LuftBlick TN 2025007

Uncertainty Validation of PGN Direct Sun NO2

Project **QA4EO-2 WP2330**

Version 1 - October 31, 2025

Prepared by

Name	Institution
Manuel Gebetsberger	LuftBlick[1]
Martin Tiefengraber	LuftBlick[1]
Alexander Cede	LuftBlick[1]

[1] *LuftBlick Earth Observation Technologies, Innsbruck, Austria*

Contents

1 Document change record

Version	Date	Section	Notes/Changes
1	October 31, 2025	All	First version setup

2 Introduction

2.1 Acronyms and Abbreviations

AMF	Air mass factor
BIC	Bayesian information criterion
BSS	Blick Software Suite
CDF	Cumulative distribution function
CRPS	Continuous ranked probability score
CRPSS	Continuous ranked probability skill score
GAM	Generalized additive model
MAE	Mean absolute error
NO_2	Nitrogen dioxide
PGN	Pandonia Global Network
PIT	Probability integral transform
QA4EO	Quality Assurance for Earth Observation
SABAT	Smooth approximation of a baseline truth
SZA	Solar zenith angle
UVF	Uncertainty validation framework
WP	Work package
wrms	Normalized rms of fitting residuals weighted with independent uncertainty

2.2 Summary

This document is the final report of WP 2330 of the ESA project QA4EO-2, performing an uncertainty validation of PNG's direct sun total column NO_2 product (retrieval version rnvs3p1-8).

Reliable uncertainty quantification is essential for the scientific integrity of atmospheric trace gas retrievals. This work package presents an Uncertainty Validation Framework (UVF) developed within the QA4EO project to assess the uncertainty reporting of Pandora direct sun total column NO_2 products. The framework is based on the Smooth Approximation of a Baseline Truth (SABAT) approach, which employs a Generalized Additive Model (GAM) to separate shared atmospheric variability from instrument-specific systematic offsets. By estimating a smooth diurnal baseline common to all co-located Pandora instruments, and individual intercepts, the method enables the identification of systematic biases and the evaluation of reported uncertainty components.

Results from multiple sites (Figure: Ratio of optimized versus reported combined uncertainty) demonstrate that the reported combined uncertainty tends to be underestimated in polluted environments (e.g., Seoul-SNU) and slightly overestimated in remote locations (e.g., Izaña). The analysis reveals that the systematic component (basically calibration uncertainty) is the dominant contributor to combined uncertainty and primarily responsible to describe the observed mismatch between instruments. The UVF thus provides a statistically consistent, data-driven means to validate and refine Pandora uncertainty reporting, enhancing the reliability of PGN data products and offering a scalable validation framework applicable to other tracegases and sensor systems following the same requirements.

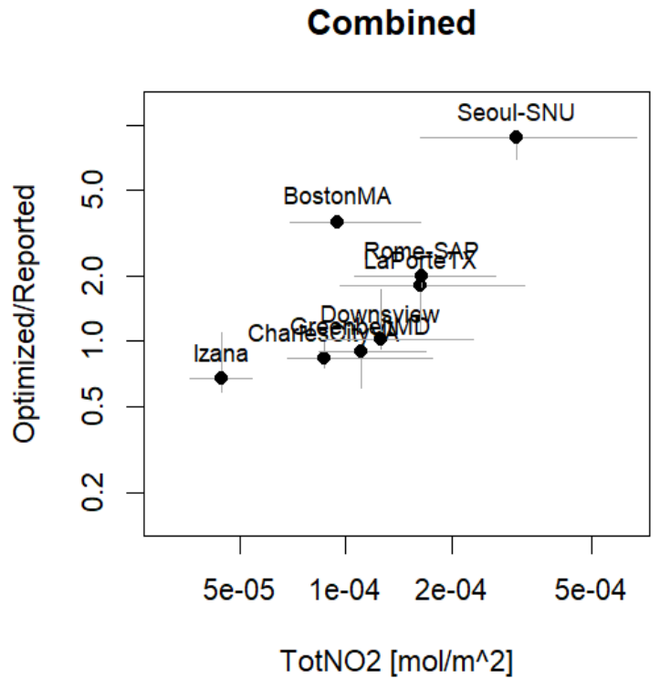

Figure: Ratio of optimized versus reported combined uncertainty as a function of median NO_2

3 Datasets

3.1 Requirements

The basic requirement to perform uncertainty validation is having co-located measurements from at least two instruments. Due to the developed SABAT approach, measurements do not have to be taken at exactly the same time, but at least at the same location in order to derive a baseline truth. The more instruments are part of the validation, the more realistic the baseline amount can be estimated. The second requirement is a standardized uncertainty reporting (see 4.1 Uncertainty components) nomenclature that is part of the Blick Software Suite (BSS) since processor p1-8, and based on the collaborative work with the National Physical Laboratory.

Due to non-existence of co-located direct sun measurements with proper uncertainty reporting next to PGN instruments, the dataset collection and study of this WP is limited to co-located Pandoras only.

3.2 Selected locations and datasets

Location (PGN Name)	Instruments	Start	End
BostonMA	Pandora153s1 Pandora155s1	2020-04-23	2021-04-26
CharlesCityVA	Pandora31s1 Pandora58s1	2024-05-24	2025-01-29
Downsview	Pandora103s1 Pandora104s1	2018-06-27	2024-06-18
GreenbeltMD	Pandora2s1 Pandroa30s1	2022-11-09	2025-07-27
Izana	Pandora101s1 Pandora121s1 Pandora209s1	2023-09-23	2025-07-25
LaPorteTX	Pandora58s1 Pandora63s1	2021-10-07	2022-01-23
Rome-SAP	Pandora117s1 Pandora138s1	2020-07-22	2020-09-30
Seoul-SNU	Pandora149s1 Pandora163s1	2020-11-03	2021-10-26

Table: Locations selected for uncertainty validation

The following figures provide an overview of the location-specific time series analyzed in this study. The data are based on retrieval version rvns3p1-8, as detailed in the accompanying ReadME document. To ensure a robust estimation of the baseline amounts, only high-quality data were included in the analysis.
Each figure displays daily median values as dots, while the vertical bars represent the 5th and 95th percentiles, illustrating the variability of the measurements within each day for the investigated instruments at the respective site.

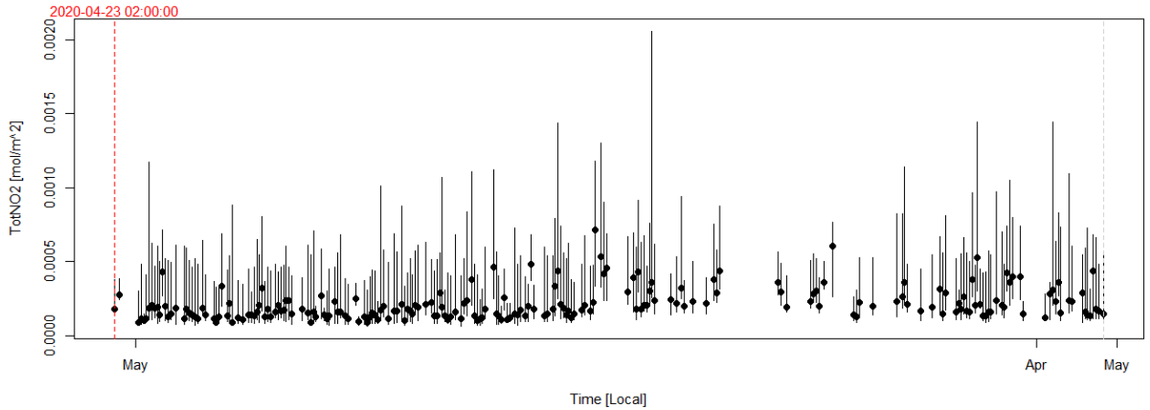

Figure: Overview BostonMA

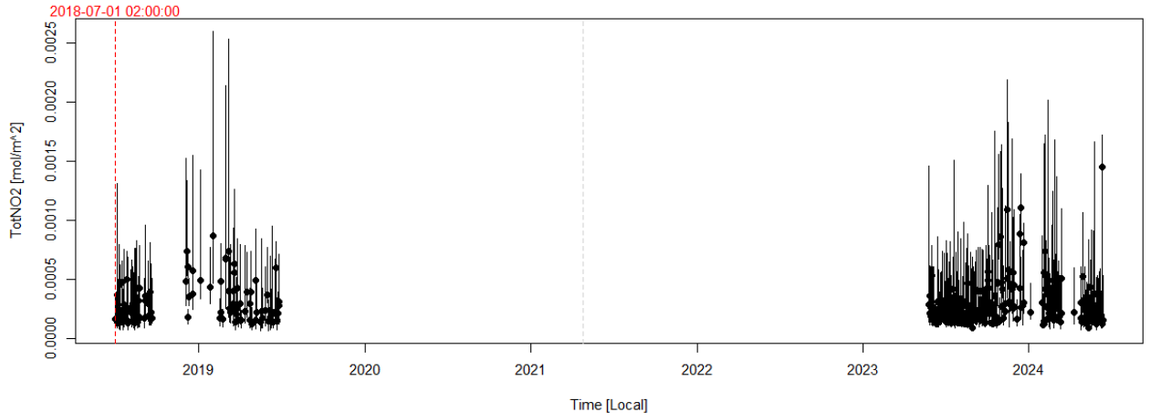

Figure: Overview Downsview

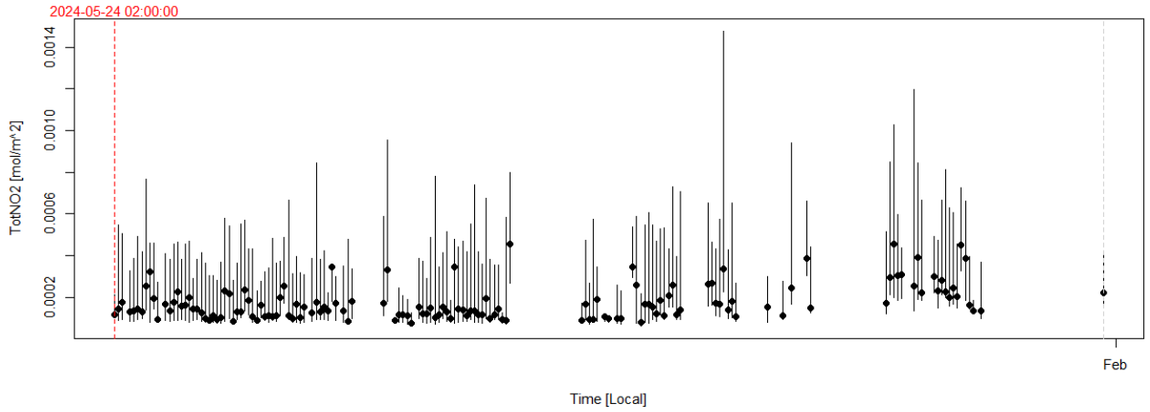

Figure: Overview CharlesCityVA

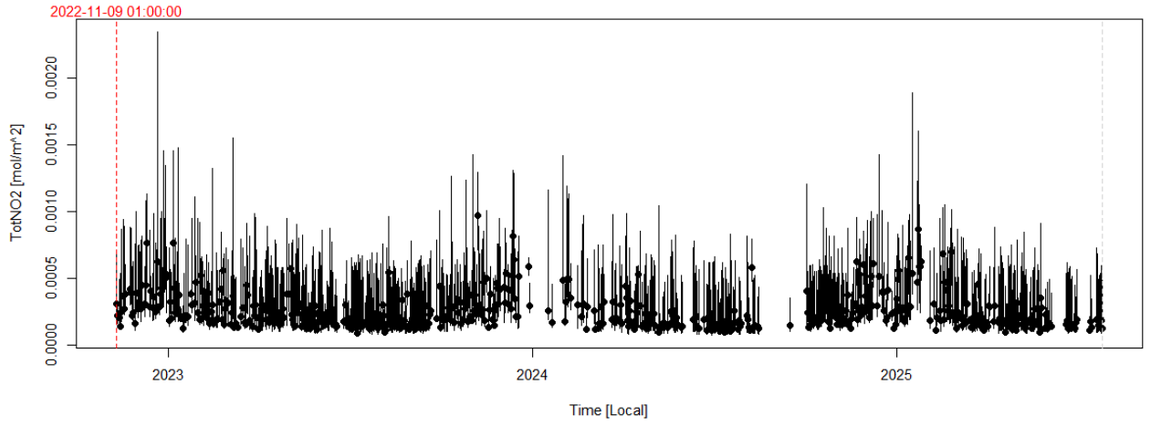

Figure: Overview GreenbeltMD

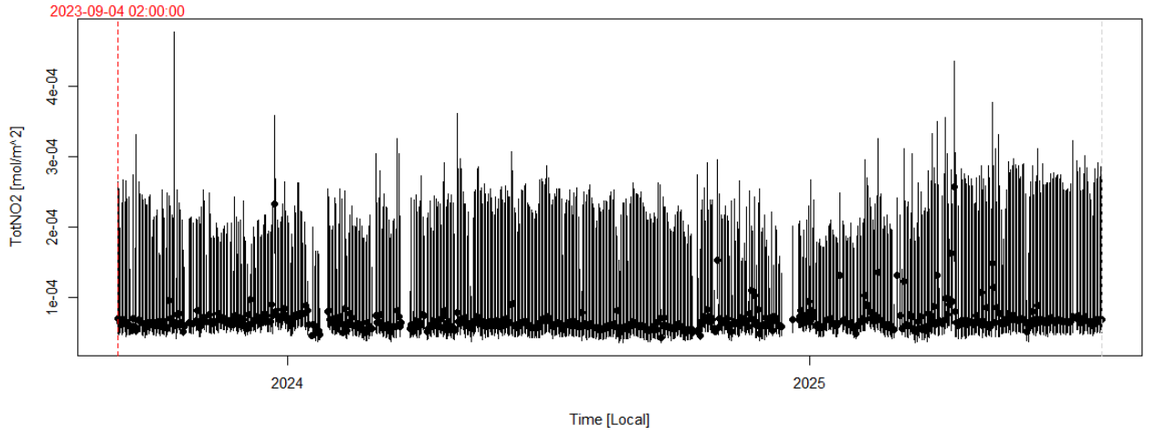

Figure: Overview Izana

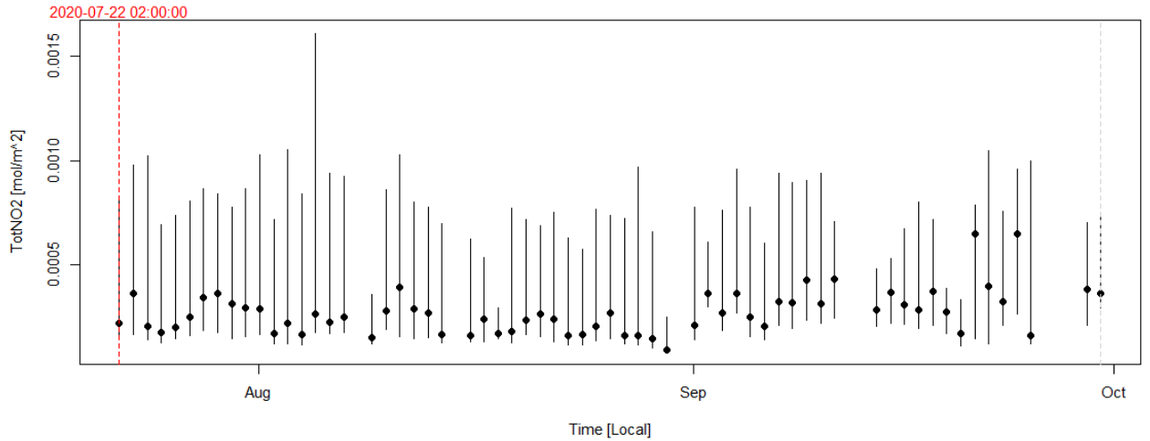

Figure: Overview Rome-SAP

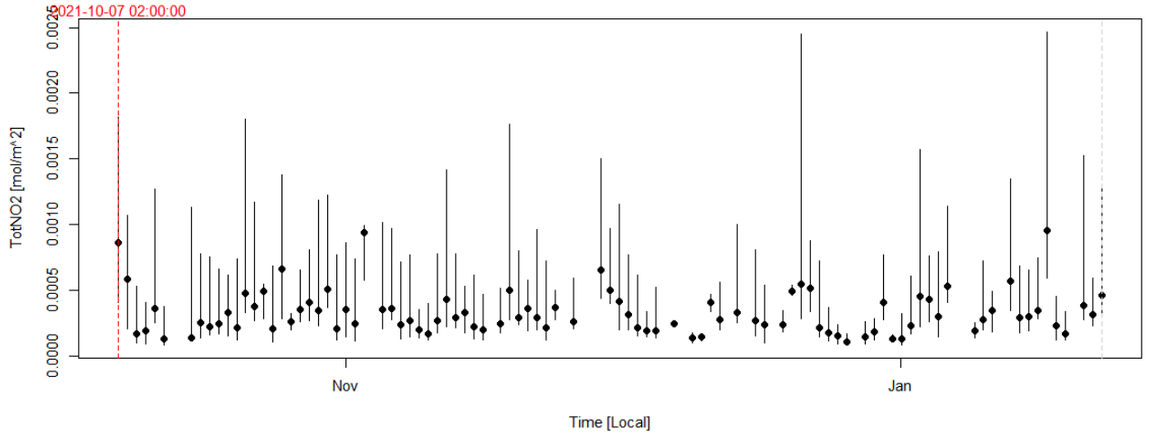

Figure: Overview LaPorteTX

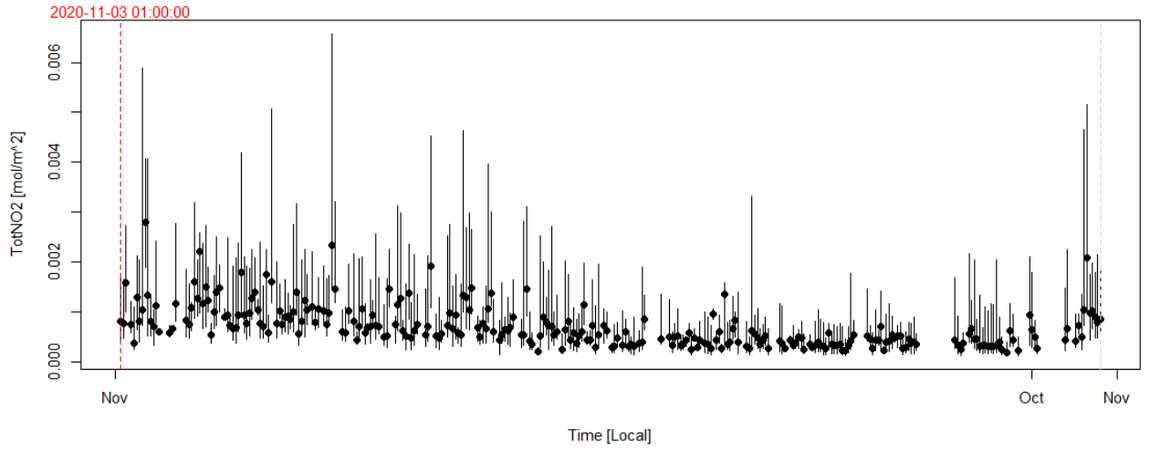

Figure: Overview Seoul-SNU

4 Methodology

4.1 Uncertainty components

The Blick Software Suite of the Pandonia Global Network (PGN), as used within the uncertainty validation framework, provides a standardized approach for reporting retrieval uncertainties. The methodology follows the definitions outlined in the PGN Data Products Readme.

In this framework, the combined reported uncertainty associated with each retrieved data point is composed of three distinct components, which differ in their temporal correlation characteristics:

1. **Random/ Independent Uncertainty** (U_{random})

   Represents random, uncorrelated noise contributions. These uncertainties have a correlation length in time of zero, meaning they vary independently between individual measurements. An example is the photon noise propagated into a retrieved total column value, which is uncorrelated between different measurement times.

2. **Systematic / Common Uncertainty** ($U_{systematic}$)

   Corresponds to systematic errors that are fully correlated in time (correlation length is infinite). Such errors affect all retrievals using the same reference or calibration data identically. For instance, an offset in the assumed slant column of the reference spectrum propagates as a common uncertainty to all subsequently retrieved columns.

3. **Mixed / Structured Uncertainty** (U_{mixed})

   Describes partially correlated uncertainties, where the correlation length is finite. These typically arise from slowly varying model or input parameter mismatches that affect a series of temporally close measurements in a similar way but are not correlated over longer time spans. A typical example is an error due to a mismatch between the assumed and the true effective temperature of a trace gas, which introduces a consistent bias over short timescales but not across different days.

The combined uncertainty (U) reported for each measurement combines these components in quadrature according to:

$$U = \sqrt{U_{random}^{2} + U_{systematic}^{2} + U_{mixed}^{2}}$$

This reporting scheme ensures that both random and systematic effects are transparently represented and enables a consistent interpretation of uncertainty across the PGN data products.

4.2 Smooth approximation of a baseline truth (SABAT)

Uncertainty validation can only be done by comparing to a true value. Such a truth is hardly given for atmospheric trace gases. Therefore, one approach is to create something close to this, called a smooth approximation of a baseline truth. By using the “wisdom of the crowd” concept, we make use of co-located instruments which are supposed to deliver the same gas amounts. The SABAT approach was originally introduced for inter-comparison of co-located Pandora instruments (Tiefengraber et. al, 2022), but can be used towards uncertainty validation within the Uncertainty validation framework.

In the SABAT approach, the true atmospheric signal is represented by a *shared smooth function* over time, which is common to all instruments under comparison. This shared function captures the diurnal variability of the observed quantity (slant column amounts) and is modeled using a Generalized Additive Model (Hastie T. and R. Tibshirani, 1986)

Instrument-specific calibration offsets are modeled as intercepts, allowing for a flexible description of systematic differences between instruments. Mathematically, the expectation value of the measured slant column y_i is expressed by a Gaussian distribution:

$$y_i \sim N(\mu_i, \sigma), with\ \mu_i = \beta_0 + \beta_i + s(x_i)$$

where $s(x_i)$ denotes the smooth diurnal effect estimated from all instruments *i* jointly, and β_i are the instrument-specific intercepts reflecting systematic biases. Hereby, the first dataset serves as the reference with intercept β_0. The baseline truth is then defined as the shared smooth term plus the mean intercept across all instruments:

$$\mu_{baseline}(x) = s(x) + \overline{\beta}, where\ \overline{\beta} = \frac{1}{m}\sum_{i=1}^{m}\beta_i$$

This *baseline* represents the best statistical approximation of the true atmospheric signal without assuming any instrument as the absolute reference. The smoothness of $s(x)$is optimized using the Bayesian Information Criterion (BIC) to avoid overfitting while retaining relevant diurnal structures. The BIC penalizes model complexity and is defined as:

$$BIC = -2 \cdot log(L) + log(n) \cdot npar$$

where *L* is the likelihood, n the number of observations, and $npar$ the number of estimated parameters.

Figure: BIC optimization example illustrates a clear optimum for an example day at Seoul-SNU. The corresponding smooth diurnal effect $s(x)$ is presented by Figure: Smooth diurnal effect example.

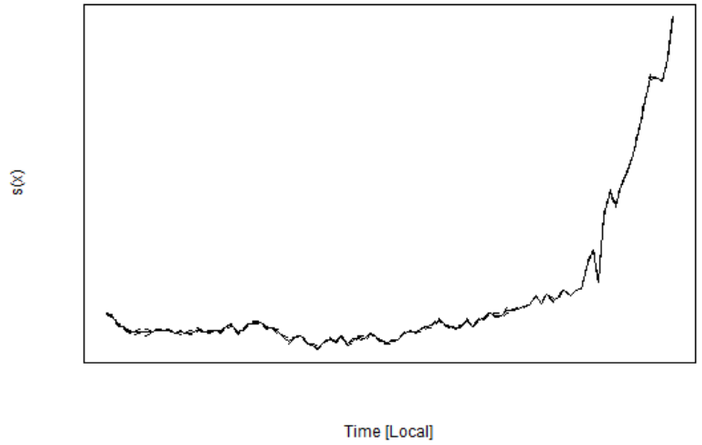

Figure: Smooth diurnal effect example

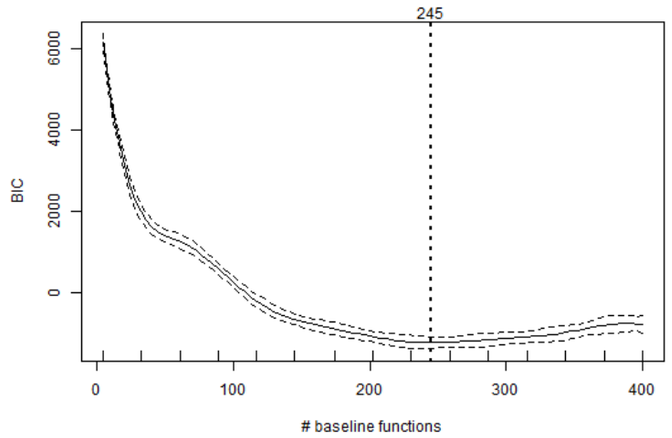

Figure: BIC optimization example with optimum shown by vertical line

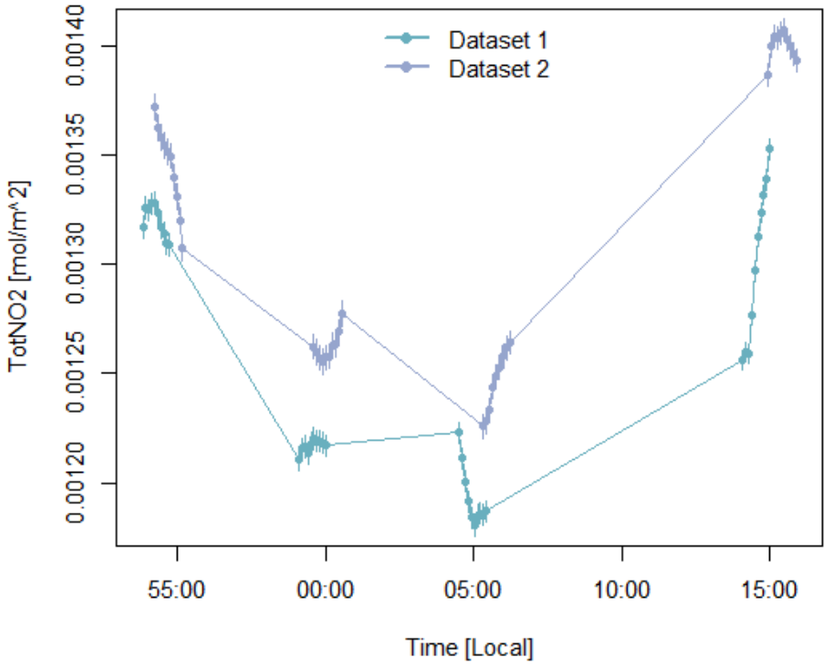

Figure: SABAT daily example at Seoul-SNU before bias correction

The zoomed-in period of Figure: SABAT daily example at Seoul-SNU before bias correction shows two datasets which measure roughly at the same time and with overlapping periods. The uncertainty bars present the combined uncertainty as given in Section 4.1 Uncertainty components. By using the SABAT results, the bias-corrected datasets can be brought into agreement, as illustrated in Figure: SABAT daily example at Seoul-SNU after bias correction. Since this systematic difference should be reflected by the systematic uncertainty component, only random and mixed components are left in the figure.

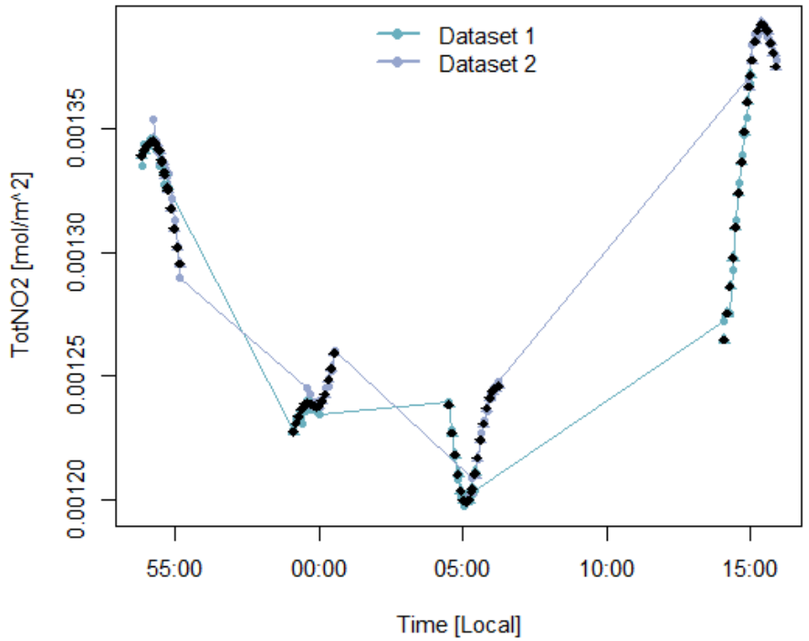

Figure: SABAT daily example at Seoul-SNU after bias correction with the smooth daily effect / baseline amount in black. Combined uncertainty consists of random and mixed components only.

4.3 Metrics to evaluate the uncertainty reporting

4.3.1 Probability integral transform (PIT)

The Probability Integral Transform (PIT) is a diagnostic tool, originally used to assess the calibration of probabilistic weather forecasts (Anderson J., 1996, Hamill T., 2001). It is based on the principle that, for a continuous random variable Y with cumulative distribution function (CDF) $F_Y(y)$, the

transformed variable $Z=F_Y(Y)$ is uniformly distributed on the interval [0,1], provided that F_Y correctly represents the true underlying distribution of Y. In practice, when a probabilistic model provides a full distribution $F_t(y)$ for an observed value y_t, the PIT value is computed as $PIT_t=F_t(y_t)$.

By aggregating PIT values over a series of distributions, one can evaluate the overall reliability (or calibration) of the uncertainty reporting of an instrument. The typical expected visual expressions of the PIT are shown in Figure: PIT characteristics.

The shape of the PIT histogram can be interpreted as follows, independent of the chosen binning:

- If the uncertainty reporting is statistically consistent, the PIT histogram will be approximately uniform.
- A U-shaped histogram indicates underdispersion (the predictive distributions are too narrow).
- An inverted U-shape suggests overdispersion (the predictive distributions are too wide).
- Skewed histograms reveal systematic bias.

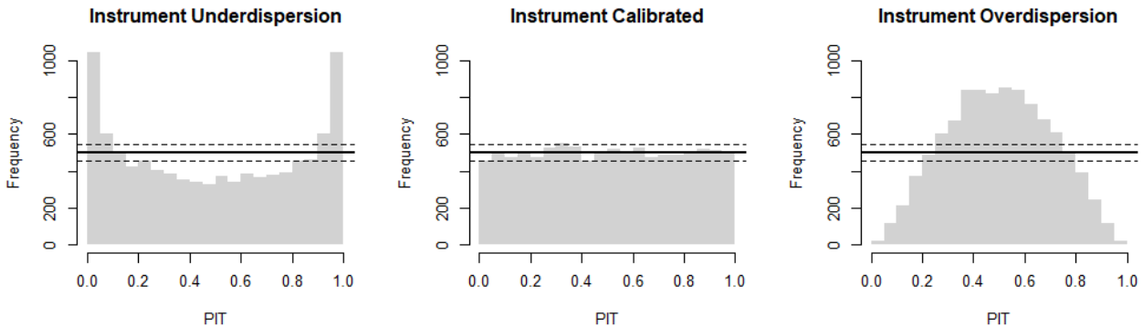

Figure: PIT characteristics

The PIT histograms include consistency bars, which indicate the expected statistical range for a perfectly uniform distribution given the sample size and binning. Values falling outside these bounds reflect statistically significant deviations from calibration (Gebetsberger et. al 2018).

4.3.2 Continuous ranked probability score (CRPS) and skill score (CRPSS)

While the PIT is just a visual representation of how well a reported distribution matches the true realization, it cannot quantitatively evaluate the uncertainty reporting.

The Continuous Ranked Probability Score (CRPS) is a proper scoring rule used to evaluate the accuracy of probabilistic forecasts for continuous variables, and therefore able to quantify the reporting by a single number. It measures the distance between the predicted cumulative distribution function (CDF) F(y) and the observed outcome y_{obs}, thus providing a single summary metric that reflects both calibration and sharpness of the forecast distribution (Gneiting et. al 2007). Hereby, sharpness refers to the width of the distribution. The sharper a distribution, the smaller the standard deviation. Mathematically, the CRPS is defined as

$$CRPS(F, y_{obs}) = \int_{-\infty}^{\infty} [F(y) - 1\{y \geq y_{obs}\}]^2 dy,$$

where $1\{y \geq y_{obs}\}$ is an indicator function that equals 1 if $y \geq y_{obs}$ and 0 otherwise. The CRPS can be interpreted as the integrated squared difference between the forecast CDF and the empirical step function at the observation y_{obs}. A smaller CRPS indicates a better probabilistic forecast, with zero being the ideal score when the forecast perfectly matches the observation. For deterministic forecasts (where F(y) is a step function centered at a single value), the CRPS reduces to the mean absolute error (MAE), making it a natural extension of classical point forecast evaluation to the probabilistic domain.

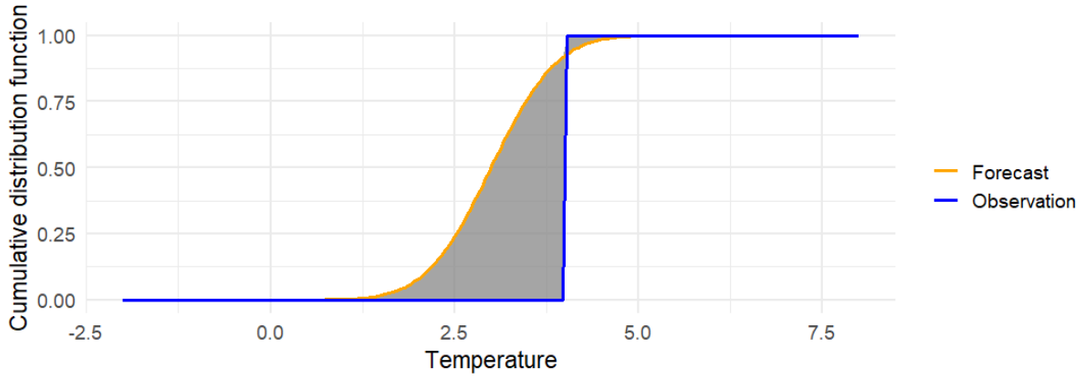

Figure: CRPS example (grey area) of a temperature forecast distribution (orange) comparing the to the observed truth (blue).

Figure: CRPS example illustrates this concept by a probabilistic temperature forecast that is compared to the true observation. The CRPS for this individual forecast can be interpreted as the grey area. Ideally, the probabilistic forecast would provide the expectation value of the observation with a standard deviation going towards zero, in order to have the smallest area, or CRPS respectively.

The CRPS is widely used in meteorology, hydrology, and ensemble forecasting, as it provides a holistic assessment of probabilistic forecast performance without requiring arbitrary threshold selection. In the context of this study, the CRPS is used as an optimization criteria to adjust the reported uncertainty towards a reporting with minimum CRPS .

To quantitatively assess the performance of the optimized uncertainty components, the Continuous Ranked Probability Skill Score (CRPSS) can be applied. The CRPSS provides a relative measure of predictive skill by comparing the mean Continuous Ranked Probability Score (CRPS) of the optimized against an original reported uncertainty. It is defined as

$$CRPSS = 1 - \frac{\overline{CRPS_{optimized}}}{\overline{CRPS_{reported}}},$$

A CRPSS value close to 1 indicates that the optimized uncertainty description reproduces the observed distribution well and outperforms the reference, whereas values near 0 or negative values imply little to no improvement.

4.3.3 CRPS optimization simulations

To evaluate the behavior of the Continuous Ranked Probability Score (CRPS) as a function of reported uncertainty, synthetic data were generated under controlled conditions where the true uncertainty exhibits a solar zenith angle (SZA) dependency. Each simulated observation y was drawn from a Gaussian distribution with a fixed mean μ and a standard deviation σ_{true} defined as

$$\sigma_{true} = 5e^{-5} + 1e^{-4} \cdot SZA^2,$$

where SZA is uniformly sampled between 20° and 80°. To mimic systematic misreporting, the initially reported uncertainties were scaled versions of the truth (σ_{true}), representing either underdispersive ($\sigma_{reported} = 0.3\,\sigma_{true}$) or overdispersive ($\sigma_{reported} = 1.5\,\sigma_{true}$) conditions.

The CRPS for each Gaussian forecast was computed using the analytical form

$$CRPS(y, \mu, \sigma) = \sigma[z(2\Phi(z) - 1) + 2\phi(z) - 1/\sqrt{\pi}],$$

with $z\frac{y-\mu}{\sigma}$, and Φ denoting the standard normal cumulative distribution and probability density functions, respectively.

To correct the misreported uncertainties, a parametric optimization was applied by minimizing the mean CRPS over all samples. The reported uncertainty $\sigma_{reported}$ was adjusted using a smooth scaling function $\alpha(SZA)$ represented as a natural spline basis:

$$\sigma_{scaled}(SZA) = \sigma_{reported} \cdot e^{B(SZA)\beta} + offset,$$

where $B(SZA)$ is the spline basis matrix and β are the fitted coefficients. The parameters β and the offset term was optimized using the BFGS algorithm to minimize the CRPS loss function.

Figure: Simulated underdispersive uncertainty reporting and Figure: Simulated overdispersive uncertainty reporting illustrate the results for the underdispersive and overdispersive cases, respectively. Each figure consists of:

1. Three PIT histograms showing the distributions for
   - the *truth* (using σ_{true}),
   - the *reported* uncertainty, and
   - the *optimized* uncertainty after CRPS minimization.
2. Uncertainty as a function of SZA, comparing the true, reported, and optimized values.

The optimized uncertainty curve reproduces the SZA-dependent pattern of the true uncertainty much more accurately than the initially reported one, demonstrating that CRPS minimization successfully detects and corrects SZA-dependent miscalibration in both scenarios. The PIT histogram based on the optimized uncertainties is nearly flat and closely matches the distribution obtained from the true uncertainties, indicating a proper uncertainty reporting.

In contrast, the PIT histograms derived from the underdispersive and overdispersive reported uncertainties exhibit pronounced U-shaped and inverted U-shaped patterns, respectively, consistent with under- and overestimation of variability.

Overall, these results confirm that CRPS-based optimization provides an effective framework for adjusting reported uncertainties in retrieval products. By explicitly incorporating SZA as a predictor, the method captures systematic dependencies and enhances the probabilistic reliability and consistency of the reported uncertainty estimates.

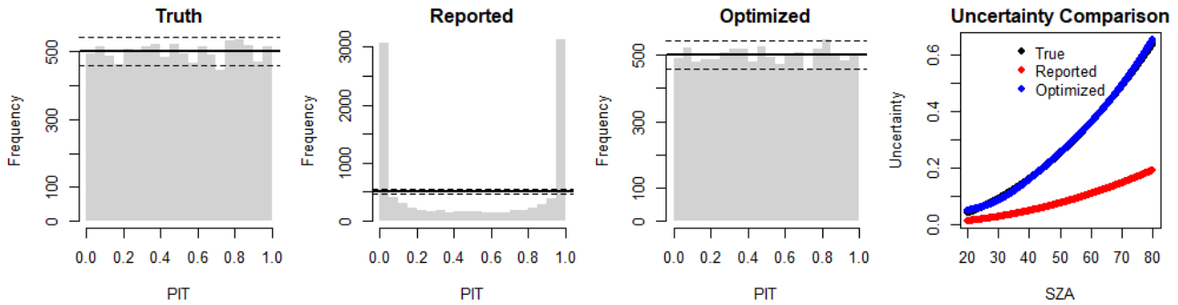

Figure: Simulated underdispersive uncertainty reporting with CPRS optimization

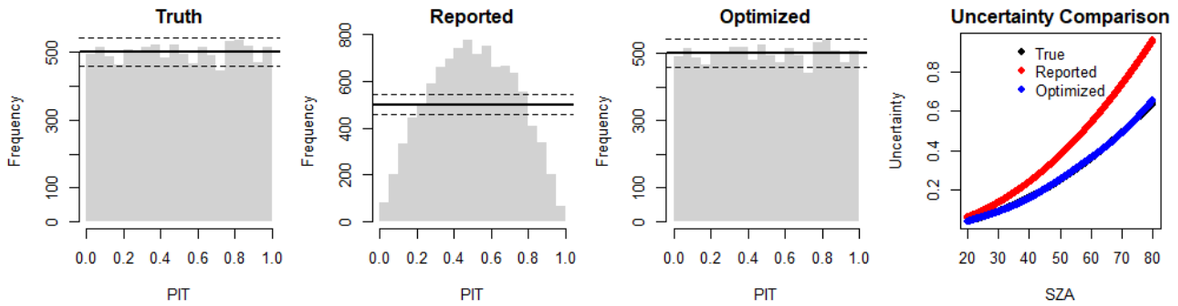

Figure: Simulated overdispersive uncertainty reporting with CPRS optimization

4.4 Uncertainty validation framework (UVF)

The Uncertainty Validation Framework (UVF) builds on the SABAT approach to systematically evaluate and optimize the reported uncertainty components of the Pandora direct sun total column products. The method separates and sequentially validates the systematic, random, and mixed uncertainty contributions by combining the squared uncertainty components.

In the first step, the systematic daily differ*ence* between the datasets is applied using the *intercept* obtained from the SABAT regression. Hereby, the maximum difference of all datasets being part of the regression is used as the observed daily bias. To quantify the quality of the reported systematic uncertainty, the ratio between the systematic daily difference and the reported systematic uncertainty can be used, as this ratio expresses how well the reported uncertainty envelope covers the observed variability:

$$R_{sys} = observed\ daily\ bias\ /\ U_{systematic}$$

Values of R_{sys} > 1 indicates that the reported systematic uncertainty is too low, as the uncertainty reporting does not reflect the true observed difference. Contrary, values of R_{sys} < 1 suggests an overestimation. This ratio can therefore be directly used to scale and optimize the systematic uncertainty component.

In the second step, after correcting for the systematic bias, the remaining variation of the datasets is investigated. Under ideal conditions, the remaining variations would just be random noise, and therefore be solely described by the random uncertainty reporting. However, as we expect structured patterns to occur, the remaining uncertainty is a combination of the random plus mixed components.

These remaining uncertainty components can be visually validated using the PIT, and can be quantified using the CRPS optimization. The CRPS minimization, performed as a function of the solar zenith angle (SZA), ensures that the uncertainty reporting captures SZA-dependent effects and yields a statistically consistent predictive distribution. Hereby we can also make use of the ratio approach

$$R_{mix} = optimized_{mixed} / U_{mixed}$$

Where R_{mix} reports how much the mixed uncertainty component (U_{mixed}) needs to be optimized, in the same way as for R_{sys}.

Finally, the new optimized combined total uncertainty is derived as the quadratic sum of the optimized components:

$$U_{optimized} = \sqrt{optimized_{systematic}^{2} + optimized_{mixed}^{2}}$$

5 Uncertainty validation and conclusions

The validation of the uncertainty reporting for Pandora direct sun total column NO_2 was conducted using the Uncertainty Validation Framework (UVF). The UVF approach allows decomposition and quantitative validation of the individual uncertainty components — namely, independent (random), mixed (structured), and common (systematic) uncertainties — through optimization against co-located Pandora observations.

Across multiple sites, the comparison between reported and optimized uncertainties revealed a systematic underestimation of the common (systematic) component for high-NO_2 environments such as Seoul-SNU. Comparing the observed systematic difference with the reported systematic uncertainty by the ratio R_{sys}, the systematic uncertainty would need to be adjusted by a factor of 7-9 for the two datasets, as shown in Figure: Systematic uncertainty evaluation for Seoul-SNU .

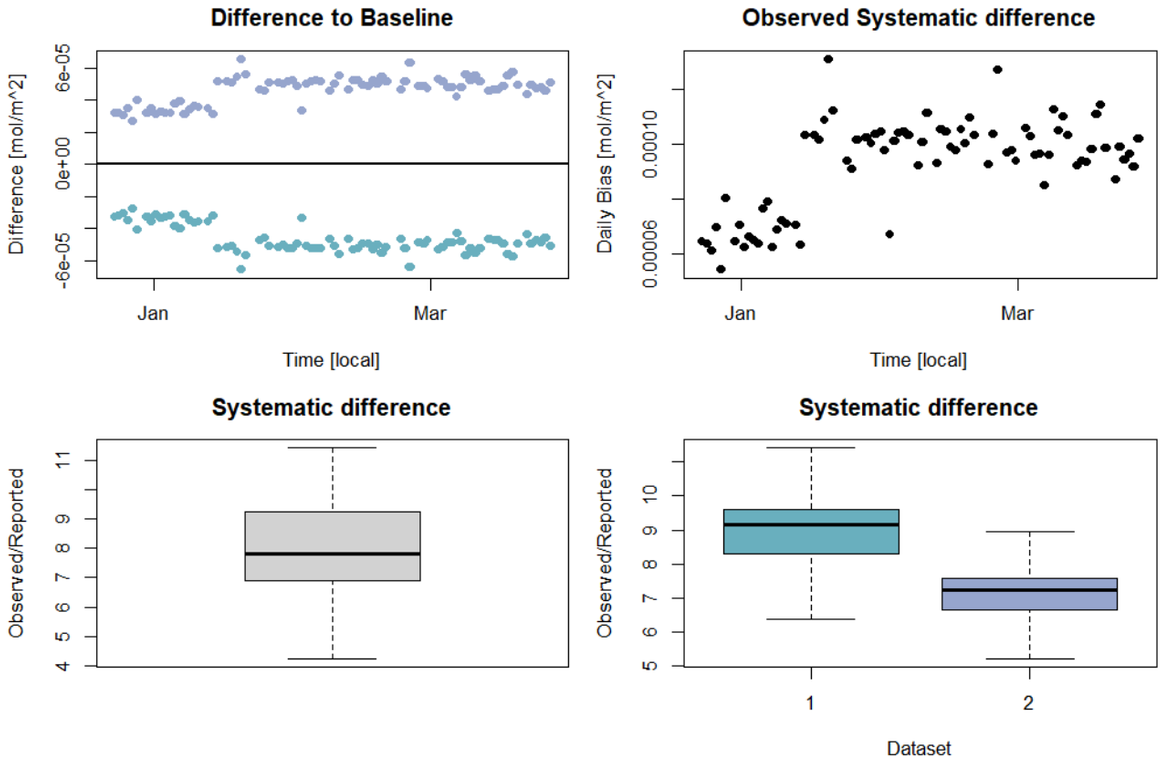

Figure: Systematic uncertainty evaluation for Seoul-SNU in the period January to March 2021 for the two datasets (green and blue)

After removing the systematic difference, the residual variations were still not fully covered by the remaining combined uncertainty, covering random and mixed uncertainty components. The corresponding PIT histogram showed a typical U-shape which indicates a too low uncertainty reporting (Figure: Mixed uncertainty for Seoul-SNU). This visual finding was also supported by the CRPS optimization, which suggests an increased mixed component.

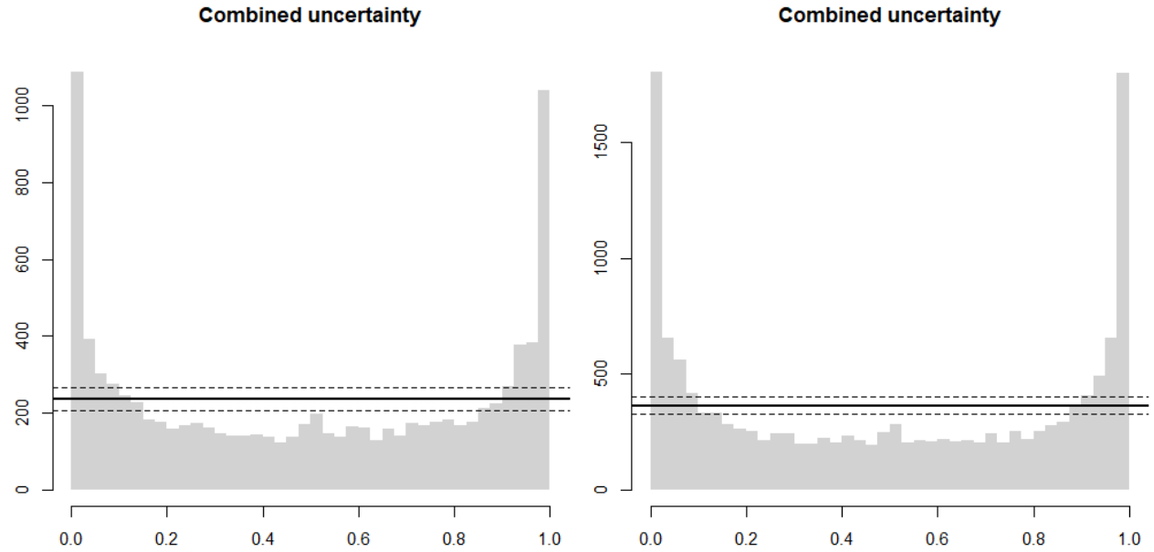

Figure: Mixed uncertainty for Seoul-SNU for the two datasets

The R_{mix} values larger than 1 suggests a minor increase of 1.6 to 2.2 on the median, which is SZA dependent, as shown in Figure: Optimized mixed uncertainty for Seoul-SNU by SZA bins. These two findings indicate that, in particular, calibration-related sources of uncertainty (e.g., slant column and effective gas temperature in the reference spectra) might currently be underestimated in the Pandora retrieval chain, where the uncertainty reporting does not reflect the observed differences between datasets.

However, Figure: Combined uncertainty for Seoul-SNU visualizes the original and optimized combined uncertainty reporting. Even this huge scaling factor of 7-9 for systematic, and up to 2.2 for mixed components, does not blow up the uncertainty artificially. Large scale and even small scale variations are still distinguishable, but the uncertainty reporting obtains much more reliability about most likely actual values of the gas amount.

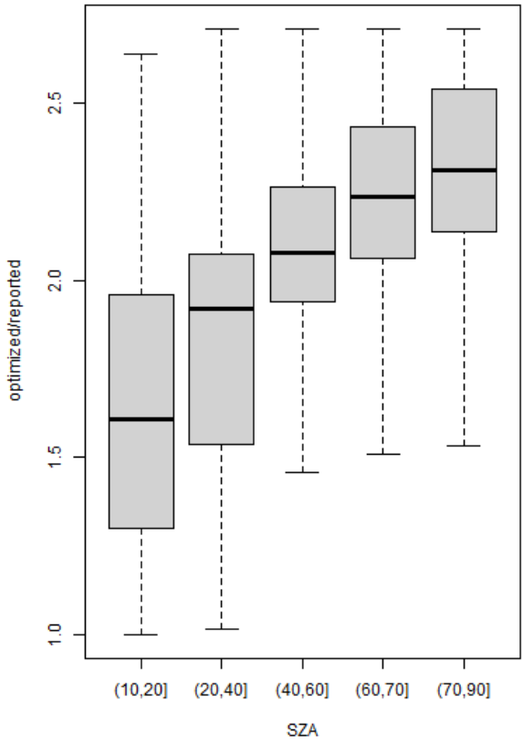

Figure: Optimized mixed uncertainty ratio for Seoul-SNU by SZA bins

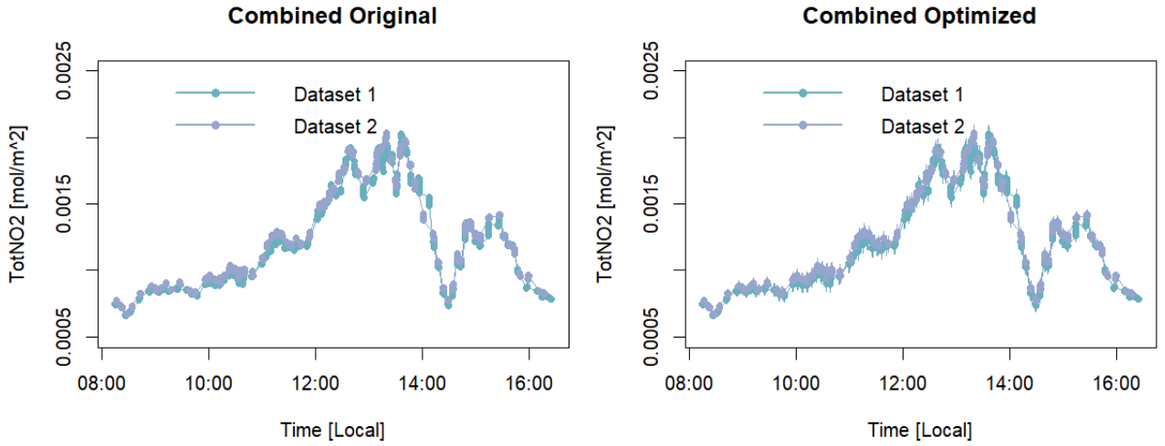

Figure: Combined uncertainty for Seoul-SNU 2020-11-12

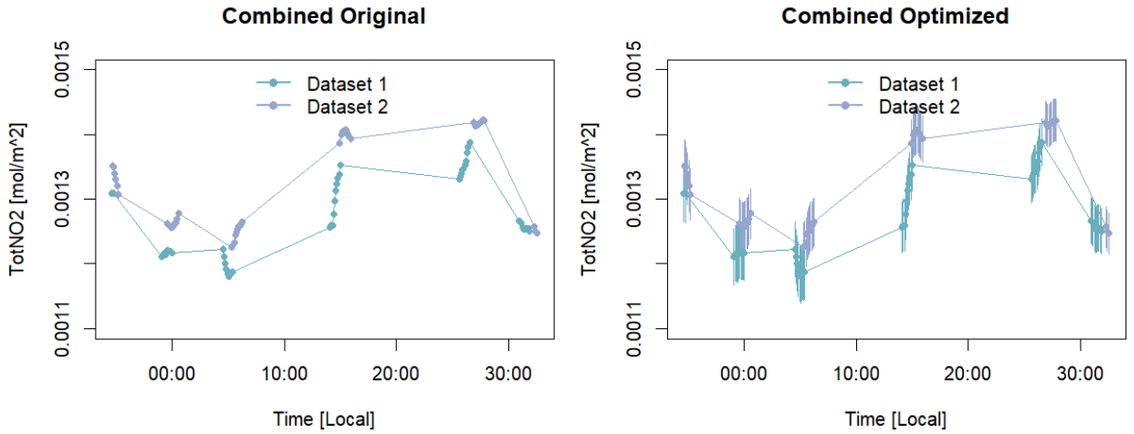

Figure: Combined uncertainty for Seoul-SNU zoomed in between 15:00 and 15:30 local time

This improved reliability can be visually assessed by comparing the PIT histograms before and after the optimization, as shown in Figure: PIT for originally and optimized uncertainty with a strongly reduced U-shape for both datasets. The W-shape, visible for both the originally reported and the optimized situation, can be explained by a slightly in-proper distributional assumption of the evaluated observation. E.g. if the observations are based on a heavier-tailed distribution such as the logistic or student-t, evaluating a PIT using the Gaussian assumption can lead to W-shaped patterns (Gebetsberger et. al 2018). This shape even remains after the optimization, but much less expressed on the tails of the distribution.

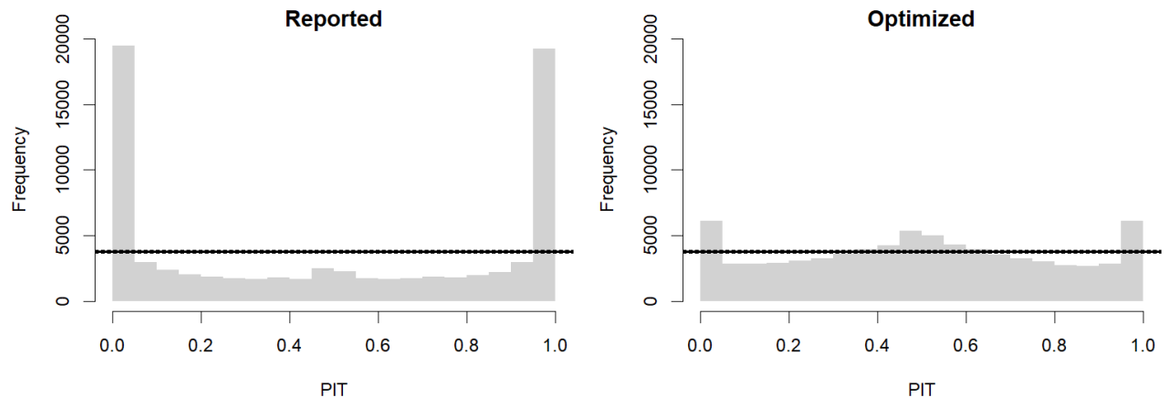

Figure: PIT for originally reported (left) and optimized uncertainty (right).

This improved PIT shape is also quantified in the probabilistic validation metric, the CRPS, as summarized in Table: CRPS and CRPSS summary. The CRPS for Seoul-SNU improved by a factor of 12.62, or expressed as a skill score of CRPSS by 0.92. While Seoul-SNU depicts a highly polluted site that also describes one of the more extreme cases of the uncertainty reporting, improvements but of lower magnitude were found for all other sites. Remote sites such as Izaña, show the smallest improvement of a factor of 1.52, which translates to a skill score of 0.34.

Location	CRPS_reported	CRPS_optimized	Improvement	CRPSS
Izana	1.37×10^{-6}	9.00×10^{-7}	1.52	0.342
LaPorteTX	4.14×10^{-6}	2.49×10^{-6}	1.66	0.399
Downsview	2.62×10^{-6}	1.57×10^{-6}	1.67	0.401
CharlesCityVA	1.61×10^{-6}	9.39×10^{-7}	1.72	0.418
GreenbeltMD	2.34×10^{-6}	1.11×10^{-6}	2.11	0.527
Rome-SAP	5.17×10^{-6}	2.37×10^{-6}	2.18	0.541
BostonMA	6.10×10^{-6}	2.24×10^{-6}	2.72	0.633
Seoul-SNU	2.24×10^{-5}	1.78×10^{-6}	12.62	0.921

Table: CRPS and CRPSS summary sorted by improvement

Figure: Optimized uncertainty components summarizes the validated uncertainty components for the investigated sites, as a function of their median NO_2 amounts. In contrast to Seoul-SNU, at remote sites such as Izaña, the reported systematic uncertainty appeared even slightly overestimated. This site-dependent behavior suggests that the uncertainty parameterization in the retrieval algorithm may not fully capture the variability in measurement conditions or instrument configurations across the network.

The mixed (structured) component, which captures retrieval and algorithm-related uncertainties, was generally found to be slightly underestimated, as seen by R_{mix} > 1 in the left graphic of Figure: Optimized uncertainty components for all locations investigated. However, its overall contribution to

the combined uncertainty is minor compared to the systematic component, and thus its effect is largely masked in the combined uncertainty analysis.

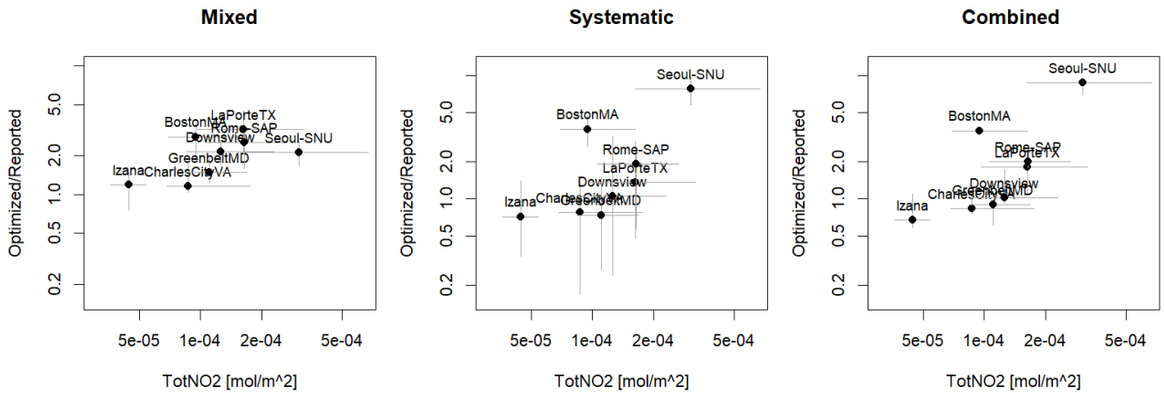

Figure: Optimized uncertainty components as a function of the median total NO_2 amounts for mixed (left) and systematic (middle) components, and the combined uncertainty (right).

The combined uncertainty—representing the sum of independent, mixed, and systematic contributions—showed notable site dependence. At polluted sites, the combined uncertainty was underestimated, driven primarily by the systematic component. Conversely, for remote and cleaner sites, a slight overestimation was observed. Nevertheless, even with an increased mixed component, the remote sites showed an overall decreased uncertainty, due to the higher magnitude of the systematic component.

Satellite validation processes are currently not affected by the identified underestimations, as the associated uncertainty reporting is typically much higher for the satellite. However, the improved quantification of the PGN uncertainty enhances the overall reliability and interpretability of Pandora data products. The UVF approach successfully demonstrated its capability to validate and adjust individual uncertainty components, thereby potentially strengthening confidence in the reported uncertainty metrics. However, it must be noted that the UVF is used as a diagnostic tool to highlight the problematic components, in order to allow a re-investigation of the algorithm.

Further work will focus on investigating the underlying causes of the underestimated systematic uncertainty. Potential contributing factors include unaccounted calibration drifts, environmental dependencies, or limitations in the current modeling of reference spectra. Extending the framework to additional atmospheric constituents beyond NO_2 will provide a broader assessment of the Pandora network's uncertainty performance.