National Physical Laboratory

7th Sentinel-2 Validation Team Meeting
13 – 15 October 2025

Pixel-level validation of land cover maps - an outlook from the perspective of metrology

Anna Pustogvar[1,2], Andrew Thompson[1]

[1] National Physical Laboratory, UK
[2] University of Leicester, UK

Quality assessment of land cover (LC) maps

- Quality assessment of LC maps is coordinated by CEOS WGCV LPV LC focus group
- In September 2025, the LC focus group released a new Good Practice Validation Protocol
- This protocol is largely focused on LC map's quality assessment based on confusion matrices

Main limitation: confusion matrix provides no information on per-pixel quality of LC maps

Confusion matrix

Land cover map \ Reference dataset	A	B	C	D	Σ
A	p_{AA}	p_{AB}	p_{AC}	p_{AD}	p_{A+}
B	p_{BA}	p_{BB}	p_{BC}	p_{BD}	p_{B+}
C	p_{CA}	p_{CB}	p_{CC}	p_{CD}	p_{C+}
D	p_{DA}	p_{DB}	p_{DC}	p_{DD}	p_{D+}
Σ	p_{+A}	p_{+B}	p_{+C}	p_{+D}	1

$K = \#$ of classes

$$\textbf{Overall accuracy (OA)} = \sum_{k=1}^{K} p_{kk} \times 100$$

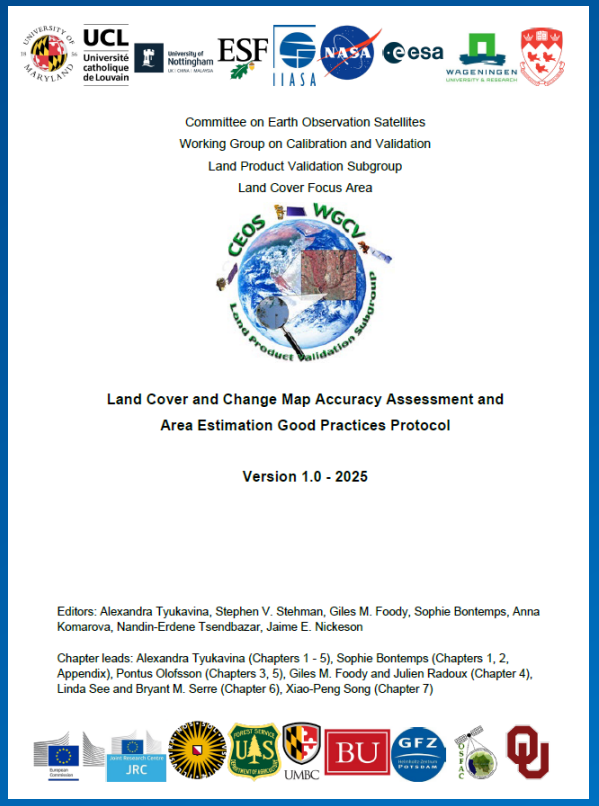
$$\textbf{User's accuracy (UA)} = \frac{p_{kk}}{p_{k+}} \times 100$$

$$\text{Commision errors} = 100 - \text{UA}$$

$$\textbf{Producer's accuracy (PA)} = \frac{p_{kk}}{p_{+k}} \times 100$$

$$\text{Omission errors} = 100 - \text{PA}$$

- Indicators of per-pixel quality (specifically, metrologically rigorous per-pixel uncertainties) are needed for propagation to downstream applications of LC maps


Committee on Earth Observation Satellites
Working Group on Calibration and Validation
Land Product Validation Subgroup
Land Cover Focus Area

Land Cover and Change Map Accuracy Assessment and Area Estimation Good Practices Protocol

Version 1.0 - 2025

Editors: Alexandra Tyukavina, Stephen V. Stehman, Giles M. Foody, Sophie Bontemps, Anna Komarova, Nandin-Erdene Tsendbazar, Jaime E. Nickeson

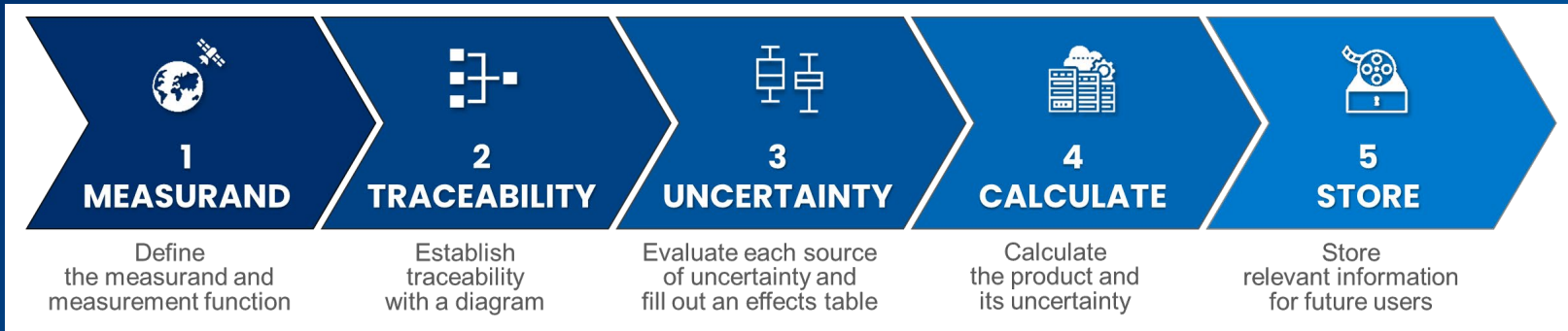
Chapter leads: Alexandra Tyukavina (Chapters 1 - 5), Sophie Bontemps (Chapters 1, 2, Appendix), Pontus Olofsson (Chapters 3, 5), Giles M. Foody and Julien Radoux (Chapter 4), Linda See and Bryant M. Serre (Chapter 6), Xiao-Peng Song (Chapter 7)

https://lpvs.gsfc.nasa.gov/PDF/CEOS_WGCV_LPV_Land_Cover_protocol_Sept2025_V1.pdf

Metrologically rigorous uncertainties and their validation

- For more typical EO products (i.e., quantitative products such as surface reflectance, land or sea surface temperature), the estimation of per-pixel metrologically rigorous uncertainties and their validation can be performed by following the QA4EO guidelines

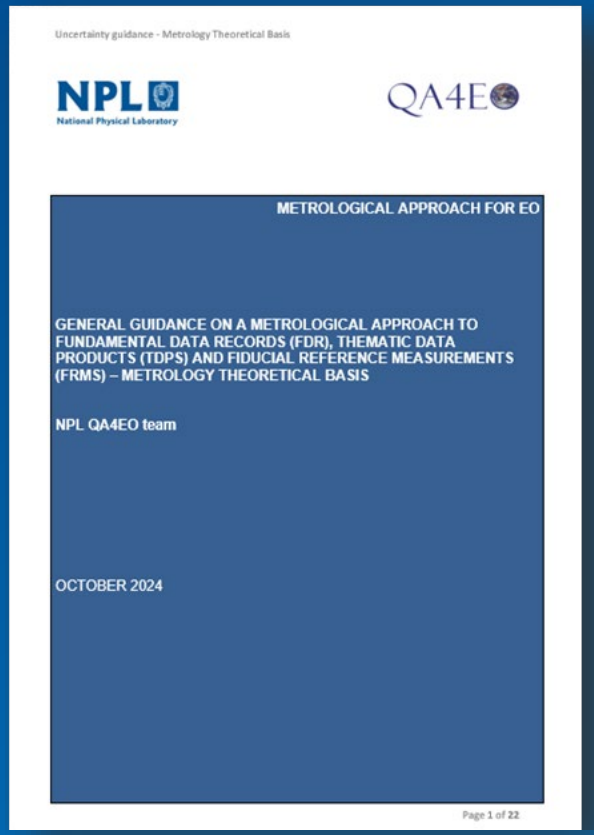
Uncertainty estimation



Validation of uncertainties

$$E_N = \frac{|\rho_1 - \rho_2|}{k\sqrt{u_1^2 + u_2^2 + u_{\text{comp}}^2}}$$

Not directly applicable to categorical variables


Uncertainty guidance - Metrology Theoretical Basis

NPL National Physical Laboratory

QA4EO

METROLOGICAL APPROACH FOR EO

GENERAL GUIDANCE ON A METROLOGICAL APPROACH TO FUNDAMENTAL DATA RECORDS (FDR), THEMATIC DATA PRODUCTS (TDPS) AND FIDUCIAL REFERENCE MEASUREMENTS (FRMS) – METROLOGY THEORETICAL BASIS

NPL QA4EO team

OCTOBER 2024

Page 1 of 22

https://qa4eo.org/docs/2_Metrology_Document.pdf

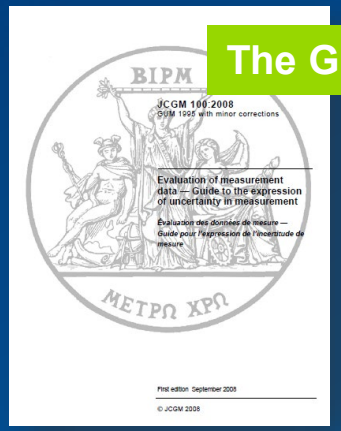
Categorical variables in metrology

- QA4EO guidelines represent the adaptations of the authoritative metrological documents for EO domain

The VIM 3rd edition

- is focused on quantities
- provides a definition of categorical variables (i.e., nominal properties)
- clarifies that the term measurement (and therefore measurement uncertainty) does not apply to categorical variables
- does not provide an alternative uncertainty term that would be applicable to categorical variables

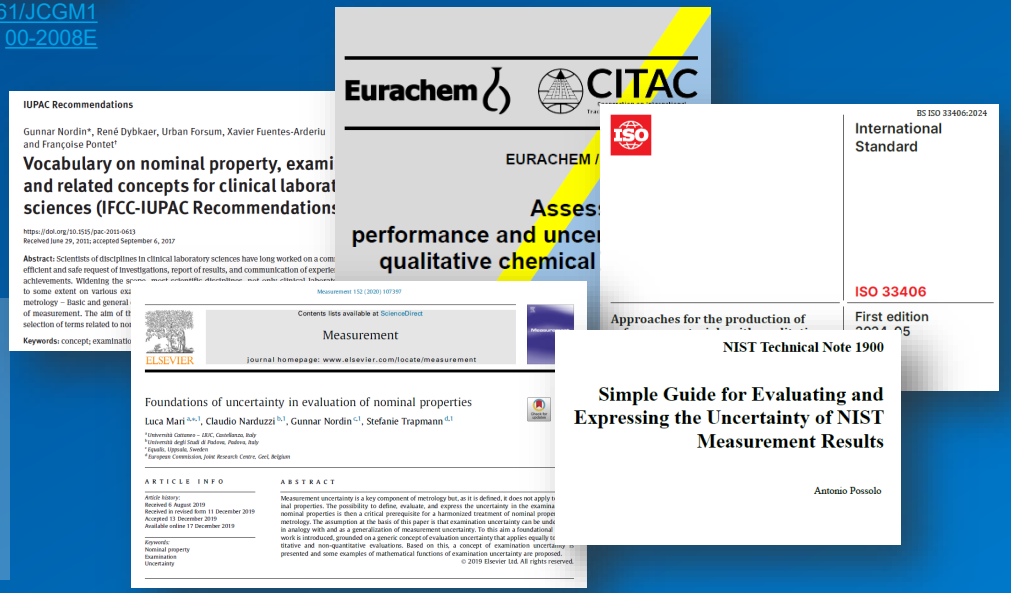
https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf



The GUM

- is focused on quantities
- only exception - GUM-4 on Conformity assessment (a special case of categorical variables)

https://doi.org/10.59161/JCGM100-2008E



NB the interest towards categorical variables has been growing in the metrology community in recent years:

- the 4th VIM is expected to have an entire chapter on categorical variables
- uncertainty and other metrological concepts have been discussed in several international standards and metrology-related publications

Categorical variables in EO

Due to the lack of metrological guidance from the VIM and the GUM (and the QA4EO) for categorical EO products, we need to start with answering the basic questions:

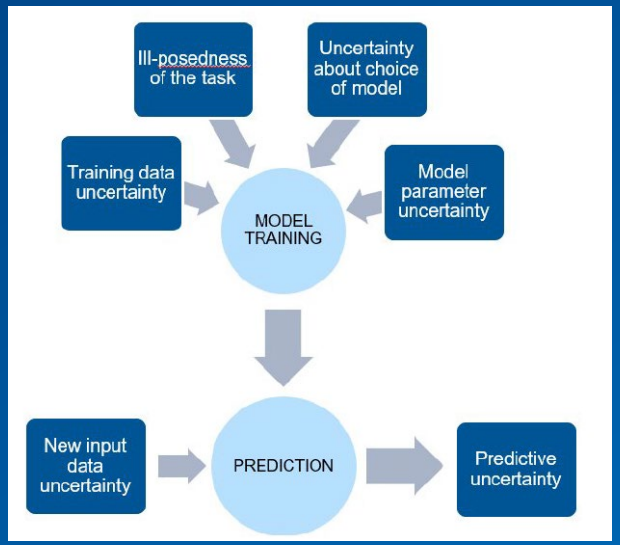
Q1: What are uncertainties for categorical variables?

Q2: To what extent are the principles of uncertainty estimation and validation developed for quantities transferable to categorical variables?

Categorical variables in EO

- **Q1: What are uncertainties for categorical variables?**
  - Uncertainty can be fully expressed by Probability Mass Function (PMF)

- **Q2: To what extent are the principles of uncertainty estimation and validation developed for quantities transferable to categorical variables?**
  - Similarly to quantities, the combined uncertainty for categorical variables can be obtained by propagating all contributing sources of uncertainty



submitted to Metrologia in March 2025

A metrological framework for uncertainty evaluation in machine learning classification models

Samuel Bilson[1], Maurice Cox[1], Anna Pustogvar[2,3], and Andrew Thompson[1]

[1]Department of Data Science, National Physical Laboratory, Teddington, UK
[2]Department of Thermal & Radiometric Metrology, National Physical Laboratory, Teddington, UK
[3]School of Geography, Geology & The Environment, University of Leicester, Leicester, UK

https://arxiv.org/abs/2504.03359

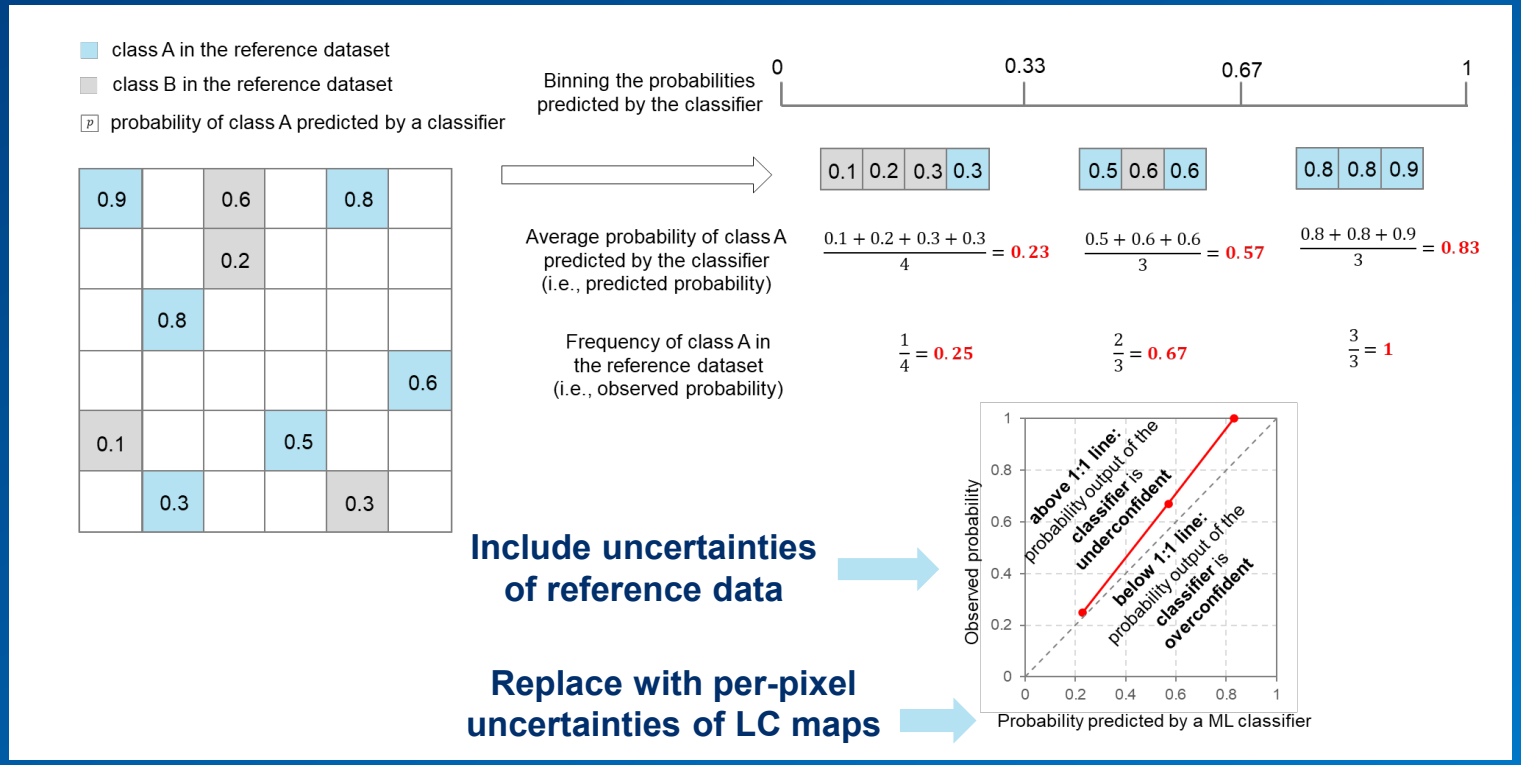
Categorical variables in EO

- **Q2: To what extent are the principles of uncertainty estimation and validation developed for quantities transferable to categorical variables?**

Quantitative variables →
Equivalence ratio

$$E_N = \frac{|\rho_1 - \rho_2|}{k\sqrt{u_1^2 + u_2^2 + u_{\text{comp}}^2}}$$

Categorical variables →
Adaptation of "classifier calibration" method



ESA Met4EO (Metrology for Earth Observation) project

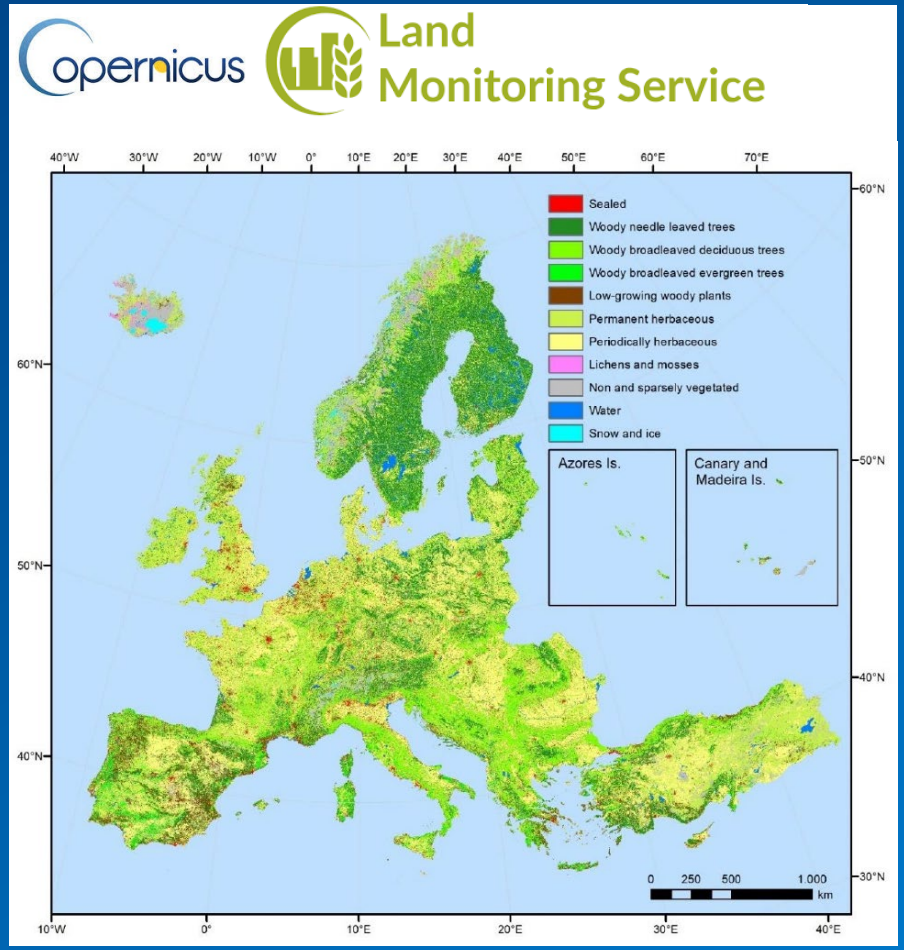
- A new ESA project delivered by NPL, with the objective of developing innovative concepts for assessing the quality of various EO products following the principles of metrology. Including:
  - WP1: AI for Cal/Val
  - WP2: Metrology Tools, Training and Support
  - WP3: Metrology for Land Cover Classification
  - WP4: CEOS-PVP
  - WP5: Uncertainty Analysis for Altimetry

- The three objectives of WP3 is
  - to implement a prototype framework for metrologically rigorous estimation of per-pixel uncertainties of LC maps
  - to implement a prototype framework for the validation of these uncertainties
  - to define the pathway for applying the FRM concept to LC maps

ESA Met4EO project

LC classification

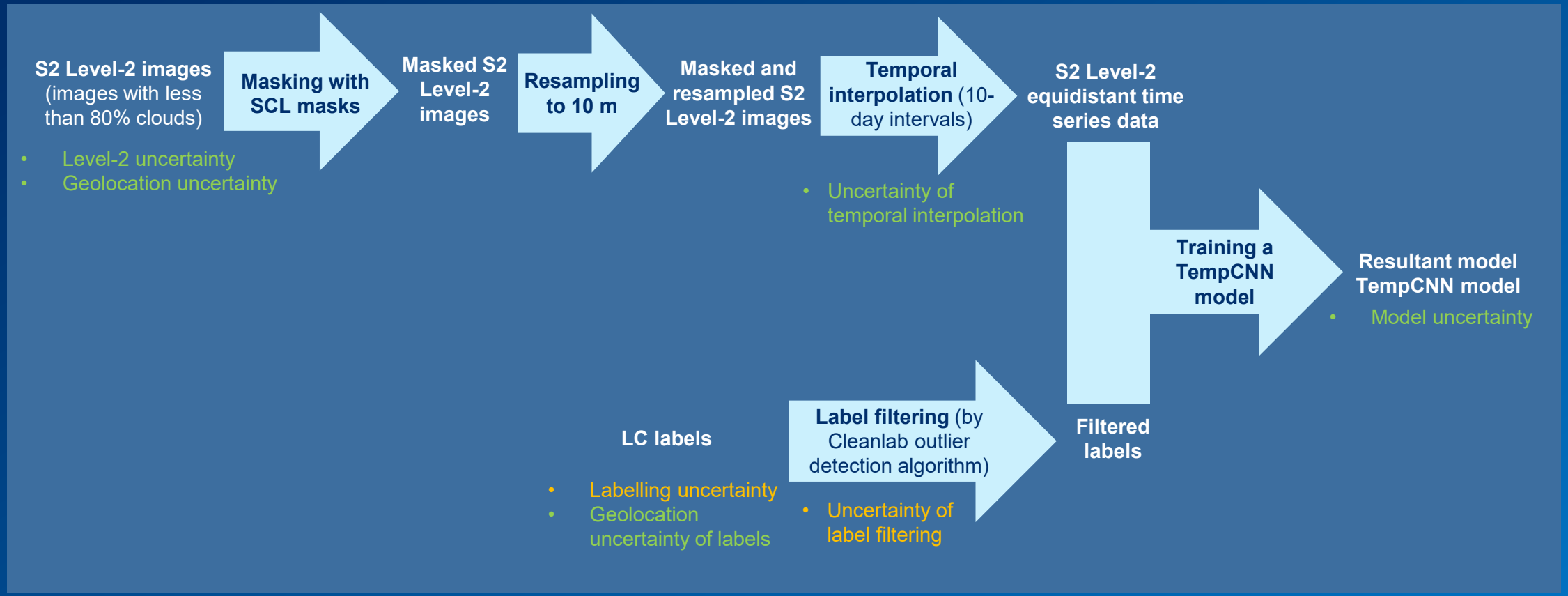
- We are using the processing chain of CLMS CLCplus Backbone LC product as a foundation (with some modifications)
  - developed by Copernicus Land Monitoring Service (CLMS)
  - high resolution (10 m) product of European land cover
  - is produced
    - based on a time series of Sentinel-2
    - 8 bands and 4 spectral indices
      - B1, B3, B4, B5, B8, B9, B11, B12
      - NDVI, NDWI, NDMI, NBR
    - by Temporal Convolutional Neural Network (TempCNN) classifier
    - training data: LUCAS, existing CLMS LC products, photo-interpretation of VHR, etc.
  - 11 basic LC classes



ESA Met4EO project

NPL

CLCplus Backbone LC product (training stage)

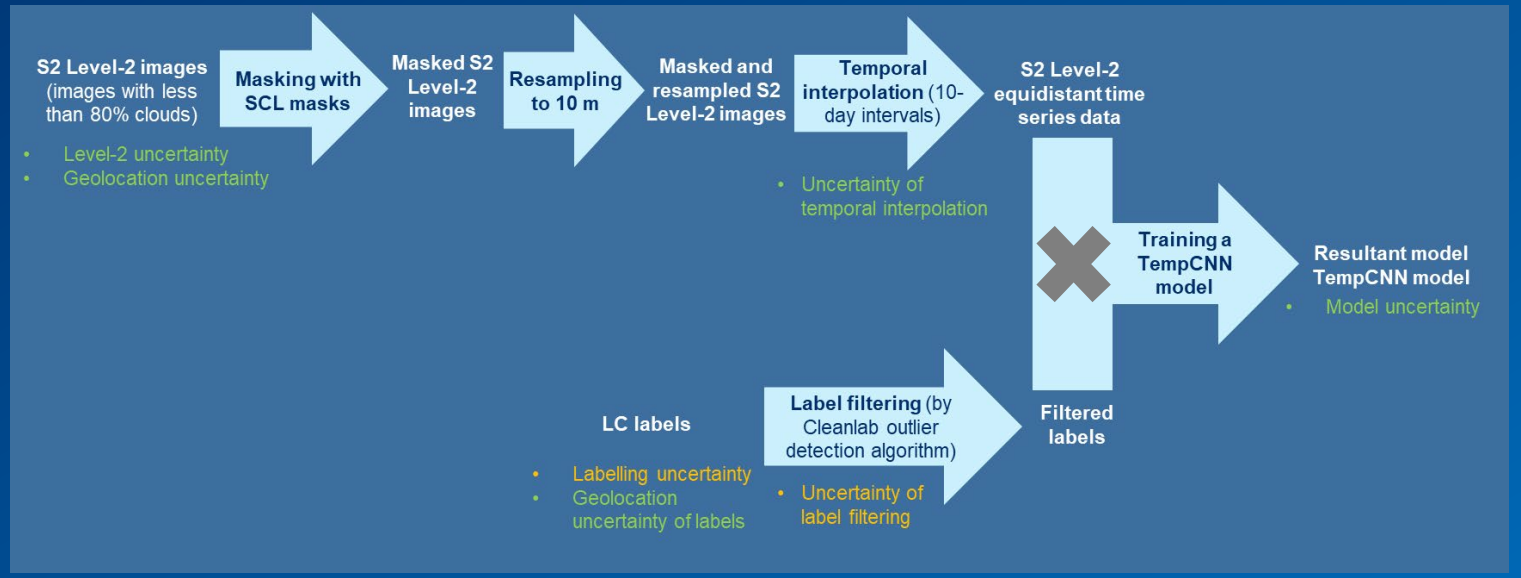


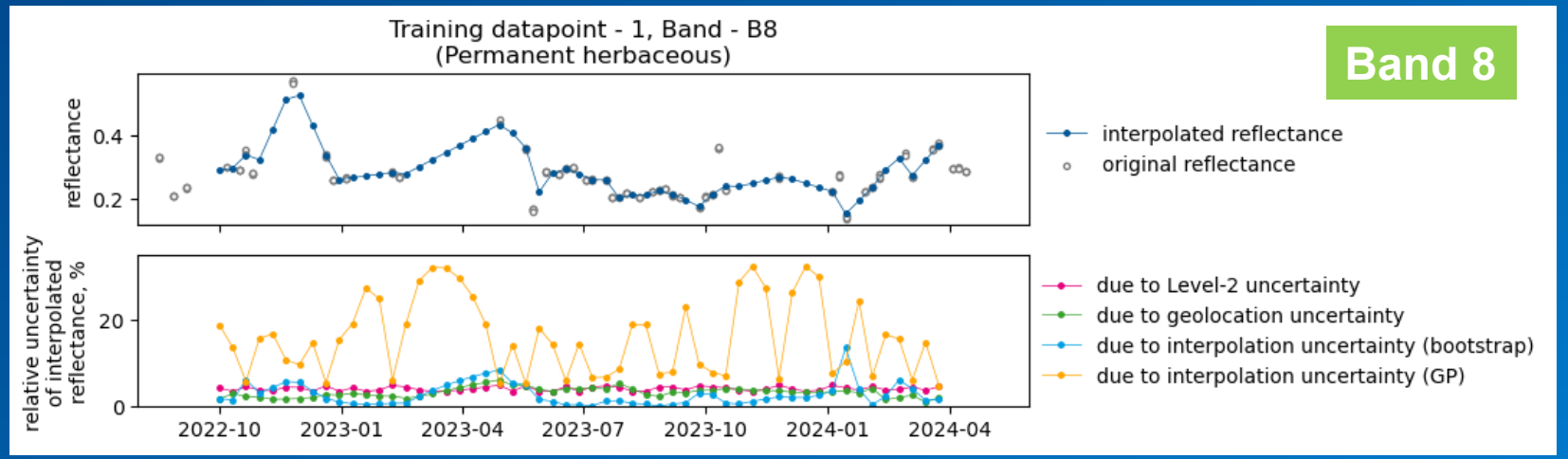
ESA Met4EO project

NPL

Uncertainty propagation framework

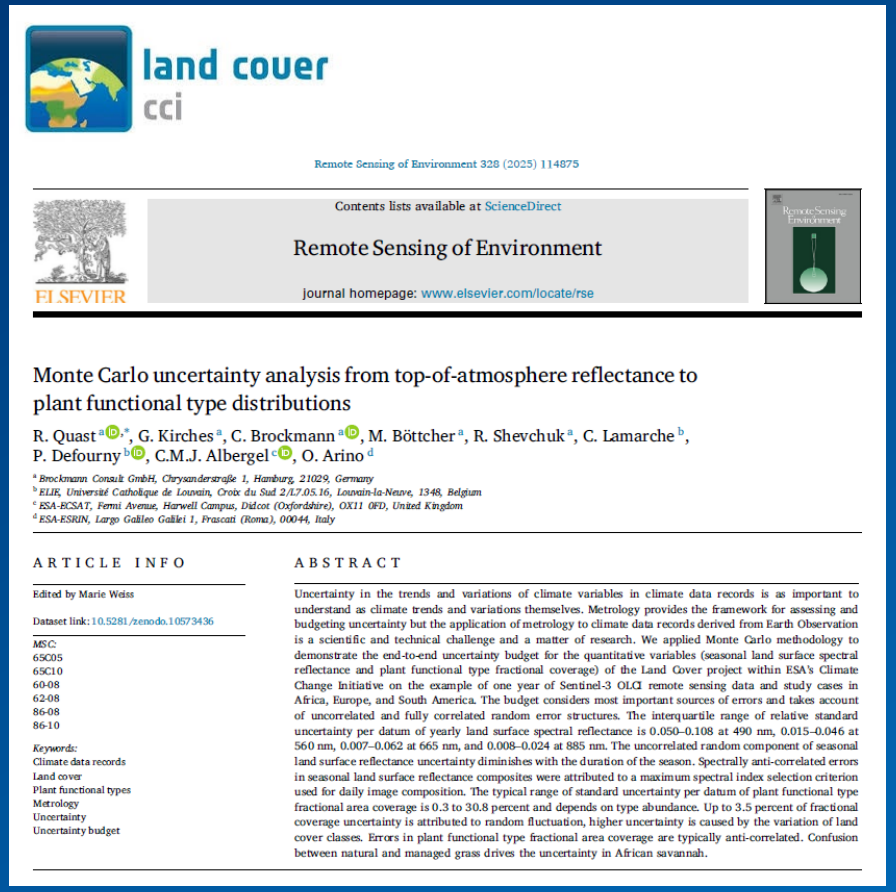
- Sentinel-2 Level-2 uncertainty
  - based on SIAC atmospheric correction
- Geolocation uncertainty
  - Sentinel-2 positional accuracy
  - positional accuracy of a typical GPS sensor
- Temporal interpolation uncertainty
  - Gaussian Processes (being refined)
  - Bootstrap resampling (being refined)





Other efforts on bringing metrology to LC mapping

- Noteworthy efforts in the same direction – estimating metrologically rigorous per-pixel uncertainties of LC maps – are currently being made by the ESA CCI MRLC team



land cover cci

Remote Sensing of Environment 328 (2025) 114875

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

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]

[a] Brockmann Consult GmbH, Chrysanderstraße 1, Hamburg, 21029, Germany
[b] ELIE, Université Catholique de Louvain, Croix du Sud 2/L7.05.16, Louvain-la-Neuve, 1348, Belgium
[c] ESA-ECSAT, Fermi Avenue, Harwell Campus, Didcot (Oxfordshire), OX11 0FD, United Kingdom
[d] ESA-ESRIN, Largo Galileo Galilei 1, Frascati (Roma), 00044, Italy

ARTICLE INFO

Edited by Marie Weiss

Dataset link: 10.5281/zenodo.10573436

MSC:
65C05
65C10
60-08
62-08
86-08
86-10

Keywords:
Climate data records
Land cover
Plant functional types
Metrology
Uncertainty
Uncertainty budget

ABSTRACT

Uncertainty in the trends and variations of climate variables in climate data records is as important to understand as climate trends and variations themselves. Metrology provides the framework for assessing and budgeting uncertainty but the application of metrology to climate data records derived from Earth Observation is a scientific and technical challenge and a matter of research. We applied Monte Carlo methodology to demonstrate the end-to-end uncertainty budget for the quantitative variables (seasonal land surface spectral reflectance and plant functional type fractional coverage) of the Land Cover project within ESA's Climate Change Initiative on the example of one year of Sentinel-3 OLCI remote sensing data and study cases in Africa, Europe, and South America. The budget considers most important sources of errors and takes account of uncorrelated and fully correlated random error structures. The interquartile range of relative standard uncertainty per datum of yearly land surface spectral reflectance is 0.050–0.108 at 490 nm, 0.015–0.046 at 560 nm, 0.007–0.062 at 665 nm, and 0.008–0.024 at 885 nm. The uncorrelated random component of seasonal land surface reflectance uncertainty diminishes with the duration of the season. Spectrally anti-correlated errors in seasonal land surface reflectance composites were attributed to a maximum spectral index selection criterion used for daily image composition. The typical range of standard uncertainty per datum of plant functional type fractional area coverage is 0.3 to 30.8 percent and depends on type abundance. Up to 3.5 percent of fractional coverage uncertainty is attributed to random fluctuation, higher uncertainty is caused by the variation of land cover classes. Errors in plant functional type fractional area coverage are typically anti-correlated. Confusion between natural and managed grass drives the uncertainty in African savannah.

https://doi.org/10.1016/j.rse.2025.114875

Conclusions

- Land cover represents a categorical variable
- Uncertainty evaluation and validation of uncertainties for such variables are currently not covered by the GUM and the VIM (and consequently by the QA4EO)
- In this presentation, we showed the ongoing efforts dedicated to addressing this gap

NPL

National Physical Laboratory

npl.co.uk