Towards Enhanced Data Reliability

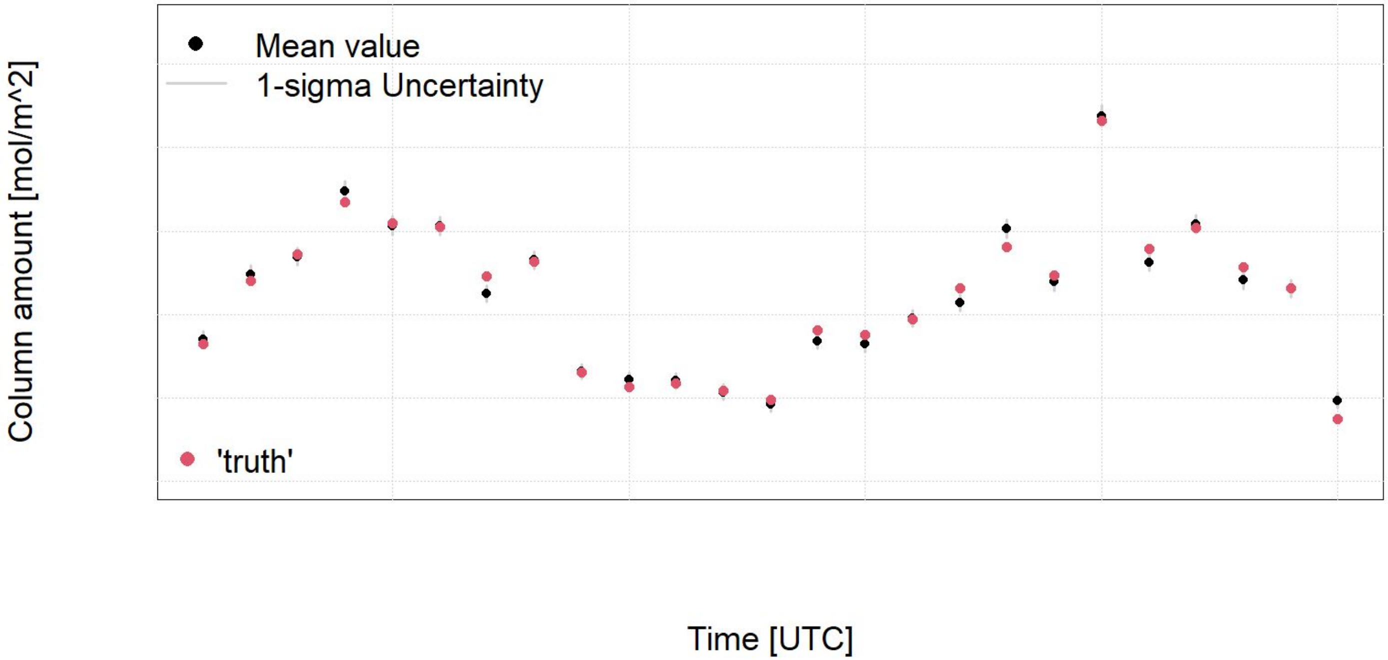
Introducing a Probabilistic Method for Uncertainty Validation in PGN Data Products

QA4EO WP2128

Manuel Gebetsberger, LuftBlick

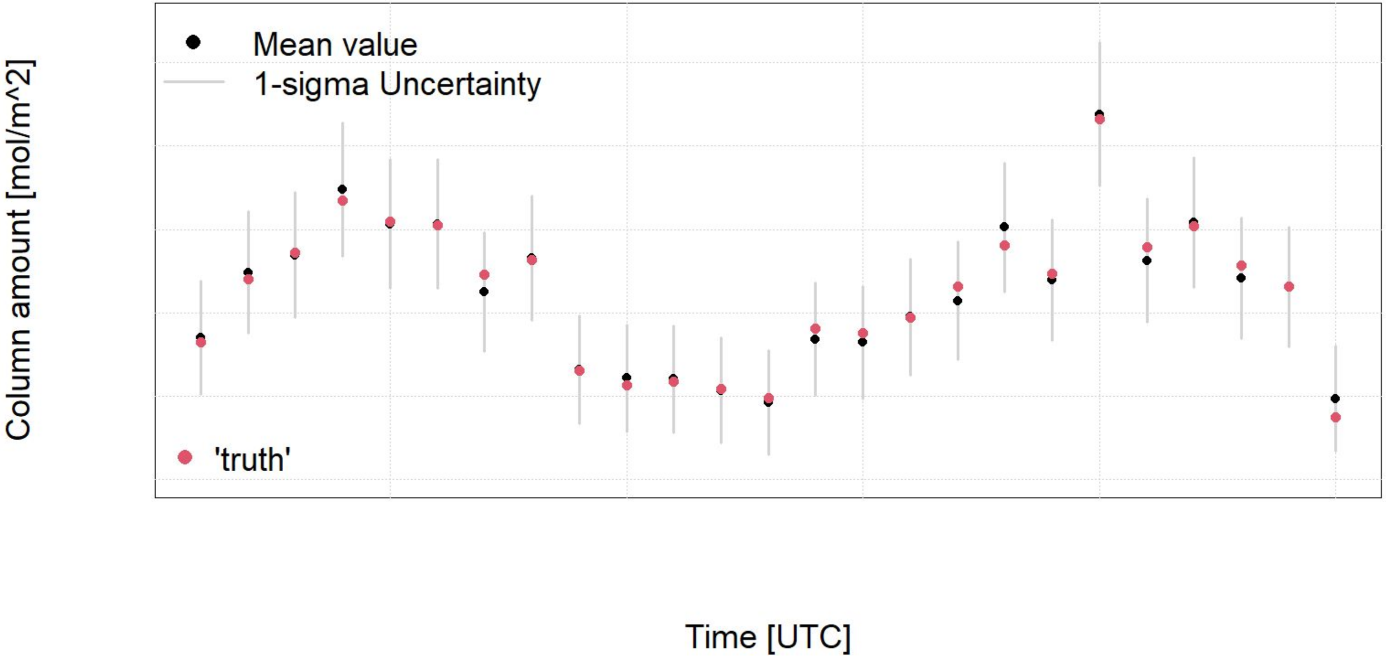
IDEAS-QA4EO Cal/Val Workshop #5, Thessaloniki/Greece, 2024-06-11

What is a meaningful uncertainty?



Assuming an observation can be used as the “truth”: Is the uncertainty reporting for the column amount (e.g, direct sun total column NO2) representative?

What is a meaningful uncertainty?

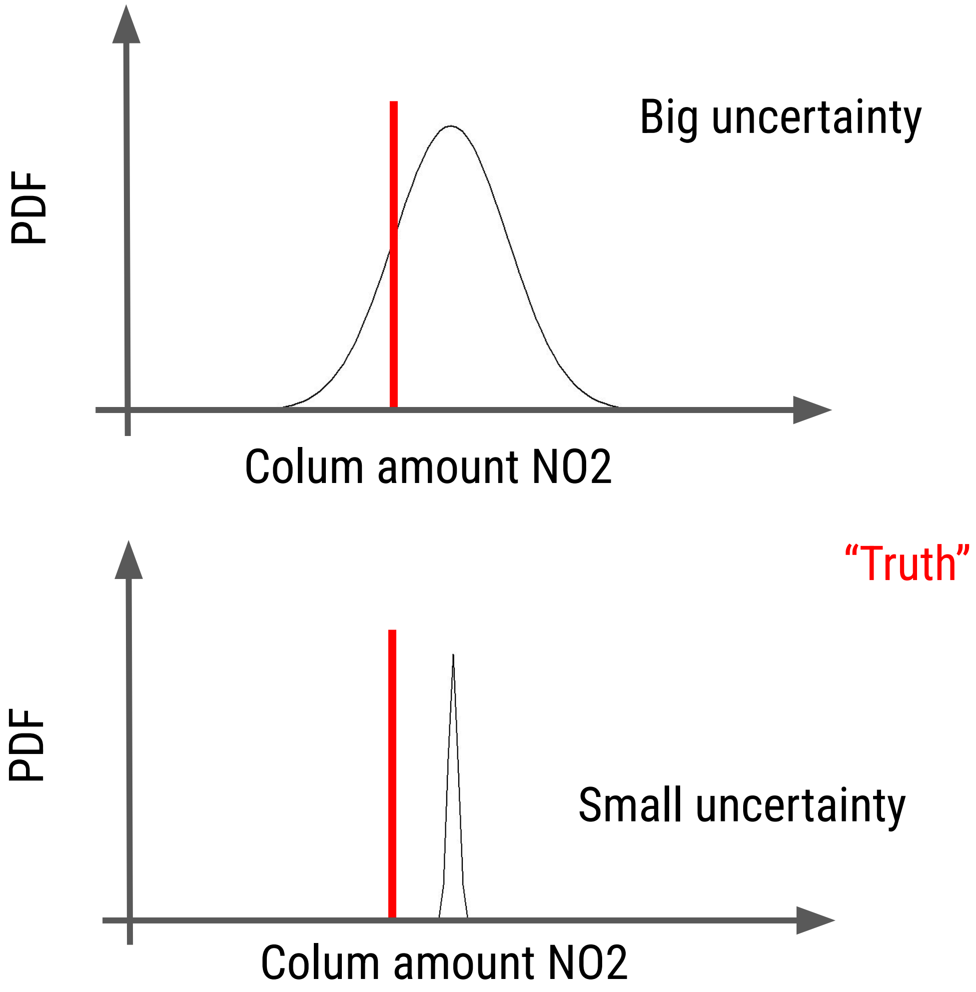


Assuming an observation can be used as the “truth”: Is the uncertainty reporting for the column amount (e.g, direct sun total column NO2) representative? Should it be bigger?

Goal

- **Provide meaningful uncertainty (Probability Density Function PDF)** reporting to users: meet statistical consistency in the reporting with respect having small uncertainty intervals (done in previous QA4EO projects mostly).

- Build a **framework for evaluating uncertainty** components and test its applicability. (current work package)

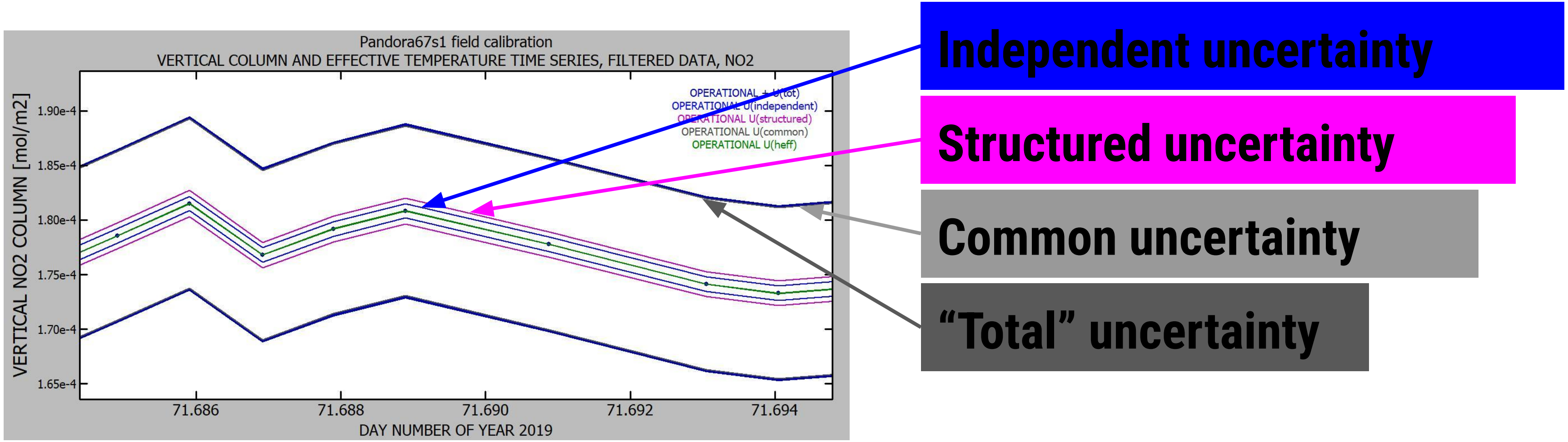


Which ‘truth’ to be used for uncertainty validation?

- Sometimes no verification source available -> we don’t know the truth.
- Possibility to use co-located Pandoras, independently calibrated to obtain “smooth approximation of baseline truth” (SABAT), which is described by a shared daily effect:
    - Get shared daily variations via regression approach
    - Remove systematic differences

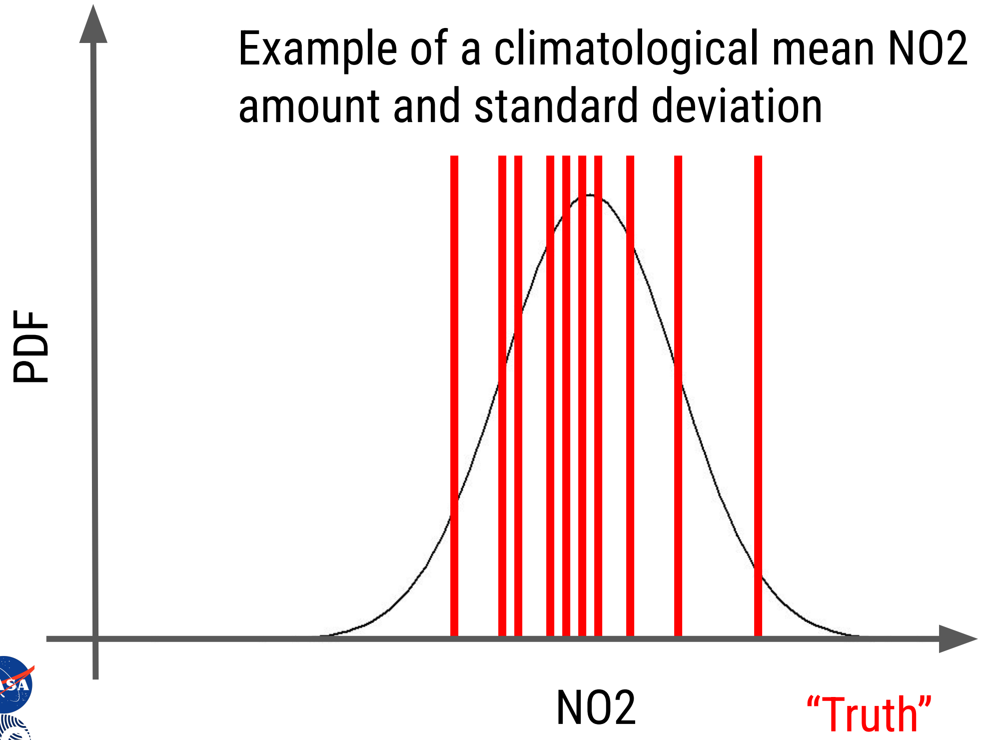
PGN Data Uncertainty Components

PGN uses uncertainty nomenclature based on core principles of metrological traceability (in collaboration with UK National Physics Laboratory, NPL).

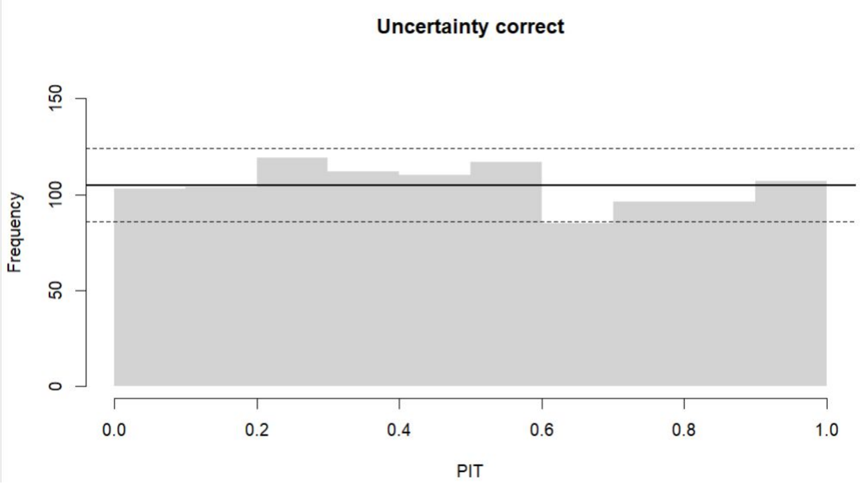


Validation Method: Probability Integral Transform (PIT)

Evaluate the reported Probability Density Function (PDF), e.g. Gaussian assumption, in terms of the observation (truth).

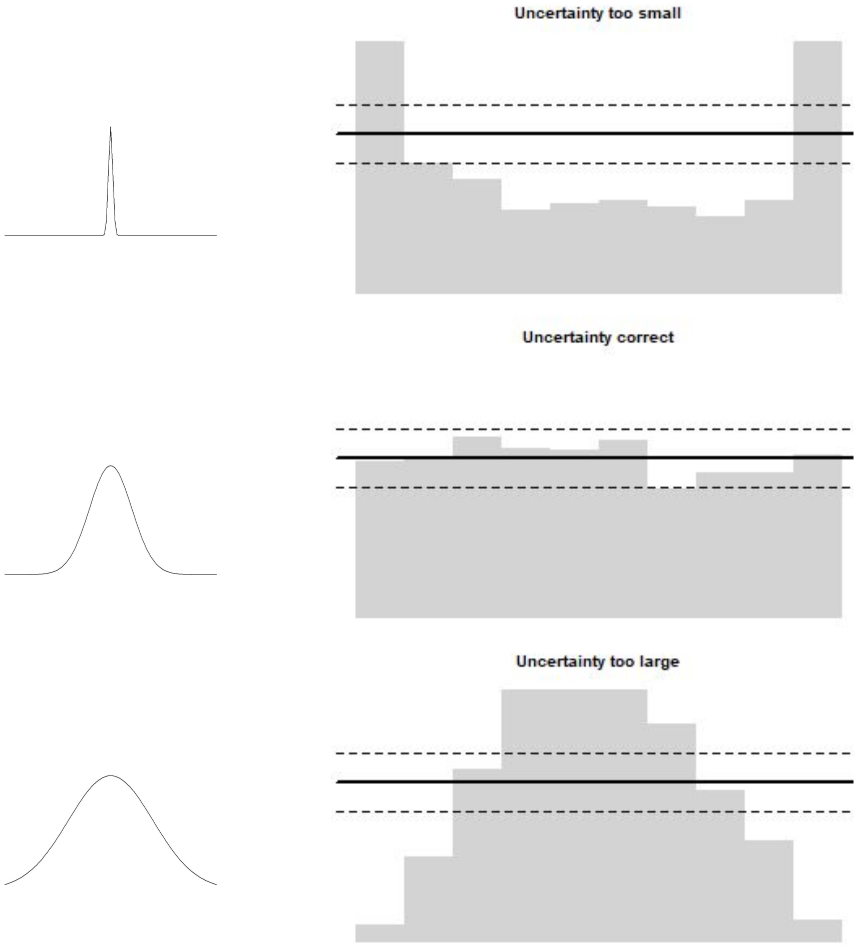


Quantile-binning the PDF from -infinity to infinity and count at which quantiles the observation falls into.



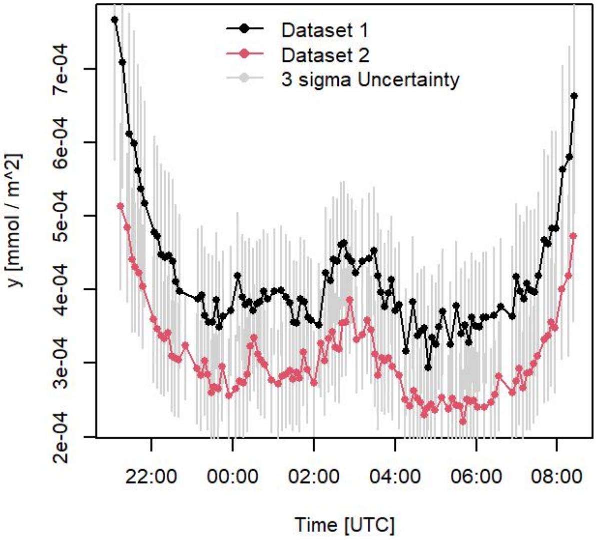
Uncertainty Closure Strategy

- Obtain systematic difference to evaluate “Common Uncertainty”
- After difference is removed, quantify the reported uncertainty components by remaining variations around baseline amount with PIT

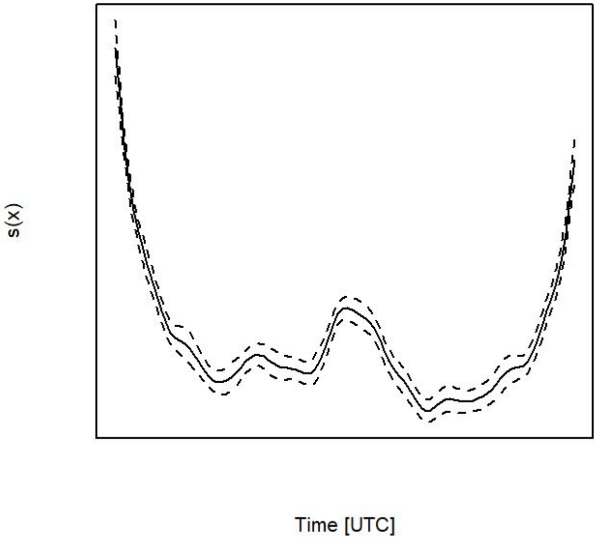


Uncertainty Closure Demonstration

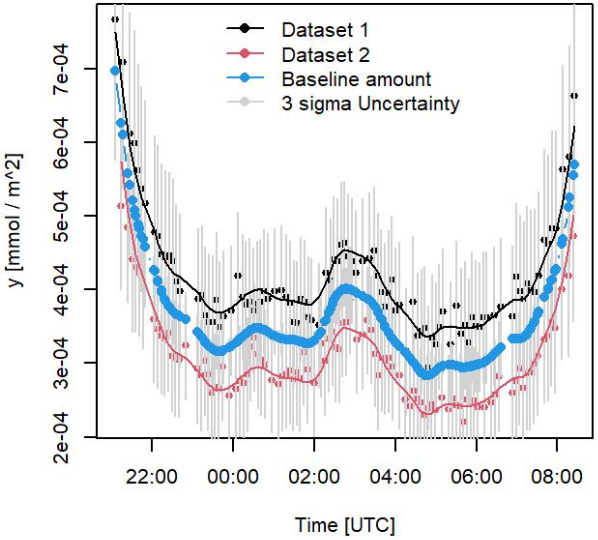
Slant columns



Shared Daily effect

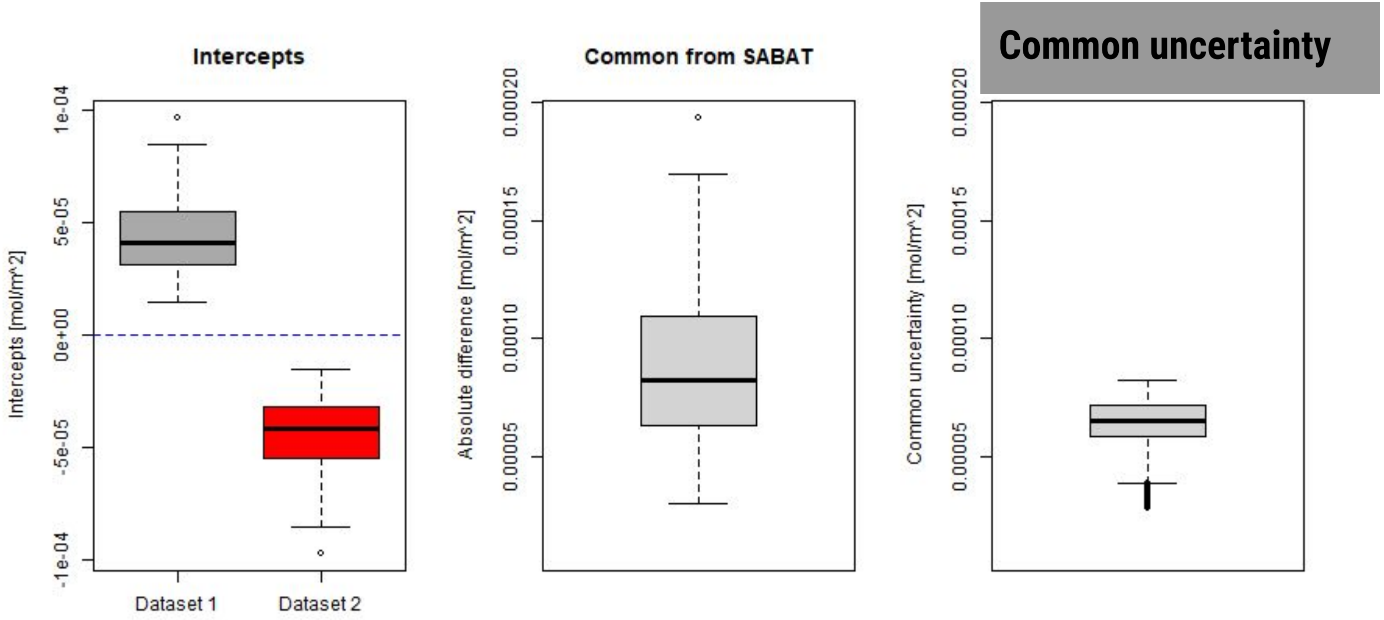


Bias reduced baseline amounts



Obtain the systematic bias to the baseline by the regression intercept, and repeat this for each day.

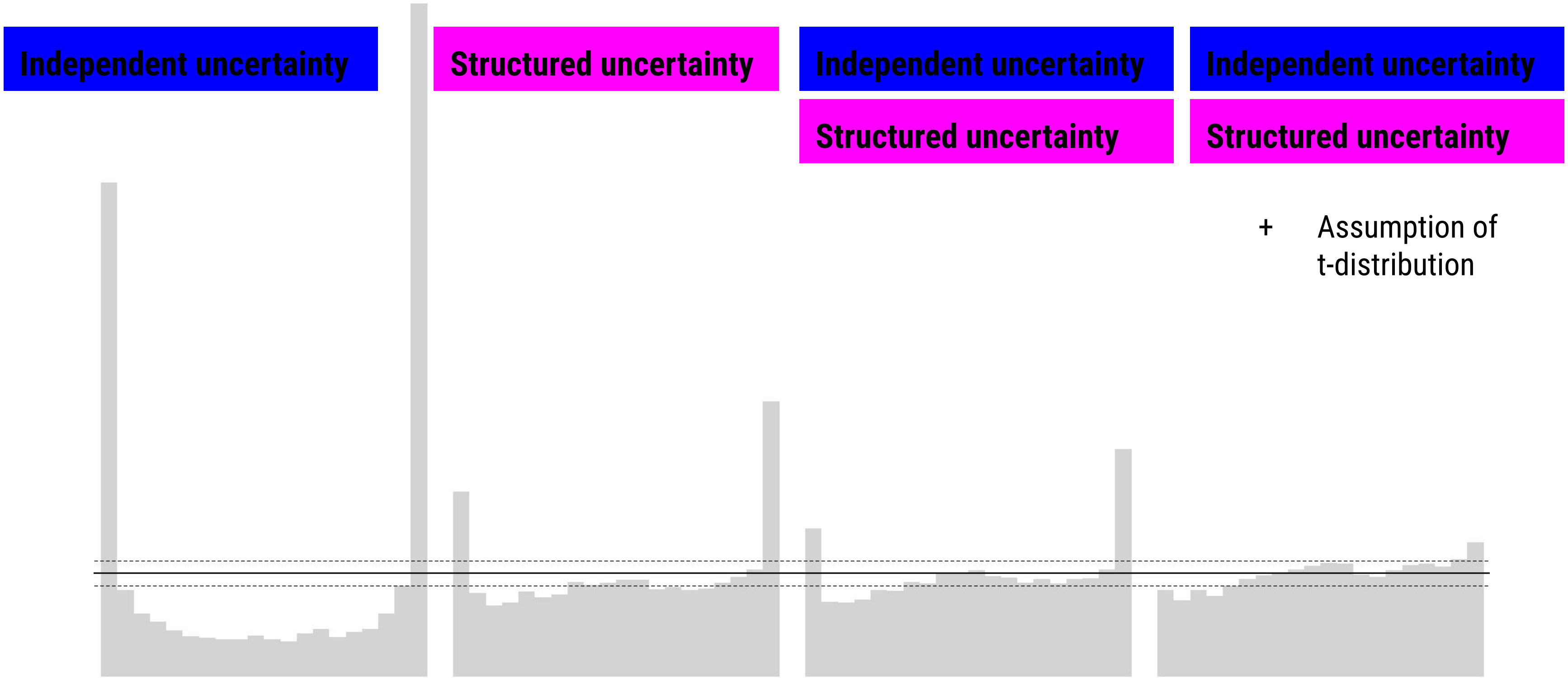
Common Uncertainty



Empirical common uncertainty (from SABAT), described by the maximum range of the intercepts,is comparable to the reported common uncertainty component reported by the Blick Software Suite (BSS)

Uncertainty Closure and Distributional Assumption

Evaluate uncertainty components on the residuals of intercept-corrected datasets



Conclusions

Key facts:

- **Uncertainty closure** framework demonstrated
- More datasets easily integratable to increase statistics.
- **Extendable** to compare Pandora with **external datasets** (e.g. Brewer) or uncertainties of other instrumentation (e.g., EUBREWNET, AERONET)

Open questions:

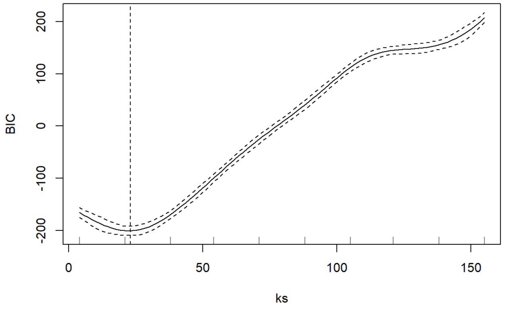
- wrong assumption of Gaussian residuals due to “extremes”. It is not said, that the underlying process is purely Gaussian, but can indeed be t-distributed, or even logistic to account for heavier distributional tails.

Next step:

- Apply uncertainty validation to PGN direct sun NO2 product.

Which 'truth' to be used for uncertainty validation?

- No verification source available -> we don't know the truth.
- Possibility to use co-located Pandoras, independently calibrated to obtain "smooth approximation of baseline truth" (SABAT)
- Remove systematic biases
- Get shared daily variations



Finding the optimum fitting parameters of how much to follow the variations