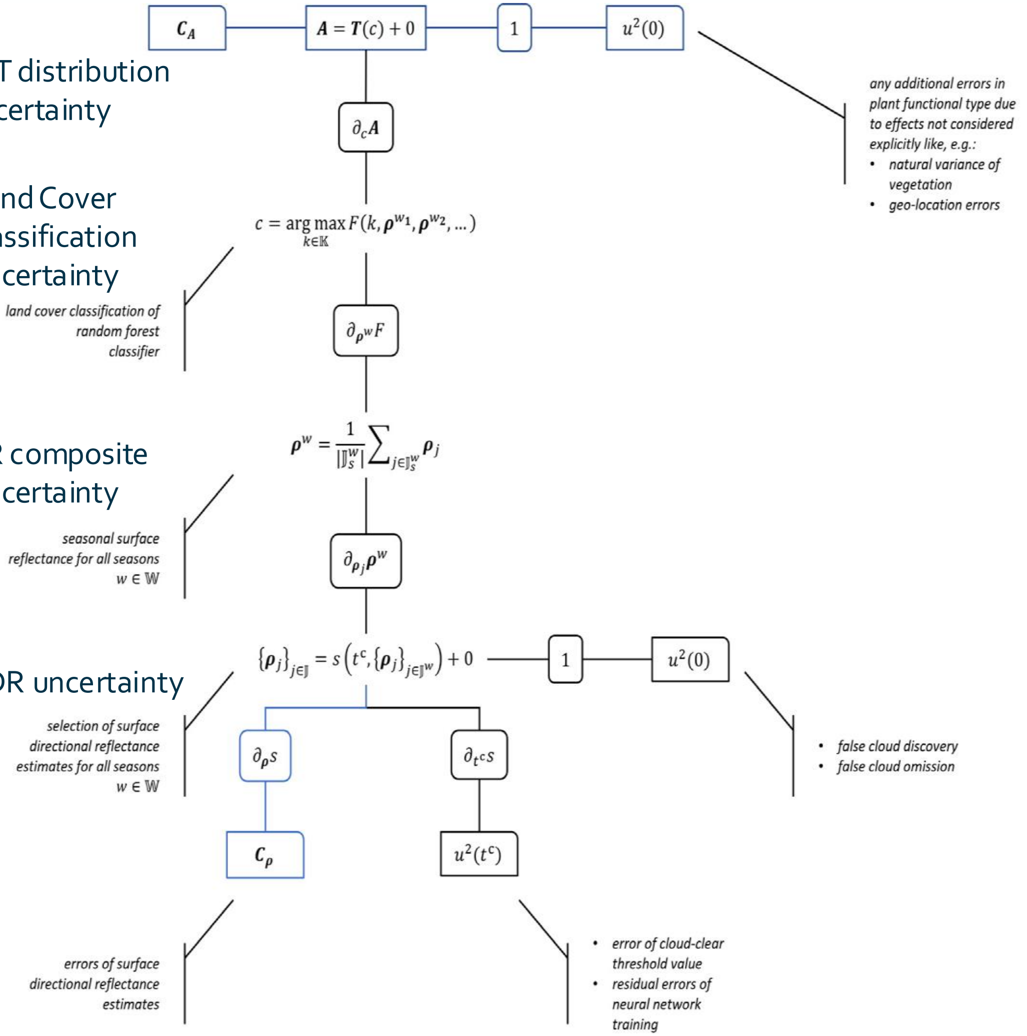
Surface reflectance uncertainty using estimates of measurands, associated uncertainties and their correlation structures

PFT distribution uncertainty

Land Cover classification uncertainty

SR composite uncertainty

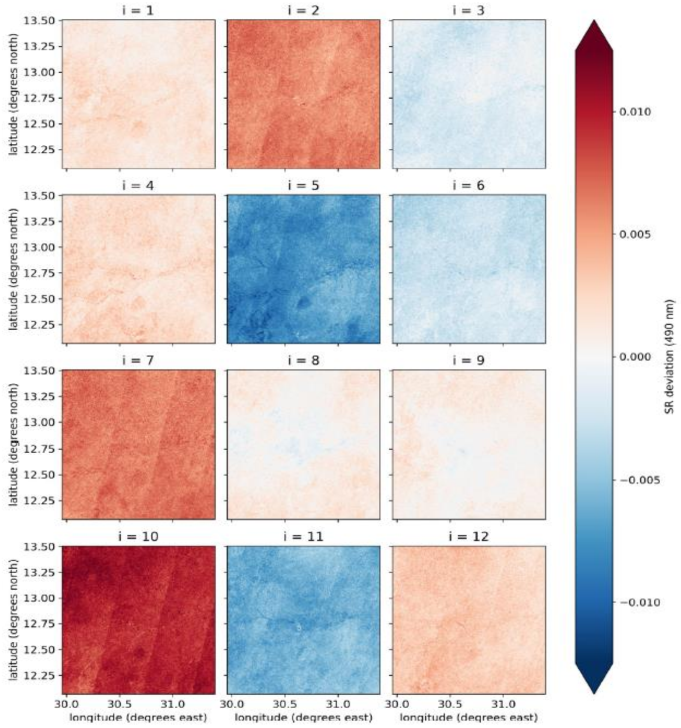
SDR uncertainty



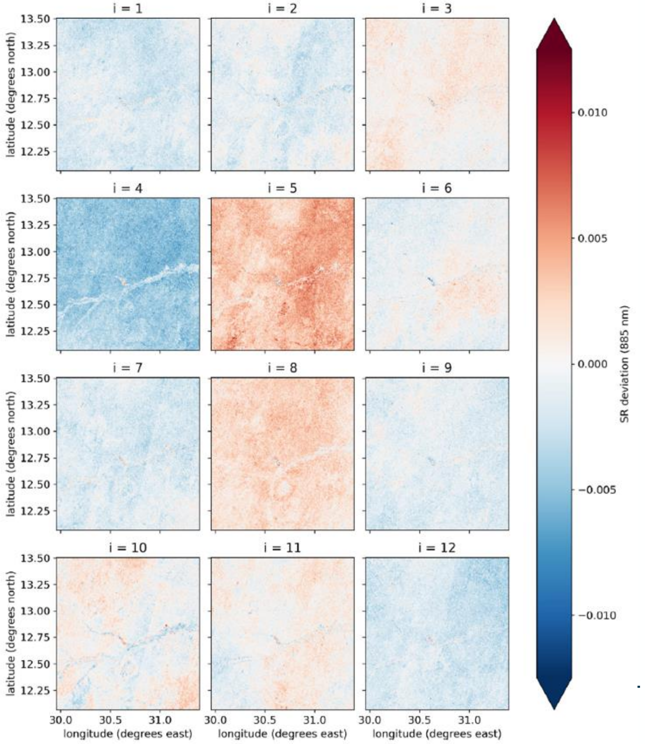
African site

Example : surface reflectance deviation of 12 variants from the ensemble mean Monte Carlo variants of annual reflectance composite

490 nm band



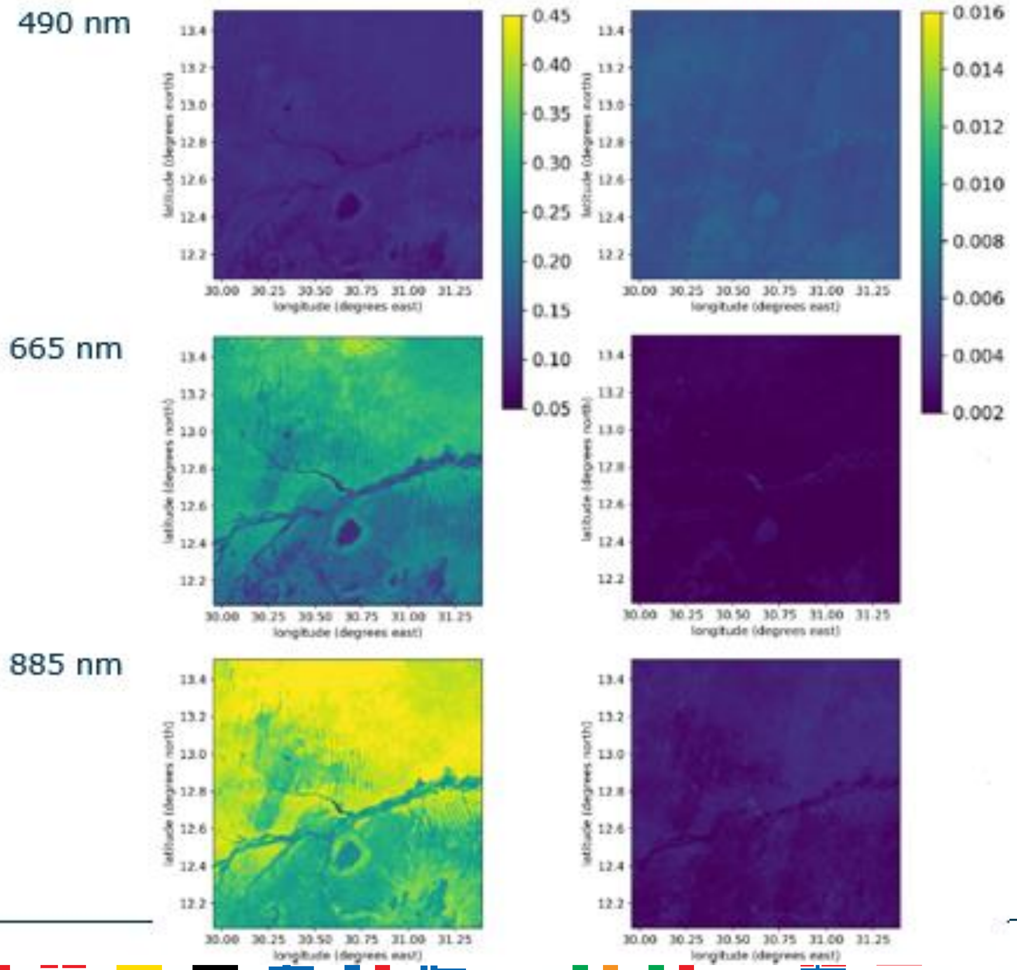
885 nm band



End-to-end uncertainty from TOA reflectance to Plant Functional Type (PFT) distribution

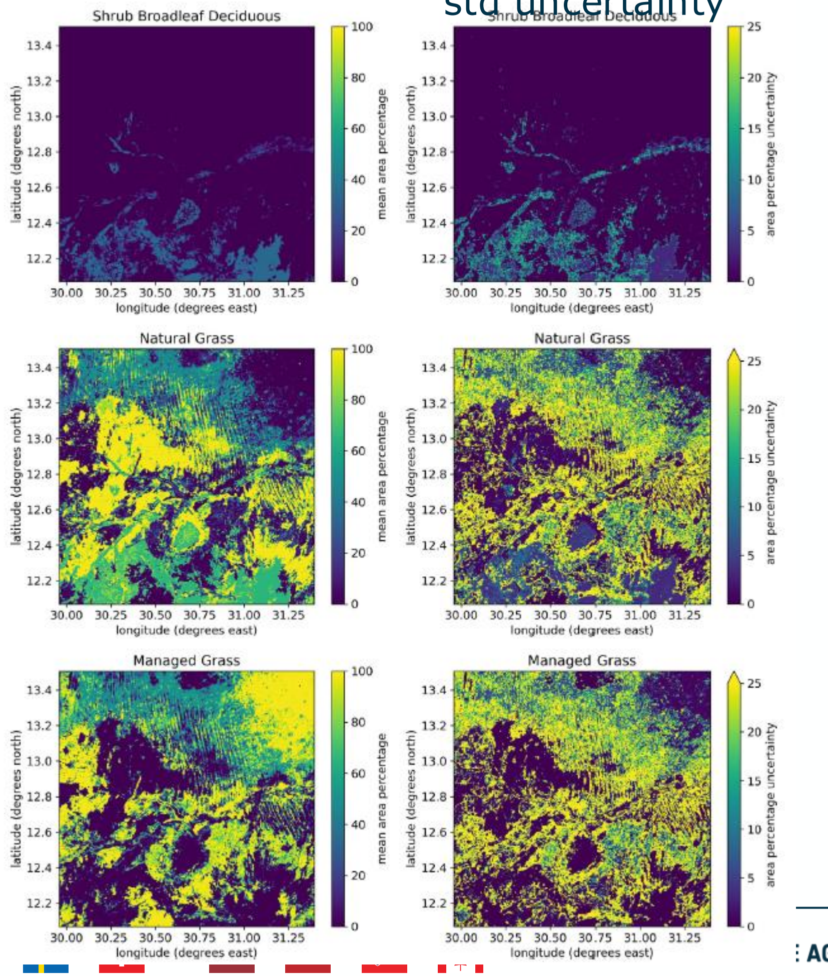
Annual surface reflectance (SR) composite uncertainty

SR composite | Std uncertainty



Mean and standard uncertainty of area percentages of most abundant PFTs

PFT area percentage | PFT area percentage std uncertainty



Seeding Questions

- On what aspects of uncertainty quantification can the CCI projects work together?
- Are there aspects of working with uncertainty data that should be universal across ECVs e.g. uncertainties on trend estimates? Stability analysis?
- Is there scope to have a unified framework for dealing with sensor resolution in CDRs?
    - Do we need something different for classification algorithms where resolution changes the answer?
    - To what degree should we compensate for temporal changes in sensor resolution in the uncertainty budget?
- Can we calculate metrological uncertainties when using machine learning algorithms?
    - What is a useful uncertainty estimate for machine learning applications e.g. downscaling?
- How do we use uncertainty information to inform observation-model comparisons?
- How well do data users understand uncertainties? How well do we understand the uncertainty information that data users need? How do we make stronger links between data providers and data users such that we close the gap? Does this need to be application specific?