Uncertainty-Decomposed Vegetation Trait Retrieval for Sentinel-3:

A Copernicus-Ready Bayesian Framework

Jochem Verrelst, Pablo Reyes-Muñoz, Emma De Clerck, Yuxin Zhang, José Luis García Soria, Miguel Morata, Carolina Tenjo

Outline of the presentation:

- Many retrieval methods — different assumptions
- Validation data remains sparse and geographically limited
- Operational users need confidence information
- Uncertainty-aware retrievals are emerging
- But uncertainty maps alone are hard to interpret…

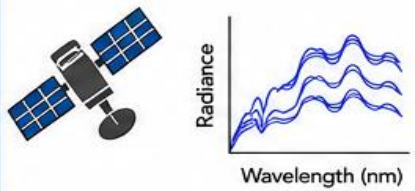
- **Optical remote sensing data**
- **Hybrid retrieval**
- **Uncertainties in EO processing chain**
- **Aleatoric** vs **epistemic** uncertainty: **Methods**
- **Benchmarking Sentinel-3 vegetation traits uncertainty strategy**
- Towards **reference-free trustworthiness** protocol
- **Conclusions**

Sources of uncertainty in the hyperspectral vegetation retrieval chain

Different sources arise at each stage of the processing chain and propagate to the final vegetation products.

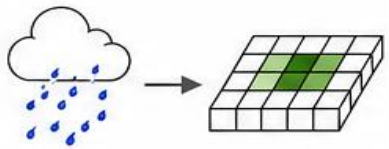
L1: Top-of-atmosphere radiance

Raw hyperspectral radiance measurements



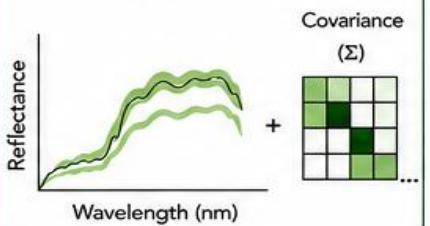
Atmospheric correction

Removal of atmospheric effects and adjacency



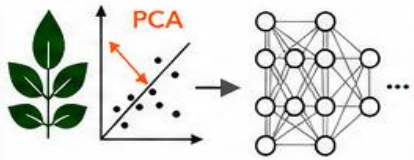
L2A: Surface reflectance

Atmospherically corrected surface reflectance



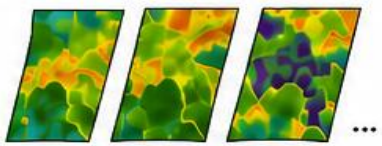
Retrieval models

Transformation of reflectance (e.g., PCA) and inversion to vegetation variables



L2B: Vegetation products

Biophysical variables (e.g., LAI, Cab, Cw, FAPAR, N)



1. Radiometric measurement uncertainties

- Instrument noise (SNR)
- Radiometric calibration uncertainty
- Spectral response function uncertainty
- Detector nonlinearity
- Stray light
- Correlated measurement errors (spectral covariance)

2. Atmospheric & geometric uncertainties

Atmospheric

- Aerosol optical thickness
- Water vapour, ozone
- Gas absorption
- Adjacency effects
- Scattering assumptions

Geometric & illumination

- Sun–sensor geometry
- Topography (slope, aspect)
- BRDF / anisotropy
- Co-registration errors

3. Propagation to surface reflectance

- Residual atmospheric correction errors
- Spectrally correlated uncertainties
- Increased sensitivity in absorption features

4. Retrieval-model uncertainties

- Model structural assumptions (e.g., RTM simplifications)
- Parameter / prior uncertainties
- Training data limitations and representativeness
- Feature–space representation (e.g., PCA dimensionality reduction)
- Model architecture and hyperparameters (ML)
- Extrapolation / out-of-distribution conditions
- Hybrid model coupling and internal inconsistencies

5. Reference-data uncertainties

- Field sampling errors
- Instrument and laboratory measurement errors
- Destructive measurements
- Spatial support mismatch
- Temporal representativeness
- Scale and aggregation differences

Propagation and interaction

Uncertainties propagate through the chain and interact in complex ways, including intermediate feature transformations used in hyperspectral retrieval. Their relative importance varies with wavelength, scene conditions, vegetation type, and sensor characteristics.

Impact on final products

The combination of all sources determines the reliability, interpretability, and transferability of hyperspectral vegetation products.

UNCERTAINTY TYPES: ALEATORIC vs EPISTEMIC UNCERTAINTY

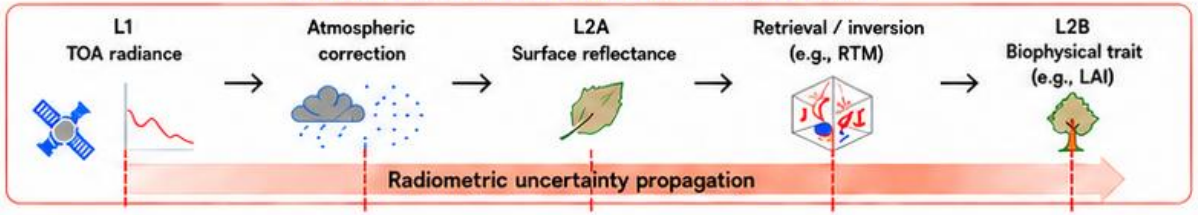
Uncertainty in EO model predictions comes from two fundamental sources.

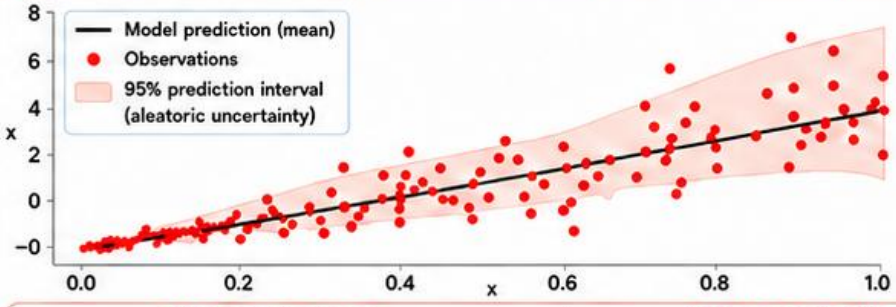
A ALEATORIC UNCERTAINTY

Observation / radiometric uncertainty

Aleatoric uncertainty originates from radiometric uncertainty and observation conditions and propagates through the EO processing chain (sensor → atmosphere → surface → retrieval). It is irreducible.

PROPAGATION THROUGH THE EO PROCESSING CHAIN





KEY MESSAGE

Propagates through the EO chain.

Sensor noise, calibration uncertainty, atmospheric correction, BRDF effects, mixed pixels, and surface heterogeneity.

CHARACTERISTICS

- Propagated from L1 radiometric uncertainty
- Includes sensor, atmosphere, and canopy variability
- Present in all retrievals
- Quantified through uncertainty propagation (Monte Carlo, analytical / Jacobian)

EXAMPLES (SOURCES)

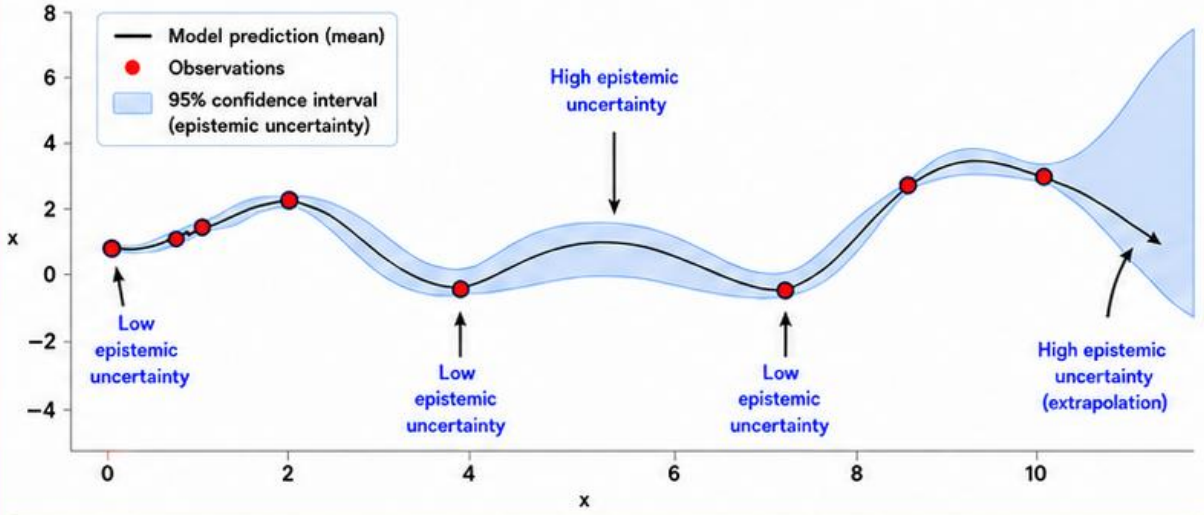
- Sensor noise / SNR
- Atmospheric correction residuals
- BRDF and observation geometry
- Mixed pixels / canopy heterogeneity

B EPISTEMIC UNCERTAINTY

Model (knowledge) uncertainty

Epistemic uncertainty arises from limited knowledge, model assumptions, structure errors or scarce data. It is reducible with more/better data or improved models. It is reducible.

Can be reduced by collecting more/better data or improving the model.



CHARACTERISTICS

- Due to limited data, model structure, parameter uncertainty
- Large in regions with little or no data
- Quantified by the spread of the mean prediction across plausible models

EXAMPLES (SOURCES)

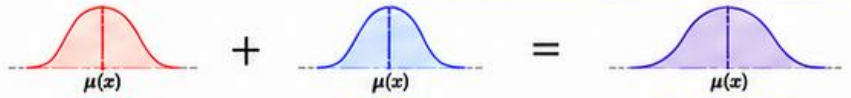
- Uncertain model parameters
- Missing physical processes
- Limited training data
- Structural model errors

KEY DIFFERENCES

Aspect	Aleatoric Uncertainty	Epistemic Uncertainty
Source	Radiometric uncertainty, noise, natural variability	Limited knowledge, model assumptions
Reducible?	No (irreducible)	Yes (reducible)
Behavior with more data	Remains the same	Decreases
Where it appears	Everywhere (all observations and predictions)	Large in data-sparse / extrapolation regions

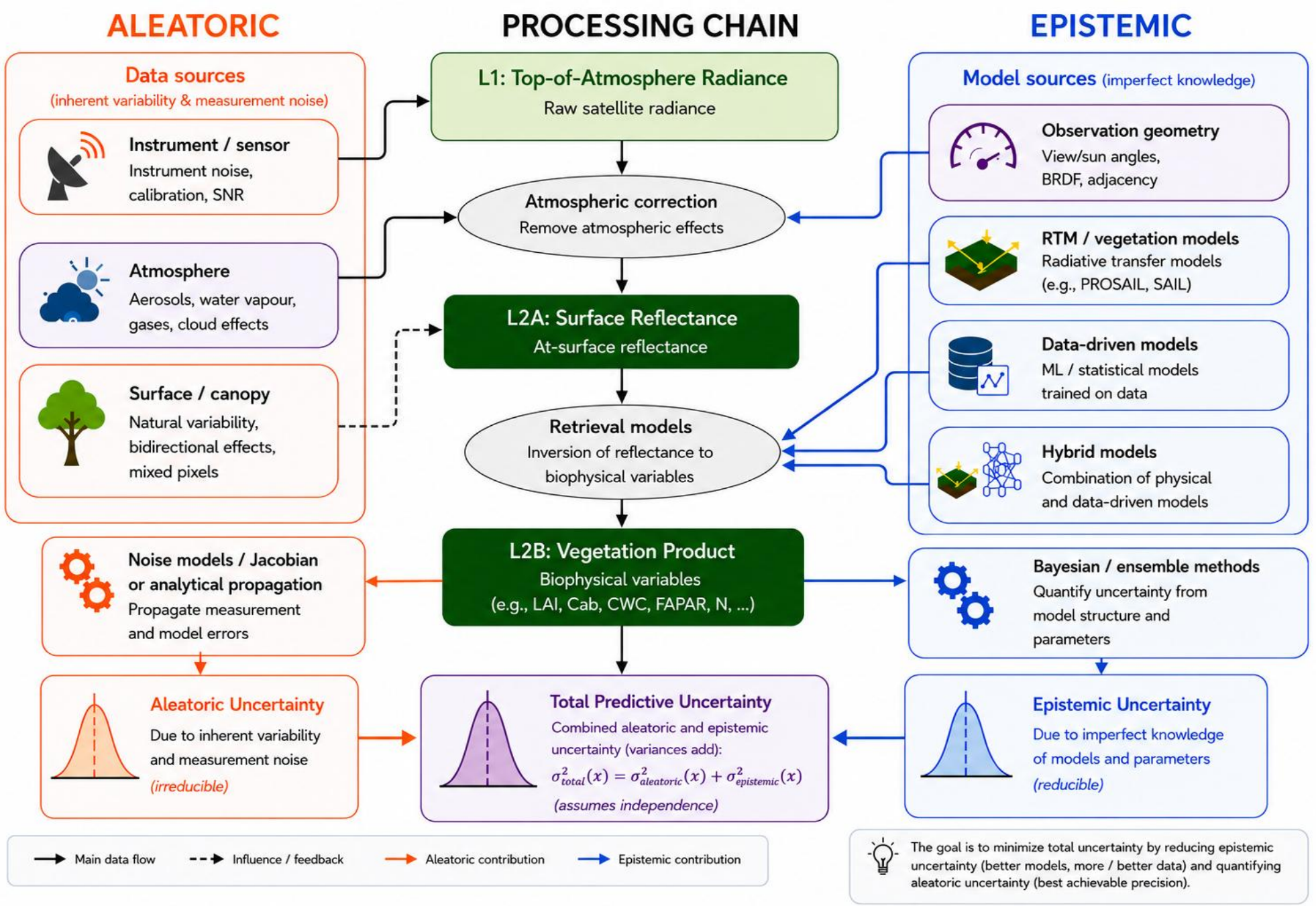
HOW THEY COMBINE

Assuming independence, variances add: $\sigma^2_{total}(x) = \sigma^2_{aleatoric}(x) + \sigma^2_{epistemic}(x)$



TAKE-HOME MESSAGE

Aleatoric uncertainty sets the best achievable precision (the noise floor). Epistemic uncertainty sets the potential for improvement.



Verrelst, J., García-Soria, J. L., Reyes-Muñoz, P., De Clerck, E., Morata, M., & Rivera-Caicedo, J. P. (2026). **Epistemic and aleatoric uncertainty in optical vegetation trait retrieval: Concepts, Methods, and Outlook**. *ISPRS Journal of Photogrammetry and Remote Sensing*, *234*, 20-45.

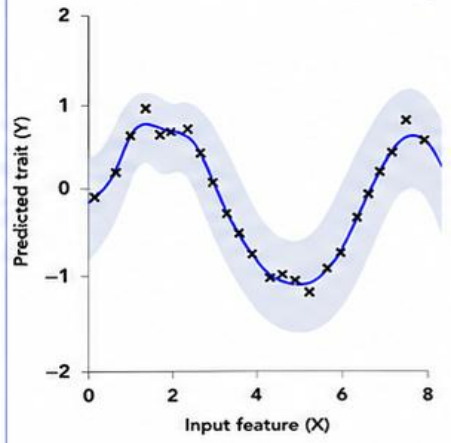
ML methods with epistemic uncertainty: conceptual examples

Representative Uncertainty-Aware Machine Learning Regression Algorithms

Comparison of representative probabilistic machine learning regression algorithms for estimating predictive uncertainty in hyperspectral vegetation retrieval.

These algorithms quantify **epistemic uncertainty** (reducible in principle) and provide uncertainty-aware predictions.
They are designed to capture nonlinearity and account for uncertainty in the data and/or model.

(a) Gaussian Process Regression (GPR)



MECHANISM

Places a Gaussian prior over functions. Analytical posterior mean and variance at any input based on a covariance kernel. Uncertainty grows away from data.

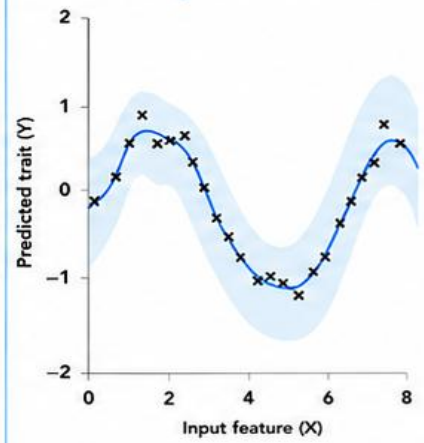
STRENGTHS

- Well-calibrated uncertainty
- Captures model uncertainty naturally
- Works well with small datasets

LIMITATIONS

- Computational cost $\mathcal{O}(n^3)$
- Sensitive to kernel choice
- Scalability to large datasets

(b) Bayesian Neural Networks (BNNs)



MECHANISM

Treats network weights and biases as random variables with prior distributions. Bayesian inference (e.g., variational inference or MCMC) approximates the posterior over weights. Predictions integrate over this posterior.

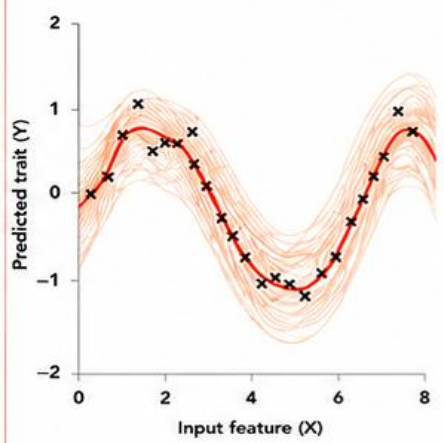
STRENGTHS

- Principled Bayesian treatment
- Captures complex nonlinear relationships
- Provides full predictive distributions

LIMITATIONS

- Computationally intensive
- Requires approximate inference
- Sensitive to architecture & priors

(c) Monte Carlo Dropout



MECHANISM

Dropout kept active at test time. Multiple stochastic forward passes approximate a Bayesian neural network posterior.

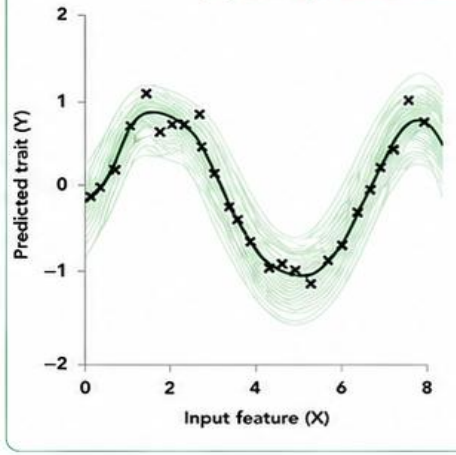
STRENGTHS

- Scales to large & complex data
- Compatible with existing DNNs
- Simple to implement in deep nets

LIMITATIONS

- Approximation quality depends on dropout rate
- May be miscalibrated
- Sensitive to network design

(d) Deep Ensembles



MECHANISM

Train multiple deep networks with different initializations / data splits. Uncertainty from variability across ensemble members.

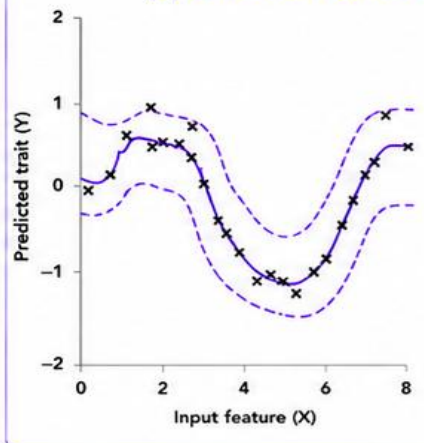
STRENGTHS

- Strong predictive performance
- Well-calibrated uncertainty
- Simple and effective

LIMITATIONS

- Computationally expensive
- Large storage requirements
- Less efficient for very large ensembles

(e) Quantile Random Forest (QRF)



MECHANISM

Directly estimates conditional quantiles (e.g., 5%, 50%, 95%). Uncertainty from the empirical distribution of responses in leaf nodes.

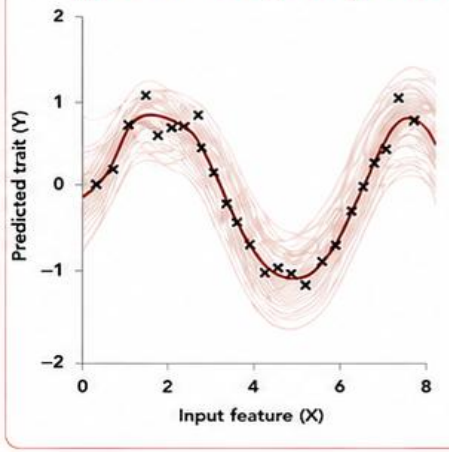
STRENGTHS

- Distribution-free, non-parametric
- Captures asymmetric uncertainty
- Works well with tabular data

LIMITATIONS

- Less smooth intervals
- Poor extrapolation
- Needs many data in leaves

(f) Kernel Ridge Regression + Bootstrapping



MECHANISM

Kernel ridge regression for mean prediction. Bootstrap resampling of data to approximate sampling distribution of predictions.

STRENGTHS

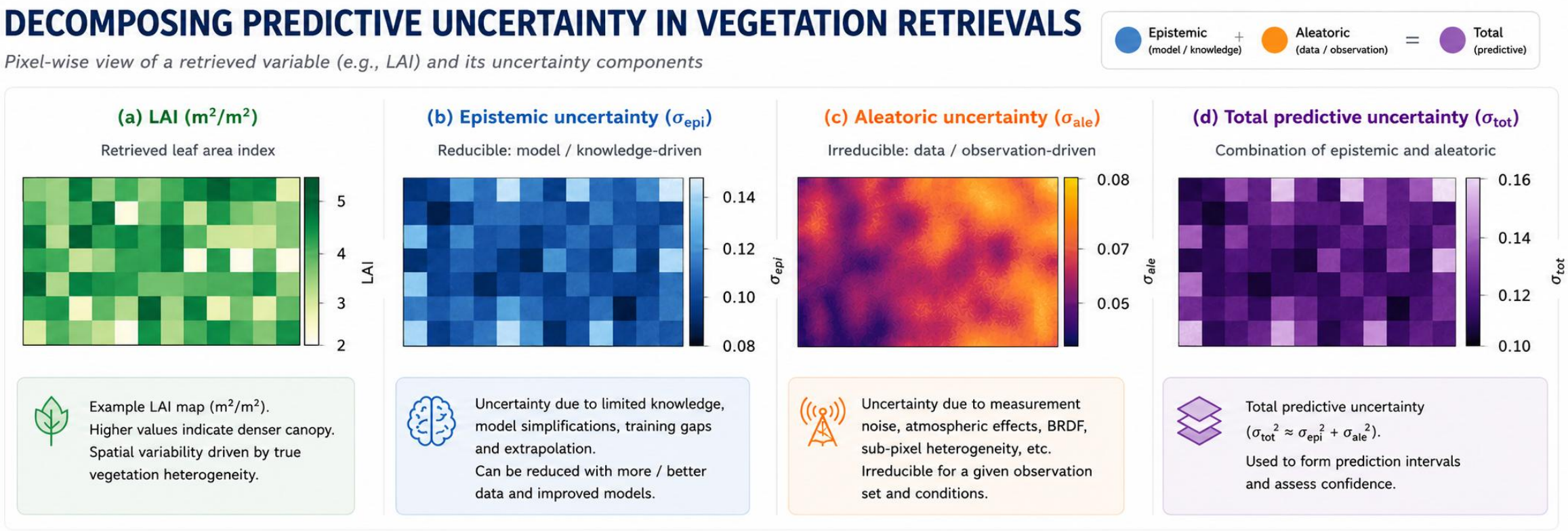
- Simple & model-agnostic add-on
- Provides uncertainty estimates
- Works with small/medium datasets

LIMITATIONS

- Captures total uncertainty (mix)
- Bootstrapping can be unstable
- Computational cost with large B

Moving towards L2B: UQ in vegetation traits retrievals

- **L1 → L2A:** tracing and propagation of uncertainty sources into surface reflectance products (**primarily aleatoric uncertainty**)
- **L2B vegetation retrievals:** current studies largely focus on **epistemic uncertainty**, often neglecting propagated aleatoric uncertainty
- **Both uncertainty types are required to obtain: total predictive uncertainty**
- **No operational end-to-end UQ framework currently exists in L2B:** a clear need remains for **benchmarking and integrating aleatoric and epistemic uncertainty quantification**



GBOV-SYN-GPR

→ GPR **empirical** models trained on upscaled **GBOV reference data**

🛰️ Sentinel-3 SYNERGY (SY_2_SYN)

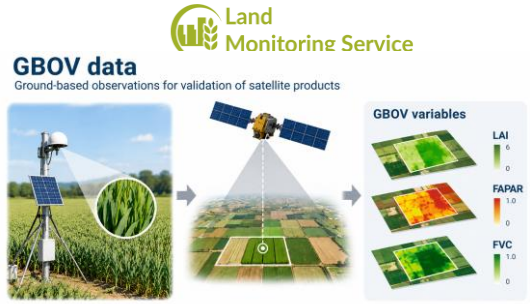
- OLCI + SLSTR fusion at 300 m
- BRDF-normalized surface reflectance
- Improved basis for vegetation trait retrieval

🌿 Vegetation Traits Need Reliability

- **LAI, FAPAR, FVC**
- Users require **uncertainty-aware products**

📏 GBOV as Training Resource

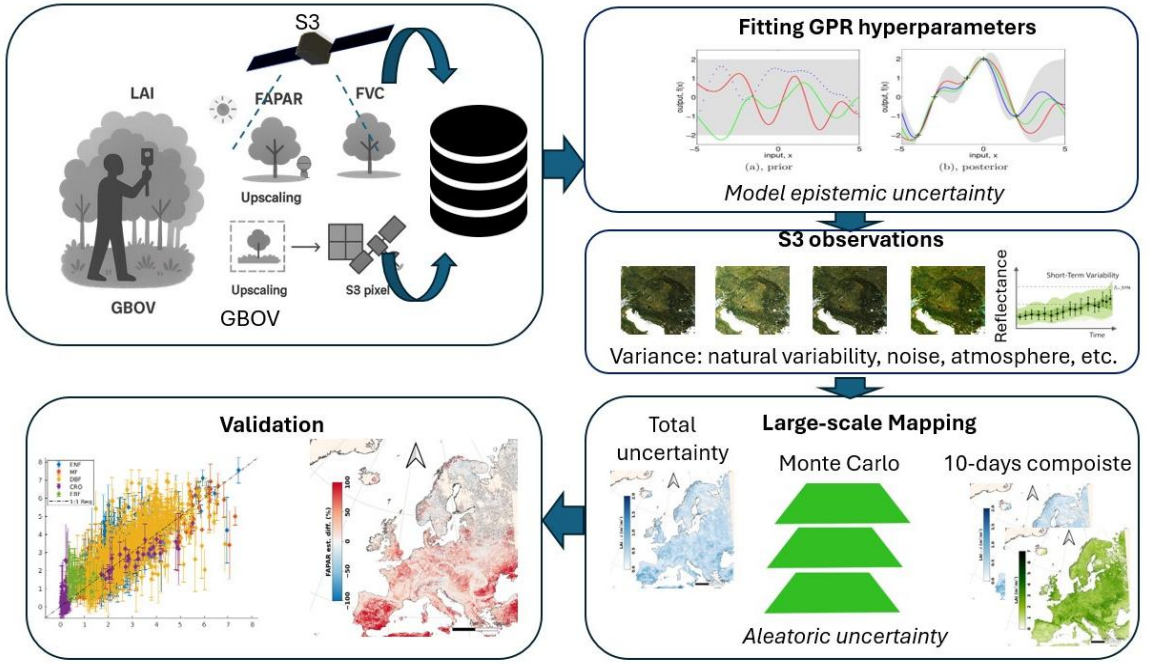
- Harmonized, upscaled in situ reference data
- Multi-biome, multi-year coverage
- Enables empirical model calibration

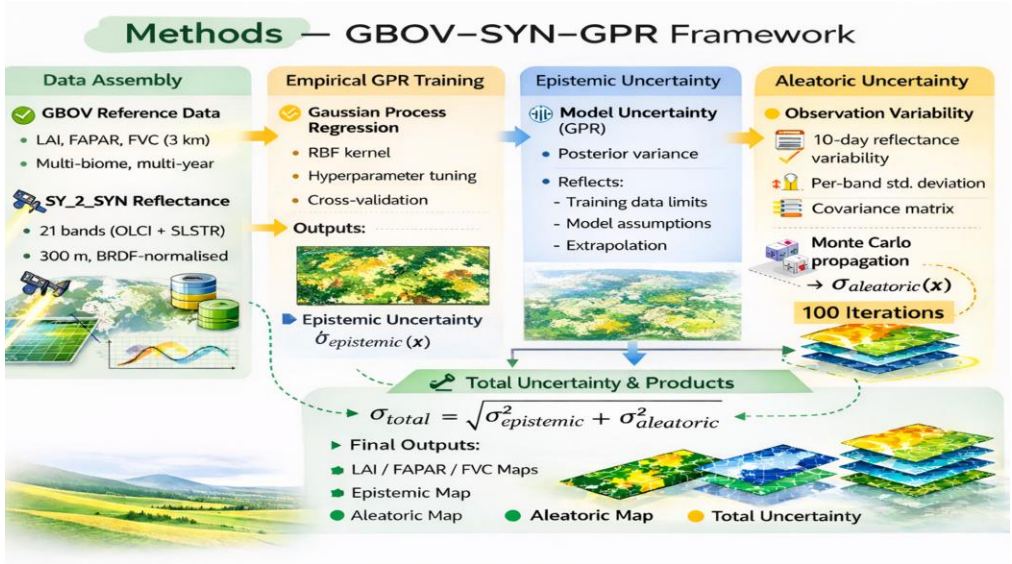


🤖 Uncertainty propagation

- Train **GPR models** directly on GBOV + SY_2_SYN
- Explicit decomposition of:
  - 🌫️ **Aleatoric** uncertainty: **radiometric measurement (noise)-based**
  - 🧠 **Epistemic** uncertainty: **model-based**

Workflow



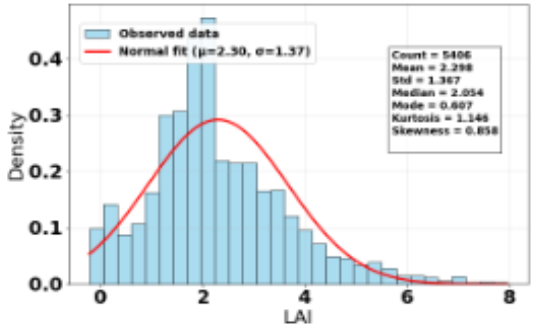


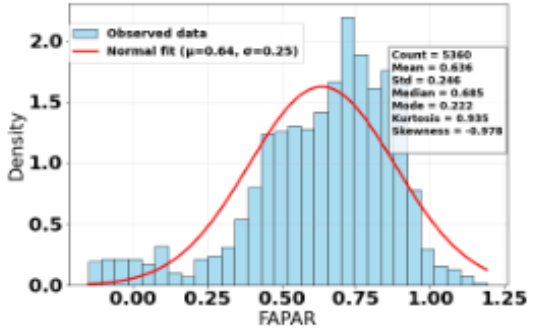
🎯 Goal: Deliver scalable, uncertainty-quantified vegetation trait products from SY_2_SYN.

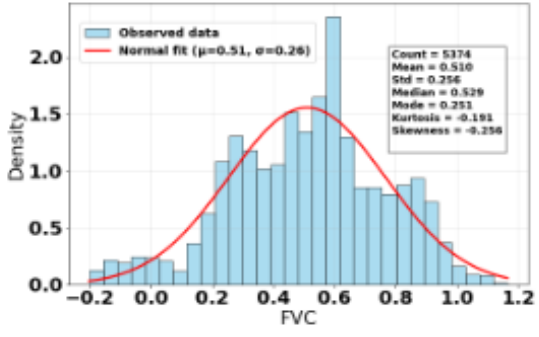
Reyes-Muñoz, Pablo, et al. "**GBOV-trained Gaussian process models for Sentinel-3 SYN vegetation traits with decomposed uncertainty quantification**." *ISPRS Journal of Photogrammetry and Remote Sensing* 242 (2026): 156-183.

GBOV data distribution

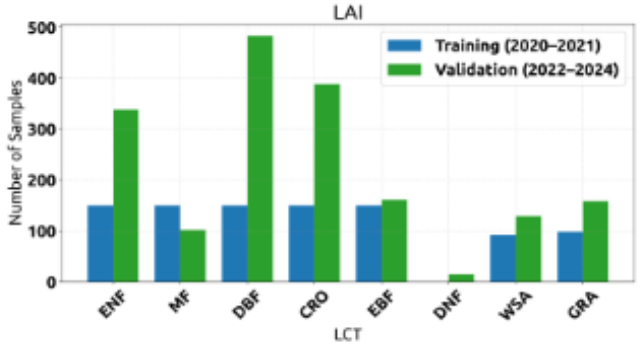
All data: #5350 (2022–2024)

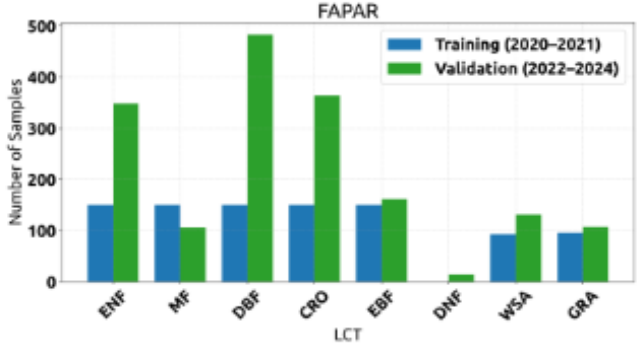


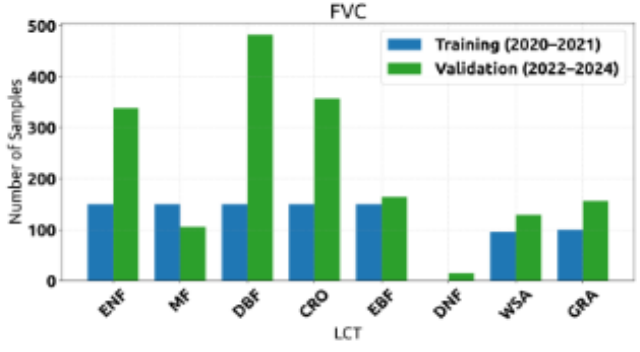


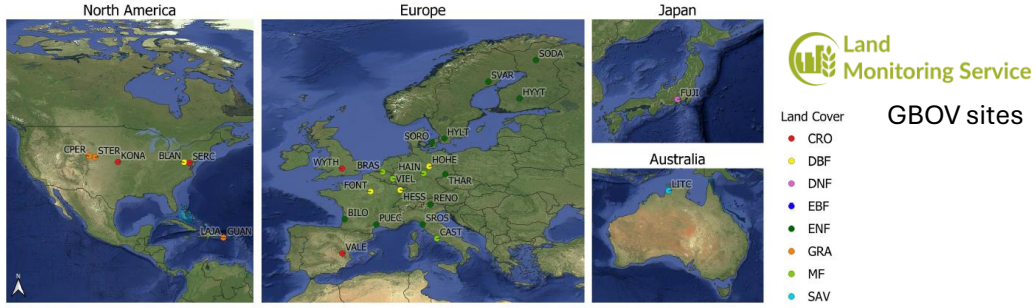


Spread training (#3650) /testing (#1700) per PFT

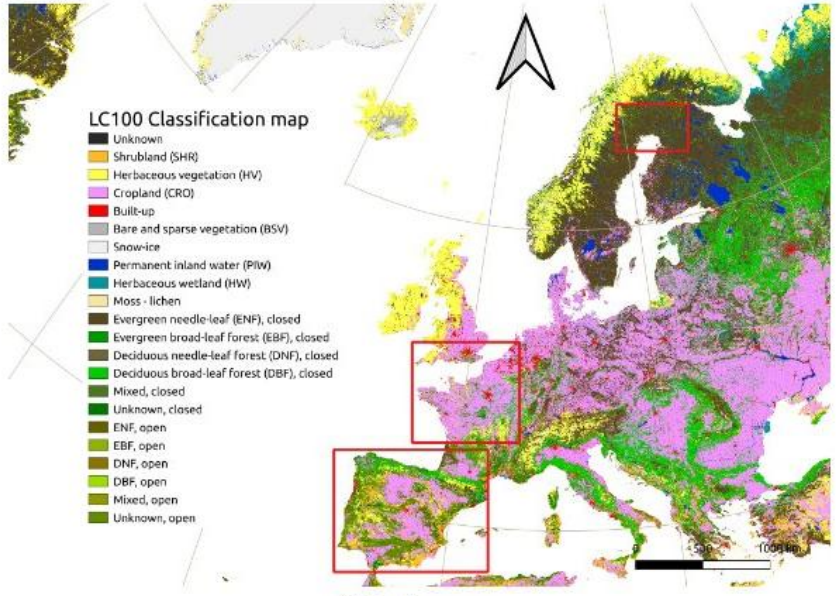








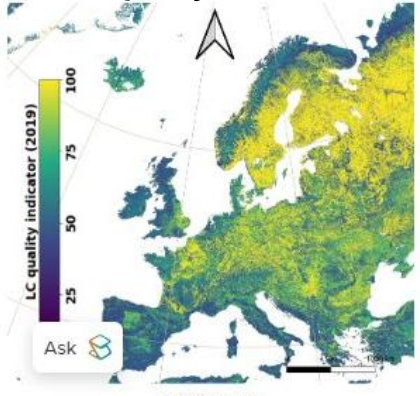
GBOV sites



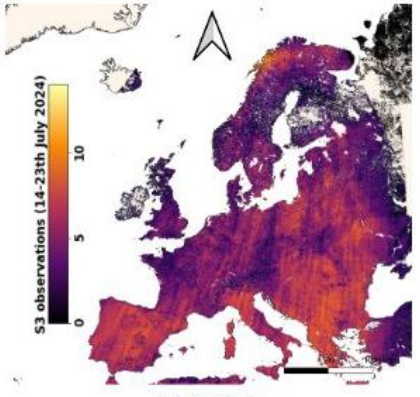
(a) Land cover map

Selected areas for aleatoric uncertainty propagation

Quality indicator



(b) Detail A

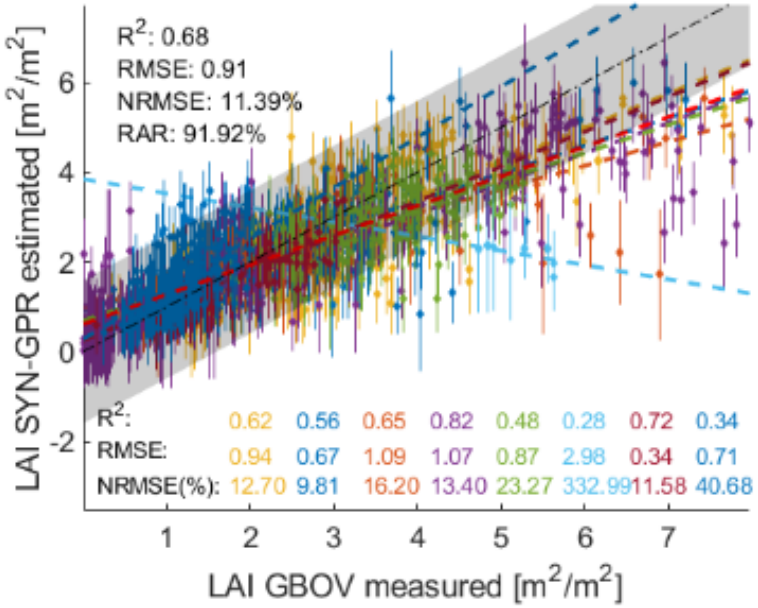


(c) Detail B

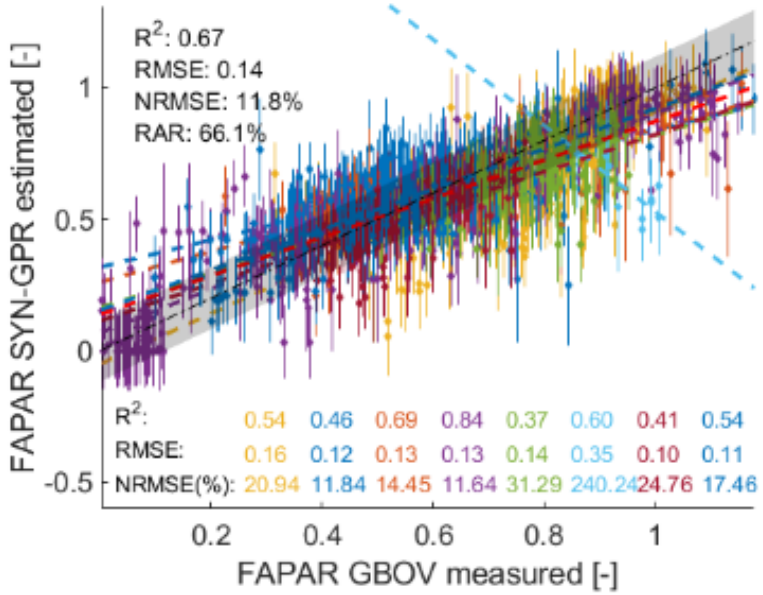
Number of cloud-free S3 observations per pixel within the 10 day temporal range 14-23 July 2024

Validation against independent GBOV data (#1700)

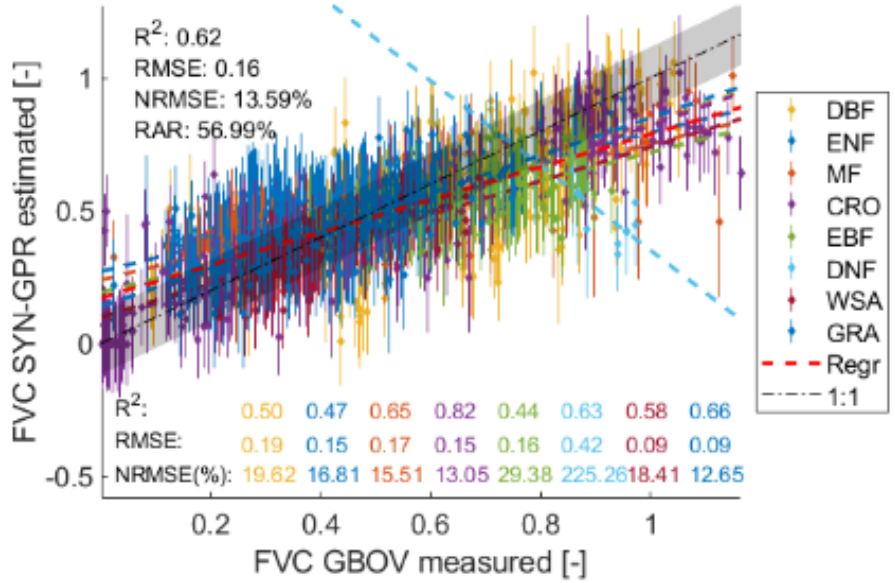
LAI (μ, σ)



FAPAR (μ, σ)



FVC (μ, σ)



📈 Strong Overall Agreement

- **LAI:** $R^2 \approx 0.68$ (RMSE ≈ 0.91)
- **FAPAR:** $R^2 \approx 0.67$ (RMSE ≈ 0.14)
- **FVC:** $R^2 \approx 0.62$ (RMSE ≈ 0.16)

🌎 Robust Across Biomes

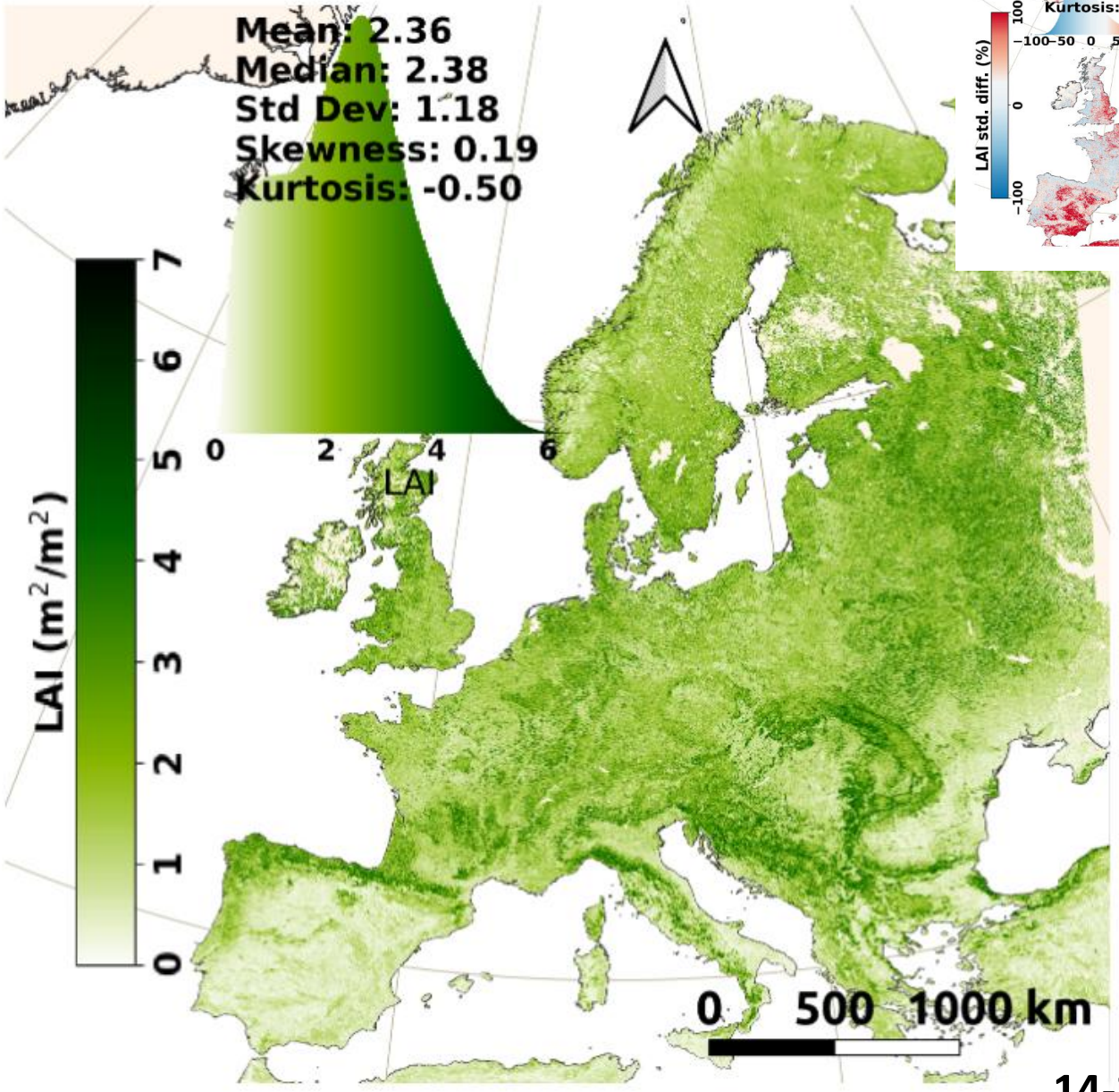
- Good performance in forests (DBF, ENF, MF)
- Larger spread in croplands & grasslands
- Increased variability at high vegetation density

⚖️ Expected Saturation Behaviour

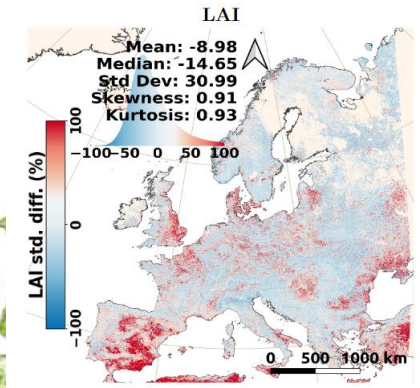
- Slight underestimation at high values
- Typical reflectance-based saturation effects

- **Epistemic uncertainty is Consistent**

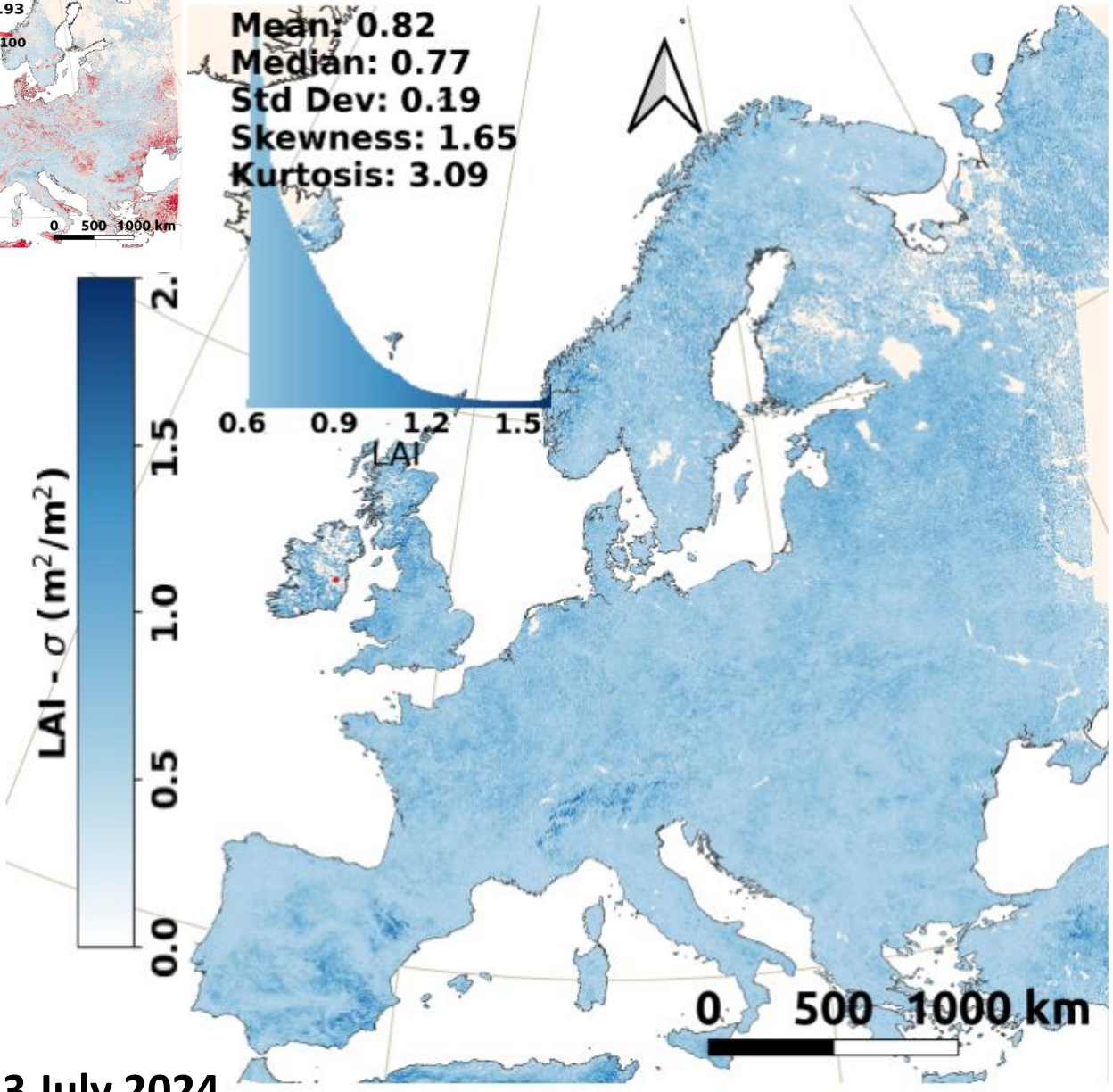
LAI (μ)



Comparison vs CLMS

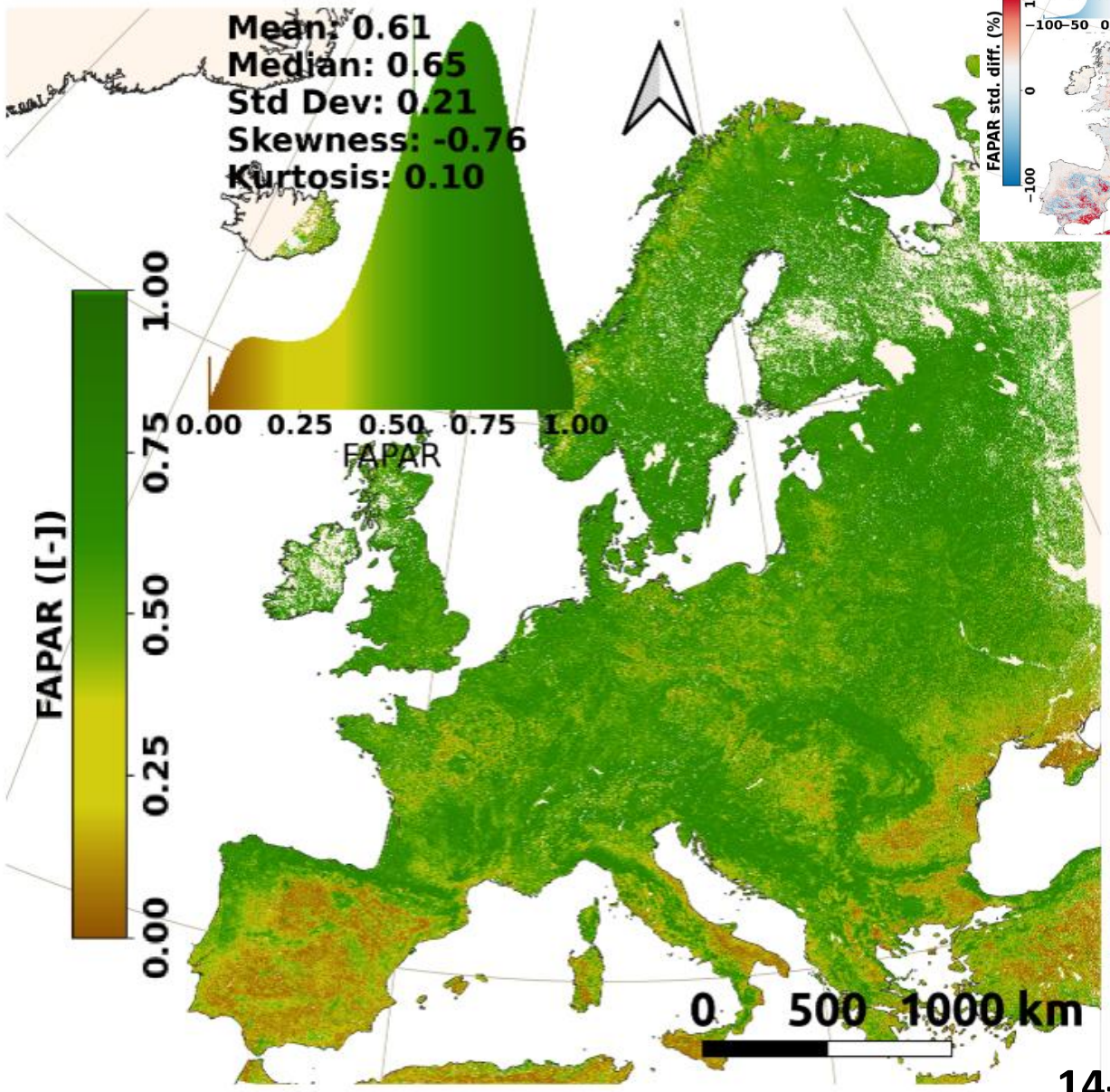


Epistemic LAI (σ)

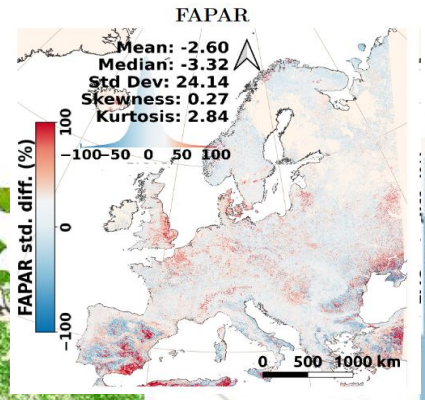


14-23 July 2024

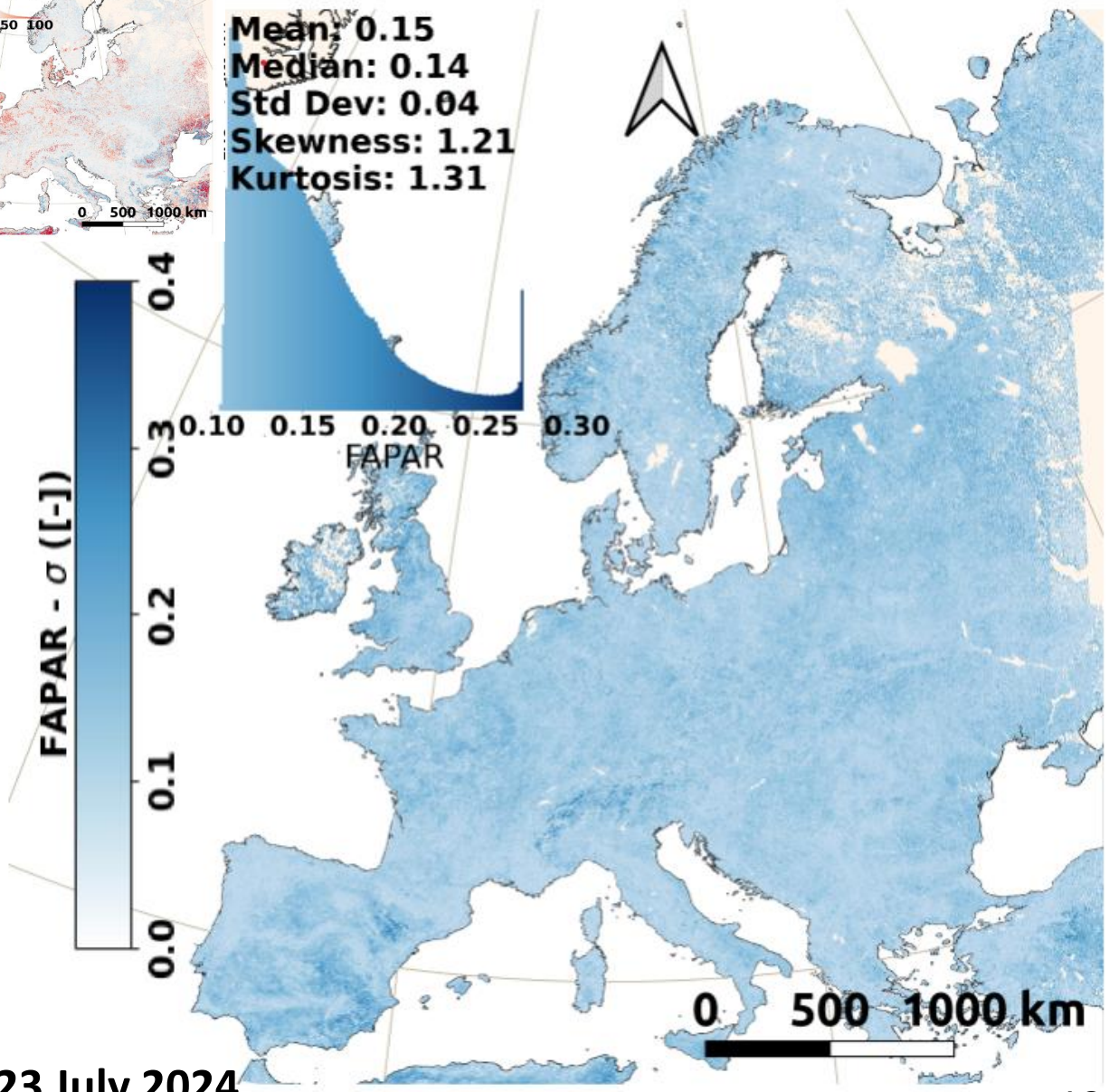
FAPAR (μ)



Comparison vs CLMS

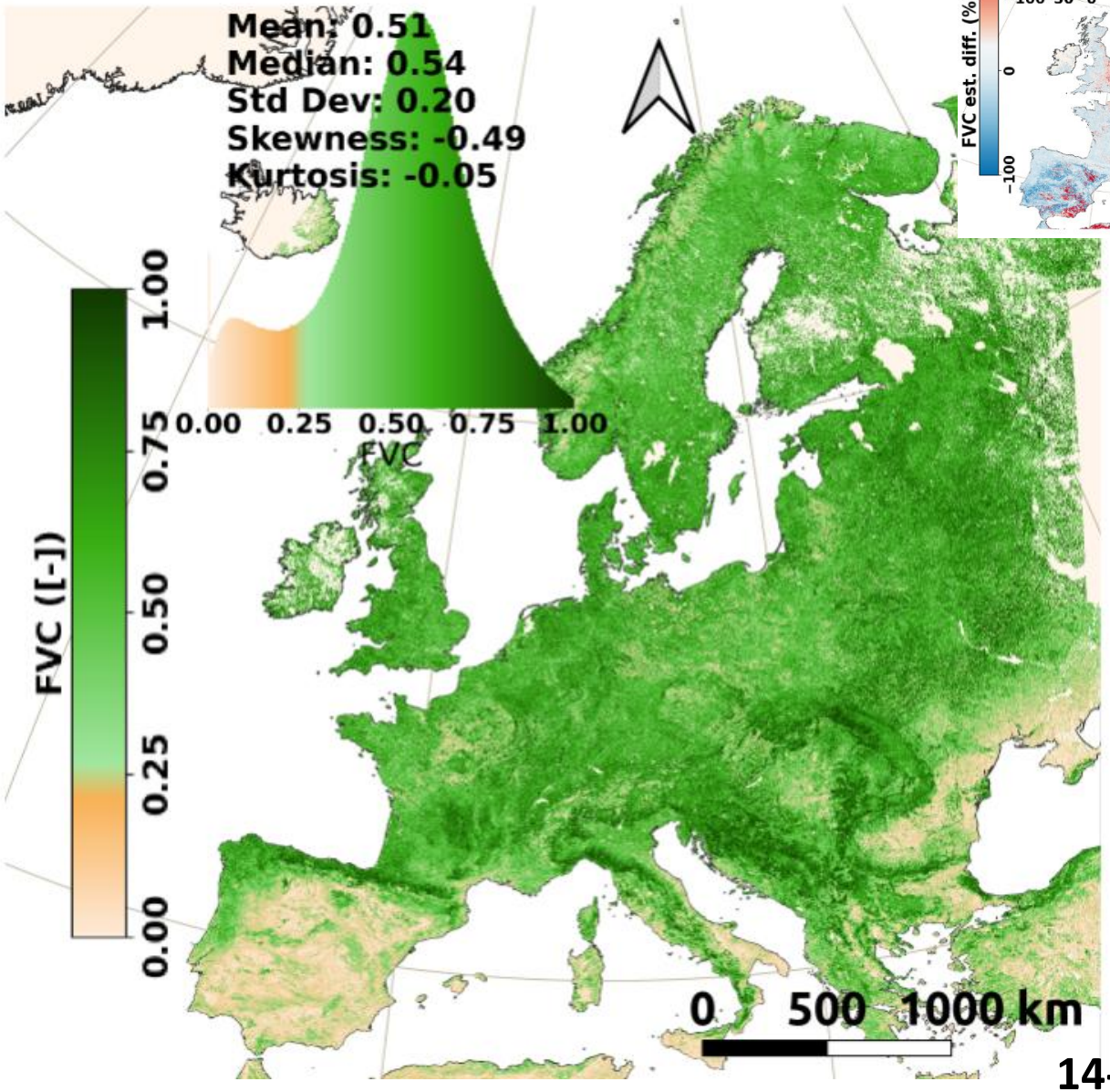


FAPAR (σ)

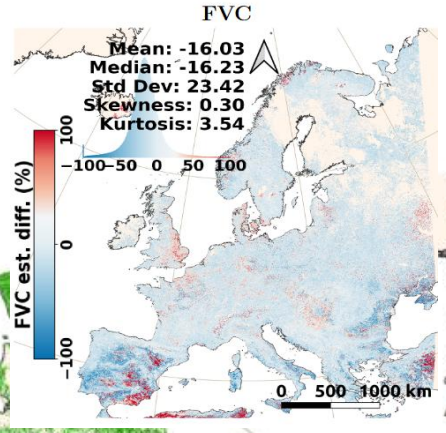


14-23 July 2024

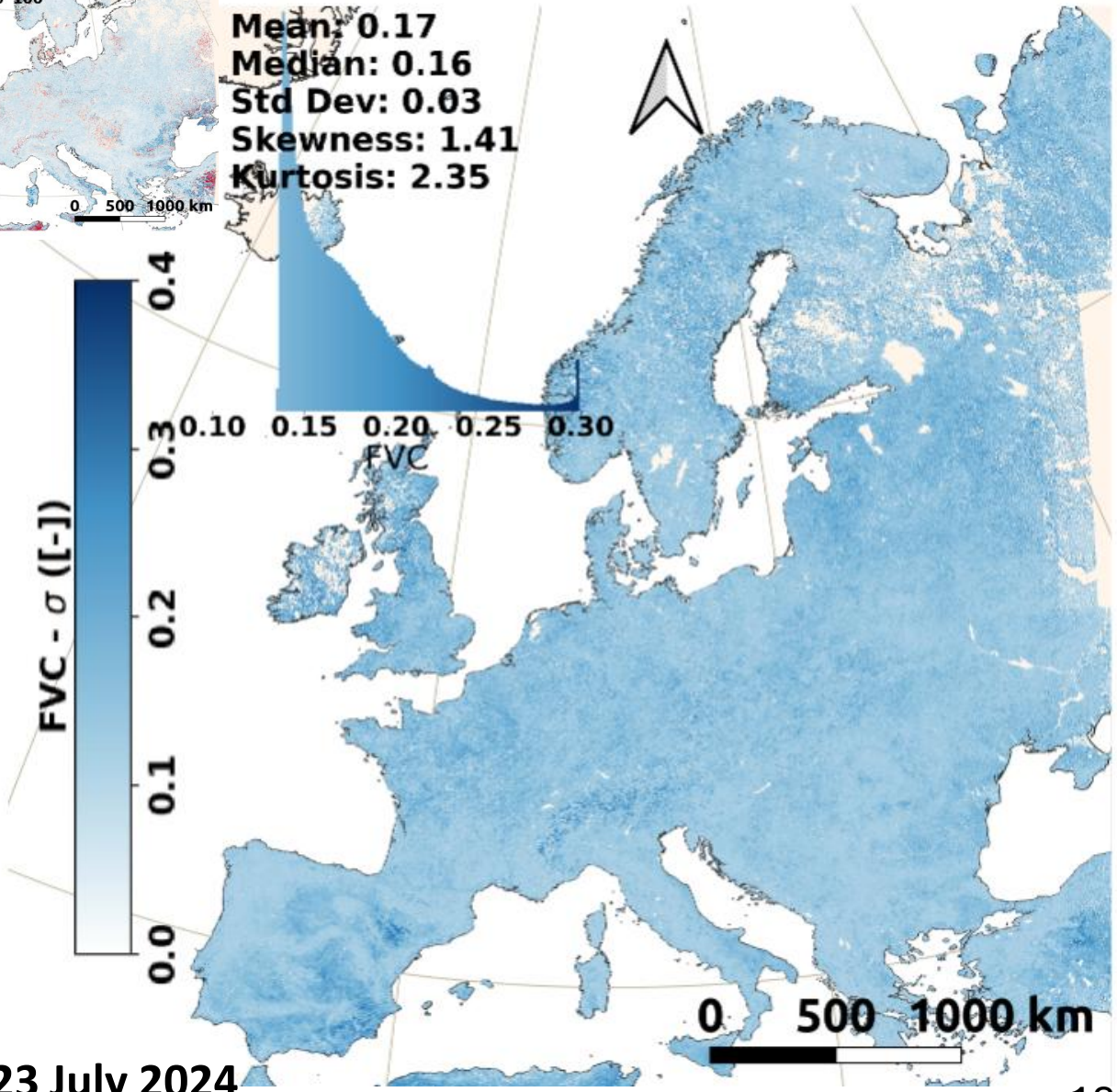
FVC (μ)



Comparison vs CLMS

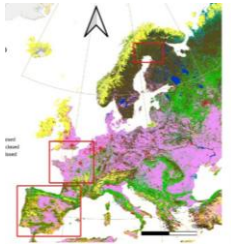


FVC (σ)



14-23 July 2024

Aleatoric uncertainty



🛰️ Reflectance Variability (10-day window)

- Multiple SY_2_SYN acquisitions per pixel
- Per-band **standard deviation**
- Construct reflectance **covariance matrix (Σ_refl)**

🎲 Monte Carlo Sampling

- Sample reflectance vectors from: $\mathcal{N}(\mu_{refl}, \Sigma_{refl})$
- **100 perturbations per pixel**

🤖 Propagation Through GPR

- Each sample → trait prediction
- Ensemble of LAI / FAPAR / FVC outputs

📊 Aleatoric Variance

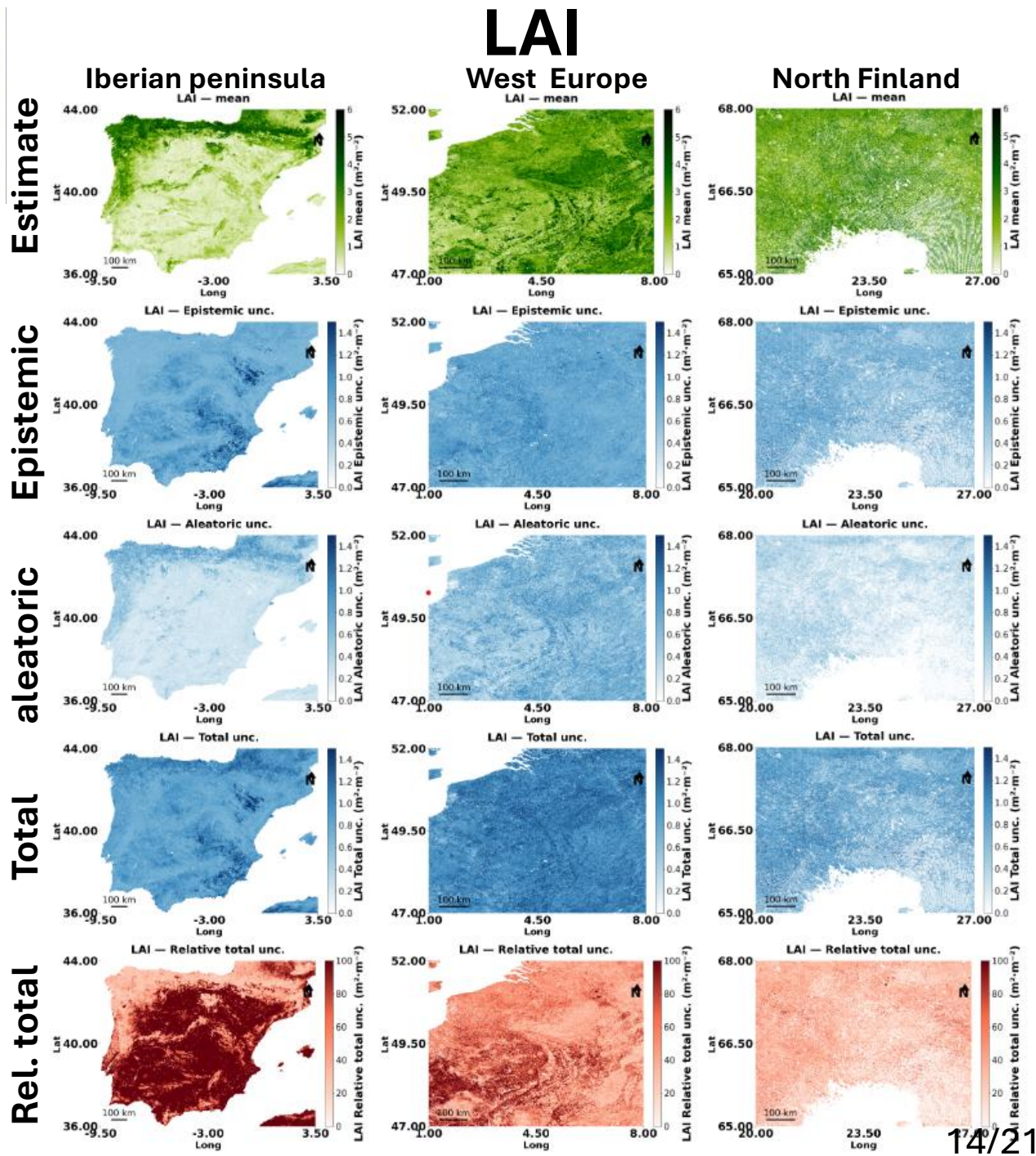
$$\sigma^2_{aleatoric} = Var(f(x_k))$$

- Pixel-wise variance of MC predictions
- Captures observation-driven uncertainty

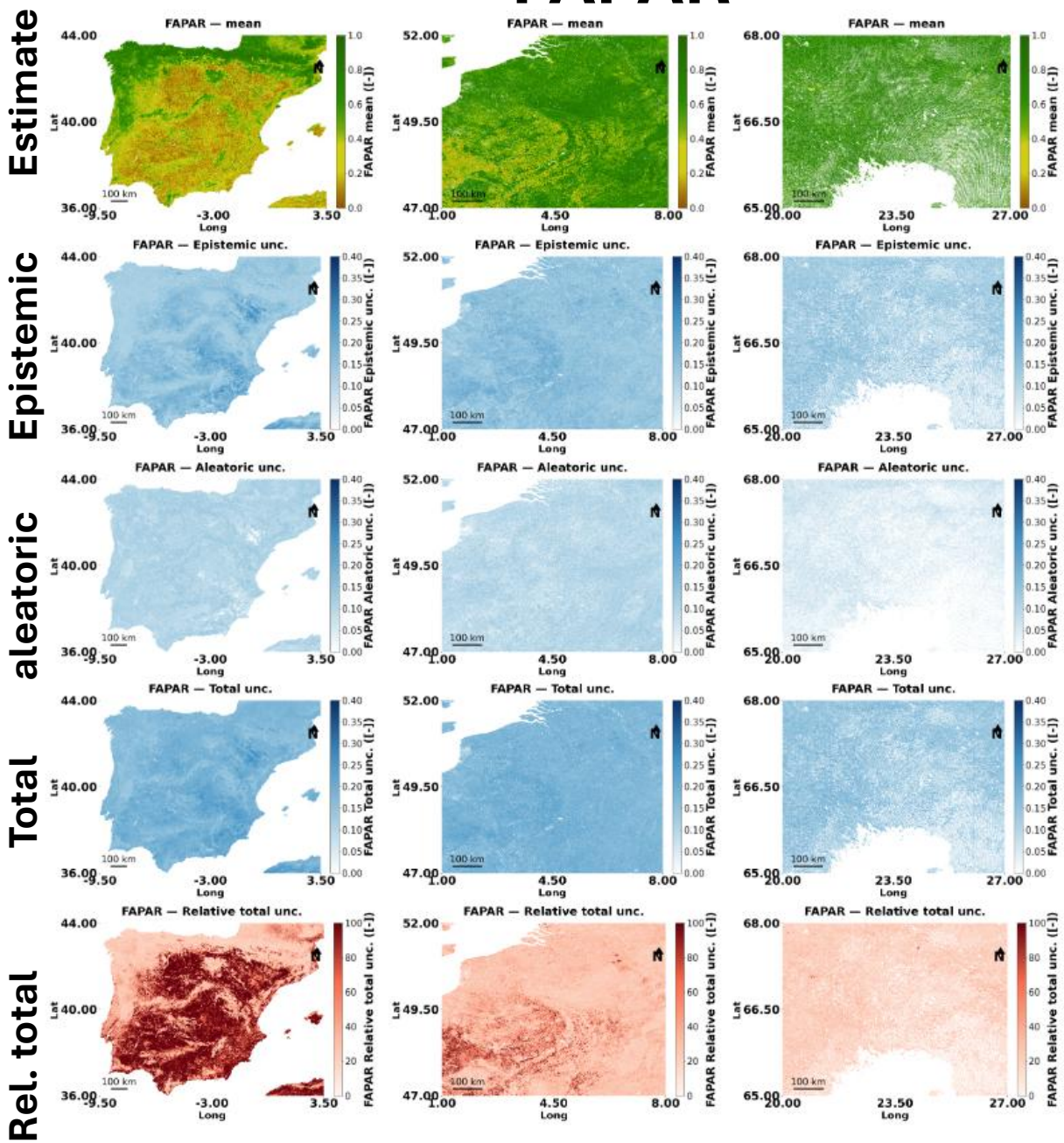
⚖️ Total Uncertainty

- Combined pixel-wise uncertainty

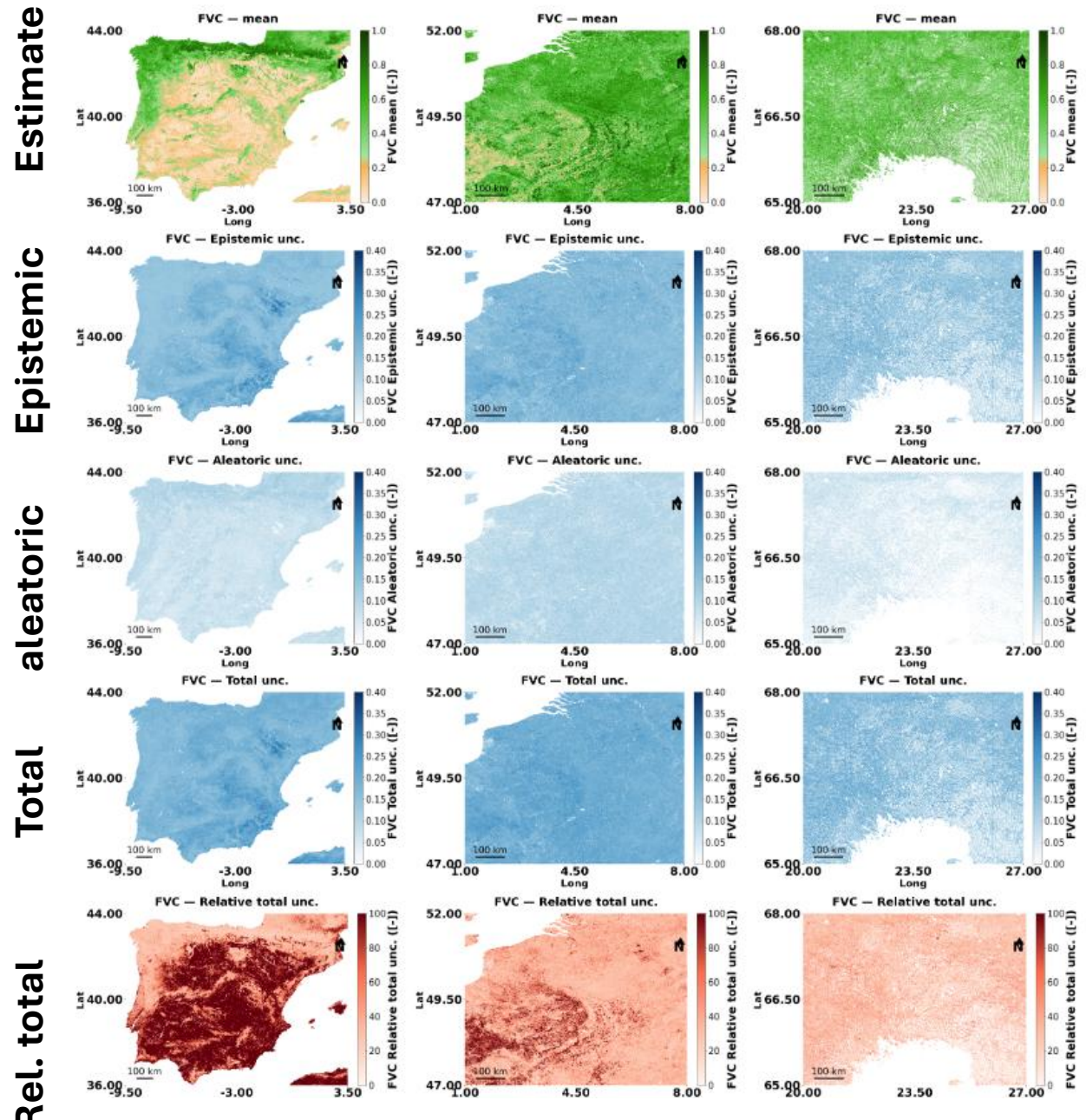
$$\sigma_{total} = \sqrt{\sigma^2_{epistemic} + \sigma^2_{aleatoric}}$$



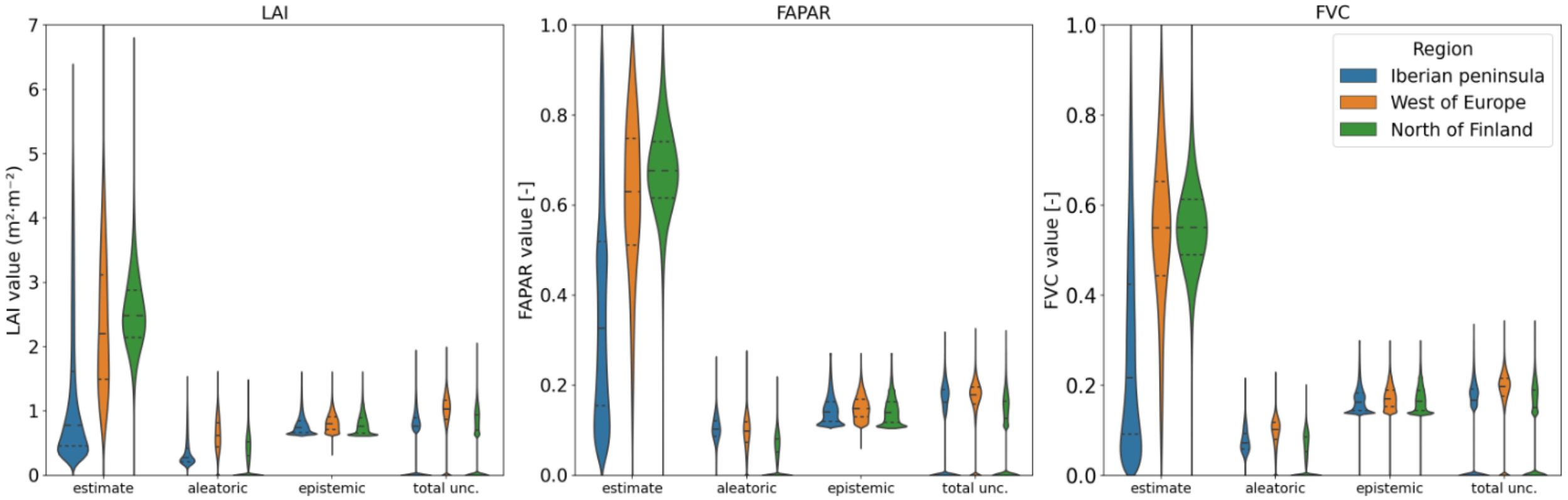
FAPAR


Estimate
Epistemic
aleatoric
Total
Rel. total
FAPAR — mean
FAPAR — Epistemic unc.
FAPAR — Aleatoric unc.
FAPAR — Total unc.
FAPAR — Relative total unc.
FAPAR mean ([-])
FAPAR Epistemic unc. ([-])
FAPAR Aleatoric unc. ([-])
FAPAR Total unc. ([-])
FAPAR Relative total unc. ([-])
Lat
Long
100 km

LAI


Estimate
Epistemic
aleatoric
Total
Rel. total
FVC — mean
FVC — Epistemic unc.
FVC — Aleatoric unc.
FVC — Total unc.
FVC — Relative total unc.
FVC mean ([-])
FVC Epistemic unc. ([-])
FVC Aleatoric unc. ([-])
FVC Total unc. ([-])
FVC Relative total unc. ([-])
Lat
Long
100 km

Epistemic vs aleatoric uncertainty



- 🌍 **Clear Regional Gradient in Trait Estimates**
- 🌊 **Iberian Peninsula:** lowest LAI / FAPAR / FVC
- 🌳 **West of Europe:** intermediate values
- 🎄 **North of Finland:** highest vegetation levels
- ➡️ Retrieval captures expected biogeographical contrast

🌫️ **Aleatoric Uncertainty is Smaller**

- Lower magnitude across all traits
- Slightly higher in heterogeneous regions (Iberia)
- Reflects observation-driven variability

🧠 **Epistemic Uncertainty Dominates**

- Generally larger than aleatoric component
- Indicates model/training limitations remain primary source
- More pronounced at higher trait values

$$\sigma_{total} = \sqrt{\sigma^2_{epistemic} + \sigma^2_{aleatoric}}$$

⚖️ **Total Uncertainty Follows Epistemic Pattern**

- Total ≈ strongly driven by epistemic term
- Regional differences preserved in total uncertainty

EVALUATION vs TRUSTWORTHINESS

Two complementary perspectives on uncertainty-aware vegetation retrievals

EVALUATION
(needs reference data)

Goal: Are uncertainty estimates statistically **correct**?

Accuracy metrics
RMSE, Bias, R^2

Reliability (calibration)
Reliability diagram, PIT

Uncertainty–error consistency
Consistency analysis

Decision-theoretic assessment
Risk–coverage analysis

Answers: How well do the uncertainties match the actual errors?

≠

Different goals, different questions, complementary information.

TRUSTWORTHINESS
(no reference data)

Goal: Can retrievals be **trusted operationally** in this scene?

Uncertainty decomposition
Aleatoric (data) + Epistemic (model)

Training support (OOD)
PCA–Mahalanobis distance to training distribution

Alea–epi interpretation
Why uncertainty occurs (e.g., noise, saturation, extrapolation, mix)

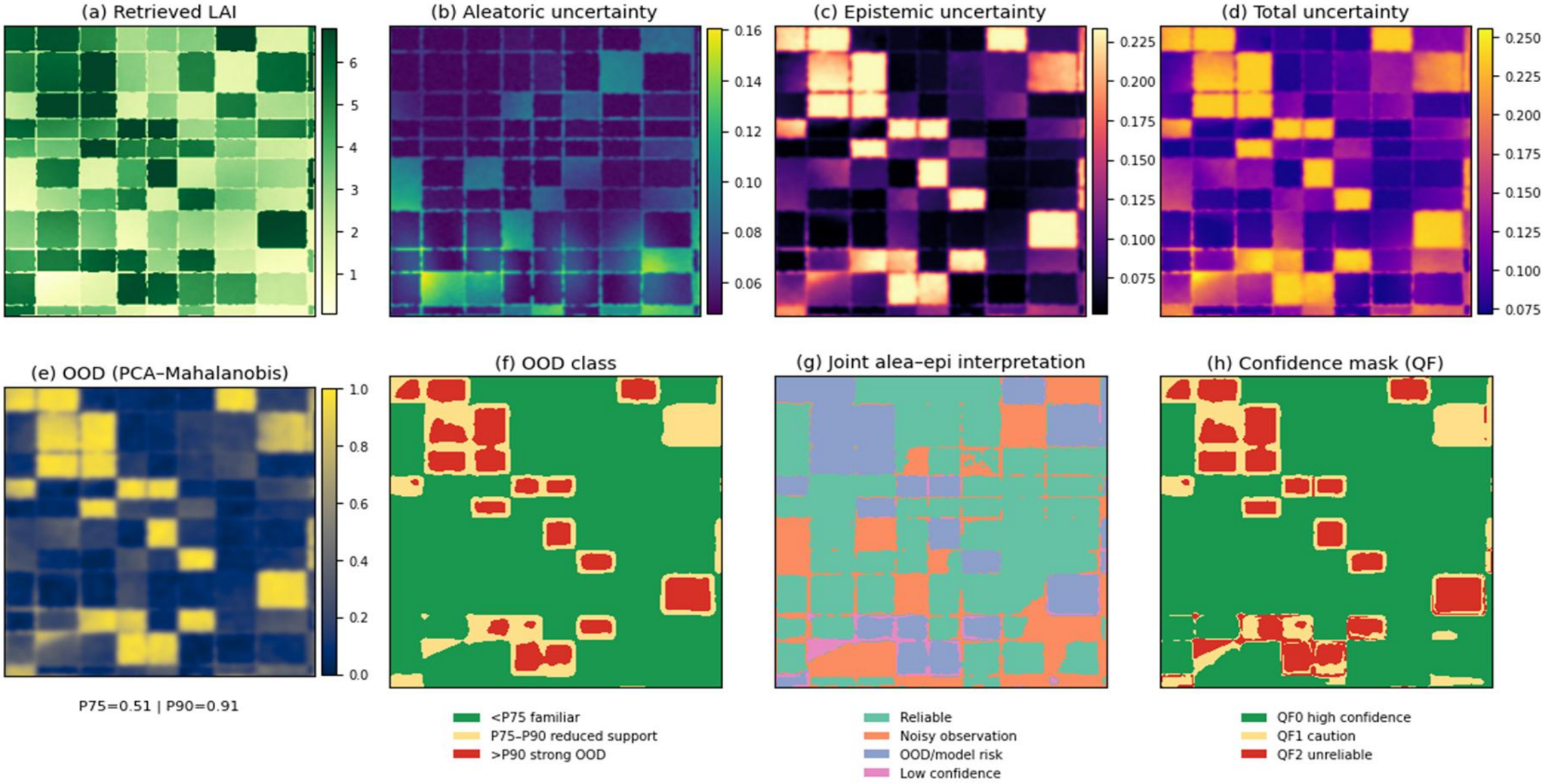
Confidence mask (QF)
Operational summary
QF0 (high) – QF2 (low)

Answers: How much can we trust the retrievals in this scene, and where?

Evaluation tells us "how well". | **Trustworthiness tells us "how much we can trust".**

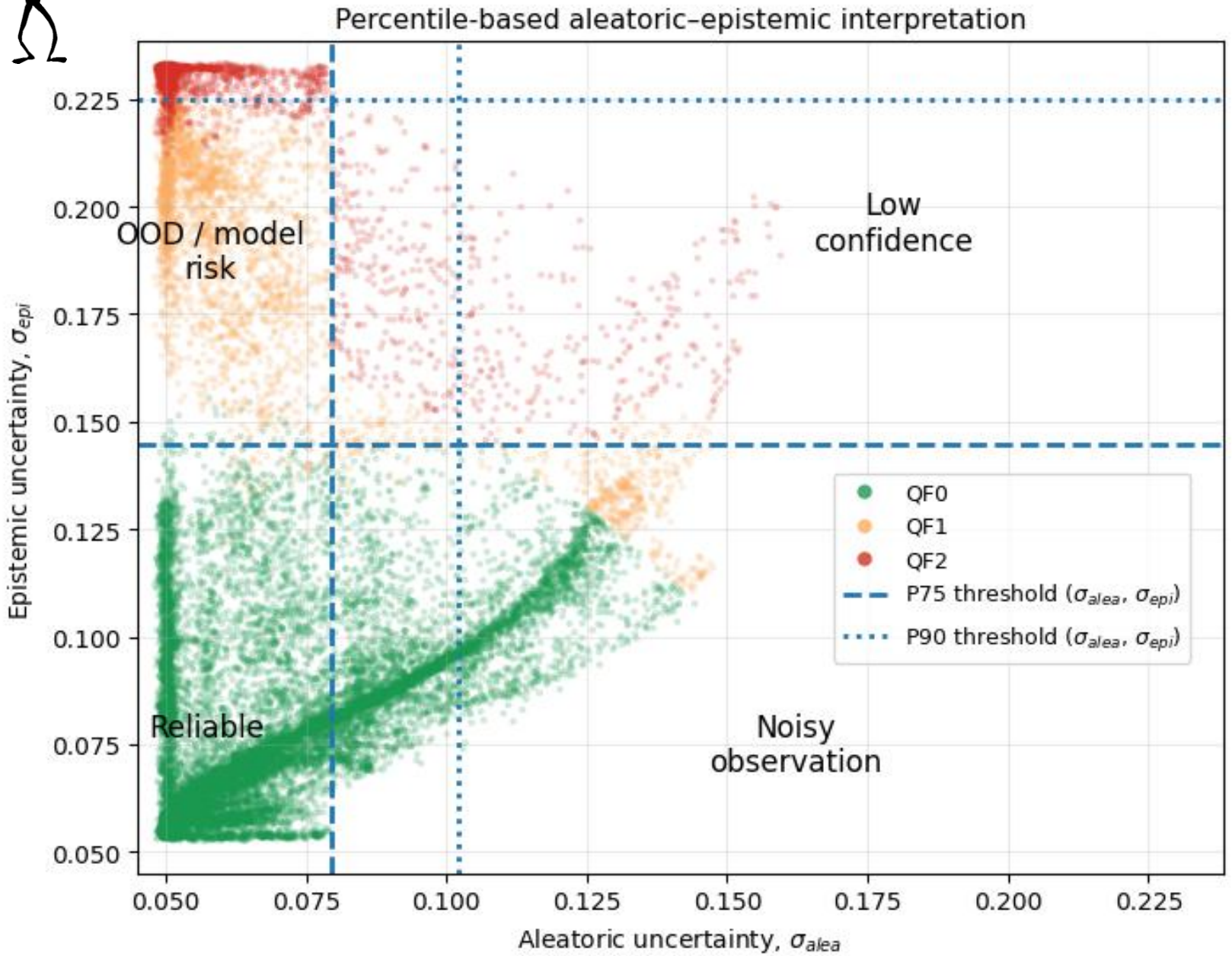
Together, they enable transparent, defensible and actionable vegetation products.

From uncertainty towards trustworthiness...



Uncertainty quantification is only a first step! We need to provide <u>trustworthiness</u>

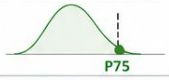
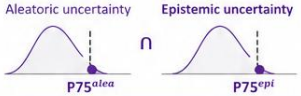
How to interpret L2B products?



HOW THE TRUSTWORTHINESS SCORE (S) IS CALCULATED

S is an integer score ($S \in \{0, 1, 2, 3, 4, 5\}$) obtained by summing five binary indicators.

1) FIVE WARNING INDICATORS

U_{75} Total uncertainty exceeds P75	$I(U > P75_U)$		1 if true, 0 otherwise
+ U_{90} Total uncertainty exceeds P90	$I(U > P90_U)$		1 if true, 0 otherwise
+ OOD_{75} OOD (Mahalanobis) exceeds P75	$I(M > P75_M)$		1 if true, 0 otherwise
+ OOD_{90} OOD (Mahalanobis) exceeds P90	$I(M > P90_M)$		1 if true, 0 otherwise
+ AE Aleatoric–epistemic interaction (both high)	$I(\sigma_{alea} > P75_{alea} \cap \sigma_{epi} > P75^{epi})$		1 if true, 0 otherwise

2) TRUSTWORTHINESS SCORE (S)

$$S = U_{75} + U_{90} + OOD_{75} + OOD_{90} + AE = \sum_{k=1}^{5} I_k$$

- $I(\cdot)$ is an indicator function: 1 if the condition is true, 0 otherwise.
- Each component contributes 0 or 1 point.
- Therefore, $S \in \{0, 1, 2, 3, 4, 5\}$ **(an integer score)**.

Why include AE?

AE identifies pixels where both observations are noisy (high aleatoric) and the model is uncertain (high epistemic) at the same time. This combination poses a higher risk to retrieval trustworthiness than either component alone.

3) QUALITY FLAG (QF) MAPPING

Score (S)	QF	Meaning	Interpretation
$S = 0$	QF0 (high confidence)	●	All indicators are normal. Retrieval generally trustworthy.
$1 \leq S \leq 2$	QF1 (use with caution)	●	At least one warning signal (elevated uncertainty and/or OOD). Retrieval likely useful, but uncertainty deserves attention.
$S \geq 3$	QF2 (potentially unreliable)	●	Multiple warning signals simultaneously (high uncertainty, OOD risk, and/or high **alea & epi**). Retrieval potentially unreliable; interpret with caution or exclude from downstream analysis.

4) EXAMPLE OF SCORING

Indicator	Condition	Value (0/1)
U_{75}	$U > P75_U$	1
U_{90}	$U > P90_U$	0
OOD_{75}	$M > P75_M$	1
OOD_{90}	$M > P90_M$	1
AE	$\sigma_{alea} > P75_{alea} \cap \sigma_{epi} > P75^{epi}$	0
Sum (S)	1 + 0 + 1 + 1 + 0 = 3	3

5) RESULTING QUALITY FLAG

$S = 3 \Rightarrow$ **QF2**

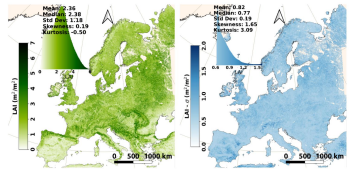
Multiple warning signals detected.

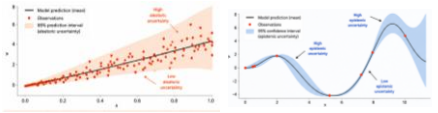
Retrieval potentially unreliable; interpret with caution or exclude from downstream analysis.

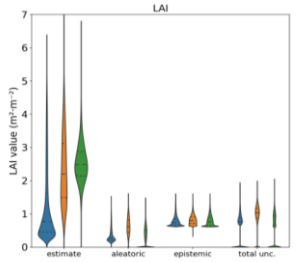
The scoring is hierarchical: exceedance of more extreme thresholds (P90) adds additional evidence of reduced trustworthiness beyond elevated (P75) conditions.

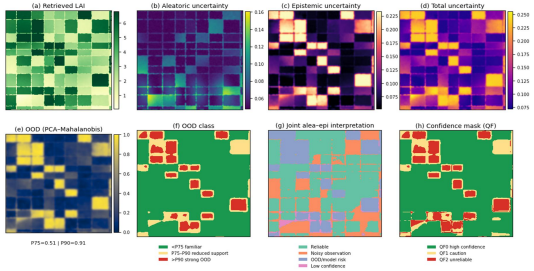
Certainty in conclusions

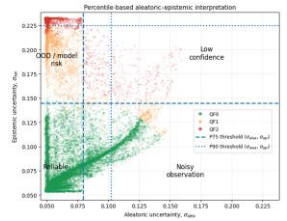
1. **Uncertainty is an essential part of vegetation trait retrieval** — an L2B product should provide the **estimate together with its uncertainty**.
2. **Aleatoric and epistemic uncertainty describe different limitations** and should be quantified separately before being combined into **total predictive uncertainty**.
3. **For the Sentinel-3 GBOV experiment, epistemic uncertainty is the dominant contribution**, pointing to model/training-data limitations rather than observation noise as the main uncertainty source.
4. **Uncertainty alone does not guarantee trustworthiness** — OOD/applicability diagnostics provide complementary information on whether a prediction is made within a sufficiently represented domain.
5. **Next step: operational confidence information** combining **prediction + decomposed uncertainty + OOD/applicability** into intuitive **quality/confidence flags** for large-scale Copernicus vegetation products.











Towards vegetation products that tell us not only *what* we retrieve, but also *how much we can trust it*.

Implemented into ARTMO
https://artmotoolbox.com/

Special Issue

Uncertainty Quantification and Propagation in Vegetation Earth Observation

Guest Editors

Dr. Jochem Verrelst

Dr. Luke A. Brown

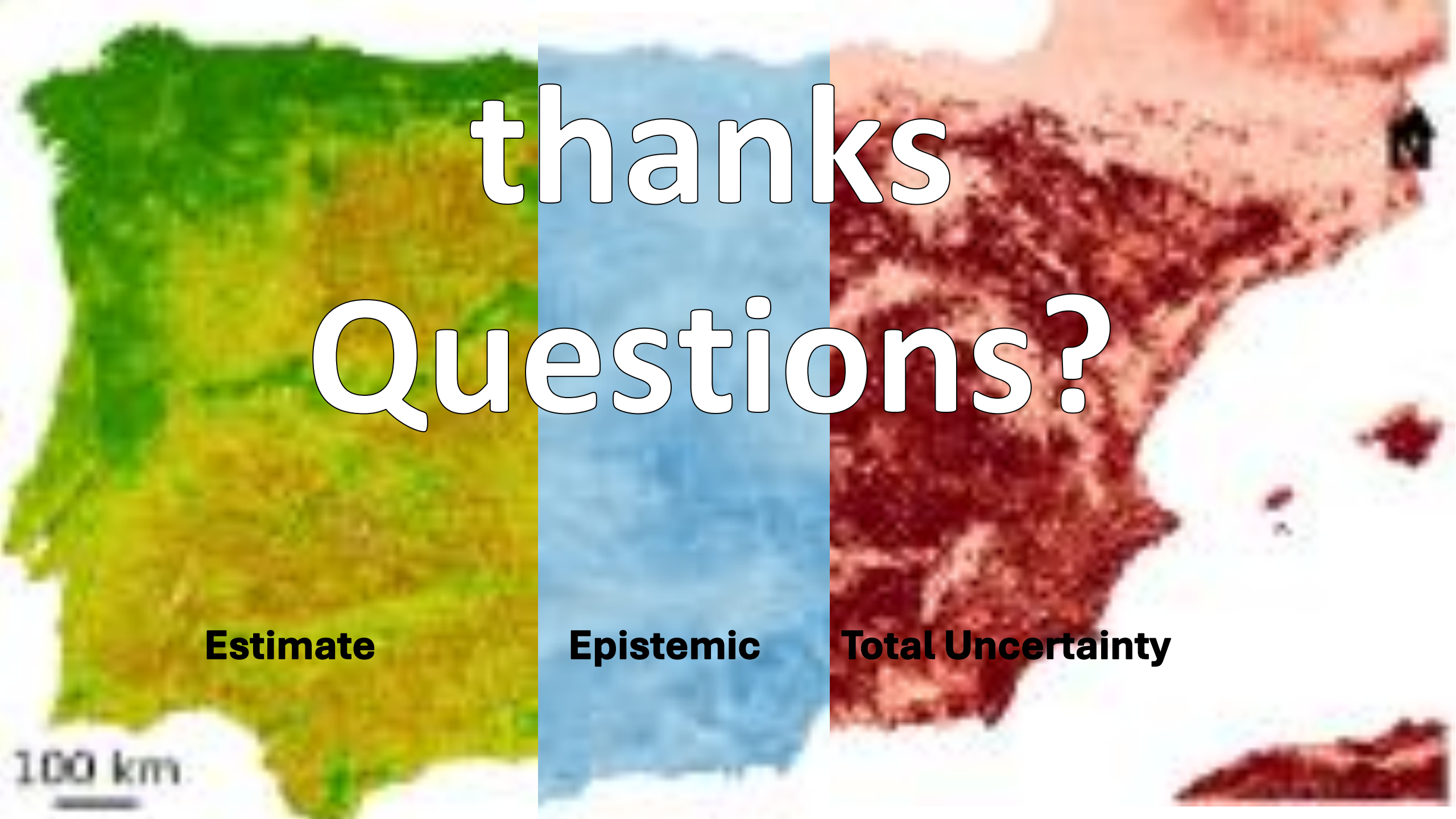
Deadline

28 February 2027

remote sensing

IMPACT FACTOR 4.3

CITESCORE 9.4


thanks
Questions?
Estimate
Epistemic
Total Uncertainty
100 km