https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

Quantifying the impact of input data-induced dataset shift on machine learning model applications: A case study of regional reactive nitrogen wet deposition

Yan Zhang[1], Jiani Tan[1,2#], Qing Mu[3], Joshua S. Fu[4], Li Li[1#]

[1]School of Environmental and Chemical Engineering, Shanghai University, Shanghai, 200444, China

[2]Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Ministry of Ecology and Environment, Shanghai Academy of Environmental Sciences, Shanghai, 200233, P.R. China

[3]Department of Health and Environmental Sciences, School of Science, Xi'an Jiaotong-Liverpool University, 111 Ren'ai Road, Suzhou, 215123, China

[4]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, 37996, USA

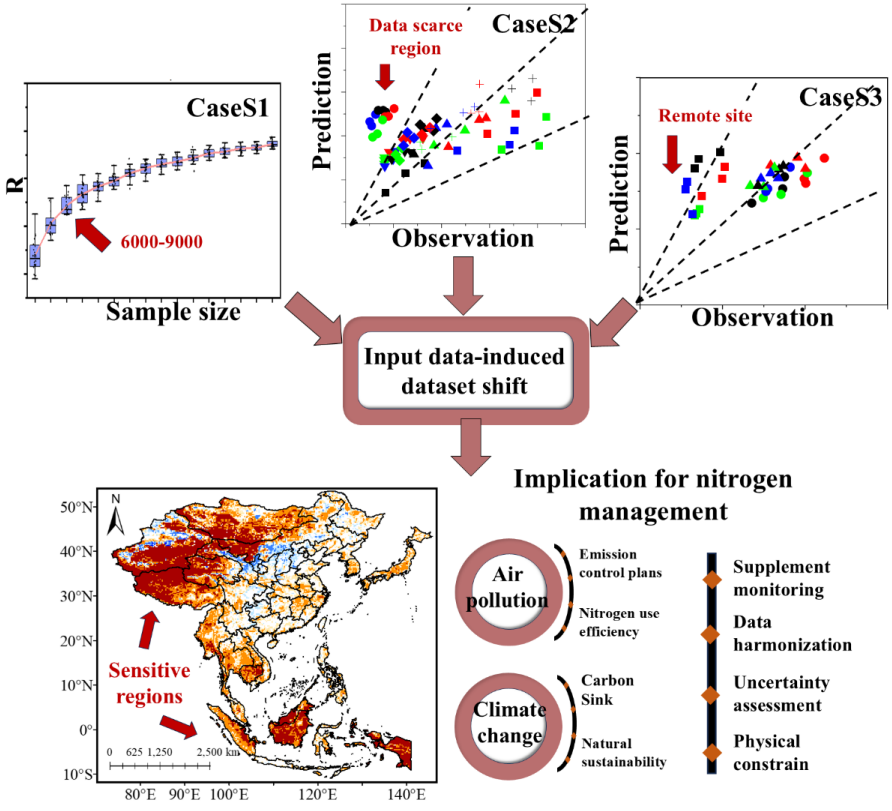
Correspondence to: Jiani Tan(Jiani-tan@shu.edu.cn) and Li Li (lily@shu.edu.cn)

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

Abstract: Machine learning (ML) has been extensively applied to studies on spatial distribution characteristics of atmospheric composition, but quantitative assessments of uncertainties arising from input data properties are still lacking. To address this gap, we conducted a case study on wet deposition of reactive nitrogen (D_{wet}). The Extreme Gradient Boosting (XGBoost) model was employed to predict D_{wet} in East Asia and Southeast Asia (SEA) with a compiled dataset from multiple sources. We quantified the impacts of input data characteristics on model performance and D_{welt} estimations via three sensitivity experiments. Key findings revealed that: (1) Sample size below 6,000–9,000 led to a maximum 12% accuracy loss, while exceeding this threshold provided marginal performance improvement. (2) Uneven spatial distribution of monitoring sites caused 9–51% deviations from baseline performance, with >50% variability in D_{wet} estimations in data-scarce regions (e.g., western China and SEA). (3) Imbalanced site types lead to insufficient representation of remote sites, resulted in 9–40% overall accuracy loss and a high risk of severe overestimation (100%) in remote areas. The bias was attributed to both data range shifts and altered feature-target relationships (e.g., NH_3 emission vs. D_{wet}-NH_4^+). Additionally, inconsistencies among multi-source datasets and limitations of ML structure further introduced uncertainties. This study quantified previously unaddressed input data-induced uncertainties in ML-based N_r deposition research, providing critical insights for reliable application of ML-derived data in N_r management. The proposed uncertainty assessment framework is also applicable to other ML-based geospatial interpolation tasks facing data scarcity challenges.

Keywords: Nitrogen deposition, Machine learning, Uncertainty quantification, Model robustness, Sample-to-Feature Ratio, Sample selection, Imbalanced data

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026



Graphic Abstract

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

1. Introduction

Machine learning (ML) demonstrated superior capability than traditional methods (e.g., linear regression), due to data and hypothesis-driven approach and capability of describing complex relationships among variables (Zhu et al., 2023; Zhong et al., 2021). Zhu et al. (2023) and Zhong et al. (2021) reviewed common issues related to ML application in environmental science, which covered a broad range of topics related to data (i.e., data size, data enrichment, and feature selection) and model (i.e., model selection, model development and evaluation). Each category could pose considerable impacts on model interpretation and data reliability, yet only limited number of studies (24%) have included uncertainty assessment (Zhu et al., 2023). In addition, more attention has been paid to model optimization, for instance developing new model architectures (Zhao et al., 2022; Chen et al., 2023; Zhou et al., 2023), modifying existing models (Li et al., 2020a), and implementing ensemble techniques (Rui Li, 2021; Lu et al., 2020), while uncertainties arising from input data properties are still lacking. As machine learning approach gains increasing popularity in addressing environmental science problems (Guo et al., 2025), uncertainty assessment, especially quantitative investigation becomes a demanding and hot topic (Heid et al., 2023; Chua et al., 2023).

ML-based regional quantification of reactive nitrogen (N_r) depositions has been widely applied to assess environmental impacts and sustainability (Zhao et al., 2022; Chen et al., 2023; Yixin Guo, 2024; Yu et al., 2024; Shang et al., 2024), making data quality of crucial importance. The issues related to data inputs have been mentioned in most ML-base N_r deposition studies but have lacked in-depth discussion. For instance, Li et al. (2020a) constructed the wet NH_4^+ deposition for China with space-time ML models and claimed the number of sites used for model training could greatly affect their model robustness. Zhou et al. (2023) built China's nitrogen deposition with the Random Forest (RF) model and pointed out that insufficient data are available to assess the model's accuracy in remote regions. Zhao et al. (2022) estimated China's bulk N_r deposition by using the Generalized Additive Model (GAM) and claimed that using urban and rural sites for training the model may lacking representation for background regions. According to these studies, regional-scale studies on N_r deposition faces challenges of data scarcity (Tan et al., 2022), which leads to issues in ML development, evaluation, and application, like data insufficiency and

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

dataset shift (i.e., discrepancies between training and testing conditions (Adams, 2009), a common issue in ML-based environmental studies). To our knowledge, no systematic investigation has been conducted to quantify these influences. And lacking such uncertainty assessment may lead to overconfidence in data reliability (Zhu et al., 2025), potentially misinforming critical environmental policies and ecological risk assessments.

To fill this gap, this study presents a quantitative examination on the uncertainties in ML application on studying N_r wet deposition (D_{wet}), with particular focus on input data-induced variations. Figure 1 presents the research framework. We constructed the ML model for predicting regional D_{wet} values for East Asia (EA) and Southeast Asia (SEA) with the extreme gradient boosting (XGBoost) model (section 3.1). We then conducted three sensitivity tests (sections 3.2 and 3.3) to examine the impacts of sample size, spatial distributions, and site types of model inputs on model robustness. We further quantified the influences of these factors on the estimation of regional D_{wet} amounts (section 3.4). Through the findings, we addressed several key questions concerning uncertainty quantification raised in previous research but remain unquantified (section 4.1). Additionally, we delineated some potential sources of uncertainty uncovered in our investigation (section 4.2) and discussed the potential risks of applying ML-derived N_r deposition data and findings for N_r management (section 4.3). While the findings specifically pertain to ML-based studies on N_r deposition, the proposed framework is broadly applicable to ML applications targeting analogous topics, particularly those involving data scarcity challenges and geospatial interpolation tasks.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

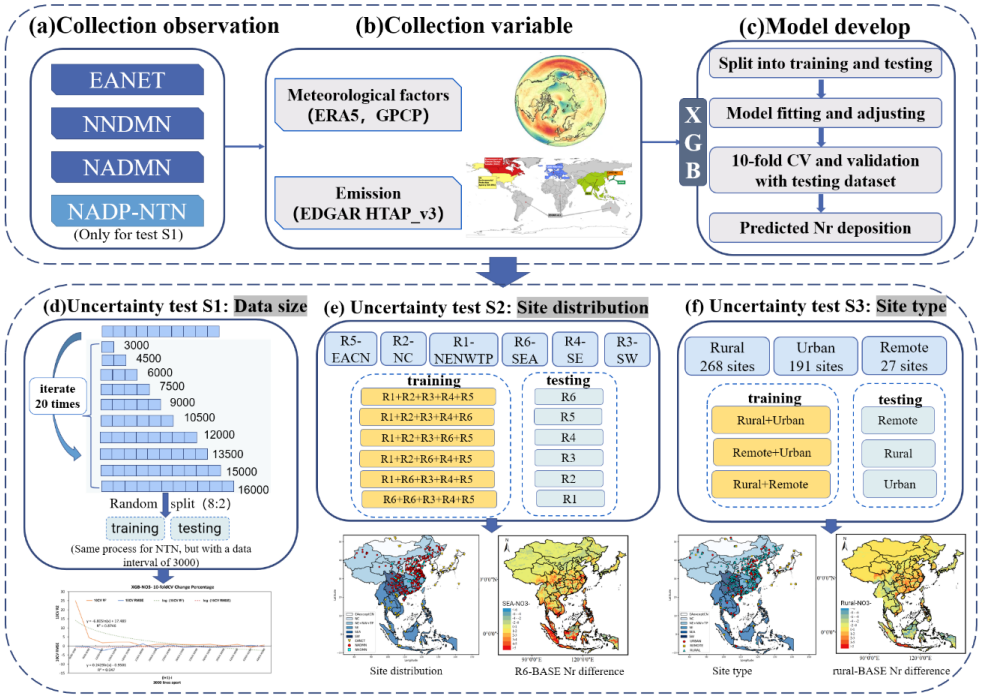


Figure 1. Framework of this study. (a-b) Collection of observational data as targets (a) and features (b) for model development. (c) Process for model development. (d-f) Sensitivity tests. Case S1: impact of sample size on model performance (d), Case S2: influence of spatial distribution of observational sites on model performance and prediction of regional deposition amounts (e), and Case S3: effect of site types (urban, rural, remote) on model robustness and deposition predictions (f).

2. Data and Methods

2.1 Collection and treatment of ground monitoring data

This study utilized ground-based measurements of NO_3^- and NH_4^+ wet deposition from monitoring networks for model development and evaluation, including (1) the National Acid Deposition Monitoring Network (NADMN), which comprises 407 sites of wet-only measurements for the year 2010 (Tan et al., 2022; Li et al., 2019); (2) the Acid Deposition Monitoring Network in East Asia (EANET) network, which provides 57 sites of wet-only measurements for 2000-2018 (link: https://www.eanet.asia/, last access: 2025/10/14); (3) and the Nationwide Nitrogen Deposition Monitoring Network (NNDMN),

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

which provides bulk measurements (wet deposition plus part of dry deposition) at 32 sites for 2010-2015 (Xu et al., 2019). We converted the values to wet-only deposition using the ratio (0.7) following (Yu et al., 2019). In the sensitivity test on influence from sample size, we used the deposition data from the National Trends Network (NTN) of National Atmospheric Deposition Program (NADP) for U.S. (link: https://nadp.slh.wisc.edu/networks/, last access:2025/10/14) as supporting information. It provides wet-only measurements at 207 sites for 2001-2018.

All datasets were subjected to the following quality control procedures: (1) Completeness check: Records with missing data were first excluded to avoid biases from incomplete entries. (2) Unit standardization: Data from multiple sources were uniformly converted to kg N ha^{-1} $month^{-1}$. (3) Outlier detection and removal: Outliers were defined as values exceeding the threshold of "mean ± 3 standard deviations (SD)". These values were screened and excluded to prevent skewing the dataset. The final processed dataset contains up to 16,000 records (Table 1), providing a robust foundation for subsequent model training, validation, and evaluation.

The target variable used in developing the ML model is the concentration of NO_3^- and NH_4^+ ions in precipitation (C_{wet}-NO_3^- and C_{wet}-NH_4^+). We focused on concentration rather than deposition, since the uncertainty in predicting precipitation has been relatively well studied by chemical transport models (Itahashi et al., 2021; Liu et al., 2021; Zhao et al., 2017) and machine learning models (Barrera-Animas et al., 2022; Ding and Peng, 2020; Peng et al., 2017b; Peng et al., 2017a; Peng et al., 2019). The C_{wet} value is calculated as equation (1).

$$C_{wet} = \frac{D_{wet}}{Prep} \quad (1)$$

Where, C_{wet} is the concentration of ions in precipitation (mg N L^{-1}), D_{wet} is the wet deposition flux (kg N ha^{-1} mon^{-1}), and *Prep* is the accumulated precipitation (mm mon^{-1}).

Table 1. Observational datasets used in this study

Datasets	Numbers of sites	Site type	Time coverage	Valid data number	Reference
NADMN	407 sites	Rural: 244, Urban: 163	2010, monthly	4130 for NH4+ 4446 for NO3-	(Tan et al., 2022)
NNDMN	32 sites	Rural: 19, Urban: 6, Remote: 7	2010-2015, monthly	1554 for NH4+ 1436 for NO3-	(Xu et al., 2019)

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

EANET	57 sites	Rural: 15, Urban: 22, Remote: 20	2000-2018, monthly	10577 for NH4+ 10728 for NO3-	https://www.eanet.asia/
NTN	207 sites	/	2001-2018, monthly	48178 for NH4+ 48370 for NO3-	https://nadp.slh.wisc.edu/networks/

2.2 Development of ML model

This study utilized the EXtreme Gradient Boosting (XGBoost) models for predicting C_{wet} values. This model is an efficient ML algorithm based on the gradient boosting framework proposed by (Chen and Guestrin, 2016). It employed innovative techniques including regularized objective functions, weighted quantile sketching, and parallel computing to enhance both predictive accuracy and computational efficiency of decision tree ensembles. The XGBoost model has been widely applied in atmospheric and environmental research, such as atmospheric pollutant concentration forecasting (Chen et al., 2019) and air quality assessment (Tao et al., 2024; Zhang et al., 2022; Li et al., 2020b).

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 (2000-2020) from anthropogenic sources for main air pollutants, including NO_2, $PM_{2.5}$, SO_2, NMVOC, and NH_3; (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 (D2M), 2-meter temperature (T_2M), 10-meter wind speed (WS_{10}), and boundary layer height (BLH); and (3) geographical information, including the latitude and longitude of monitoring sites. All data were spatially aligned to a uniform 0.25° × 0.25° resolution grid to ensure consistency.

It should be noted that some previous studies used satellite data (i.e., vertical column density (VCD)) as features. We therefore compared the models' performance between using emission data and satellite data for 2008-2010. The results showed a minimal difference in model performance between the two (Table S1). However, satellite data could not fully cover our study period (for instance, NH_3 VCD are available from 2008

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

to 2020). Therefore, we used emissions of NO_x and NH_3, which are the main precursors of C_{wet}-NO_3^- and C_{wet}-NH_4^+ as features for model development.

The dataset was partitioned into training and testing subsets following an 8:2 ratio. We adopted the "CVT+Grid Search" pathway for model development (Zhu et al., 2023). In the pretraining-to-validation step, we used a ten-fold cross-validation (10-fold CV) design and grid search method to obtain the optimal hyperparameters. The training dataset was evenly divided into 10 folds, with the models iteratively trained on 9 folds (pretraining) and validated on the remaining fold (validation). The pretraining process is used to perform feature engineering, and the validation dataset is used to tune hyperparameters. This process was repeated 10 times to ensure each data subset was used as validation data once. The average performance is reported as 10-fold CV performance. The optimal hyperparameters used are listed in Table S2. The models with the optimal structure were then evaluated in the independent test on the testing dataset.

2.3 Base case and sensitivity tests

The development of the ML model depends largely on the structure of inputs (Zhu et al., 2023). We concluded two types of dataset shift problems in ML development: (1) sample selection bias, which occurs when the distribution of training data fails to match that of the test scenario. For D_{wet}, over 80% of monitoring data clustered in eastern China (Fig. 2a), where air pollution emissions are high. The insufficient representation of the low emission areas (e.g., western China and EACN regions) could induce severe spatial sampling biases, and (2) imbalanced data bias in multiclass prediction, when data points for some classes are much rarer compared with others. For D_{wet}, the overwhelming number of sites in urban and rural areas (>80%) over background areas (natural vegetation) is highly likely to cause this kind of problem (Fig. 2b and Table 1).

This study examines the models' sensitivity to the above-mentioned dataset shift issues by comparing results from Base and three sensitivity experiments (Cases S1-3). The numbers of data used for model training and testing are listed in Table 2.

(1) In the Base case, we used the whole dataset for model training and testing.
(2) Sensitivity case 1 (Case S1) aims at examining the impacts from sample size. We used different data subsets to train and evaluate the model, with sample sizes increasing from

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

3,000 to 16,000 samples in increments of 1,500. For a rigorous and fair assessment, we employed a random sampling strategy: for each subset, the model was trained and tested 20 times, with a random split of training and testing data applied in each iteration. Results are reported as the mean and variability (range) of model performance across these 20 replicate tests.

(3) Sensitivity case 2 (Case S2) aims at investigating the sample selection bias caused by the unevenness in spatial distributions of measurements. The monitoring sites were divided into 6 categories according to the regions (Fig. 2a), including North China (NC), Northeastern China, Northwestern China and Tibetan Plateau (NENWTP), Southeast China (SE), Southwest China (SW), East Asia except China (EACN) and SEA. In each iteration, we used data of 5 regions for model training and the remaining one for model evaluation.

(4) Sensitivity case 3 (Case S3) is designed to assess the data imbalance bias that stems from the uneven distribution of site types. Urban, rural and remote sites show different characteristics in N_r depositions (e.g., seasonal variations) (Tan et al., 2022). However, the numbers of urban (191) and rural (268) sites largely outweighed that of remote sites (27 sites) (Fig. 2b and Table 1). We hypothesized that the model would exhibit reduced accuracy for remote sites, which represent 'rare events' in the dataset. To examine this, we used data of 2 site types for model training and the remaining for model evaluation.

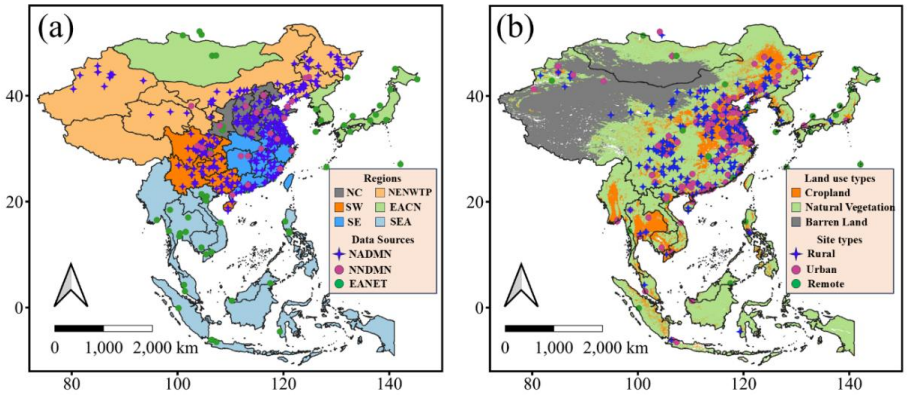


Figure 2. Study domain and distributions of monitoring sites. (a) Domains of six regions and locations of EANET, NADMN and NNDMN monitoring sites. (b) Land use

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

types of research domain and monitoring sites. The land use type came from(Liu et al., 2020).

Table 2. Sample sizes in the Base case and sensitivity tests.

Cases	Scenarios	Training dataset			Testing dataset		
		Data	Sample sizes for C_{wet}-NH_4^+	Sample sizes for C_{wet}-NO_3^-	Data	Sample sizes for C_{wet}-NH_4^+	Sample sizes for C_{wet}-NO_3^-
Base		Random 80%	13009	13288	Remaining 20%	3252	3322
Case S1		Random 80%	Increases from 2400 to 12800		Remaining 20%	Increases from 600 to 3200	
Case S2	SEA	The remaining five regions	11881	12143	SEA	4380	4467
	EACN		11844	12168	EACN	4417	4442
	NENWTP		15121	15389	NENWTP	1140	1221
	NC		14336	14701	NC	1925	1909
	SW		14768	15057	SW	1493	1553
	SE		13355	13592	SE	2906	3018
Case S3	REMOTE	Rural and Urban	12294	12591	Remote	3967	4019
	URBAN	Rural and Remote	10192	10481	Urban	6069	6129
	RURAL	Remote and Urban	10036	10148	Rural	6225	6462

2.4 Metrics for performance evaluation

We employed four metrics for model evaluation: (1) correlation coefficient (R) to validate predictive power, (2) root mean square error (RMSE) for absolute error magnitude, (3) mean absolute error (MAE) for robust absolute deviation analysis, and (4) normalized mean bias (NMB) for systematic bias assessment. The calculations of the metrics are shown in equations (2-5).

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i-Y_i)^2}{N}} \quad (3)$$

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

$$MAE = \frac{\sum_{i=1}^{N}|Y_i - X_i|}{N} \quad (4)$$

$$NMB = \frac{\Sigma_{i=1}^{N}(Y_i - X_i)}{\Sigma_{i=1}^{N} X_i} \quad (5)$$

where $\boldsymbol{X_i}$ and $\boldsymbol{Y_i}$ represent the observation and model prediction, respectively. $\overline{\boldsymbol{X}}$ and $\overline{\boldsymbol{Y}}$ denote the corresponding mean values. N is the number of data pairs.

To quantify the influences of dataset shift issues on model performance, we calculated the Symmetric Mean Absolute Percentage Error (sMAPE) on the testing data (Eq. 6). This metric could eliminate the influence from data scaling, in case the C_{wet} values varied largely among different scenarios. Then, we compared the sMAPE values between sensitivity cases and the Base case (Eq.7). The Deviation index is used to quantify the influences on model performance.

$$sMAPE = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|Y_i - X_i|}{(|Y_i| + |X_i|)/2} \quad (6)$$

$$\text{Deviation } (Si) = \frac{\text{sMAPE(Si)} - \text{sMAPE(BASE)}}{sMAPE(BASE)} \times 100\% \quad (7)$$

Where $\mathbf{S_i}$ denotes the results of sensitivity case i.

3. Results

3.1 ML model performance in Base case

Figure 3 demonstrates the model performance on predicting C_{wet} values. For C_{wet}-NH_4^+, the model achieved a R value of 0.74 and RMSE of 0.74 mg N L^{-1} in the independent test (Fig. 3c). The predictions agree well with the measurements at the range of 0-1 mg N L^{-1}, where 95% of the data are concentrated. However, it tends to underestimate the values larger than 1 mg N L^{-1}, especially after the values reached 4 mg N L^{-1}. This high-value underestimation was also found in the development of models (Fig. 3b). The model achieved similar performances for predicting C_{wet}-NO_3^- (Figs. 3d-f), with a consistent performance between the 10-fold CV and independent test, and a tendency of underestimating high values.

Figure 3g-h show a comparison between this study and previous ML-based studies on N_r deposition (Zhao et al., 2022; Lu et al., 2020; Chen et al., 2023; Zhou et al., 2023; Li et al., 2020a; Rui Li, 2021). To note that the cited studies all used D_{wet} as targets. For a fair

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

comparison, we also tested the model performance when using D_{wet} instead of C_{wet} as target. The model obtained 7-14% higher R values in both 10-fold CV validations and independent tests than those for C_{wet} (Figs. 3g-h, Table S3). This is probably due to the higher correlation between D_{wet} and Prep (R>0.4) than that between C_{wet} and Prep (R = - 0.1) (Fig. S1). Compared with other studies, our models showed similar accuracy (R value) with others (lines in Figs. 3g-h). We also compared the estimations on China's N_r deposition (boxes in Figs. 3g-h). We estimated 7.76 kg N ha^{-1} yr^{-1} and 4.8 kg N ha^{-1} yr^{-1} of D_{wet}-NO_3^- and D_{wet}-NH_4^+ for China (average of years 2000-2020), respectively, fell within the ranges of other studies.

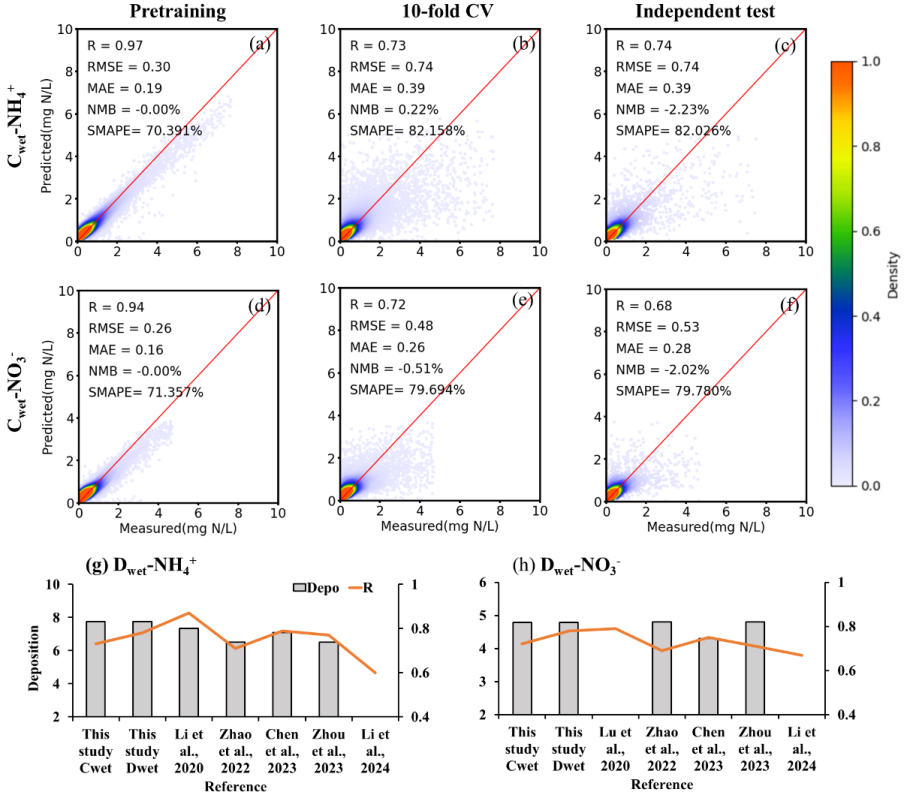


Figure 3. Model performance evaluation on C_{wet}-NO_3^- and C_{wet}-NH_4^+. (a-f) Scatter plots of observed and predicted C_{wet} values in model pretraining, 10-fold CV and independent tests. (g-h) Comparison between this study with other ML-based studies on model performance (R value) and on China's N_r deposition amounts.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

3.2 Influence of sample size on ML performance

The Sample-to-Feature Ratio (SFR), defined as the ratio of training sample size to feature size, serves as a critical indicator for dataset adequacy in ML applications. Simple classification problems typically require SFR⩾10, and complex models demand higher ratios (Zhu et al., 2023). In this section, we test the model sensitivity to a maximum SFR value of over 800 (13,000 samples with 16 features) and a minimum SFR value of 180 (3,000 samples), both exceeding the high-confidence threshold (SFR ⩾ 100) (Zhu et al., 2023). The model performances were assessed based on both model accuracy (indicated by R and RMSE values), and stability and reproducibility (denoted by the variations of R and RMSE in 20 iterations).

Figure 4 shows the model accuracies in predicting C_{wet}-NH_4^+ in the 10-fold CV test when trained with different subsets of data. To avoid bias from data selection, we repeat the evaluation process 20 times with random data selection and present the results with the box plots. Both the R and the RMSE values showed continuous improvement in model accuracies as sample sizes increase (Figs. 4a-b). The R and RMSE values improved by 12% and 9% from 3,000 samples to 16,000 samples, respectively. The ranges of 20 replicate tests also became smaller with increases in sample size, indicating higher model stability.

We plot the changes in R (ΔR, defined as the difference in R between consecutive sample size groups) and RMSE values (ΔRMSE, same definition) for every 1500 increase in the sample sizes (Fig. 4c). The ΔR values decreased almost linearly with increases in sample sizes with a slope of -0.27. We notice a large ΔR occur between 4,500 to 6,000 samples, after which the ΔR values are below 1%. Similar trend was found in ΔRMSE value, with faster changes before reaching 6,000 samples. For C_{wet}-NO_3^- (Fig. 4d and Fig. S2), the relationship between accuracy improvement and sample sizes followed a similar pattern, with 6,000 samples also identified as the “turning point”.

To further investigate how sample size affects the model inputs, we calculated the maximum and mean values of measured C_{wet}-NH_4^+ at each sample size and plot the distributions of 100-time random iterates (Fig. S3). For C_{wet}-NH_4^+, we found that the maximum and mean values in subsets are close to those in whole dataset after getting more than 6,000 samples. For C_{wet}-NO_3^-, the sample size for a stable capture of maximum value

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

is about 9,000 samples. To confirm the generalizability of this finding, we conducted an independent verification using an extensive dataset from NADP, increasing the total sample size to 48,000 (from the original Base case dataset) (Fig. S4). The results show similar shapes in the relationship between model performance and sample sizes. The slopes in ΔR (-0.24 for C_{wet}-NH_4^+; -0.28 for C_{wet}-NO_3^-) were highly consistent to those with base datasets (-0.28 for C_{wet}-NH_4^+; -0.34 for C_{wet}-NO_3^-).

Overall, our results suggest that sample size could significantly affect the ML model accuracy for predicting wet deposition by up to 12%. We found some "turning points" at which the improvements slowed down, despite the continuous improvement in model performance with increasing sample size. This turning point suggests that 6,000 samples are sufficient to balance model performance and data collection cost for C_{wet} prediction. And we suggest this value as the minimum requirement to achieve stable model performance for robust outcomes.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

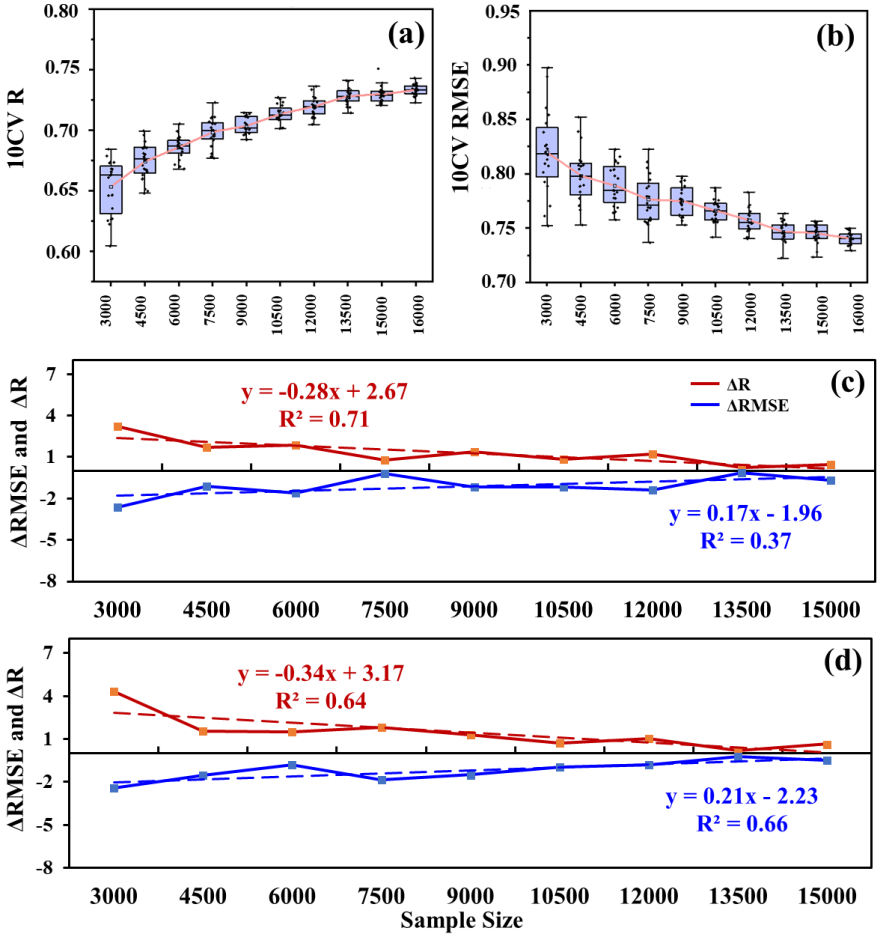


Figure 4. Influence of sample size on ML performance. (a-b) Model performance (R and RMSE) with increasing sample sizes for predicting C_{wet}-NH_4^+. The boxplots represent the results from 20 iterations with random data selections. (c-d) Percentage changes in R (ΔR) and RMSE (ΔRMSE) between consecutive sample size groups for C_{wet}-NH_4^+ (c) and C_{wet}-NO_3^-(d) (unit: %).

3.3 Sensitivity of ML performance to dataset shift in inputs

3.3.1 Influence from spatial distribution of measurements

Figure 5 presents the model performance for C_{wet}-NH_4^+ under six S2 scenarios. The model performed well in Cases S2-SE, SWand SEA (Figs. 5a-b,f). The sMAPE values in the SE, SW and SEA regions were 98%, 97% and 100%, respectively, slightly higher than that in the baseline (82%, model performance using the full dataset in Base case). The

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

model biases were higher in Cases S2-NC and NENWTP (Fig. 5c-d). The sMAPE values exceed 110% for both the NC (113%) and NENWTP (115%) regions, resulting in almost 40% deviations from the baseline in model performance. The ML model failed to catch several peak values, especially in the NC region. As a result, the model predicted trend showed a weak increasing trend (slope = 0.0009) while the observed value revealed a weak decrease (slope = -0.003). The model performed a large overestimation (NMB = 230%) in the EACN region, resulting in the highest sMAPE values (124%) and the highest deviation from baseline (52%) among all regions.

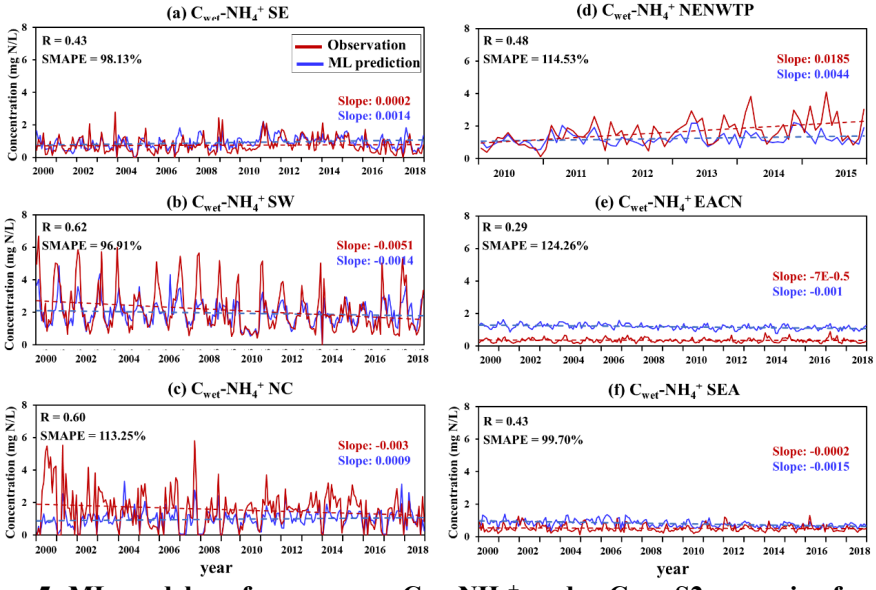


Figure 5. ML model performance on C_{wet}-NH_4^+ under Case S2 scenarios for 2000-2018 (unit: mg N L^{-1}). The evaluation for the NENWTP region was conducted for 2010-2015 due to lack of observational data.

To identify the causes of weakened model performance in certain cases, we analyzed the model performance on monthly and inter-annual variations under six S2 scenarios (Fig. 6). For Case S2-NC, the model underestimated the C_{wet} values by 40-50% during June-August (block in Fig. 6a), when observed values are the highest throughout the year (Fig. 6c). Similarly, the model has largely underestimated the values in 2000 by about 70% (Fig. 6b), when observed values were ~120% higher than the 2000-2018 average (Fig. 6d). For Case S2-EACN, the model overestimated the C_{wet}-NH_4^+ values

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

by >200% in all four seasons. According to the observations in this region, the C_{wet}-NH_4^+ is the lowest among all regions and about 60% lower than domain-average. This bias was caused by insufficient model training of values > 2 mg N L^{-1} and low values < 0.5 mg N L^{-1} (Fig. S5). Similar problem is also found in the model performance on C_{wet}-NO_3^- (Figs. S6-7), leading to 9-51% decreases in model performance.

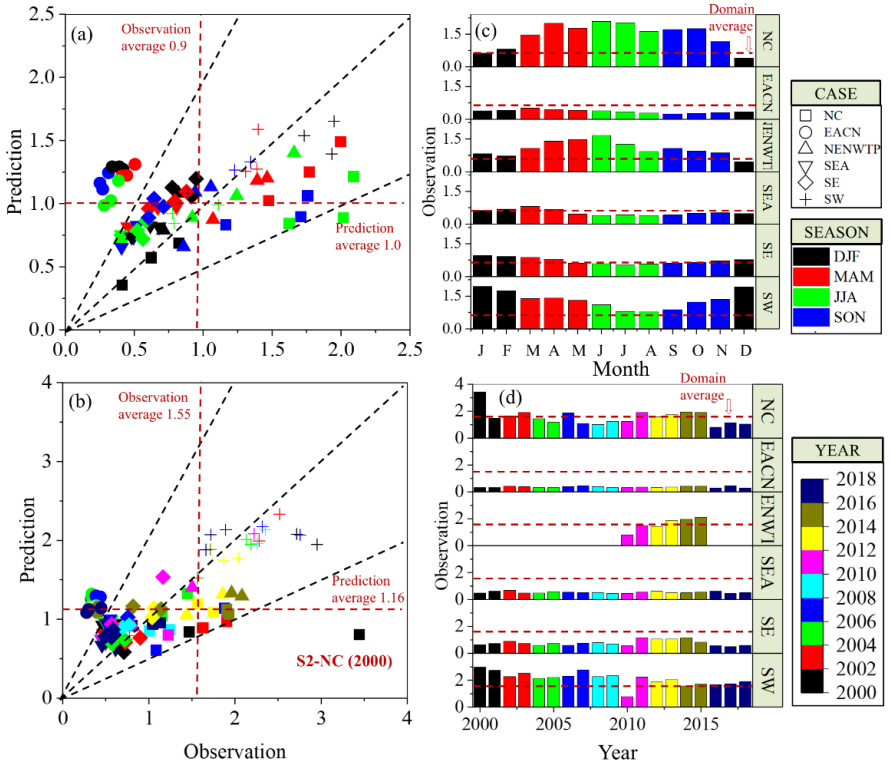


Figure 6. Comparison of ML prediction with observations under Case S2 scenarios for C_{wet}-NH_4^+ (unit: mg N L^{-1}). (a-b) Scatter plots of model performance on seasonal variations (a) and inter-annual variations (b). (c-d) Distribution of observed values in four seasons (c) and during 2000-2018 (d).

Overall, the Case S2 test showed ML's weakness in predicting the high- and low-end ranges of data, resulting in 9-51 % decreases in model accuracy as indicated by sMAPE values. This issue occurs especially when the distribution of training data has largely mismatched with that of the prediction scenario (S2-NC 2018). And the level of this influence depends on the difference between the training and testing environments.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

3.2.3 Influence from imbalanced number of site types

Figure 7 presents the model performance for C_{wet}-NH_4^+ under three S3 scenarios. The model performed best in Case S3-Urban (Fig.7a). The sMAPE values at the urban sites were 89%, slightly higher than that in the baseline (82%). The modelled inter-annual trend (slope = 0.0003) is also close to that of observation (0.0013). In Case S3-Rural, the model predictions were close to observations in all year except 2011-2015 (Figs. 7b, 7e). The good performance is due to a similar data range and distribution between rural and urban sites (Fig. 8a), which provide sufficient data for model training. The bias during 2011-2015 stemmed from the fact that observed rural C_{wet} values from NNDMN (available in 2010-2015) were much higher than those from EANET (Fig. S8).

In Case S3-Remote, the model largely overestimated the values by 109% (NMB) and 115% (sMAPE), indicating a 40% Deviation from the Base case in model accuracy. The observations at remote sites (averagely 0.37 mg N L^{-1}) are only ~40% of those at rural (0.88 mg N L^{-1}) and urban (0.81 mg N L^{-1}) sites. The model failed to predict the values at remote sites by giving a very similar estimation to urban and rural sites (Fig. 8a), which is a sample selection issue, as discussed in section 3.3.1. We also found that the R value (0.36) in this case decreased sharply from the baseline (0.74), Case S3-urban (0.65) and Case S3-rural (0.62). To explore the reason, we plotted the observed and predicted relationship between NH_3 emission and C_{wet}-NH_4^+ at three site types (Fig. 8b). The rural and urban sites showed a marginal change in C_{wet}-NH_4^+ as the N emission rises (slope=0.01) and the model has well reproduced this relationship. While remote sites showed a clear increasing trend of C_{wet}-NH_4^+ with emission (slope=1.23), but the model predicted a contrast and weak relationship (slope=-0.7). In this case, the ML model encountered not only a shift in data range, but also a potential change in the relationship between features (NH_3 emission) and target (C_{wet}-NH_4^+).

For C_{wet}-NO_3^-, the deviation from baseline is much smaller (9-23%). The relationship between C_{wet} and emission is well reproduced in Case S3-Remote for NO_3^-, although the C_{wet} values are slightly overestimated (Figs. 8c-d and Fig. S9). The emission estimation of NO_x is believed to be more robustness that of NH_3, due to the high uncertainty in industrial and agricultural NH_3 emissions (Liu et al., 2025; Crippa et al., 2023). In addition, the relationship between observed D_{wet}-NO_3^- and NO_x emission are generally simpler than that

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

between D_{wet}-NH_4^+ and NH_3 emission (e.g., linear in studies in CONUS US (Tan et al., 2020) and mainland China (Liu et al., 2022)).

Overall, Case S3 tests showed a high risk of severe (>100%) overestimation over remote areas when the number of remote sites is much less than the other two in model training. The bias is attributed to dataset shift in data distribution and feature-target relationship. The combined effect is 9-40% decreases in overall model accuracy according to sMAPE values.

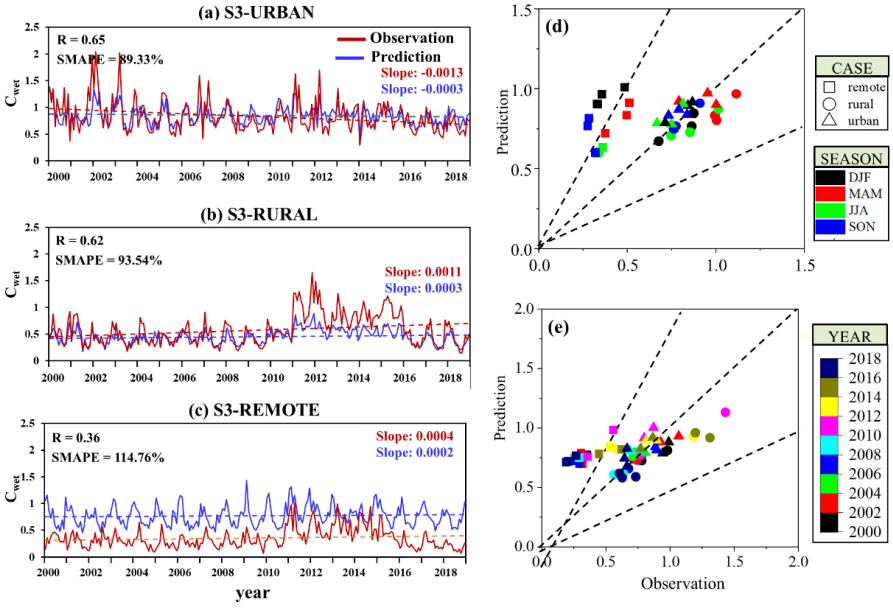


Figure 7. ML model performance on C_{wet}-NH_4^+ under Case S3 scenarios for 2000-2018 (unit: mg N L^{-1}). (a-c) Comparison between model prediction and observations. (d-e) Scatter plots of model performance on seasonal variations (d) and inter-annual variations (e).

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

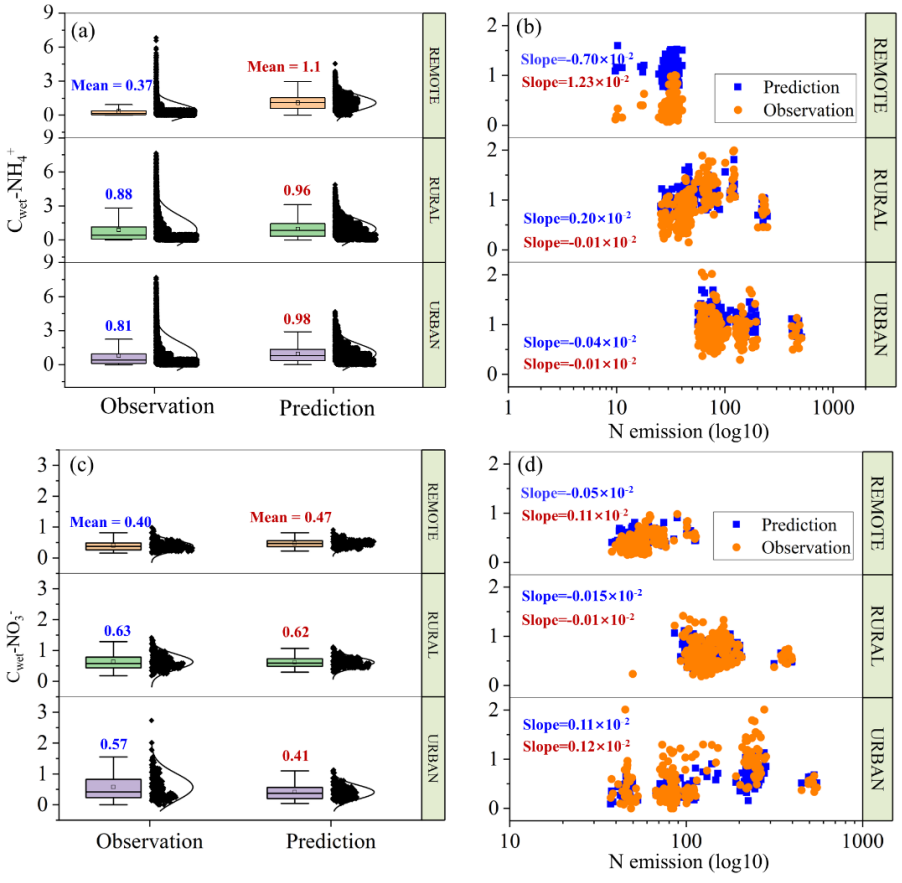


Figure 8 Comparison of ML predictions with observations under Case S3 scenarios for C_{wet}-NH_4^+ (a-b) and C_{wet}-NO_3^- (c-d) (unit: mg N L^{-1}). Comparison of observation and prediction in data distribution (a and c) and relationship between N_r emission and C_{wet} values (b and d). The values are monthly average values during 2000-2018. The N emission is the sum of NO_x and NH_3 emissions derived from HTAPv3 (unit: tons N mon^{-1}).

3.4 Impact on estimations of nitrogen deposition amounts

Figures 9-10 show the influences of each sensitivity scenario on the estimations during 2000-2020 for D_{wet}-NH_4^+. The average D_{wet} estimated in the Base case was about 9.0 kg N ha^{-1} yr^{-1} and the uncertainty (presented by one standard deviation of the difference between sensitivity scenarios and Base case) was about 0.9 kg N ha^{-1} yr^{-1}, equivalent to 10% of estimations in the Base case (Figs. 9a-b). The level of influence increased from north to

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

south, with the highest uncertainty found in the SEA region (4-7 kg N ha^{-1} yr^{-1}), followed by southern China (2-4 kg N ha^{-1} yr^{-1}) and northern China (<1 kg N ha^{-1} yr^{-1}).

The influences from site distribution (Case S2) ranged from -5% to 22%, causing -0.5 to 2.5 kg N ha^{-1} yr^{-1} deviation from the baseline (Fig. 10a). Case S2-SEA contributed the largest bias, causing an average 21% overestimation on D_{wet}-NH_4^+. The impacts are widely found in western China, northern EACN and whole SEA region (Fig. 9c), where the level of impacts reached 50%. The impact of Case S2-EACN is relatively small (0-2%) on domain-average. But the influence is significant in western China (>50%), EACN (20-50%) and SEA (>20%) regions (Fig. 9d).

The influence from imbalanced site type (Case S3) ranged from -14% to 10%, leading to -1.5 to 1 kg N ha^{-1} yr^{-1} deviation from the baseline (Fig. 10b). The absence of rural sites in model training contributed to a 12-14% underestimation in domain-average NH_4^+ wet deposition. The impact is found in large areas of western China and SEA region, especially in Malay Archipelago (>20%) (Fig. 9e). Although the model biases were highest in Case S3-Remote (Fig. 7), the depositions at rural sites were much higher than remote sites and therefore the impact on domain-average deposition is higher. The contribution of Case S3-Remote was about a 6-8% overestimation on domain-average. However, the impacts could reach >50% on high sensitivity areas in western China and SEA regions (Fig. 9f).

For D_{wet}-NO_3^-, the uncertainties ranged 0.54-0.67 kg N ha^{-1} yr^{-1}, equivalent to 9-11% of base estimation (Figs. 10c-d and Figs. S10-11). The influences from Case S2 (site spatial distribution) ranged from -13% to 19%, causing -0.9 to 1.3 kg N ha^{-1} yr^{-1} deviation from the baseline, while the influence regarding Case S3 (imbalanced site types) ranged from -8% to 8%, leading to -0.5 to 0.5 kg N ha^{-1} yr^{-1} deviation. Compared to the condition in D_{wet}-NH_4^+, there is a considerable increase in the impacts from Case S3-NC, mainly impact on the prediction for south part of Indonesia (Fig. S10h).

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

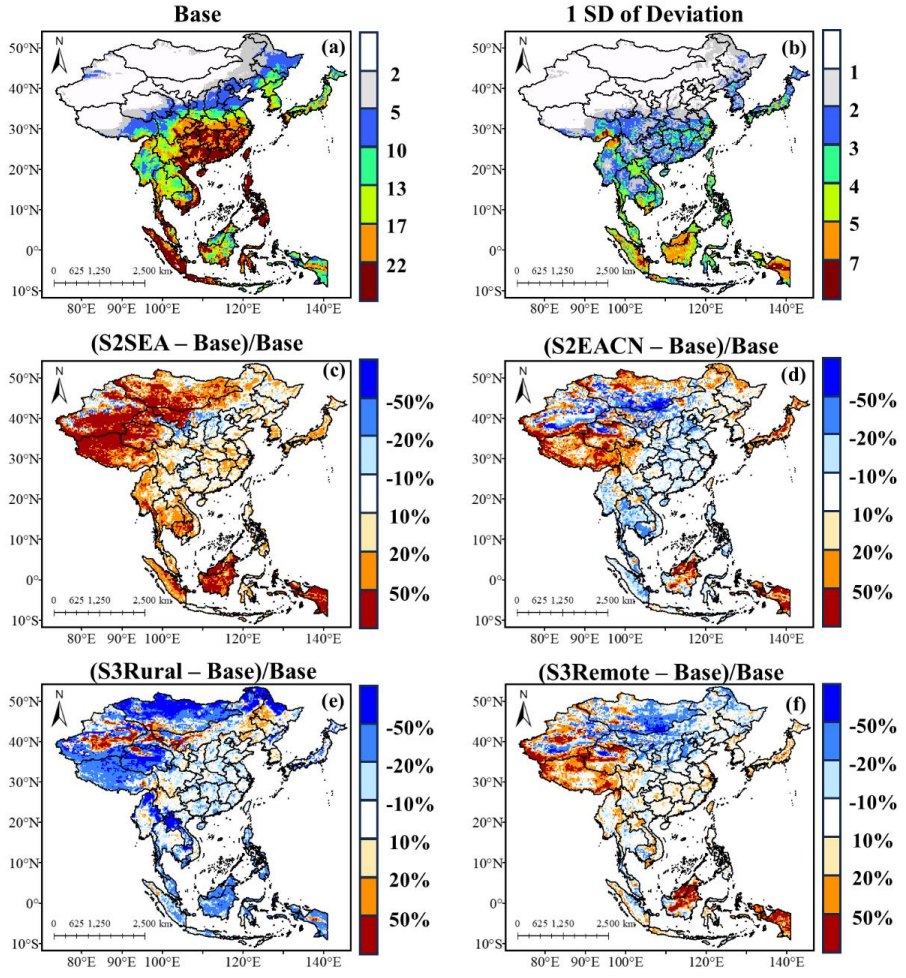


Figure 9. Estimated depositions during 2000-2020 and deviations in sensitivity scenarios for D_{wet}-NH_4^+. (a) Estimated spatial distribution of D_{wet}-NH_4^+ (unit: kg N ha^{-1} yr^{-1}), (b) One standard deviation of the difference between sensitivity scenarios and Base case (unit: kg N ha^{-1} yr^{-1}), (c-f) Percentage differences between sensitivity cases and Base case for Cases S2-EACN, S2-SEA, S3-Rural and S3-Remote (unit: %).

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

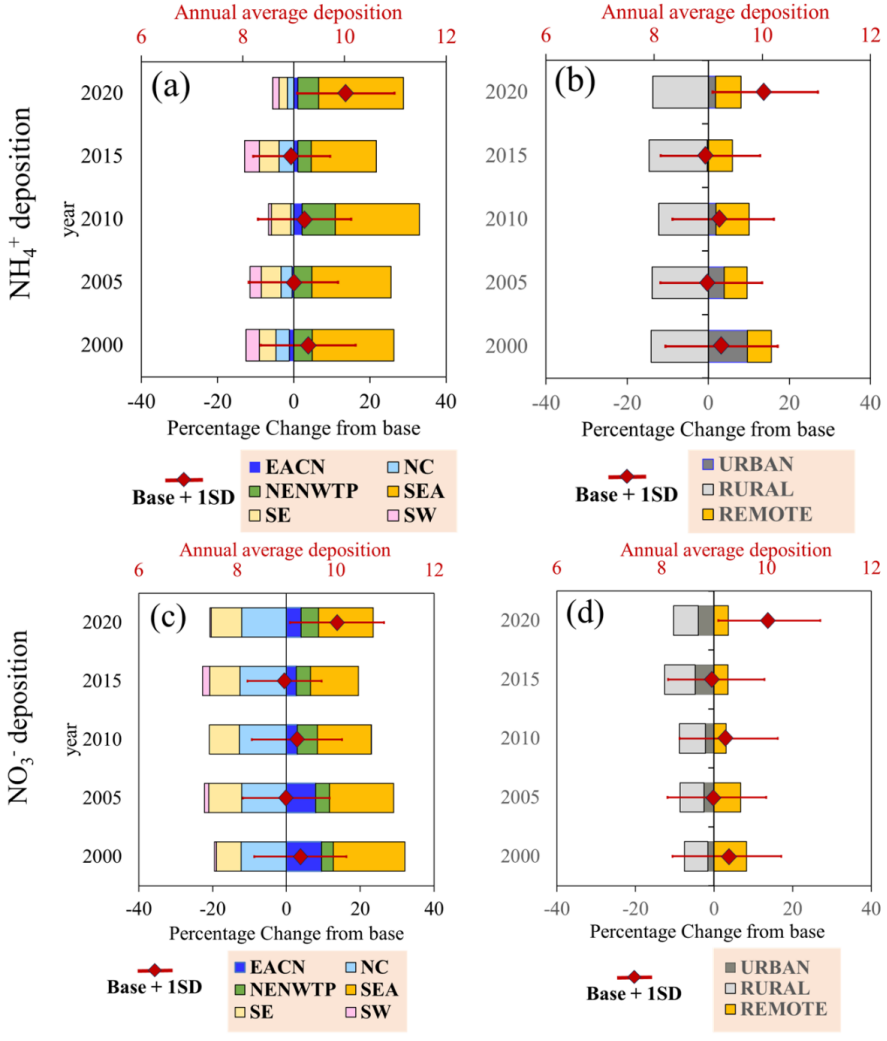


Figure10. Estimated depositions during 2000-2020 and deviations in sensitivity scenarios for D_{wet}-NH_4^+ and D_{wet}-NO_3^-. The red lines indicate the N_r deposition estimated in the Base case and uncertainties (presented by one standard deviation of the difference between sensitivity scenarios and Base case) (unit: kg N ha^{-1} yr^{-1}). The bars present the percentage differences in deposition amount estimated in sensitivity cases compared to the Base case (unit: %).

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

4. Discussion and conclusion

4.1 Key findings and contributions to previous studies

In the process of ML, various sources of uncertainty can arise, including inherent noise within the data, ambiguity or variance in model parameters, appropriateness of model selection (e.g., model structure), and uncertainty introduced by extrapolation (Jalaian et al., 2019). This study focused on quantifying the impacts from input dataset on ML performance and outcome data reliability, covering the following questions mentioned in previous studies yet not quantified.

(1) **Data sufficiency**. Li et al. (2020a) proposed that the size of data can affect the robustness of models, whereas this study further quantified the critical sample size threshold (6,000-9,000 samples). In our Case S1 (section 3.2), we found a sharp improvement in model performance as sample size increased to 6,000–9,000, with a maximum 12% accuracy loss when sample size was below this threshold. Once the data size exceeds 9,000, collecting additional data may not significantly benefit the application on N_r deposition.

(2) **Uneven spatial distribution**. Zhu et al. (2025) and Ge et al., (2024) argued that the uneven distribution of observation points can introduce bias into ML models, particularly in areas with few or no site observations. In our Case S2 (section 3.3.1), we found 9-51% deviations from baseline ML performance due to model's weakness in predicting the high- and low- end ranges of data. In the extreme case, when building the ML model with China's datasets (high deposition values), the model would predict a >100% overestimation at EACN sites (e.g., Japan, Korea) and 50% overestimation at SEA sites. The overall influence on ML-based N_r deposition ranged from -13% to 22% on domain-average and >50% on data-scarce regions (western China, EACN and SEA).

(3) **Site type imbalance**. Zhao et al. (2022) and Zhou et al. (2023) mentioned that model evaluations on remote site type were inadequate, thus training models may not fully align with the deposition conditions observed at remote areas. In the S3-Remote scenario, we found a high risk of severe (>100%) overestimation over remote areas when training models with data from urban and rural sites. The total deviation from baseline performance reached 9-40%, causing -14% to 10% uncertainty in N_r deposition estimates on domain-

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

482 average. In extreme cases, the impact of lacking information from remote sites is
483 considerable high (>50%) on remote areas in western China and SEA regions.

484 Our finding points to a higher SFR requirement than what is suggested by Zhu et al.
485 (2023). According to the number of available data from each data source (Table 1), it is
486 very hard to reach this threshold by using single-source dataset. And most of the previous
487 studied choose to use a combination of multiple data sources without data harmonization
488 process (Zhao et al., 2022; Lu et al., 2020; Chen et al., 2023; Zhou et al., 2023; Li et al.,
489 2020a; Rui Li, 2021). However, different datasets were found to be inconsistent in many
490 aspects regarding data sampling and quality (Tan et al., 2022), and the lack of standards
491 and harmonization is commonly found in monitoring data (Hassani et al., 2025). We found
492 several aspects of inconsistency in the datasets that weakens the model performance.

493 (1) Inconsistency among different data sources. Significant systematic discrepancies
494 exist between different monitoring networks (Tan et al., 2022). Different data sources
495 showed inconsistency in the data distributions (i.e., data range and monthly variations)
496 (Figs. S12-13). In addition, the regional representativeness of sites may strongly affect our
497 interpretation. Site density varies markedly across regions, being highest in the SE and NC
498 regions (approximately 1 site per million km²) and lowest in the EACN and SEA regions
499 (0.1 sites per million km²) (Table S4). These findings highlight the urgency of data
500 harmonization for multi-source dataset integration, which was rarely emphasized in
501 previous ML-based N_r deposition studies.

502 (2) A shift in feature-target relationship. The ML tried to develop a relationship
503 between emission and meteorology (as feature) and C_{wet} (as target). During this process,
504 point-to-point relationship between emission and deposition is very important, as
505 emissions are the major precursors. Using the HTAP emission inventory as inputs, which
506 only includes anthropogenic emissions (including agriculture), was a potential reason for
507 failure to build the source-sink relationship at remote sites. To explore this, we replaced the
508 NH_3 emission with satellite NH_3 VCD in Case S3-Remote for C_{wet}-NH_4^+. The model
509 performance improved with R value increased from 0.36 to 0.58 and sMAPE value
510 decreased from 114.8% to 104.5%. The predicted relationship between N_r emission and
511 C_{wet}-NH_4^+ (slope=0.39) was also in better agreement with that observed (slope=1.23) (Fig.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

S14). However, the availability of satellite data (e.g., NH_3 from Infrared Atmospheric Sounding Interferometer (IASI) satellite is only available during 2007-2021 and contains missing values) imposes constraints on its study time and precludes the capability to generate predictions for future conditions.

4.2 Other potential sources of uncertainties

This study cannot provide a systematic evaluation of the uncertainties in ML-based N_r deposition studies since this issue is far too extensive. Besides the above-mentioned issues, we listed some potential sources of uncertainty for further studies.

(1) Impact from input data quality, such as high missing rates in satellite-derived NH_4^+ data (Shang et al., 2024) and the smoothing effect of meteorological data at 0.25° resolution on local-scale phenomena (e.g. mountain sites (Muñoz-Sabater et al., 2021)), along with the inherent challenge of global-scale inventory data in quantifying small-region variations (Ma et al., 2025) (e.g. employing global emissions inventory to quantify regional surface NH_3 concentrations). In addition, data treatment such as applying a fixed wet/dry deposition conversion factor (e.g., 0.7) may underestimate actual deviations in arid regions[1].

(2) Uncertainty from model selection. The performance of different fitting algorithms is context-dependent (Di et al., 2020). Model structural deficiencies, such as the inability of conventional machine learning approaches to adequately capture nonlinear chemistry-meteorology coupling could also introduce uncertainty (Peng et al., 2024). It is suggested to integrate higher-resolution datasets with physics-constrained modeling frameworks to mitigate some of the dataset shift issues. For instance, precipitation, regional transport and nonlinearity in aerosol formation are the major driver of source-sink relationship for China (Zhao et al., 2022). The XGBoost models were developed based on a grid-to-grid relationship between features and targets and thus couldn't account for regional transport processes. In addition, integrating results from chemical transport models, which contains the secondary formation mechanisms of atmospheric species, could improve accuracy on aerosol concentration (Fu et al., 2022).

This study did not quantify the uncertainty in dry deposition. The amount of dry deposition measurements (only available from the EANET network) is insufficient to conduct such study. In addition, wet deposition fluxes derived from measurements of

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

chemical concentrations in rainwater and precipitation depth are considered more accurate than those for dry deposition fluxes, since dry deposition detection contains substantial uncertainty (Fu et al., 2022).

4.3 Implication for applying ML-derived data for N_r management

ML-derived N_r deposition data has become a valuable tool for regional-scale N_r management, but its application need to be guided by a clear understanding of the uncertainty boundaries. The acceptability of its uncertainty (9–50% in key scenarios) depends on the specific management objectives, and attention is required for the following three core application areas.

First, studies on terrestrial ecosystem protection and carbon-nitrogen cycle research. Accurate N_r wet deposition estimation in remote areas is demanding and non-negotiable, even though monitoring sites are concentrated in urban/rural regions. N_r deposition directly regulates plant productivity, soil carbon fixation, and soil greenhouse gas (GHG) fluxes, which are critical for evaluating ecosystem resilience (Sun and Fernie, 2023; Feng et al., 2023; Liao et al., 2024; Lu et al., 2021; Deng et al., 2020; Chen et al., 2020). This is particularly urgent for threshold-based assessments such as critical loads (CLs) of N_r for sensitive ecosystems (e.g., tropical forests, wetlands) (Zhao et al., 2017; Watmough, 2024; Pavlovic et al., 2023). The remote areas faced >20% variability in D_{wet} estimations due to lacking sufficient information from remote sites (Case S3-Remote) and a high risk of >50% overestimation caused by data scarcity (Case S2-EACN and SEA). These uncertainties could distort the quantification of N_r deposition impacts on terrestrial carbon sequestration (Shang et al., 2024; Lu et al., 2021) and ultimately affect the reliability of environmental protection policies guided by such studies (e.g., policies for acid deposition control in China (Xie et al., 2024)).

Secondly, evaluation on N_r emission control efficiency. Emission reduction benefits are typically calculated via the relative changes between N_r emissions and deposition (Zhao et al., 2022). Therefore, deviations in the baseline deposition amounts (up to 50%) could largely distribute the effectiveness assessment of emission cuts. Additionally, long-term future projection requires extra caution. Significant changes in predictive variables (e.g., extreme weather altering precipitation under climate change may invalidate the established

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

feature-target relationships of ML models, leading to unreliable trend predictions (Hoesung Lee 2024)). N_r deposition also serves as an "additional nitrogen input" for agricultural ecosystems. About 34% of N_r deposition is utilized by crops in China, resulting in 25% of global agricultural soil N_2O emissions (Zhang et al., 2024; Yang et al., 2021). Even small biases in deposition estimation could distort nitrogen use efficiency (NUE) calculations and misguide fertilizer reduction strategies.

To maximize the reliability of ML-derived data in N_r management, we propose four actionable suggestions:

(1) Supplement monitoring in high-uncertainty areas: Prioritize short-term intensive monitoring (or long-term site establishment) in regions with high uncertainty (e.g., western China, SEA remote areas) to improve the representativeness of training data and reduce spatial sampling bias.

(2) Standardize cross-regional data collaboration: To address the inconsistencies between datasets, it is suggested to establish a transnational/transregional collaborative platform to coordinate N_r deposition monitoring activities, integrate multi-source data (e.g., NADMN, EANET), and implement unified quality control (QC) and data harmonization processes.

(3) Highlight uncertainty in product outputs: Explicitly report of uncertainty ranges and sensitive areas (e.g., on remote sites) when releasing ML-derived N_r products, to avoid blind application.

(4) Combine ML with physics-constrained frameworks: For critical management objectives (e.g., CL assessments, long-term projection), integrate ML models with physics-based frameworks (i.e., chemical transport models) to constrain nonlinear processes (i.e., regional transport, aerosol secondary formation) that ML alone cannot capture.

Data and code availability

The current version of XGboost code is available from GitHub website https://github.com/dmlc/xgboost (last access: 15 December 2025) under an Apache-2 license. The raw data used in this study are publicly available. The the NADMN dataset can be downloaded from (Li et al., 2019). The EANET dataset can be downloaded from

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

https://www.eanet.asia/ (last access: 2025/10/14). The NNDMN dataset can be downloaded from (Xu et al., 2019). The NTN can be downloaded from NADP's website (https://nadp.slh.wisc.edu/networks/, last access:2025/10/14). The HTAP v3 emission inventory can be downloaded from https://edgar.jrc.ec.europa.eu/dataset_htap_v3 (last access: 2025/10/14). The ERA5-Land reanalysis data can be downloaded from https://cds.climate.copernicus.eu/ (last access: 2025/10/14).

The exact codes and the corresponding datasets are archived on Zenodo under (https://zenodo.org/records/18092206, Tan et al., 2025). We used the ArcMap v10.801 to generate the spatial distributions of nitrogen deposition (Fig. 2 and Fig. 9). We used the Origin 2021 to generate the scatter plots and box plots (Fig. 6, Fig. 7(d-e), Fig. 8). This study does not produce new data set.

Author contribution

YZ: data curation, formal analysis, and writing (original draft); JNT: conceptualization, formal analysis, funding acquisition, methodology, and writing (review and editing); QM and JSF: writing (review and editing); LL: conceptualization, supervision and writing (review and editing).

Competing interests:

The authors declare that they have no conflict of interest.

Acknowledgements

We acknowledge the support by Shanghai Technical Service Center of Science and Engineering Computing, Shanghai University.

Financial Support

This work is supported the National Key Research and Development Program of China (No. 2023YFC3708603, No. 2023YFC3708601) and Shanghai Pujiang Program (No.23PJ1403100). We acknowledge the support by Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Ministry of Ecology and Environment (No. 2025080164) and Jing-Jin-Ji Regional Integrated Environmental Improvement - National Science and Technology Major Project (2025ZD1202005, 2025ZD1202004).

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

References

Adams, N.: Dataset Shift in Machine Learning, Journal of the Royal Statistical Society Series A: Statistics in Society, 173, 274-274, 10.1111/j.1467-985X.2009.00624_10.x, 2009.

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., and Akanbi, L. A.: Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting, Machine Learning with Applications, 7, 10.1016/j.mlwa.2021.100204, 2022.

Chen, M., Chang, L., Zhang, J., Guo, F., Vymazal, J., He, Q., and Chen, Y.: Global nitrogen input on wetland ecosystem: The driving mechanism of soil labile carbon and nitrogen on greenhouse gas emissions, Environmental Science and Ecotechnology, 4, 10.1016/j.ese.2020.100063, 2020.

Chen, S., Chen, B., Wang, S., Sun, L., Shi, H., Liu, Z., Wang, Q., Li, H., Zhu, T., Li, D., Xia, Y., Zhao, Z., Wang, L., and Wang, L.: Spatiotemporal variations of atmospheric nitrogen deposition in China during 2008–2020, Atmospheric Environment, 315, 10.1016/j.atmosenv.2023.120120, 2023.

Chen, T. and Guestrin, C.: XGBoost： A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 10.1145/2939672.2939785, 2016.

Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Yang, J., Chen, P.-Y., Ou, C.-Q., and Guo, Y.: Extreme gradient boosting model to estimate PM2.5 concentrations with missing-filled satellite data in China, Atmospheric Environment, 202, 180-189, 10.1016/j.atmosenv.2019.01.027, 2019.

Chua, M., Kim, D., Choi, J., Lee, N. G., Deshpande, V., Schwab, J., Lev, M. H., Gonzalez, R. G., Gee, M. S., and Do, S.: Tackling prediction uncertainty in machine learning for healthcare, Nature Biomedical Engineering, 7, 711-718, 10.1038/s41551-022-00988-x, 2023.

Crippa, M., Guizzardi, D., Butler, T., Keating, T., Wu, R., Kaminski, J., Kuenen, J., Kurokawa, J., Chatani, S., Morikawa, T., Pouliot, G., Racine, J., Moran, M. D., Klimont, Z., Manseau, P. M., Mashayekhi, R., Henderson, B. H., Smith, S. J., Suchyta, H., Muntean, M., Solazzo, E., Banja, M., Schaaf, E., Pagani, F., Woo, J.-H., Kim, J., Monforti-Ferrario, F., Pisoni, E., Zhang, J., Niemi, D., Sassi, M., Ansari, T., and Foley, K.: The HTAP_v3 emission mosaic: merging regional and global monthly emissions (2000–2018) to support air quality modelling and policies, Earth System Science Data, 15, 2667-2694, 10.5194/essd-15-2667-2023, 2023.

Deng, L., Huang, C., Kim, D. G., Shangguan, Z., Wang, K., Song, X., and Peng, C.: Soil GHG fluxes are altered by N deposition: New data indicate lower N stimulation of the N2O flux and greater stimulation of the calculated C pools, Global Change Biology, 26, 2613-2629, 10.1111/gcb.14970, 2020.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L. J., and Schwartz, J.: Assessing NO(2) Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging, Environ Sci Technol, 54, 1372-1384, 10.1021/acs.est.9b03358, 2020.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

674 Ding, Y. and Peng, S.: Spatiotemporal Trends and Attribution of Drought across China
675 from 1901–2100, Sustainability, 12, 10.3390/su12020477, 2020.
676 Feng, H., Guo, J., Peng, C., Kneeshaw, D., Roberge, G., Pan, C., Ma, X., Zhou, D., and
677 Wang, W.: Nitrogen addition promotes terrestrial plants to allocate more biomass to
678 aboveground organs: A global meta-analysis, Global Change Biology, 29, 3970-3989,
679 10.1111/gcb.16731, 2023.
680 Fu, J. S., Carmichael, G. R., Dentener, F., Aas, W., Andersson, C., Barrie, L. A., Cole, A.,
681 Galy-Lacaux, C., Geddes, J., Itahashi, S., Kanakidou, M., Labrador, L., Paulot, F.,
682 Schwede, D., Tan, J., and Vet, R.: Improving Estimates of Sulfur, Nitrogen, and Ozone
683 Total Deposition through Multi-Model and Measurement-Model Fusion Approaches,
684 Environmental Science & Technology, 56, 2134-2142, 10.1021/acs.est.1c05929, 2022.
685 Ge, X., Xie, D., Mulder, J., and Duan, L.: Reevaluating the Drivers of Fertilizer-Induced
686 N2O Emission: Insights from Interpretable Machine Learning, Environmental Science
687 & Technology, 58, 15672-15680, 10.1021/acs.est.4c04574, 2024.
688 Guo, Z., Huang, L., Yan, J., Zhang, H., Jia, X., Li, M., and Li, H.: Artificial intelligence
689 technology in environmental research and health: Development and prospects,
690 Environment International, 203, 10.1016/j.envint.2025.109788, 2025.
691 Hassani, A., Salamalikis, V., Schneider, P., Stebel, K., and Castell, N.: A scalable
692 framework for harmonizing, standardization, and correcting crowd-sourced low-cost
693 sensor PM2.5 data across Europe, Journal of Environmental Management, 380,
694 10.1016/j.jenvman.2025.125100, 2025.
695 Heid, E., McGill, C. J., Vermeire, F. H., and Green, W. H.: Characterizing Uncertainty in
696 Machine Learning for Chemistry, Journal of Chemical Information and Modeling, 63,
697 4012-4029, 10.1021/acs.jcim.3c00373, 2023.
698 Hoesung Lee , K. C., Dipak Dasgupta , Gerhard Krinner , Aditi Mukherji , Peter Thorne ,
699 Christopher Trisos , José Romero , Paulina Aldunce , Ko Barrett , Gabriel Blanco ,
700 William W. L. Cheung, Sarah L. Connors , Fatima Denton , Aïda Diongue-Niang ,
701 David Dodman , Matthias Garschagen , Oliver Geden , Bronwyn Hayward ,
702 Christopher Jones , Frank Jotzo , Thelma Krug , Rodel Lasco , June-Yi Lee , Valérie
703 Masson-Delmotte , Malte Meinshausen , Katja Mintenbeck , Abdalah Mokssit ,
704 Friederike E. L. Otto , Minal Pathak , Anna Pirani , Elvira Poloczanska , Hans-Otto
705 Pörtner , Aromar Revi , Debra C. Roberts , Joyashree Roy , Alex C. Ruane , Jim Skea ,
706 Priyadarshi R. Shukla , Raphael Slade , Aimée Slangen , Youba Sokona , Anna A.
707 Sörensson , Melinda Tignor , Detlef van Vuuren , Yi-Ming Wei , Harald Winkler ,
708 Panmao Zhai , Zinta Zommers Climate Change 2023: Synthesis report summary for
709 policymakers, Intergovernmental Panel on Climate Change, 10.59327/ipcc/ar6-
710 9789291691647, 2024.
711 Itahashi, S., Ge, B., Sato, K., Wang, Z., Kurokawa, J., Tan, J., Huang, K., Fu, J. S., Wang,
712 X., Yamaji, K., Nagashima, T., Li, J., Kajino, M., Carmichael, G. R., and Wang, Z.:
713 Insights into seasonal variation of wet deposition over southeast Asia via precipitation
714 adjustment from the findings of MICS-Asia III, Atmospheric Chemistry and Physics,
715 21, 8709-8734, 10.5194/acp-21-8709-2021, 2021.
716 Jalaian, B., Lee, M., and Russell, S.: Uncertain Context: Uncertainty Quantification in
717 Machine Learning, AI Magazine, 40, 40-49, 10.1609/aimag.v40i4.4812, 2019.
718 Li, R., Cui, L., Fu, H., Zhao, Y., Zhou, W., and Chen, J.: Satellite-Based Estimates of Wet
719 Ammonium (NH4-N) Deposition Fluxes Across China during 2011–2016 Using a

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

Space–Time Ensemble Model, Environmental Science & Technology, 54, 13419-13428, 10.1021/acs.est.0c03547, 2020a.
Li, R., Cui, L., Hongbo, F., Li, J., Zhao, Y., and Chen, J.: Satellite-based estimation of full-coverage ozone (O3) concentration and health effect assessment across Hainan Island, Journal of Cleaner Production, 244, 10.1016/j.jclepro.2019.118773, 2020b.
Li, R., Cui, L., Zhao, Y., Zhang, Z., Sun, T., Li, J., Zhou, W., Meng, Y., Huang, K., and Fu, H.: Wet deposition of inorganic ions in 320 cities across China: spatio-temporal variation, source apportionment, and dominant factors, Atmos. Chem. Phys., 19, 11043-11070, 10.5194/acp-19-11043-2019, 2019.
Liao, L., Wang, J., Dijkstra, F. A., Lei, S., Zhang, L., Wang, X., Liu, G., and Zhang, C.: Nitrogen enrichment stimulates rhizosphere multi-element cycling genes via mediating plant biomass and root exudates, Soil Biology and Biochemistry, 190, 10.1016/j.soilbio.2023.109306, 2024.
Liu, H., Gong, P., Wang, J., Clinton, N., Bai, Y., and Liang, S.: Annual dynamics of global land cover and its long-term changes from 1982 to 2015, Earth System Science Data, 12, 1217-1243, 10.5194/essd-12-1217-2020, 2020.
Liu, L., Yang, Y., Xi, R., Zhang, X., Xu, W., Liu, X., Li, Y., Liu, P., and Wang, Z.: Global Wet-Reduced Nitrogen Deposition Derived From Combining Satellite Measurements With Output From a Chemistry Transport Model, Journal of Geophysical Research: Atmospheres, 126, 10.1029/2020jd033977, 2021.
Liu, M., Shang, F., Lu, X., Huang, X., Song, Y., Liu, B., Zhang, Q., Liu, X., Cao, J., Xu, T., Wang, T., Xu, Z., Xu, W., Liao, W., Kang, L., Cai, X., Zhang, H., Dai, Y., and Zhu, T.: Unexpected response of nitrogen deposition to nitrogen oxide controls and implications for land carbon sink, Nature Communications, 13, 10.1038/s41467-022-30854-y, 2022.
Liu, Q., Chen, Y., Xu, P., Shen, H., Mai, Z., Zhang, R., Guo, P., Zheng, Z., Luan, T., and Tao, S.: Vertical Profile Corrections Explain Satellite–Inventory Ammonia Discrepancies and Reveal Concentrated Agricultural Sources in China, Environmental Science & Technology, 59, 21609-21621, 10.1021/acs.est.5c08278, 2025.
Lu, X., Yuan, D., Chen, Y., Fung, J. C. H., Li, W., and Lau, A. K. H.: Estimations of Long-Term nss-SO42– and NO3– Wet Depositions over East Asia by Use of Ensemble Machine-Learning Method, Environmental science & technology, 54, 11118-11126, 10.1021/acs.est.0c01068, 2020.
Lu, X., Vitousek, P. M., Mao, Q., Gilliam, F. S., Luo, Y., Turner, B. L., Zhou, G., and Mo, J.: Nitrogen deposition accelerates soil carbon sequestration in tropical forests, Proceedings of the National Academy of Sciences, 118, 10.1073/pnas.2020790118, 2021.
Ma, J., Shi, H., Zhu, Y., Li, R., Wang, S., Lu, N., Yao, Y., Bian, Z., and Huang, K.: The Evolution of Global Surface Ammonia Concentrations during 2001–2019: Magnitudes, Patterns, and Drivers, Environmental Science & Technology, 59, 5066-5079, 10.1021/acs.est.4c14020, 2025.
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349-4383, 10.5194/essd-13-4349-2021, 2021.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

766 Pavlovic, N. R., Chang, S. Y., Huang, J., Craig, K., Clark, C., Horn, K., and Driscoll, C.
767 T.: Empirical nitrogen and sulfur critical loads of U.S. tree species and their
768 uncertainties with machine learning, Sci Total Environ, 857, 159252,
769 10.1016/j.scitotenv.2022.159252, 2023.
770 Peng, S., Ding, Y., Liu, W., and Li, Z.: 1 km monthly temperature and precipitation
771 dataset for China from 1901 to 2017, Earth System Science Data, 11, 1931-1946,
772 10.5194/essd-11-1931-2019, 2019.
773 Peng, S., Gang, C., Cao, Y., and Chen, Y.: Assessment of climate change trends over the
774 Loess Plateau in China from 1901 to 2100, International Journal of Climatology, 38,
775 2250-2264, 10.1002/joc.5331, 2017a.
776 Peng, S., Ding, Y., Wen, Z., Chen, Y., Cao, Y., and Ren, J.: Spatiotemporal change and
777 trend analysis of potential evapotranspiration over the Loess Plateau of China during
778 2011–2100, Agricultural and Forest Meteorology, 233, 183-194,
779 10.1016/j.agrformet.2016.11.129, 2017b.
780 Peng, Z., Zhang, B., Wang, D., Niu, X., Sun, J., Xu, H., Cao, J., and Shen, Z.: Application
781 of machine learning in atmospheric pollution research: A state-of-art review, Science
782 of The Total Environment, 910, 10.1016/j.scitotenv.2023.168588, 2024.
783 Rui Li, L. C., iYilong Zhao,Wenhui Zhou,Hongbo Fu: Long-term trends of ambient
784 nitrate (NO3-) concentrations across China based on ensemble machine-learning
785 models, Earth System Science Data, 13, 2147-2163, 10.5194/essd-13-2147-2021,
786 2021.
787 Shang, F., Liu, M., Song, Y., Lu, X., Zhang, Q., Matsui, H., Liu, L., Ding, A., Huang, X.,
788 Liu, X., Cao, J., Wang, Z., Dai, Y., Kang, L., Cai, X., Zhang, H., and Zhu, T.:
789 Substantial nitrogen abatement accompanying decarbonization suppresses terrestrial
790 carbon sinks in China, Nature Communications, 15, 10.1038/s41467-024-52152-5,
791 2024.
792 Sun, Y. and Fernie, A. R.: Plant secondary metabolism in a fluctuating world: climate
793 change perspectives, Trends in plant science, 10.1016/j.tplants.2023.11.008, 2023.
794 Tan, J., Fu, J. S., and Seinfeld, J. H.: Ammonia emission abatement does not fully control
795 reduced forms of nitrogen deposition, Proceedings of the National Academy of
796 Sciences, 117, 9771-9775, 10.1073/pnas.1920068117, 2020.
797 Tan, J., Su, H., Itahashi, S., Tao, W., Wang, S., Li, R., Fu, H., Huang, K., Fu, J. S., and
798 Cheng, Y.: Quantifying the wet deposition of reactive nitrogen over China: Synthesis
799 of observations and models, Science of The Total Environment, 851,
800 10.1016/j.scitotenv.2022.158007, 2022.
801 Tan, J., Zhang, Y., Mu, Q., Fu, J., & Li, L.: Python codes for "Dataset shift induced
802 machine learning uncertainty: a case study for nitrogen deposition. Zenodo [software].
803 https://doi.org/10.5281/zenodo.18092206. 2025.
804 Tao, C., Zhang, Y., Zhang, X., Guan, X., Peng, Y., Wang, G., Zhang, Q., Ren, Y., Zhao,
805 X., Zhao, R., Wang, Q., and Wang, W.: Discrepant Global Surface Ozone Responses to
806 Emission- and Heatwave-Induced Regime Shifts, Environ Sci Technol,
807 10.1021/acs.est.4c08422, 2024.
808 Watmough, S. A.: Critical loads for alkalization in terrestrial ecosystems, Science of The
809 Total Environment, 927, 10.1016/j.scitotenv.2024.171967, 2024.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

Xie, D., Ge, X., Duan, L., and Mulder, J.: Effects of acid deposition control in China: a review based on responses of subtropical forests, Frontiers of Environmental Science & Engineering, 18, 10.1007/s11783-024-1837-4, 2024.

Xu, W., Zhang, L., and Liu, X.: A database of atmospheric nitrogen concentration and deposition from the nationwide monitoring network in China, Scientific Data, 6, 10.1038/s41597-019-0061-2, 2019.

Yang, Y., Liu, L., Zhang, F., Zhang, X., Xu, W., Liu, X., Wang, Z., and Xie, Y.: Soil Nitrous Oxide Emissions by Atmospheric Nitrogen Deposition over Global Agricultural Systems, Environmental Science & Technology, 55, 4420-4429, 10.1021/acs.est.0c08004, 2021.

Yixin Guo, H. Z., Wilfried Winiwarter,Jinfeng Chang,Xiaolin Wang,Mi Zhou,Petr Havlik,David Leclere,Da Pan,David Kanter,Lin Zhang: Aspirational nitrogen interventions accelerate air pollution abatement and ecosystem protection, Science Advances, 10, 10.1126/sciadv.ado0112, 2024.

Yu, G., Jia, Y., He, N., Zhu, J., Chen, Z., Wang, Q., Piao, S., Liu, X., He, H., Guo, X., Wen, Z., Li, P., Ding, G., and Goulding, K.: Stabilization of atmospheric nitrogen deposition in China over the past decade, Nature Geoscience, 12, 424-429, 10.1038/s41561-019-0352-4, 2019.

Yu, Q., Ge, X., Zheng, H., Xing, J., Duan, L., Lv, D., Ding, D., Dong, Z., Sun, Y., Maximilian, P., Xie, D., Zhao, Y., Zhao, B., Wang, S., Mulder, J., Larssen, T., and Hao, J.: A probe into the acid deposition mitigation path in China over the last four decades and beyond, National Science Review, 11, 10.1093/nsr/nwae007, 2024.

Zhang, B., Zhang, Y., and Jiang, X.: Feature selection for global tropospheric ozone prediction based on the BO-XGBoost-RFE algorithm, Scientific Reports, 12, 10.1038/s41598-022-13498-2, 2022.

Zhang, T., Wu, X., Li, C., and Cheng, C.: Crop-Specific Emission Projection Suggests Peaking of Agricultural N2O by the Middle Century, Environmental Science & Technology, 58, 22967-22979, 10.1021/acs.est.4c06592, 2024.

Zhao, Y., Zhang, L., Chen, Y., Liu, X., Xu, W., Pan, Y., and Duan, L.: Atmospheric nitrogen deposition to China: A model analysis on nitrogen budget and critical load exceedance, Atmospheric Environment, 153, 32-40, 10.1016/j.atmosenv.2017.01.018, 2017.

Zhao, Y., Xi, M., Zhang, Q., Dong, Z., Ma, M., Zhou, K., Xu, W., Xing, J., Zheng, B., Wen, Z., Liu, X., Nielsen, C. P., Liu, Y., Pan, Y., and Zhang, L.: Decline in bulk deposition of air pollutants in China lags behind reductions in emissions, Nature Geoscience, 15, 190-195, 10.1038/s41561-022-00899-1, 2022.

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J. J., and Zhang, H.: Machine Learning: New Ideas and Tools in Environmental Science and Engineering, Environ Sci Technol, 55, 12741-12754, 10.1021/acs.est.1c01339, 2021.

Zhou, K., Xu, W., Zhang, L., Ma, M., Liu, X., and Zhao, Y.: Estimating nitrogen and sulfur deposition across China during 2005 to 2020 based on multiple statistical models, Atmospheric Chemistry and Physics, 23, 8531-8551, 10.5194/acp-23-8531-2023, 2023.

https://doi.org/10.5194/egusphere-2025-6160
Preprint. Discussion started: 20 February 2026

855 Zhu, J., Jia, Y., Yu, G., Wang, Q., He, N., Chen, Z., He, H., Zhu, X., Li, P., Zhang, F., Liu,
856 X., Goulding, K., Fowler, D., and Vitousek, P.: Changing patterns of global nitrogen
857 deposition driven by socio-economic development, Nature Communications, 16,
858 10.1038/s41467-024-55606-y, 2025.
859 Zhu, J. J., Yang, M., and Ren, Z. J.: Machine Learning in Environmental Research:
860 Common Pitfalls and Best Practices, Environ Sci Technol, 57, 17671-17689,
861 10.1021/acs.est.3c00026, 2023.

862