Response to reviewers' comments

Reviewer#2

This manuscript tackles a real and under-quantified problem—input-data-induced dataset shift in ML-based reactive nitrogen wet deposition—and the three-case sensitivity design (sample size, spatial distribution, site-type imbalance) is well-conceived and clearly executed. The findings are useful for the community and the proposed framework has broader applicability. However, several issues need to be resolved before acceptance.

Response: We appreciate the opportunity to revise our manuscript and are grateful for the insightful comments provided by reviewer#2. In the following, we have provided detailed responses to each of the reviewer's comments. **Our responses (in blue) and the corresponding edits in the manuscript (in red) are shown below.** All the page and line numbers mentioned below are referred to the revised manuscript.

1. Lines 148–150: The regional deposition estimates in Section 3.4 and the China-average values in Section 3.1 are reported for 2000–2020, and the emission features are stated to cover 2000–2020. However, the observational targets end in 2018 (EANET 2000–2018; NADMN 2010; NNDMN 2010–2015), and the authors' own cited HTAP_v3 reference (Crippa et al., 2023) explicitly covers 2000–2018. Please reconcile how predictions and emission inputs were extended to 2019–2020, or correct the stated period to 2000–2018 throughout.

Response: We appreciate the reviewer's careful identification of the inconsistency in temporal coverage across datasets. In the revised manuscript, we made the following revisions and added additional uncertainty discussion.

Line 161-165: It provides monthly emission gridmaps (2000–2018) from anthropogenic sources for main air pollutants, including NO_2, $PM_{2.5}$, SO_2, NMVOC, and NH_3. For predictions covering 2019–2020, we held anthropogenic emissions constant at the 2018 level while adopting time-varying meteorological fields corresponding to 2019–2020.

We added a dedicated statement in Section 4.2 to elaborate associated uncertainties caused by using static emissions for 2019-2020 simulation.

Line 585-589: Holding anthropogenic emissions fixed at the 2018 level for 2019–2020 simulations potentially introduce estimation biases in the regional deposition amounts. However, the core focus of this work lies in evaluating relative discrepancies between sensitivity simulations and the baseline case, which largely weakens the impacts of temporally invariant emissions on our major conclusions.

2. Lines 126–135 and 408–413: The reconstruction of Dwet from the predicted Cwet is underspecified, which undermines the deposition-amount results. The authors predict concentration (Cwet) rather than deposition to avoid precipitation-related uncertainty, yet the headline deposition estimates and sensitivity impacts in Section 3.4 (Figs. 9–10) are expressed as Dwet. Since Eq. (1) implies Dwet = Cwet × Prep, the deposition estimates must reintroduce a precipitation product. The manuscript does not state which precipitation dataset was used for this back-conversion, nor whether its uncertainty is propagated into the reported Dwet deviations. This is important because the paper's central deliverable is Dwet uncertainty. Please clarify the conversion procedure and its treatment of precipitation uncertainty.

Response: We greatly appreciate this insightful comment on the wet deposition calculation methodology. We have supplemented description of the precipitation product used for C_{wet}-to-D_{wet} conversion in Section 2.2 and added a limitation statement in Section 3.4.

Lines 188-197: To obtain gridded monthly precipitation for converting modelled C_{wet} to D_{wet}, we constructed a fused precipitation dataset via multiple linear regression (MLR, detailed in Tan et al., 2026). We first validated ERA5-Land TP and GPCP v2.3 precipitation against site observations. ERA5-Land overestimates precipitation notably for monthly values < 500 mm, while the 2.5° coarse GPCP product performs better in low-precipitation regions yet mismatches our 0.25° model grid. Our MLR fusion integrates ERA5-Land, GPCP, geographic coordinates and monthly timestamps, trained on 2008–2018 in-situ precipitation from NNDMN, NADMN and EANET. The fused dataset yields higher R^2, lower RMSE and NMB than raw single-source precipitation and is uniformly used for all D_{wet} calculations.

Reference: Tan, J., Zhang, Y., Liu, X., Lam, Y. F., Latif, M. T., Manomaiphiboon, K., Fu, J. S., Ooi, M. C. G., Huang, L., Wang, Y., Chen, H., Zhang, K., Mu, Q., and Li, L.: Data driven analysis on changing patterns of reactive nitrogen deposition in East and Southeast Asia, Atmos. Environ., 374, 121962, https://doi.org/10.1016/j.atmosenv.2026.121962, 2026.

This fused precipitation field is adopted for both the baseline case and all sensitivity tests. Precipitation biases remain fixed across scenarios, so the relative D_{wet} differences quantified in sensitivity analyses exclusively stem from variations in pollutant input datasets that control C_{wet}. Precipitation-related uncertainties cancel out in inter-scenario comparisons, enabling us to independently isolate the effects of input dataset shifts, the central aim of this work. In sect. 3.4, we supplemented a limitation statement.

Line 607-611: The reported D_{wet} deviations between scenarios exclude independent uncertainty originating from the precipitation product. While fixed precipitation enables controlled comparative analysis, the full absolute uncertainty of real-world D_{wet} estimates would require additional propagation of precipitation retrieval and fusion errors.

3. Lines 159–165 versus Lines 502–511: In Section 2.2 the authors justify using emission data by reporting a "minimal difference" between emission- and satellite-based models for 2008–2010. However, in Section 4.1 they show that replacing NH_3 emission with satellite NH_3 VCD in the S3-Remote scenario raises R from 0.36 to 0.58 and improves the emission–Cwet slope agreement. Since remote-site behavior is central to the paper's conclusions, the aggregate "minimal difference" finding may not hold in exactly the regime that matters most. I recommend qualifying the Section 2.2 statement to note that the equivalence is a domain-average result and does not extend to remote sites.

Response: We fully agree with the reviewer's critical comment. The inconsistency between Section 2.2 and Section 4.1 arises from conflating whole-domain average metrics with site-specific performance at remote background locations. We have revised the wording in Section 2.2 to explicitly distinguish domain-wide average results from remote-site performance, and we add cross-references between the two sections for logical consistency.

Line 182-184: The results showed a minimal difference in model performance between the two for domain-average results across all sites (Table S1). However, this overall similarity cannot be generalized to remote sites, as further discussed in Section 4.1.

Line 566-569: Despite the domain-wide consistency between emission-driven and satellite-driven models across the whole domain (Section 2.2, Table S1), this prominent performance improvement at remote sites arises from the large mismatch between the observed and modelled C_{wet}-NH_4^+-emission relationship under the emission-based model (Fig. 8b).

4. Line 358: Section-numbering error. The subsection "3.2.3 Influence from imbalanced number of site types" follows "3.3.1 Influence from spatial distribution of measurements". Given the parent heading 3.3, the site-type subsection should be numbered 3.3.2, not 3.2.3.

Response: We have corrected the subsection numbering error from 3.2.3 to 3.3.2 in the revised manuscript.

5. Lines 274–275 versus Lines 147–156, 164: Unreconciled feature count. The SFR calculation states "13,000 samples with 16 features," but the described feature set—2 emission precursors (NO_x and NH_3), 8 meteorological variables, and 2 geographic variables—sums to 12. Please reconcile the feature count (e.g., explicitly list all 16 features), since SFR is a stated basis for the sample-size conclusions.

Response: We thank the reviewer for identifying this numerical inconsistency. In the revised manuscript, we added Table S1 in the supplementary information, which listed all baseline predictive features and their data sources. We clarify that the baseline XGBoost model adopted for SFR calculation used 15 predictive features, rather than 16. Satellite NH_3 and NO_2 VCD variables are only used for additional experiment and are not incorporated into the baseline feature set. Following are the revisions in the manuscript.

Line 158-179: The targets used for developing XGBoost models were described in section 2.1. The features used encompassed three categories of gridded data: (1) air pollution emission data from Hemispheric Transport of Air Pollution (HTAP) v3 database (Crippa et al., 2023) (https://edgar.jrc.ec.europa.eu/dataset_htap_v3, last access: 2025/10/14). It provides monthly emission gridmaps (2000-2018) from anthropogenic sources for main air pollutants, including NO_2, $PM_{2.5}$, SO_2, NMVOC, and NH_3. For predictions covering 2019–2020, we held anthropogenic emissions constant at the 2018 level while adopting time-varying meteorological fields corresponding to 2019–2020. (2) meteorological parameters from ERA5-Land reanalysis data (https://cds.climate.copernicus.eu/, last access: 2025/10/14), including total column water (TCW), total column water vapor (TCWV), surface pressure (SP), total precipitation (TP), 2-meter dewpoint temperature (D_2M), 2-meter temperature (T_2M), 10-meter wind speed (WS_{10}), and boundary layer height (BLH). (3) precipitation datasets including monthly accumulated wet-deposition precipitation from in-situ monitoring (Table 1) and monthly precipitation from the Global Precipitation Climatology Project (GPCP v2.3) (https://cds.climate.copernicus.eu/, last access: 2025/10/14). Additional satellite-derived variables, including NH_3 vertical column density (VCD) from Infrared Atmospheric Sounding Interferometer (IASI, https://iasi.aeris-data.fr/, last access: 2025/10/14) and NO_2 VCD from Ozone Monitoring Instrument (OMI, https://www.temis.nl/airpollution/no2.php, last access: 2025/10/14), are only adopted for additional experiment and are not included in the baseline feature set. All data were spatially aligned to a uniform 0.25° × 0.25° resolution grid to ensure consistency. See full list of variables in Table S1.

Line 315-318：In this section, we test the model sensitivity to a maximum SFR value of over 867 (13,000 samples with 15 features) and a minimum SFR value of 200 (3,000 samples), both exceeding the high-confidence threshold (SFR ≥ 100) (Zhu et al., 2023).

Table S1 Features used for developing the XGBoost model and their sources

No	Feature	Source	Note
1	NMVOC	HTAPv3 database	
2	NOx		
3	$PM_{2.5}$		
4	SO_2		
5	NH_3		
6	BLH	ERA5-Land reanalysis data	Boundary layer height
7	D2M		2-meter dewpoint temperature
8	WS_{10}		10-meter wind speed
9	SP		Surface pressure
10	T2M		2-meter temperature
11	TCW		total column water
12	TCWV		total column water vapor
13	TP		Total precipitation
14	Prep_in-situ	In-situ measurements	Monthly accumulated precipitation
15	Prep_GPCP	GPCP dataset	Monthly accumulated precipitation
16	NO_2 VCD	OMI satellite	Only for additional experiment
17	NH_3 VCD	IASI satellite	Only for additional experiment

6.Line 226 (Equation 2): Typesetting error in the correlation coefficient formula. As printed, the denominator is a difference of two square-root terms, and the second term appears as $(Y_i - Y)^2$ rather than $(Y_i - \bar{Y})^2$. The Pearson correlation denominator should be the product of the two root-sum-of-squares terms with mean-centered Y. Please correct the equation.

Response: We have corrected the error in Equation (2).

$$R = \frac{\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i-\bar{X})^2}-\sqrt{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}} \quad (2)$$

7. Line 262 versus Line 410, plus editorial errors: Numeric clarification and typographical cleanup. The China-average Dwet-NH_4^+ is given as 4.8 kg N ha^{-1} yr^{-1} while the Base-case domain-average Dwet-NH_4^+ is ~9.0 kg N ha^{-1} yr^{-1} (Line 410); please state explicitly that these refer to China versus the full EA+SEA domain so readers do not read them as contradictory. In addition, please correct recurring typographical errors, including "Dwelt" (Line 21), "ML-base" (Line 63), "SWand"

(Lines 321, 333), and "The the NADMN" (Line 600). A careful language edit throughout is advisable.

Response: We have clarified the two distinct Dwet-NH_4^+ averages to avoid misinterpretation: 4.8 kg N ha^{-1} yr^{-1} refers to China-only, while the ~9.0 kg N ha^{-1} yr^{-1} value corresponds to the full East + Southeast Asia study domain.

Line 455-456: The average D_{wet} estimated in the Base case was about 9.0 kg N ha-1 yr-1 for the study domain (Fig. 2)

All listed typographical errors are corrected in the revised manuscript, and a full round of language editing has been conducted across the manuscript to eliminate additional grammatical and spelling mistakes.