REGIME-CONDITIONED COVERAGE OF MONTE CARLO INTERVALS FOR TROPOMI-DERIVED SURFACE NO_2 DURING WINTER STAGNATION IN THE SOFIA BASIN

Dimitar Makariev, Plamen Trenchev

Space Research and Technology Institute, Bulgarian Academy of Sciences
e-mail: dmakariev@space.bas.bg; ptrenchev@space.bas.bg

Keywords: *TROPOMI; nitrogen dioxide; Monte Carlo simulation; interval coverage; planetary boundary layer; winter stagnation; ERA5; Sobol sensitivity; Sofia Basin; air quality*

Abstract: *Aggregate interval coverage can conceal systematic failure under the meteorological conditions associated with the highest surface pollution. This study develops the winter NO_2 case proposed in the authors' earlier reviews into a reproducible uncertainty-propagation and validation protocol for the Sofia Basin. TROPOMI tropospheric NO_2 columns are converted to surface mass concentrations with an explicit mixed-layer approximation. Monte Carlo sampling propagates uncertainty in the retrieved column, ERA5 boundary-layer height, 10 m wind, and pixel-to-station representativeness. A prespecified ERA5 rule separates shallow, weak-wind stagnation from ventilated station-days; the rule is a stagnation classifier, not a direct measurement of a temperature inversion. The central 90% interval is evaluated using regime-specific coverage, median width, and the proper interval score. Day-block resampling preserves dependence among stations exposed to the same daily retrieval and meteorology. In a synthetic January 2021 verification dataset, Sofia coverage is 63.9% under stagnation and 87.7% under ventilation, a difference of −23.8 percentage points (95% block-bootstrap interval −39.2 to −7.7). Boundary-layer height contributes 61.4% of first-order output variance under stagnation and 32.2% under ventilation. These results verify the analysis workflow but do not provide evidence about January 2021 air quality. Scientific inference requires replacement of the synthetic fields with collocated TROPOMI, ERA5, and EEA observations.*

РЕЖИМНО ОБУСЛОВЕНО ПОКРИТИЕ НА МОНТЕ КАРЛО ИНТЕРВАЛИ ЗА ПРИЗЕМЕН NO_2, ИЗВЛЕЧЕН ОТ TROPOMI, ПРИ ЗИМЕН ЗАСТОЙ В СОФИЙСКАТА КОТЛОВИНА

Димитър Макариев, Пламен Тренчев

Институт за космически изследвания и технологии – Българска академия на науките
e-mail: dmakariev@space.bas.bg; ptrenchev@space.bas.bg

Ключови думи: *TROPOMI; азотен диоксид; Монте Карло симулация; покритие на интервали; планетарен граничен слой; зимен застой; ERA5; Собол анализ на чувствителността; Софийска котловина; качество на въздуха*

Резюме: *Обобщеното покритие на интервалите може да скрие систематични грешки при метеорологичните условия, свързани с най-високи приземни концентрации. Изследването развива зимния пример за NO_2 от предишните публикации на авторите във възпроизводим протокол за разпространяване на неопределеността и валидиране в Софийската котловина. Тропосферните колони NO_2 от TROPOMI се преобразуват в приземни масови концентрации чрез явно допускане за хомогенно смесване в граничния слой. Монте Карло симулацията разпространява неопределеността в сателитната колона, височината на граничния слой от ERA5, вятъра на 10 m и представителността между пиксела и станцията. Предварително зададено правило на ERA5 разделя дните на застойни и проветриви; правилото класифицира застой, а не измерва пряко температурна инверсия. Централният 90% интервал се оценява чрез покритие, медианна ширина и правилен интервален показател, като блоковото преизбиране по дни запазва зависимостта между станции с обща метеорология. В синтетичния набор за проверка за януари 2021 г. покритието в София е 63.9% при застой и 87.7% при проветряване, разлика −23.8 процентни пункта (95% блоков интервал −39.2 до −7.7). Височината на граничния слой обяснява съответно 61.4% и 32.2% от дисперсията от първи ред. Тези резултати проверяват аналитичния процес, но не са доказателство за реалното качество на въздуха през януари 2021 г. Научните изводи изискват замяна на синтетичните полета с колокирани наблюдения от TROPOMI, ERA5 и EAOC.*

1. Introduction

Tropospheric nitrogen dioxide is an indicator of combustion-related emissions and a precursor of ozone and nitrate aerosol. Sentinel-5 Precursor TROPOMI provides near-daily tropospheric NO_2vertical columns at kilometre-scale nadir resolution [1, 2]. A satellite column does not directly measure the concentration at a monitoring inlet. Column-to-surface conversion depends on the vertical NO_2profile, the retrieval averaging kernel, clouds, surface reflectance, boundary-layer depth, chemistry, and spatial representativeness [2–7]. These dependencies become critical during winter stagnation, when emissions can remain concentrated in a shallow layer and a pixel-average column may not describe the surface enhancement at a particular station.

Sofia offers a demanding case. The city lies in an intermontane basin where weak winds and winter temperature inversions impede ventilation. Lidar observations have documented shallow, multilayer nocturnal structure over Sofia, while later studies report frequent winter inversions and pollution episodes associated with weak dispersion [8–10]. A method that appears calibrated across all winter days can therefore fail on the subset of shallow-boundary-layer days most relevant to exposure and air-quality management.

Our earlier SES contribution reviewed Monte Carlo simulation and hybrid atmospheric modelling as parts of a wider intelligent-modelling toolkit [11]. A subsequent review connected uncertainty propagation, Earth-observation data, and an operational hybrid workflow [12]. The present work turns that conceptual programme into a falsifiable pilot experiment. Its central question is not whether a Monte Carlo interval is wide enough on average. It is whether a nominal 90% interval covers independent surface observations at approximately the same rate in stagnant and ventilated regimes.

The analysis is organised around two prespecified hypotheses. **H1:** coverage of the nominal 90% interval is lower on stagnant than on ventilated Sofia station-days. **H2:** the first-order Sobol index of boundary-layer height is larger under stagnation. The study makes the assumptions in the column-to-surface conversion explicit, fixes the meteorological regime rule before evaluation, and reports calibration and sharpness by regime rather than only in aggregate. The January pilot is a physically constrained synthetic verification dataset. It tests the code path, estimands, and graphics but cannot support conclusions about measured January 2021 pollution. The final inference requires collocated TROPOMI, ERA5, and EEA records.

2. Study design and data

2.1. Spatial design

The target domain comprises EEA monitoring locations within Sofia and a control location in Plovdiv. Sofia represents the basin setting in which shallow winter boundary layers and weak winds can sustain near-surface accumulation. Plovdiv is not a clean-background reference; it is an urban control outside the Sofia Basin, used to test whether an apparent coverage deficit follows the Sofia stagnation regime rather than affecting every urban station indiscriminately.

The pilot uses four stable station aliases: three Sofia locations representing two urban-background contexts and one traffic-influenced context, plus one Plovdiv control. The aliases deliberately separate the analysis code from final station identifiers. In the observational run, each alias will map to an EEA station code, coordinates, station type, inlet height, and the TROPOMI pixel-selection record. Station metadata and validated hourly NO_2 data will be obtained from the EEA air-quality download service [13].

2.2. One-month pilot period

January 2021 is used as a winter demonstrator month. Bulgarian studies document frequent winter inversions, shallow boundary-layer structure, and pollution episodes in Sofia, but the present synthetic dataset does not reproduce a specific observed episode [8–10]. Each day contributes one satellite-overpass comparison per station when quality screening succeeds. The provisional pilot contains 31 days and 124 station-days. The primary Sofia evaluation contains 93 station-days: 36 observations on 12 stagnant days and 57 observations on 19 ventilated days. The Plovdiv control contributes 31 station-days. This separation is retained throughout the coverage analysis so that the control city does not influence the Sofia estimate.

The numerical fields in the pilot data file are generated from deterministic pseudo-random draws with seed 20261017. They preserve realistic orderings: stagnant days have lower boundary-layer heights, lower winds, higher urban columns, and larger relative uncertainty. They do not reproduce an EEA or TROPOMI measurement. This distinction is retained in every data row through a data_status field.

2.3. TROPOMI and ERA5 inputs

The observational implementation will use the offline or reprocessed Sentinel-5P Level-2 NO_2 product and retain pixels with qa_value ≥ 0.75, following product guidance and validation practice [2, 3, 5, 14, 15]. Pixel values will be area-weighted when a station falls within more than one valid footprint. The comparison time will be the TROPOMI overpass. Station observations will be averaged within a symmetric temporal window, with at least 75% hourly completeness; both the window width and completeness will be recorded in the data manifest. The product-supplied column precision will enter the Monte Carlo distribution rather than being discarded after screening.

ERA5 provides hourly boundary-layer height and 10 m wind components [16]. Wind speed is calculated from the zonal and meridional components:

$$u_{10} = \sqrt{u^2 + v^2}$$

Meteorological fields are interpolated to the satellite overpass and station coordinate. ERA5 is used consistently for regime classification and for centring the input distributions; this avoids mixing classification from one meteorological product with uncertainty propagation from another.

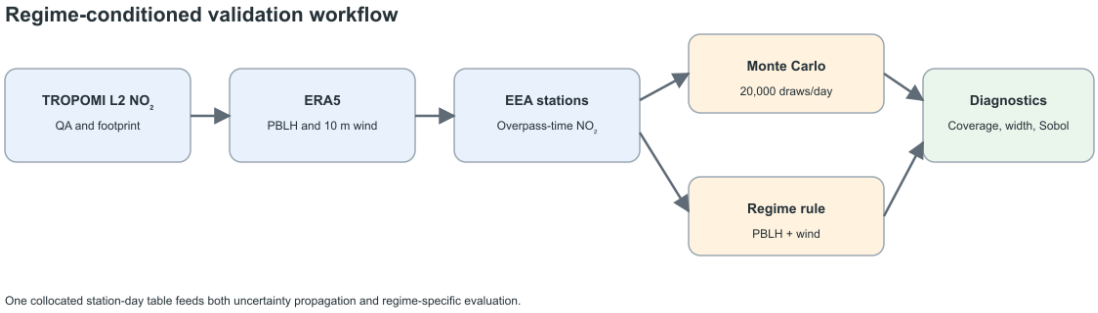


Fig. 1. End-to-end pilot workflow. The same collocated station-day record supports interval construction, regime-conditioned validation, and sensitivity analysis.

2.4. Regime definition

The primary regime rule is applied to ERA5 values interpolated to the TROPOMI overpass and is fixed before coverage evaluation:

$$stagnant\ if\ PBLH < 350\ m \wedge u_{10} < 2.0\ m\ s^{-1};\ ventilated\ otherwise$$

The threshold pair is intentionally simple and operational. It identifies the joint state of shallow mixing and weak horizontal ventilation; it does not diagnose the sign or strength of a temperature inversion from a vertical temperature profile. The paper therefore uses *stagnant* and *ventilated* for the evaluated regimes and reserves *inversion* for the physical motivation. A sensitivity appendix in the final observational analysis will repeat the calculation for boundary-layer thresholds of 300 and 400 m and wind thresholds of 1.5 and 2.5 m s^{-1}. If the inference changes materially, threshold sensitivity will be reported rather than selecting the most favourable split.

3. Monte Carlo model

3.1. Column-to-surface conversion

For day *d*, station *s*, and Monte Carlo draw *k*, the provisional surface concentration is

$$C_{d,s}^{(k)} = 10^6\, M_{NO_2}\, \frac{\Omega_{d,s}^{(k)}}{H_{d,s}^{(k)}}\, R_{d,s}^{(k)}\, V_{d,s}^{(k)}$$

where Ω is the screened TROPOMI tropospheric vertical column in mol m^{-2}, $M(NO_2)$ = 46.0055 g mol^{-1}, H is the ERA5 boundary-layer height in metres, and the factor 10^6 converts g m^{-3} to µg m^{-3}. R is a positive station-day multiplicative term representing footprint-to-inlet mismatch and unresolved local enhancement. It is not an emission inventory estimate. The dimensionless ventilation perturbation is

$$V_{d,s}^{(k)} = \left[\frac{\bar{u}_{d,s} + u_0}{u_{d,s}^{(k)} + u_0}\right]^{\beta}$$

with u_0 = 0.5 m s^{-1} and β = 0.4. The offset limits numerical instability at very low wind speed. The exponent is a pilot assumption and will be examined in sensitivity analysis rather than fitted on the evaluation month.

The term $10^6 M\Omega/H$ is a mass-conserving mixed-layer approximation. It assumes that the column relevant to the station is confined below H and uniformly distributed within that layer. This assumption is stronger than merely assuming that most urban NO_2 lies in the boundary layer because TROPOMI also senses free-tropospheric NO_2 and has vertically varying sensitivity [2, 15, 17]. Published column-to-surface methods commonly use modelled concentration-to-column ratios or learned relationships with meteorology and land-use covariates [4–7]. The expression used here is therefore a transparent stress-test baseline, not a replacement for a chemical transport model.

3.2. Input distributions

Each station-day uses 20,000 draws. In the baseline design, Ω, H, u, and R are sampled independently so that first-order Sobol indices have their standard interpretation. Dependence is physically plausible, especially between boundary-layer height, wind, and the observed column, so the final analysis will include a separate correlation sensitivity experiment; Sobol indices will not be reported for that correlated experiment. Positive inputs are sampled from lognormal distributions. For an input with arithmetic mean *x* and coefficient of variation CV, the log-scale parameters are

$$\sigma_{ln}^2 = ln(1 + CV^2), \quad \mu_{ln} = ln(x) - \frac{1}{2}\sigma_{ln}^2$$

Wind is sampled from a lower-truncated normal distribution. The provisional settings are shown in Table 1. For the observed run, the TROPOMI column CV will combine the product precision with a documented structural component for air-mass-factor and profile uncertainty [2, 3, 15]. The boundary-layer scenarios are deliberately broader under stagnation because shallow and stable layers are difficult to define consistently across observations and reanalysis [18, 19]. The R distribution represents footprint-to-inlet mismatch and unresolved local enhancement. The fixed pilot values are scenario assumptions used to test the workflow; they are not claimed as estimates for Sofia.

Table 1. Provisional Monte Carlo distributions. Final parameters will be estimated or justified against the collocated observational data.

Input	Distribution	Stagnant regime	Ventilated regime
TROPOMI NO_2 column Ω	Lognormal, centred on screened Level-2 value	CV 22%	CV 22%
Boundary-layer height H	Lognormal, centred on ERA5	CV 48%	CV 24%
10 m wind u	Truncated normal, centred on ERA5	SD 0.38 m s^{-1}	SD 0.55 m s^{-1}
Representativeness factor R	Lognormal, mean 1.0	CV 24%	CV 24%

The 5th and 95th percentiles form the central 90% uncertainty interval, and the median is the point summary. This is not a formal predictive interval unless R and the other inputs account for all material model discrepancy and observation error. Coverage against independent stations tests that assumption. Monte Carlo error is checked by repeating representative days with twice the draw count; the final workflow requires changes below 0.5 µg m^{-3} in both endpoints [20].

3.3. Coverage and sharpness

For regime r with station-day set G_r, empirical coverage is

$$\hat{p}_r = \frac{1}{n_r}\sum_{(d,s)\in G_r} \mathbb{1}\{L_{d,s} \le y_{d,s} \le U_{d,s}\}$$

where y is the station observation and [L,U] is the central 90% Monte Carlo interval. Because stations on the same day share meteorology and satellite retrieval error, uncertainty in coverage is estimated by resampling whole days within each regime. The reported 95% intervals use 20,000 day-block bootstrap samples. Coverage must be interpreted together with sharpness, summarized by the median width W = U − L. For α = 0.10, the interval score is

$$IS_\alpha(L, U; y) = (U - L) + \frac{2}{\alpha}(L - y)\,\mathbb{1}_{\{y<L\}} + \frac{2}{\alpha}(y - U)\,\mathbb{1}_{\{y>U\}}$$

Lower mean interval score is better because the score rewards narrow intervals but penalizes misses in proportion to their distance beyond the nearest endpoint [21, 22]. Coverage, median width, and mean interval score will be reported together; no single metric is treated as sufficient.

The primary estimand for H1 is the coverage contrast $\Delta\hat{p} = \hat{p}_s - \hat{p}_v$, where the subscripts denote the stagnant and ventilated regimes. Its day-block bootstrap interval is obtained by resampling days independently within each regime. Regime-specific coverage, width, and interval score

are co-primary diagnostics. Aggregate coverage is secondary. Plovdiv is reported as a control diagnostic and is not pooled with Sofia to claim transferability.

3.4. First-order Sobol indices

Variance-based sensitivity analysis decomposes model-output uncertainty into contributions from individual inputs and their interactions [23]. The first-order Sobol index is

$$S_i = \frac{Var_{X_i}(E[Y \mid X_i])}{Var(Y)}$$

where Y is the simulated surface concentration and X_i is one uncertain input. The pilot uses two independent sample matrices and 50,000 base draws per regime. Indices are calculated at representative regime conditions rather than averaged across days. The comparison of interest is the change in S_h, the first-order index of boundary-layer height, between stagnant and ventilated regimes.

4. Illustrative pilot results

4.1. Daily intervals

Figure 2 shows the complete January sequence for one Sofia urban-background alias. The shaded band is the 90% Monte Carlo interval; orange background strips mark stagnant days. Intervals widen sharply during stagnant episodes because the same absolute uncertainty in a shallow boundary layer becomes a much larger relative uncertainty in C. Several provisional station values approach or exceed the upper endpoint during these episodes, whereas most ventilated-day values remain inside the band.

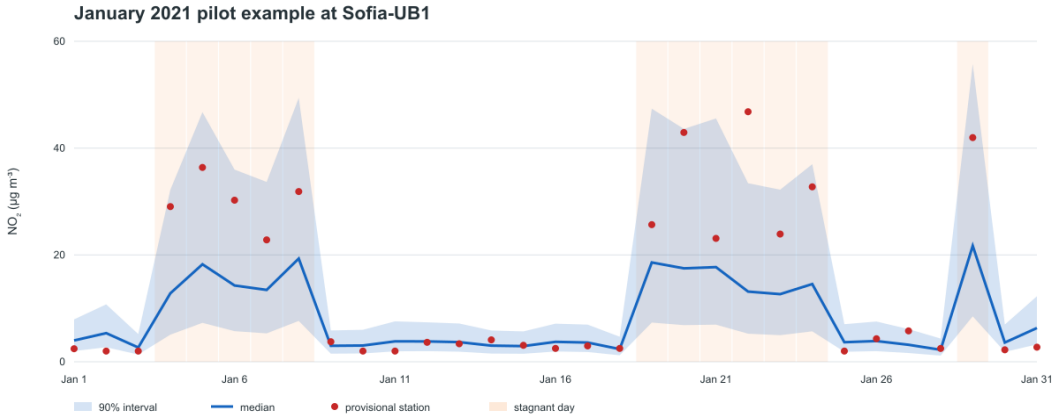


Fig. 2. Illustrative daily 90% intervals and provisional station values for Sofia-UB1 in January 2021. All numerical values are synthetic pilot data.

4.2. Regime-conditioned coverage

Across the 93 Sofia station-days, coverage is 78.5% (day-block bootstrap 95% interval 69.9–87.1%), below the nominal 90% level in the synthetic verification dataset. Coverage is 63.9% (52.8–75.0%) for the 36 stagnant station-days and 87.7% (77.2–96.5%) for the 57 ventilated station-days. The prespecified contrast is Δp = −23.8 percentage points (95% block-bootstrap interval −39.2 to −7.7). The Plovdiv control reaches 96.8% (90.3–100.0%) across 31 station-days.

The stagnant Sofia intervals are also much wider. Their median width is 38.1 µg m^{-3}, compared with 5.1 µg m^{-3} for ventilated days. The mean 90% interval score is 100.2 µg m^{-3} under stagnation and 7.2 µg m^{-3} under ventilation; the large stagnant value reflects both wide intervals and large exceedances. Wider intervals therefore do not restore calibration in the synthetic demonstrator. Table 2 reports the coverage and width results using mutually consistent analysis groups.

Table 2. Illustrative coverage and sharpness. Confidence limits resample complete days; all values are synthetic.

Evaluation group	Station-days	Days	Coverage	95% block-bootstrap interval	Median width (µg m^{-3})
Sofia, all regimes	93	31	78.5%	69.9–87.1%	7.0
Sofia, stagnant	36	12	63.9%	52.8–75.0%	38.1
Sofia, ventilated	57	19	87.7%	77.2–96.5%	5.1
Plovdiv control	31	31	96.8%	90.3–100.0%	4.8

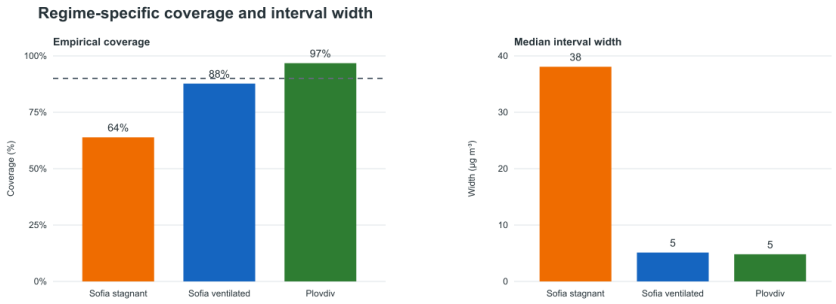


Fig. 3. Illustrative empirical coverage and interval width by regime. The dashed line marks nominal 90% coverage.

4.3. Sensitivity structure

Boundary-layer height is the dominant first-order uncertainty source under stagnation, with S_h = 0.614. Its share decreases to 0.322 in the ventilated regime. The TROPOMI column share rises from 0.134 to 0.277, and the representativeness-factor share rises from 0.154 to 0.325. The wind perturbation contributes about 0.03 in both regimes. The remaining variance is associated with interactions and Monte Carlo estimation error.

This result is consistent with the model structure: concentration varies inversely with H, and the stagnant distribution assigns H the largest relative uncertainty. It should not be interpreted as independent evidence that ERA5 boundary-layer height is erroneous. Sobol indices describe uncertainty under the chosen input distributions; changing those distributions changes the indices.

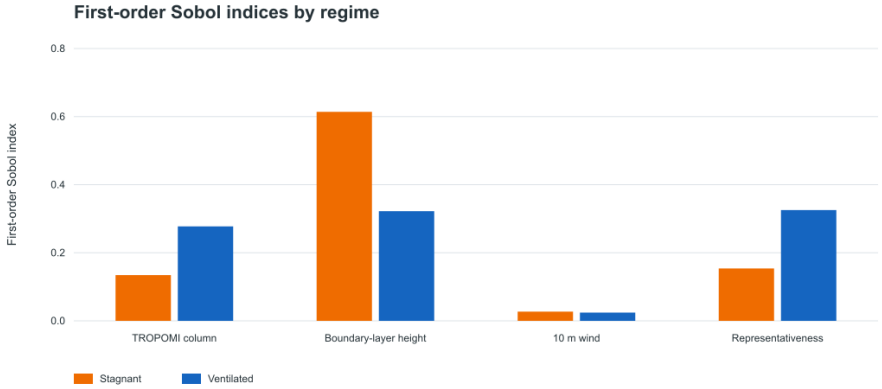


Fig. 4. First-order Sobol indices for representative stagnant and ventilated conditions in the illustrative pilot.

4.4. Surface enhancement diagnostic

The relationship between the Monte Carlo median and the provisional station value differs by regime (Fig. 5). Ventilated points cluster around the 1:1 line, while stagnant Sofia points shift above it. This pattern is the intended stress test: a homogeneous mixed-layer conversion can understate near-surface enhancement when stable stratification confines emissions to a layer shallower or more internally stratified than the ERA5 boundary-layer estimate.

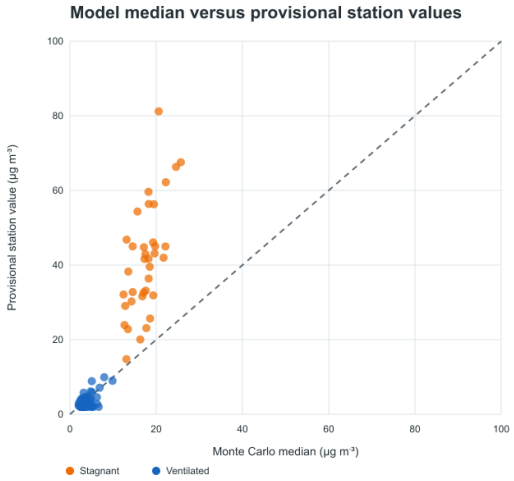


Fig. 5. Monte Carlo median versus provisional Sofia station values. The comparison diagnoses regime-dependent near-surface enhancement rather than model fit.

5. Discussion

5.1. Why aggregate coverage is insufficient

An interval can attain near-nominal overall coverage by combining overcoverage in easy conditions with undercoverage in the conditions that matter most. For air-quality applications, the cost is asymmetric: missing high concentrations during winter stagnation is more consequential than producing a conservative interval on a well-ventilated day. Regime-conditioned coverage therefore functions as a scientific diagnostic, not merely a statistical refinement.

The synthetic pilot shows how this diagnostic changes interpretation. Sofia-wide coverage of 78.5% hides a 23.8 percentage-point contrast between regimes, and the interval score confirms that stagnant failures are not explained by width alone. This pattern was deliberately embedded in the verification data to test whether the analysis recovers a known regime-dependent failure. It therefore validates the implementation, not the atmospheric hypothesis. Only the observational replacement can determine whether the same failure exists in Sofia.

5.2. Physical interpretation

The homogeneous-mixing assumption is most defensible near the early-afternoon TROPOMI overpass when convective mixing is established. It is least defensible during persistent winter inversions, complex basin circulations, and situations with strong vertical gradients. TROPOMI urban enhancements reflect wind-dependent accumulation and transport [24]. The retrieval also depends on an a priori vertical profile through the air-mass factor, while independent work demonstrates that tropospheric NO_2 has substantial vertical structure [2, 3, 15, 17]. A shallow or internally stratified layer can therefore affect both the retrieved column and its conversion to a surface value.

The high stagnant-regime S_h is meaningful within the pilot because H enters the conversion inversely and receives a larger relative uncertainty. It provides a practical prioritisation: improved boundary-layer information, such as radiosonde, ceilometer, lidar, or a higher-resolution model, should be tested before adding complexity elsewhere. Yet the Sobol result does not prove causality from H alone. Interactions among H, the retrieved profile, local emissions, and spatial representativeness may dominate individual high-pollution episodes.

5.3. Role of the Plovdiv control

Plovdiv guards against attributing every winter mismatch to Sofia topography. If both cities show the same regime-conditioned failure after observational replacement, a product-wide retrieval or conversion issue is more plausible. If the deficit is concentrated in Sofia under the prespecified stagnation rule, basin-specific mixing and representativeness become stronger explanations. A single control city cannot establish generality, but it strengthens the design compared with a Sofia-only evaluation.

5.4. Replacement with observed data

The final analysis will replace each synthetic field without changing the schema or figures. The replacement sequence is:

1. ingest screened TROPOMI Level-2 pixels and preserve processor version, orbit, footprint, QA, and column precision;
2. collocate ERA5 boundary-layer height and wind at overpass time;
3. ingest validated EEA hourly NO_2, station metadata, and temporal aggregation records;
4. rerun the fixed regime rule, Monte Carlo sampler, coverage analysis, and Sobol design;
5. publish the data manifest, exclusions, random seed, and sensitivity settings with the revised figures.

No threshold or uncertainty range should be tuned on the final evaluation month. If calibration is required, it should use a separate period, followed by a locked evaluation on held-out days.

6. Limitations

The pilot simplifies chemistry, transport, vertical structure, and spatial representativeness. It does not model NO–NO_2–O_3 partitioning, photochemical lifetime, advection between source and pixel, free-tropospheric NO_2, or subpixel emission gradients. The ventilation term is empirical. ERA5 resolution may not resolve basin circulations or the true inversion top. EEA in-situ observations sample a small volume and may represent traffic or urban-background conditions differently from a TROPOMI footprint.

The baseline Monte Carlo design assumes independent inputs. In reality, Ω, H, and u can be correlated through the same meteorological regime, and retrieval error can covary with clouds and surface reflectance. Ignoring dependence can distort both interval width and Sobol attribution. The final analysis will therefore report an independence baseline, a rank-correlation sensitivity analysis,

and the resulting change in coverage endpoints. The first-order Sobol decomposition will remain restricted to the independent-input baseline.

Coverage estimates from one month have broad confidence intervals and serial dependence. Station-days from the same day share meteorology and satellite retrieval error, so a simple binomial interval is optimistic. The final analysis should use a day-block bootstrap and should report station-specific results before pooling. The prespecified thresholds describe stagnation but do not directly diagnose an inversion from a temperature profile. A radiosonde- or lidar-based inversion label would provide a stronger physical reference.

Finally, the current numbers are synthetic. They demonstrate the planned inference and expose potential failure modes, but they are not evidence about actual January 2021 coverage. Only the observational replacement can support conclusions about TROPOMI-derived surface NO_2 in Sofia.

7. Conclusions

The proposed workflow evaluates a central 90% Monte Carlo interval separately under shallow, weak-wind stagnation and ventilated boundary layers. In the synthetic Sofia verification dataset, coverage of 78.5% separates into 63.9% under stagnation and 87.7% under ventilation, while boundary-layer height accounts for 61% and 32% of first-order output variance, respectively. The analysis correctly recovers the deliberately embedded contrast and therefore verifies the implementation. It does not yet establish that the mixed-layer conversion under-covers real surface NO_2 observations.

The scientific test begins when collocated observations replace the synthetic fields. The two locked hypotheses ask whether stagnant coverage is lower than ventilated coverage and whether boundary-layer height contributes more first-order variance under stagnation. The reporting design keeps regime definition, uncertainty propagation, calibration, sharpness, and sensitivity attribution connected to the same station-day records and prevents the Plovdiv control from altering the primary Sofia estimate.

Data and code availability

The synthetic station-day table, random seed, and generated figures accompany the working manuscript in the SES 2026 project folder. The observational submission will archive the processing code, the data manifest, station aliases mapped to public EEA identifiers, TROPOMI processor versions, and all exclusion decisions in a citable repository, subject to the access conditions of the source services.

Acknowledgements

The authors acknowledge the European Space Agency and Copernicus programme for Sentinel-5P TROPOMI products, the Copernicus Climate Change Service for ERA5, and the European Environment Agency and Bulgarian monitoring authorities for air-quality data infrastructure. The final manuscript will add any project, computing, and data-provider acknowledgements required by the observational dataset.

References:

1. J. P. Veefkind et al., “TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications,” *Remote Sensing of Environment*, vol. 120, pp. 70–83, 2012. https://doi.org/10.1016/j.rse.2011.09.027
2. J. van Geffen et al., “Sentinel-5P TROPOMI NO_2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data,” *Atmospheric Measurement Techniques*, vol. 15, pp. 2037–2060, 2022. https://doi.org/10.5194/amt-15-2037-2022
3. T. Verhoelst et al., “Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO_2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks,” *Atmospheric Measurement Techniques*, vol. 14, pp. 481–510, 2021. https://doi.org/10.5194/amt-14-481-2021
4. V. Balamurugan, J. Chen, A. Wenzel, and F. N. Keutsch, “Spatiotemporal modeling of air pollutant concentrations in Germany using machine learning,” *Atmospheric Chemistry and Physics*, vol. 23, pp. 10267–10285, 2023. https://doi.org/10.5194/acp-23-10267-2023
5. D. Griffin et al., “Development and validation of satellite-derived surface NO_2 estimates using machine learning versus traditional approaches in North America,” *Atmospheric Measurement Techniques*, vol. 18, pp. 7497–7511, 2025. https://doi.org/10.5194/amt-18-7497-2025
6. X. Zhao et al., “Retrieval of total column and surface NO_2 from Pandora zenith-sky measurements,” *Atmospheric Chemistry and Physics*, vol. 19, pp. 10619–10642, 2019. https://doi.org/10.5194/acp-19-10619-2019

7. S. Acker, T. Holloway, and M. Harkey, “Satellite detection of NO_2 distributions using TROPOMI and TEMPO and comparison with ground-based concentration measurements,” *Atmospheric Chemistry and Physics*, vol. 25, pp. 8271–8288, 2025. https://doi.org/10.5194/acp-25-8271-2025
8. K. Ivanov and R. Evgeniev, “Temperature inversions and their relation to the air pollution in the city of Sofia,” *Journal of Environmental Protection and Ecology*, vol. 23, no. 2, pp. 441–453, 2022.
9. E. Hristova et al., “Black Carbon in Bulgaria: Observed and Modelled Concentrations in Two Cities for Two Months,” *Atmosphere*, vol. 13, article 213, 2022. https://doi.org/10.3390/atmos13020213
10. I. Kolev, P. Savov, B. Kaprielov, O. Parvanov, and V. Simeonov, “Lidar observation of the nocturnal boundary layer formation over Sofia, Bulgaria,” *Atmospheric Environment*, vol. 34, no. 19, pp. 3223–3235, 2000. https://doi.org/10.1016/S1352-2310(99)00490-2
11. D. Makariev and P. Trenchev, “Advanced Computational Approaches for Data-Driven and Intelligent Modeling of Atmospheric Processes,” in *Proceedings of the Twenty-first International Scientific Conference Space, Ecology, Safety, SES 2025*, Sofia, Bulgaria, 2025. https://space.bas.bg/SES/archive/SES%202025_DOKLADI/3_Remote%20Sensing/9_Makariev.pdf
12. D. Makariev and P. Trenchev, “Review of Contemporary Intelligent Methods for Atmospheric Process Modeling and a Proposed Hybrid Framework Integrating Earth System Digital Twins with Copernicus Observia AI,” *Aerospace Research in Bulgaria*, vol. 38, 2026. https://journal.space.bas.bg/arhiv/n%2038/Articles/25_Makariev.pdf
13. European Environment Agency, “Air Quality download service,” accessed 17 October 2026. https://www.eea.europa.eu/en/datahub/datahubitem-view/datahubitem-view/778ef9f5-6293-4846-badd-56a29c70880d
14. European Space Agency, “Copernicus Sentinel-5P TROPOMI Level 2 Nitrogen Dioxide total column products, Version 02,” 2021. https://doi.org/10.5270/S5P-9bnp8q8
15. J. H. G. M. van Geffen, H. J. Eskes, K. F. Boersma, and J. P. Veefkind, “TROPOMI ATBD of the total and tropospheric NO_2 data products,” S5P-KNMI-L2-0005-RP, issue 2.9.0, 2025. https://sentiwiki.copernicus.eu/__attachments/1673595/S5P-KNMI-L2-0005-RP%20-%20Sentinel-5P%20TROPOMI%20ATBD%20NO2%20data%20products%202025-2.90.pdf
16. H. Hersbach et al., “The ERA5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, pp. 1999–2049, 2020. https://doi.org/10.1002/qj.3803
17. R. P. Horner, E. A. Marais, N. Wei, R. G. Ryan, and V. Shah, “Vertical profiles of global tropospheric nitrogen dioxide (NO_2) obtained by cloud slicing the TROPOspheric Monitoring Instrument (TROPOMI),” *Atmospheric Chemistry and Physics*, vol. 24, pp. 13047–13064, 2024. https://doi.org/10.5194/acp-24-13047-2024
18. V. A. Sinclair et al., “Boundary-layer height and surface stability at Hyytiälä, Finland, in ERA5 and observations,” *Atmospheric Measurement Techniques*, vol. 15, pp. 3075–3103, 2022. https://doi.org/10.5194/amt-15-3075-2022
19. F. Madonna, D. Summa, P. Di Girolamo, F. Marra, Y. Wang, and M. Rosoldi, “Assessment of Trends and Uncertainties in the Atmospheric Boundary Layer Height Estimated Using Radiosounding Observations over Europe,” *Atmosphere*, vol. 12, article 301, 2021. https://doi.org/10.3390/atmos12030301
20. T. Roy and H. Gupta, “How certain are our uncertainty bounds? Accounting for sample variability in Monte Carlo-based uncertainty estimates,” *Environmental Modelling & Software*, vol. 136, article 104931, 2021. https://doi.org/10.1016/j.envsoft.2020.104931
21. T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *Journal of the American Statistical Association*, vol. 102, pp. 359–378, 2007. https://doi.org/10.1198/016214506000001437
22. T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B*, vol. 69, no. 2, pp. 243–268, 2007. https://doi.org/10.1111/j.1467-9868.2007.00587.x
23. A. Saltelli et al., “Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index,” *Computer Physics Communications*, vol. 181, pp. 259–270, 2010. https://doi.org/10.1016/j.cpc.2009.09.018
24. A. Lorente et al., “Quantification of nitrogen oxides emissions from build-up of pollution over Paris with TROPOMI,” *Scientific Reports*, vol. 9, article 20033, 2019. https://doi.org/10.1038/s41598-019-56428-5