https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

Measurement report: A multi-year, high-resolution, county-scale atmospheric NH_3 dataset from the North China Plain

Jiyang Lyu[1#], Yu Niu[1#], Fanlei Meng[1], Wen Xu[1]*

[1]College of Resources and Environmental Sciences; National Observation and Research Station of Agriculture Green Development (Quzhou, Hebei), China Agricultural University, Beijing, 100193, China

*Correspondence to: W. Xu (wenxu@cau.edu.cn); Tel: + 86 18518151688

# Contributed equally to this work.

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

Abstract

Ammonia (NH_3) is a critical precursor to fine particulate matter ($PM_{2.5}$) pollution. However, long-term observations at the county scale with high temporal resolution remain scarce. We present a four-year (2021–2025) hourly NH_3 dataset from 10 monitoring sites across Quzhou County, a typical intensive agricultural county on the North China Plain. To address missing values caused by instrumental interruptions, we evaluated three machine learning models and selected XGBoost for data imputation. The reconstructed dataset reveals that the annual mean NH_3 concentration in Quzhou County exhibited an overall decreasing trend following an initial rise, increasing from 31 ppb in 2021–2022 to a peak of 36.8 ppb in 2022–2023, and subsequently declining to 26.8 ppb by 2024–2025. A persistent north–south gradient highlights substantial spatial heterogeneity, with mean concentrations ranging from 20.9 ppb at northern cropland sites to 52.8 ppb at southern livestock hotspots. Temporally, NH_3 exhibits a bimodal seasonal cycle peaking in March and June, and a diurnal maximum between 07:00 and 10:00 local time. SHAP analysis identified water vapor pressure, air temperature, and wind speed as the primary meteorological controls. NH_3 concentrations were elevated when vapor pressure exceeded 1.03 kPa and air temperature surpassed 12.4 °C, and were suppressed when wind speed exceeded 1.12 m s^{-1}. This dataset provides a robust observational foundation for evaluating emission reductions, validating satellite products, and informing air quality models.

Keywords: Atmospheric ammonia; Machine learning; Spatiotemporal features; Driving factor

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

1 Introduction

Ammonia (NH_3), the dominant alkaline gas in the atmosphere, is a key precursor of secondary inorganic aerosols (SIA) and plays an important role in the formation of fine particulate matter ($PM_{2.5}$) and haze (Zhang et al., 2017). In China, a series of clean air policies, including the Action Plan on Air Pollution Prevention and Control (State Council, 2013) and the Action Plan for Continuous Improvement of Air Quality (State Council, 2023), have substantially reduced emissions of acidic gases such as sulfur dioxide (SO_2) and nitrogen oxides (NO_x). However, SIA remains a major component of severe pollution episodes, and NH_3 is central to its formation (Zhai et al., 2021). Consequently, NH_3 has become a critical constraint on further air quality improvements (Liu et al., 2023; Ti et al., 2022). The North China Plain (NCP), one of China's most intensive agricultural regions, is also a global NH_3 emission hotspot due to extensive fertilizer use and livestock production (Sun et al., 2025; Zhang et al., 2018). Clarifying the spatiotemporal patterns of NH_3 in this region and their driving mechanisms is therefore essential for developing effective NH_3 abatement and haze mitigation strategies.

Acquiring long-term, continuous NH_3 observations is essential for evaluating the effectiveness of regional emission reductions, yet such records remain difficult to obtain through routine field monitoring (Liu et al., 2023). Regional NH_3 monitoring still relies largely on passive samplers, whose weekly or monthly resolution cannot resolve rapid responses to diurnal meteorological variations (Lan et al., 2021; Pan et al., 2018). Although high-frequency active online instruments have been deployed more widely in recent years, long-term field operation is hindered by their vulnerability to high humidity, power interruptions, routine calibration requirements, and high maintenance costs. These challenges frequently result in measurement interruptions and substantial data gaps (Lan et al., 2024; Hua et al., 2024). Robust missing data imputation strategies are therefore critical for reconstructing continuous NH_3 records and for analyzing the atmospheric dynamics of ammonia (Arnaut et al., 2024).

Long-term, high-resolution NH_3 observations remain particularly scarce at the county scale, where agricultural management and emission-reduction measures are often implemented. Thermal infrared sounders such as IASI and CrIS, together with satellite-based modeling approaches, have expanded estimates of surface NH_3 concentrations and emissions from regional to global scales (Clarisse et al., 2009; Van Damme et al., 2014; Shephard and Cady-Pereira, 2015; Liu et al., 2022; Ma et al., 2025). However, these products provide spatial and temporal coverage different from that of continuous surface

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

networks: their sensitivity to near-surface NH_3 depends on atmospheric thermal conditions, and their footprints and overpass schedules limit direct assessment of local hourly variability (Shephard and Cady-Pereira, 2015; Liu et al., 2025; Sun et al., 2025). Satellite-derived products also require high-quality ground observations for evaluation and bias characterization. Dense, long-term ground-based networks that resolve spatial heterogeneity within mixed crop and livestock systems remain rare. This lack of reference observations limits the evaluation of satellite products and high-resolution atmospheric models and the assessment of local emission reduction measures.

In this study, we present a comprehensive, high-resolution NH_3 dataset from Quzhou County, a representative agriculturally intensive county on the NCP characterized by a winter wheat and summer maize rotation system and intensive livestock production (Feng et al., 2022; Meng et al., 2022). Leveraging the Quzhou National Agriculture Green Development Environmental Monitoring Platform, we collected hourly ambient NH_3 and meteorological data from 10 ground stations over a continuous four-year period from 13 May 2021 to 12 May 2025. With sites strategically distributed across areas dominated by intensive livestock farming, conventional cropland, and mixed land uses, this dense monitoring network provides the necessary coverage to capture county-scale NH_3 variability and to resolve spatial gradients that are typically missed by sparse regional networks. The combination of high temporal resolution (hourly) and extended multi-year coverage allows for robust characterization of diurnal, seasonal, and interannual variability, as well as the identification of meteorological threshold effects.

The objectives of this Measurement Report were to (1) evaluate machine learning methods for filling gaps in an hourly NH_3 dataset, (2) characterize multiscale spatiotemporal variability across the county, (3) quantify the contributions and nonlinear threshold responses of key meteorological drivers using interpretable machine learning, and (4) provide an open dataset for model evaluation, satellite validation, and local policy assessment.

Providing this dataset addresses the scarcity of high-resolution, long-term NH_3 measurements at the county scale, enabling more robust assessments of emission reduction strategies and advancing the scientific understanding of NH_3 dynamics in one of the world's most intensively managed agricultural landscapes.

2 Materials and methods

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

2.1 Study area and data collection

Quzhou County, Handan, Hebei Province (114.83°E-115.22°E, 36.58°N-36.95°N), a typical agriculturally intensive county on the NCP, was selected as the research area (Ren et al., 2025). The county is topographically flat with a temperate semi-humid continental monsoon climate, characterized by four distinct seasons and coincident warm and wet periods (Zhang et al., 2025). Winter wheat-summer maize rotation dominate the cropland, particularly the in the northern areas such as SYZ and WZ, which represent conventional farmlands without major proximate livestock sources (Meng et al., 2022; Zhang et al., 2018), and currently, the county sustains highly intensive livestock and poultry production, with dense breeding activities and intensive organic fertilization concentrated in the southern areas like NLY (Feng et al., 2022; Meng et al., 2022; Table S1). To capture county-scale NH_3 variability, 10 monitoring sites were established across Quzhou County (Fig. 1).

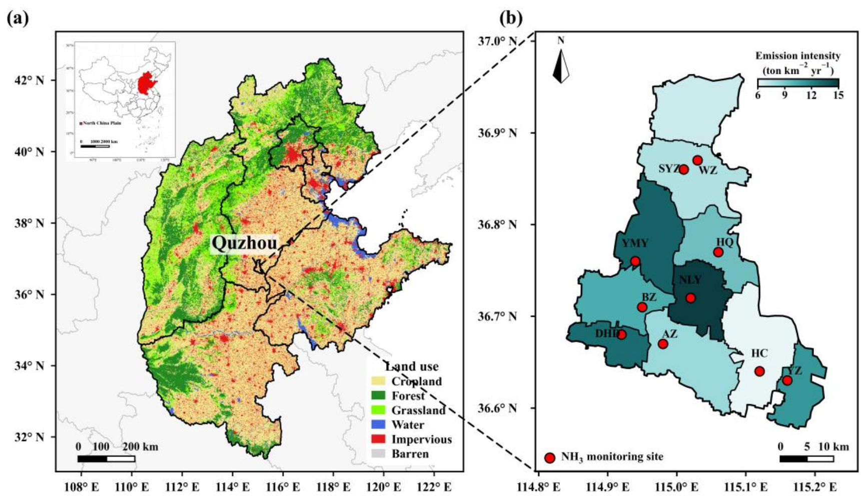

Figure 1. Study area, monitoring network, and township-level ammonia emission intensity. (a) Location of the North China Plain in China and of Quzhou County within the North China Plain. (b) Township-level ammonia emission intensity in Quzhou County (ton km^{-2} yr^{-1}; Kang et al., 2023a) and the locations of the 10 monitoring stations (red dots). Station abbreviations: SYZ (Experimental Station), WZ (Wangzhuang), NLY (Nanliyue), DHD (Dahedao), AZ (Anzhai), BZ (Baizhai), HQ (Huaiqiao), YZ (Yizhuang), HC (Houcun), and YMY (Yumiaoyuan).

The continuous observations were conducted from May 13, 2021, to May 13, 2025. Detailed configurations regarding the measured variable, Instrument model, Instrument height, and Sampling

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

frequency are summarized in Table 1. Rigorous QA/QC procedures, including monthly calibrations and regular maintenance.

Table 1. Summary of instruments and parameters for NH_3 and meteorological observations.

Measured variable	Instrument model	Instrument height	Sampling frequency
NH_3 concentration	Model 17i	2.0 m	1 min
Air temperature	HMP155A	10.0 m	30 min
Relative humidity	HMP155A	10.0 m	30 min
Wind speed	034E	10.0 m	30 min
Direction	034E	10.0 m	30 min
Atmospheric pressure	CS106	1.5 m	30 min
Vapor pressure	$e_s=0.611\times\exp\left(\frac{17.27\times T}{T+237.3}\right)$	10.0 m	30 min

Notes: Observations were conducted from 13 May 2021 – 12 May 2025. All meteorological sensors were connected to a fully automated CR1000X data logger. Rigorous zero and span calibrations for the NH_3 analyzer were performed monthly, alongside regular maintenance. The saturated water vapor pressure (e_s) is automatically calculated based on the temperature.

2.2 Data preparation

To mitigate environmental interference and sensor drift typical of online instruments operating at high frequencies, we applied rigorous QA/QC procedures before model development. Following the China Meteorological Administration's surface observation quality control specifications (China Meteorological Administration, 2020) and recent agricultural station methods (Zhang et al., 2026), and considering historical climate extremes and physical boundaries on the NCP (Nishimura et al., 2023), we performed threshold screening and internal consistency checks (Table S2). A persistence test removed records where a variable remained constant for six consecutive hours (frozen probe). After QA/QC, the overall NH_3 data missing rate was 17.98%, ranging from 6.88% to 25.88% across stations (Table S1). To maintain data continuity, missing meteorological observations were imputed using a spatially coordinated approach (Hasanpour Kashani and Dinpashoh, 2012), calculated as the arithmetic mean of the current variable from other normally operating stations. Hourly precipitation data missing because of

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

rain gauge malfunctions were filled with ERA5-Land total precipitation (Muñoz-Sabater et al., 2021). ERA5-Land precipitation was extracted from the nearest 0.1° grid cell to each station, converted from meters to millimeters per hour, and shifted from UTC to China Standard Time. ERA5-Land precipitation was evaluated against the corresponding rain gauge observations without missing values, yielding an R^2 of 0.7, a mean bias of 8.8×10^{-9} mm, an RMSE of 0.12 mm, and an MAE of 0.02 mm (Fig. S4).

To prevent numerical discontinuity at 0° and 360°, wind direction was decomposed into two orthogonal continuous variables using trigonometric functions (Grange et al., 2018). The conversion formulas are as follows:

$$WD_{\mathrm{sin}}=\sin\left(WD\times\frac{\pi}{180}\right) \tag{1}$$

$$WD_{\mathrm{cos}}=\cos\left(WD\times\frac{\pi}{180}\right) \tag{2}$$

Where WD is the original wind direction angle (in degrees), and $\frac{\pi}{180}$ is the degree-to-radian conversion coefficient. $\mathrm{WD_{sin}}$ and $\mathrm{WD_{cos}}$ represent the sine and cosine components, respectively, both ranging from [-1, 1] and were used as independent features in the machine learning models.

After QA/QC and gap filling, a feature set comprising meteorological variables (air temperature, relative humidity, vapor pressure, wind speed, wind direction sine/cosine, atmospheric pressure, precipitation), temporal variables (year, month, day, hour), and a spatial categorical variable (station ID) was assembled to reconstruct the hourly NH_3 dataset. The meteorological series showed stable seasonal behavior over the study period: summer was generally warm, humid, rainy, and associated with low air pressure, whereas winter was cold, dry, and associated with relatively high air pressure (Fig. S1).

2.3 Machine learning development

We evaluated three tree-based models (Random Forest, XGBoost, and LightGBM) for NH_3 gap filling to capture the complex nonlinear relationships between NH_3 concentrations and spatiotemporal drivers (Geng et al., 2021; Li et al., 2024). Random Forest was used to smooth random noise in the high-frequency observations via parallel averaging of multiple decision trees (Breiman, 2001). Its prediction is given by:

$$\hat{\mathrm{y}}_{\mathrm{i}}=\frac{1}{K}\sum_{k=1}^{K} f_k\left(\mathrm{x_i}\right) \tag{3}$$

where K is the number of decision trees, and $f_k(x_i)$ is the prediction from the k-th tree. Hyperparameters (node splitting randomness, tree depth) were optimized using grid search with 10-fold

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

cross-validation (Probst et al., 2019). To further enhance robustness against measurement noise, we applied the extratrees splitting rule, which randomly generates split thresholds to reduce prediction variance (Geurts et al., 2006).

XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017) are boosting models that iteratively build trees to correct residual errors. XGBoost incorporates a tree-complexity penalty and L2 regularization into its objective function:

$$Obj=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k) \tag{4}$$

$$\Omega(f_t)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{5}$$

where l is the loss function, Ω is the complexity penalty term, T is the number of leaf nodes, w_j is the weight of leaf node j, γ is the penalty for each additional leaf, and λ is the L2 regularization coefficient applied to leaf weights. Together, these terms constrain model complexity and reduce the risk of overfitting (Chen and Guestrin, 2016). LightGBM employs a histogram algorithm and leaf-wise growth for computational efficiency (Ke et al., 2017). Because XGBoost and LightGBM have large, continuous hyperparameter spaces, we used Bayesian optimization with 10-fold cross-validation (Yang and Shami, 2020), tuning maximum tree depth (5-20), learning rate (0.01-0.2), maximum leaf nodes (20-50), and minimum data volume (4-8) (Tırınk, 2025; Yang and Shami, 2020). The preprocessed dataset was randomly split into a training set and a testing set at a ratio of 75:25 (Arnaut et al., 2024; Geng et al., 2021).

2.4 Model interpretability

Although ensemble tree models are well suited to complex nonlinear relationships, their internal structures are often difficult to interpret (Lundberg et al., 2020; Stirnberg et al., 2021). To improve interpretability, we applied the SHAP (Shapley Additive exPlanations) method based on cooperative game theory (Lundberg and Lee, 2017). Using the TreeSHAP algorithm optimized for tree models (Lundberg et al., 2020), the NH_3 concentration predicted by the model can be decomposed into the marginal contribution values of each input feature. The additive explanation model is as follows:

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

$$f(x)=\phi_0+\sum_{i=1}^{M}\phi_i \quad (6)$$

where $f(x)$ represents the model predicted NH_3 concentration for the current sample. ϕ_0 represents the average baseline prediction value of the model across all samples. M represents the total number of input features. ϕ_i represents the SHAP value assigned to the i-th feature. Considering the high time cost of directly calculating Shapley values for the entire high-frequency observation dataset, this study adopted a spatiotemporal stratified random sampling strategy. 5040 representative samples were unbiasedly extracted from the test set. This approach maintained the spatiotemporal distribution characteristics (daily, monthly, and interannual variations, as well as monitoring stations) of the original data while ensuring the feasibility of SHAP calculations.

At the global level, the mean absolute SHAP value of each feature in the sampled dataset was used to quantify the relative contributions of spatiotemporal and meteorological drivers to county-scale NH_3 concentrations. SHAP dependence plots were then used to examine the local nonlinear effects of key meteorological variables, especially vapor pressure and air temperature (Stirnberg et al., 2021).

3 Results and discussion

3.1 Machine learning model performance for gap filling

To reconstruct missing hourly NH_3 observations, we compared XGBoost, Random Forest, and LightGBM (Fig. 2). All three models captured the nonlinear structure of the training data well, with XGBoost performing best (R^2 = 0.97, RMSE = 5.65 ppb, MAE = 3.41 ppb). Random Forest yielded values of 0.89, 11.64 and 5.66 ppb respectively, whereas LightGBM were 0.92, 8.85 and 5.65 ppb. Their performance diverged more clearly on the independent test set. XGBoost again gave the best overall result, with R^2 = 0.78, RMSE = 14.43 ppb and MAE = 7.5 ppb. LightGBM followed (R^2 = 0.74, RMSE = 15.46 ppb, MAE = 8.58 ppb), while Random Forest also achieved R^2 = 0.73 but with slightly larger errors (RMSE = 16.85 ppb, MAE = 8.59 ppb). These results indicate that the regularization built into XGBoost improves generalization when the model is trained on multidimensional spatiotemporal meteorological features (Chen and Guestrin, 2016).

Further analysis combining test set residue reveals significant differences in the error structures of the models within the extreme value range (Fig. 2). Random Forest tended to overestimate low

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

background concentrations and underestimate high peaks; its test set fitted line slope was 0.56 with an intercept of 13.93, reflecting the tendency of parallel averaging to smooth extremes (Pendergrass et al., 2022). For samples with observed concentrations > 60 ppb (Table S3), the MAE of Random Forest was as high as 28.52 ppb. LightGBM, by contrast, was more sensitive to local noise under high-frequency strong fluctuations, likely because of its leaf-wise growth strategy, which can increase the risk of overfitting (Tırınk, 2025). When observed concentrations exceeded 60 ppb, its maximum residual reached 406.26 ppb. XGBoost better constrained tree growth through explicit complexity penalties and therefore remained more stable for extreme values (Chen and Guestrin, 2016). For samples > 60 ppb, its maximum residual was 387.74 ppb, its MAE was 22.16 ppb, and its R^2 was 0.61, all better than those of the other two models. Overall, XGBoost provided the optimal combination of accuracy, generalization, and representation of extremes and was therefore selected for missing data imputation and hourly NH_3 reconstruction.

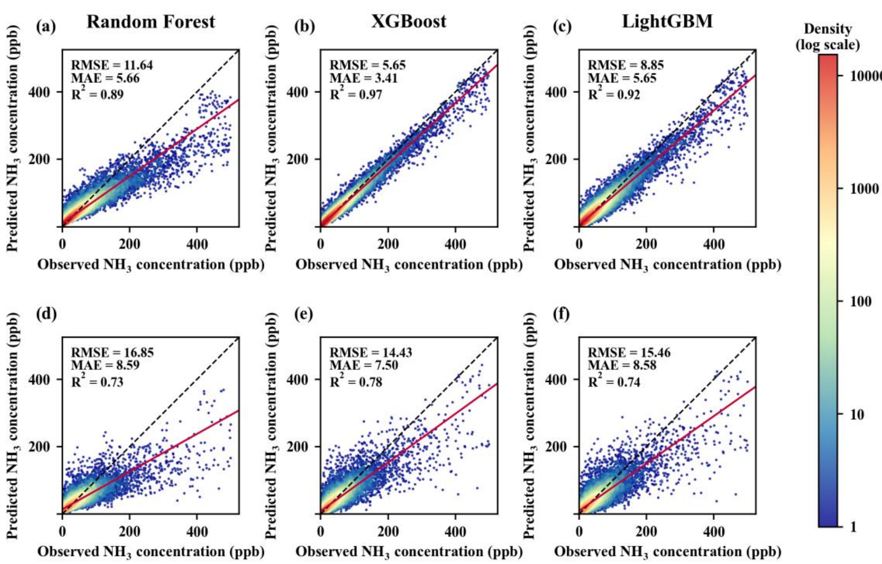

Figure 2. Training set and test set performance of Random Forest, XGBoost, and LightGBM for hourly NH_3 prediction. The top row (a–c) shows training results and the bottom row (d–f) shows test results. Point color indicates log scale density, black dashed lines are 1:1 reference, and red lines are linear fits. Each panel reports RMSE, MAE, and R^2.

3.2 Spatial heterogeneity and interannual trends of NH_3

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

Marked spatial differences were observed among stations, showing strong spatial clustering of NH_3 pollution in Quzhou County consistent with discrete local sources of contrasting intensity. The southern NLY site recorded the highest concentrations throughout the study period, increasing from 44.9 ppb in 2021-2022 to 61.4 ppb in 2022-2023, before declining to 53.9 ppb and 51.2 ppb in the following two years, with a period mean of 52.8 ppb. HC followed a similar but weaker pattern, with annual means of 36.5, 42.2, 45.9, and 33 ppb (Fig. 3). Large intensive livestock farms are densely distributed around NLY, where continuous microbial degradation of organic nitrogen in manure and rapid enzymatic hydrolysis of urea in animal urine act as substantial and persistent emission sources (Kang et al., 2023a). HC is surrounded by farmlands with high fertilization intensity, where the widespread application of nitrogenous fertilizers triggers rapid hydrolysis and subsequent strong volatilization of gaseous NH_3 from the soil surface (Kang et al., 2023b). Consequently, the elevated background NH_3 concentrations at these southern stations are driven by persistent agricultural nitrogen losses (Feng et al., 2022; Zhang et al., 2018). By contrast, SYZ and WZ consistently exhibited lower concentration. At SYZ, annual means ranged from 21.2 to 22.4 ppb during the first three years and fell to 17.6 ppb in 2024-2025, yielding a period mean of 20.9 ppb. Stations such as AZ and YMY also declined to around the regional average in 2024-2025 (Fig. 3). This north-south contrast closely matches the local agricultural land-use pattern: the south is dominated by intensive livestock and poultry production (e.g., NLY as main hotspots), whereas the north is characterized mainly by conventional cropland and research experimental fields (Zhang et al., 2018).

Analysis of the reconstructed dataset reveals an initial increase followed by a decrease in regional average NH_3 concentrations over the four years of observation. The county average rose from 31 ppb between 2021 and 2022 to a peak of 36.8 ppb between 2022 and 2023, before declining to 31.3 ppb and 26.8 ppb in the subsequent two annual periods (Fig. 3a). This pattern suggests that NH_3 pollution peaked in the middle of the observation period and subsequently weakened. The decline in the final two years coincides with the period when intensified fertilizer reduction policies and livestock manure management measures were implemented across the NCP (Liu et al., 2023; Sun et al., 2025).

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

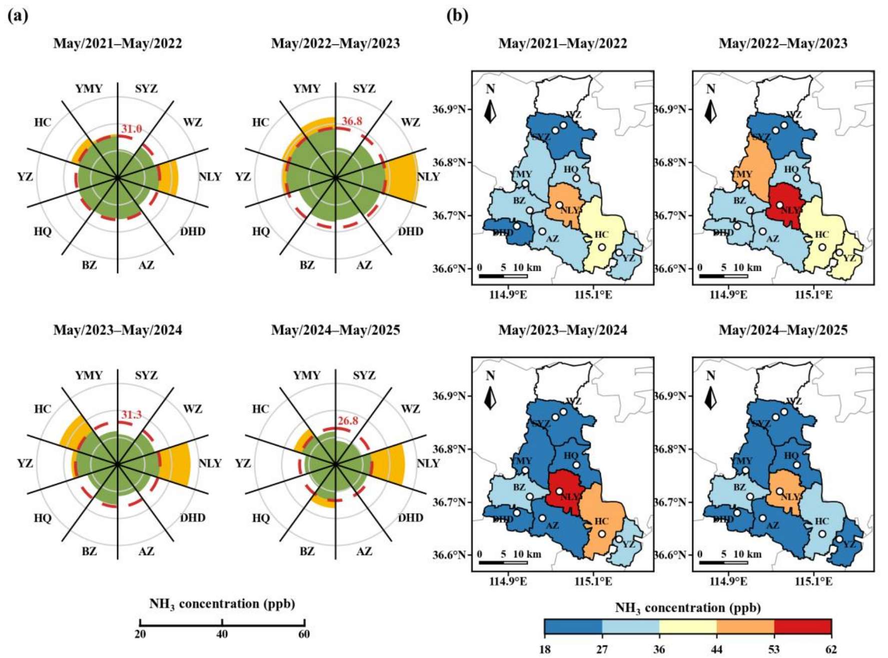

Figure 3. Interannual and spatial variation of NH_3 concentrations in Quzhou County from 13 May 2021 to 12 May 2025. (a) Station means for four monitoring years. Green radial bars extend to the county mean and orange segments show the portion of a station mean above that reference; red dashed arcs and red values denote the county mean for each period. (b) Township scale annual means for the same four periods, with open circles marking monitoring sites.

3.3 Seasonal and diurnal cycles of NH_3

The spatial heterogeneity described above becomes even clearer at seasonal and monthly scales. Combined with the peripheral time series panels in Fig. 4 and the seasonal results in the Supplementary Materials (Fig. S2), the northern region with lower concentration and most conventional agricultural sites generally exhibit a seasonal pattern with higher concentrations in summer and lower concentrations in winter. The NH_3 concentration in summer was 35.2 ppb, followed by 33.7 ppb in spring, 28.7 ppb in winter, and 28.1 ppb in autumn. However, seasonal patterns differed significantly in the southern hotspot; NLY and HC maintained high NH_3 levels in autumn and winter. The seasonal average concentrations at NLY during autumn and winter reached 57.4 ppb and 53 ppb, respectively. This indicates that, beyond cropland fertilization, intensive livestock and poultry production provides a continuous NH_3 source. Emissions from barn ventilation and prolonged manure storage likely maintain elevated background

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

concentrations outside the main fertilization seasons, thereby weakening the typical seasonal pattern expected for sources dominated by cropland (Feng et al., 2022).

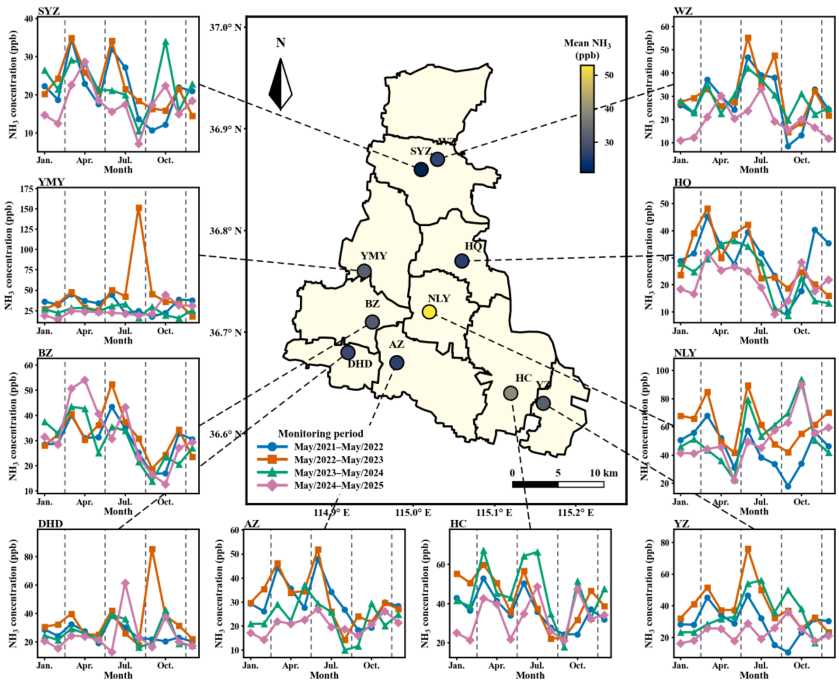

Figure 4. Monthly NH_3 concentrations at the 10 monitoring sites from 13 May 2021 to 12 May 2025. The central map shows site locations and mean concentrations over four years. Surrounding panels show monthly means for four monitoring years: 13 May 2021 to 12 May 2022, 13 May 2022 to 12 May 2023, 13 May 2023 to 12 May 2024, and 13 May 2024 to 12 May 2025. Vertical dashed lines separate the seasonal groups.

At the monthly scale, most stations showed a primary peak in June at 42.4 ppb and a secondary peak in March at 40 ppb. These two high-value months generally correspond to the topdressing period of summer maize and the basal fertilization period of winter wheat, while higher temperatures further enhance NH_3 volatilization, together producing a clear bimodal monthly pattern (Sun et al., 2025). Comparison among observation years also revealed changes in the seasonal cycle (Fig. 4), observations from 2024 to 2025 are clearly lower during spring and summer at most stations than in the previous three years. This pattern suggests that recent management practices, such as fertilizer reduction, improved fertilizer efficiency, and mechanical deep placement, may have weakened the March and June emission

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

pulses. By contrast, the sustained elevated concentrations during autumn and winter in the southern hotspot region did not decline at the same rate, indicating that NH_3 control in agriculturally intensive areas of northern China needs to extend beyond cropland fertilization to include year-round management of livestock manure (Li et al., 2024).

Hourly observations also revealed a regular diurnal cycle of NH_3 (Fig. S3). The county average diurnal curve showed a distinct early morning peak at 08:00 local time, reaching 42.9 ppb, after which the concentration declined rapidly through the day to a minimum of 23.1 ppb at 17:00. Rapid near-surface warming and the evaporation of canopy dew in the early morning likely facilitated the release of NH_3 accumulated during the night, driving the elevation of NH_3 during the morning hours (Wentworth et al., 2016). In the afternoon, stronger boundary-layer convection and turbulent exchange favored dilution of near-surface NH_3, while at night, as the boundary layer became shallower and temperature inversions developed, concentrations entered a renewed accumulation phase (Schulte et al., 2021).

3.4 Meteorological controls on NH_3 variability

Based on the SHAP global importance analysis of the XGBoost model, we further explored the driving effects of various variables on NH_3 concentration fluctuations in Quzhou County. The results indicate that spatiotemporal variables and meteorological factors jointly control NH_3 variations in the study area, with spatiotemporal factors (representing the periodicity of emission sources) contributing up to 59.3% overall. “Day” (20.5%) and “Site” (16%) ranked first and second in global importance, respectively (Fig. 5a). In the spatial dimension, the high contribution of the site variable is consistent with the spatial pattern of higher concentrations in the south and lower in the north, indicating that different underlying surfaces and emission source configurations directly affect local baseline NH_3 levels (Pan et al., 2018). Highly dense breeding areas in the south manifest as persistent intensive point sources (Kang et al., 2023a), whereas conventional farmlands in the north primarily act as seasonal non-point sources. This classification aligns with the fundamental emission mechanisms where livestock waste degradation provides a year-round constant NH_3 source, while fertilizer application triggers distinct seasonal peaks (Feng et al., 2022; Zhang et al., 2018). In the temporal dimension, the “day” feature ranks first in importance because agricultural activities such as fertilization are not uniformly distributed over time but are concentrated on specific calendar dates, leading to sharp, predictable peaks in NH_3 concentrations (Feng et al., 2022). Combined with the long-term evolution characteristics described, it is

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

plausible that measures promoted in the NCP in recent years, such as chemical fertilizer reduction, soil testing and formulated fertilization, and refined management, may be mitigating the extreme emission pulses of traditional fertilization seasons (Liu et al., 2023; Sun et al., 2025).

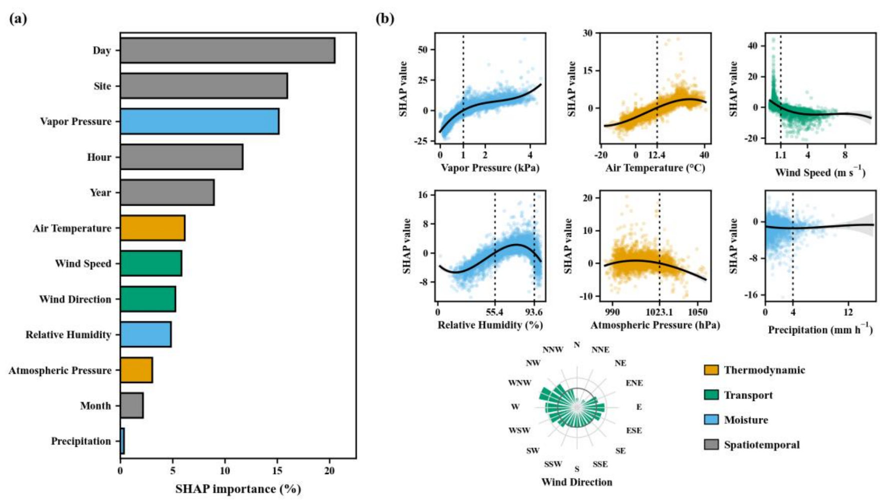

Figure 5. Global feature importance and meteorological SHAP dependence for the selected XGBoost model. (a) Relative mean absolute SHAP importance. (b) Dependence plots for vapor pressure, air temperature, wind speed, relative humidity, atmospheric pressure, precipitation, and wind direction. Vertical dashed lines indicate estimated response thresholds.

Among the meteorological factors, vapor pressure (15.2%), wind speed (5.9%), and air temperature (6.2%) were the core variables affecting NH_3 concentrations in Quzhou County (Fig. 5a). The SHAP dependence plots revealed that both thermal and moisture conditions exhibit distinct nonlinear threshold effects (Fig. 5b). When vapor pressure exceeded 1.03 kPa and relative humidity was within the range of 55.4%-93.6%, the SHAP contribution showed a significant positive increase, indicating that higher surface water vapor and ambient humidity favor the hydrolysis and volatilization of nitrogenous components in farmland urea and livestock manure (Ni and Pacholski, 2022). However, when relative humidity further increased to a saturated state (> 93.6%) or precipitation occurred, NH_3 could be scavenged via wet deposition through dissolution and conversion into particulate ammonium salts, thereby attenuating gaseous concentrations (Li et al., 2024). Furthermore, the partial dependence relationship of air temperature exhibited an exponential enhancement trend, with a critical threshold at approximately 12.4 °C. Beyond this threshold, high temperatures not only enhance the volatilization potential of liquid phase ammonium but may also promote the thermal dissociation of semi-volatile ammonium nitrate and release gaseous NH_3, thereby driving up concentration levels during warm seasons (Huo et al., 2025).

Kinetic transport processes also play a crucial regulatory role in the accumulation and dispersion of local NH_3. The SHAP curve for wind speed demonstrated that light breeze conditions are most favorable for the accumulation of NH_3 near the observation height: moderate turbulent exchange can transport NH_3 released near the surface and within the canopy to the monitoring height (Schulte et al., 2021). Once the wind speed exceeded 1.12 m s^{-1}, its SHAP contribution turned negative, indicating that stronger wind speeds are often accompanied by more significant advective transport and boundary layer mixing, which can rapidly dilute local elevated NH_3. Combined with the wind rose diagram (Fig. 5b), it can be further observed that under suitable wind directions and light breeze conditions, NH_3 emitted from southern livestock farms or centralized fertilization areas is more prone to short-distance transport, thereby intensifying the spatial clustering characteristics of pollution hotspots (Sun et al., 2025; Wang et al., 2025).

4 Conclusions

This Measurement Report presents a comprehensive hourly NH_3 dataset spanning four years (13 May 2021-12 May 2025) from 10 monitoring sites in Quzhou County, a representative intensive agricultural county on the North China Plain. To address missing values caused by instrumental interruptions, we evaluated three ensemble machine learning models, selecting XGBoost for final imputation due to its optimal predictive performance (test R^2 = 0.78, RMSE = 14.43 ppb, MAE = 7.5 ppb) and robust representation of extreme events.

Analysis of the reconstructed dataset reveals the following key features. The county-averaged NH_3 concentration increased from 31 ppb in 2021-2022 to a peak of 36.8 ppb in 2022-2023, then declined to 31.3 ppb and 26.8 ppb in the following two years. A persistent north-south gradient exists, with four-year site means ranging from 20.9 ppb at northern cropland sites to 52.8 ppb at southern livestock hotspots, reflecting contrasting land-use types and emission sources. NH_3 exhibits a bimodal seasonal cycle with peaks in June and March, and a distinct early-morning maximum at 08:00 local time. SHAP analysis identifies vapor pressure, air temperature, and wind speed as the dominant meteorological controls, with

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

positive contributions when vapor pressure exceeds 1.03 kPa or temperature exceeds 12.4 °C, and negative contributions when wind speed exceeds 1.12 m s^{-1}.

The dataset is suitable for evaluating agricultural emission reduction policies, validating satellite NH_3 retrievals (e.g., IASI, CrIS), constraining regional air quality models, and supporting health and ecosystem assessments. Limitations include the dataset's specificity to one county and potential uncertainty in gap-filled values during extended data gaps. Future work should integrate such observations with emission inventories and chemistry models.

Code availability

Code related to this paper may be requested from the authors.

Data availability

All datasets used in this study are publicly available on Figshare (https://doi.org/10.6084/m9.figshare.32232129) (Lyu et al., 2026).

Author contributions

W.X. designed the study and reviewed and edited the manuscript. J.L. analyzed the data, wrote the original draft, and revised the manuscript. Y.N. conducted the field experiment and performed the measurements. F.M. commented on the figures. All authors read and approved the final version.

Competing interests

The authors declare no competing interests.

Financial support

This work was supported by the National Key Research and Development Program of China (2023YFC3707402), the National Natural Science Foundation of China (42507544), and the China Scholarship Council (No. 20222260).

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

References:

Arnaut, F., Đurđević, V., Kolarski, A., Srećković, V. A., and Jevremović, S.: Improving air quality data reliability through bi-directional univariate imputation with the random forest algorithm, Sustainability, 16, 7629, https://doi.org/10.3390/su16177629, 2024.

Breiman, L.: Random forests, Mach. Learn., 45, 5-32, https://doi.org/10.1023/A:1010933404324, 2001.

Chen, T. and Guestrin, C.: XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794, https://doi.org/10.1145/2939672.2939785, 2016.

China Meteorological Administration: Quality control of surface meteorological observation data (Standard No. QX/T 118-2020), Meteorological Press, Beijing, 2020.

Clarisse, L., Clerbaux, C., Dentener, F., Hurtmans, D., and Coheur, P.-F.: Global ammonia distribution derived from infrared satellite observations, Nat. Geosci., 2, 479-483, https://doi.org/10.1038/ngeo551, 2009.

Feng, S., Wang, Y., Zheng, X., Wang, Y., Gao, Z., and Chen, D.: Overlooked nonagricultural and wintertime agricultural NH_3 emissions in Quzhou County, North China Plain: Evidence from 15N-stable isotopes, Environ. Sci. Technol. Lett., 9, 212-218, https://doi.org/10.1021/acs.estlett.1c00935, 2022.

Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., Peng, Y., Huang, X., He, K., and Zhang, Q.: Tracking Air Pollution in China: Near Real-Time $PM_{2.5}$ Retrievals from Multisource Data Fusion, Environ. Sci. Technol., 55, 12106-12115, https://doi.org/10.1021/acs.est.1c01863, 2021.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach. Learn., 63, 3–42, https://doi.org/10.1007/s10994-006-6226-1, 2006.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random Forest meteorological normalisation models for Swiss PM10 trend analysis, Atmos. Chem. Phys., 18, 6223-6239, https://doi.org/10.5194/acp-18-6223-2018, 2018.

Hasanpour Kashani, M. and Dinpashoh, Y.: Evaluation of efficiency of different estimation methods for missing climatological data, Stoch. Environ. Res. Risk Assess., 26, 59-71, https://doi.org/10.1007/s00477-011-0536-y, 2012.

Hua, V., Nguyen, T., Dao, M.-S., Nguyen, H. D., and Nguyen, B. T.: The impact of data imputation on air quality prediction problem, PLoS One, 19, e0306303, https://doi.org/10.1371/journal.pone.0306303, 2024.

Huo, H., Gao, Y., Sun, L., Gao, Y., Gao, H., and Yao, X.: Investigating dual character of atmospheric ammonia on particulate NH4NO3: Reducing evaporation versus promoting formation, Atmosphere, 16, 685, https://doi.org/10.3390/atmos16060685, 2025.

Kang, J., Meng, F., Liu, X., and Xu, W.: Emission inventory and distribution characteristics of anthropogenic ammonia in Quzhou county, North China plain, Huanjing Kexue, 44(1), 94–103,

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

https://doi.org/10.13227/j.hjkx.202202139, 2023a.

Kang, J., Wang, J., Heal, M. R., Goulding, K., de Vries, W., Zhao, Y., Feng, S., Zhang, X., Gu, B., Niu, X., Zhang, H., Liu, X., Cui, Z., Zhang, F., and Xu, W.: Ammonia mitigation campaign with smallholder farmers improves air quality while ensuring high cereal production, Nat. Food, 4, 751–761, https://doi.org/10.1038/s43016-023-00833-7, 2023b.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., and Liu, T. Y.: LightGBM: A highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst., 30, 3146-3154, https://proceedings.neurips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf, 2017.

Lan, Z., Lin, W., Pu, W., and Ma, Z.: Measurement report: Exploring NH_3 behavior in urban and suburban Beijing: comparison and implications, Atmos. Chem. Phys., 21, 4561-4573, https://doi.org/10.5194/acp-21-4561-2021, 2021.

Lan, Z., Zhang, X., Lin, W., Xu, X., Ma, Z., and Jin, J.: Atmospheric NH_3 in urban Beijing: long-term variations and implications for secondary inorganic aerosol control, Atmos. Chem. Phys., 24, 9355-9368, https://doi.org/10.5194/acp-24-9355-2024, 2024.

Li, D., Liu, M., Han, H., and Wang, J.: Nonlinear impacts of air pollutants and meteorological factors on $PM_{2.5}$: An interpretable GT-iFormer model with SHAP analysis, Atmosphere, 17, 266, https://doi.org/10.3390/atmos17030266, 2026.

Li, R., Gao, Y., Zhang, L., Shen, Y., Xu, T., Sun, W., and Wang, G.: Global estimates of ambient reactive nitrogen components during 2000-2100 based on the multi-stage model, Atmos. Chem. Phys., 24, 7623-7636, https://doi.org/10.5194/acp-24-7623-2024, 2024.

Liu, P., Ding, J., Liu, L., Xu, W., and Liu, X.: Estimation of surface ammonia concentrations and emissions in China from the polar-orbiting Infrared Atmospheric Sounding Interferometer and the FY-4A Geostationary Interferometric Infrared Sounder, Atmos. Chem. Phys., 22, 9099-9110, https://doi.org/10.5194/acp-22-9099-2022, 2022.

Liu, P., Li, X., Zhang, C., and Mu, Y.: It is time to reduce atmospheric pollutant emissions from agricultural and residential activities in rural China for the sustainable improvement of air quality, Environ. Sci. Technol., 57, 19102-19105, https://doi.org/10.1021/acs.est.3c08670, 2023.

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, Environ. Sci. Technol., 59, 21609-21621, https://doi.org/10.1021/acs.est.5c08278, 2025.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., and Lee, S. I.: From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2, 56-67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

Lundberg, S. M. and Lee, S. I.: A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst., 30, 4765-4774, https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf, 2017.

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

Lyu, J., Niu, Y., Meng, F., and Xu, W.: Measurement report: A multi-year, high-resolution, county-scale atmospheric NH_3 dataset from the North China Plain, Figshare [data set], https://doi.org/10.6084/m9.figshare.32232129, 2026.

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, Environ. Sci. Technol., 59, 5066-5079, https://doi.org/10.1021/acs.est.4c14020, 2025.

Meng, F., Wang, M., Strokal, M., Kroeze, C., Ma, L., Li, Y., and Zhang, F.: Nitrogen losses from food production in the North China Plain: A case study for Quzhou, Sci. Total Environ., 804, 151557, https://doi.org/10.1016/j.scitotenv.2021.151557, 2022.

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 Syst. Sci. Data, 13, 4349-4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Ni, K. and Pacholski, A. S.: Soil moisture and temperature effects on granule dissolution and urease activity of urea with and without inhibitors: An incubation study, Agriculture, 12, 2037, https://doi.org/10.3390/agriculture12122037, 2022.

Nishimura, M., Aoki, T., Niwano, M., Matoba, S., Tanikawa, T., Yamasaki, T., Yamaguchi, S., and Fujita, K.: Quality-controlled meteorological datasets from SIGMA automatic weather stations in northwest Greenland, 2012-2020, Earth Syst. Sci. Data, 15, 5207-5226, https://doi.org/10.5194/essd-15-5207-2023, 2023.

Pan, Y., Tian, S., Liu, D., Fang, Y., Zhu, X., Gao, J., Bao, H., Wu, G., Peng, P., Qiu, G., Ge, B., Carmichael, G. R., and Wang, Y.: Identifying ammonia hotspots in China using a national observation network, Environ. Sci. Technol., 52, 3926–3934, https://doi.org/10.1021/acs.est.7b05235, 2018.

Pendergrass, D. C., Zhai, S., Kim, J., Koo, J.-H., Lee, S., Bae, M., Kim, S., Liao, H., and Jacob, D. J.: Continuous mapping of fine particulate matter (PM2.5) air quality in East Asia at daily 6 × 6 km2 resolution by application of a random forest algorithm to 2011-2019 GOCI geostationary satellite data, Atmos. Meas. Tech., 15, 1075-1091, https://doi.org/10.5194/amt-15-1075-2022, 2022.

Probst, P., Wright, M. N., and Boulesteix, A. L.: Hyperparameters and tuning strategies for random forest, Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 9, e1301, https://doi.org/10.1002/widm.1301, 2019.

Ren, T., Yang, X., Zhang, W., Tang, W., Li, Y., Tian, Y., Ren, J., Yan, J., Guo, X., An, Z., and Zhang, H.: Smallholder farmers can achieve more sustainable wheat production through Consolidating Land for Uniform Practice, Front. Plant Sci., 16, 1517683, https://doi.org/10.3389/fpls.2025.1517683, 2025.

Schulte, R. B., van Zanten, M. C., Rutledge-Jonker, S., Swart, D. P. J., Wichink Kruit, R. J., Krol, M. C., van Pul, W. A. J., and Vilà-Guerau de Arellano, J.: Unraveling the diurnal atmospheric ammonia budget of a prototypical convective boundary layer, Atmos. Environ., 249, 118153, https://doi.org/10.1016/j.atmosenv.2020.118153, 2021.

Shephard, M. W. and Cady-Pereira, K. E.: Cross-track Infrared Sounder (CrIS) satellite observations of tropospheric ammonia, Atmos. Meas. Tech., 8, 1323-1336, https://doi.org/10.5194/amt-8-1323-2015, 2015.

State Council of the People's Republic of China: Action Plan on Air Pollution Prevention and Control, Beijing, China, http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm, 2013.

State Council of the People's Republic of China: Action Plan for Continuous Improvement of Air Quality, Beijing, China, https://english.www.gov.cn/archive/statecouncilgazette/202312/20/content_WS65824dc1c6d0868f4e8e2616.html, 2023.

Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., and Andersen, H.: Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning, Atmos. Chem. Phys., 21, 3919-3934, https://doi.org/10.5194/acp-21-3919-2021, 2021.

Sun, F., Cui, Y., Su, J., Shephard, M. W., Kharol, S. K., Zhang, Y., and Hu, C.: Decadal changes in atmospheric ammonia and dry deposition across China inferred from space-ground measurements and model simulations, Atmos. Chem. Phys., 25, 17907-17932, https://doi.org/10.5194/acp-25-17907-2025, 2025.

Ti, C., Xia, L., Chang, S. X., and Yan, X.: Mitigation of agricultural NH_3 emissions reduces $PM_{2.5}$ pollution in China: A finer scale analysis, J. Clean. Prod., 350, 131507, https://doi.org/10.1016/j.jclepro.2022.131507, 2022.

Tırınk, S.: Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XGBoost, LightGBM, and SVM, PLoS One, 20, e0334252, https://doi.org/10.1371/journal.pone.0334252, 2025.

Van Damme, M., Clarisse, L., Heald, C. L., Hurtmans, D., Ngadi, Y., Clerbaux, C., Dolman, A. J., Erisman, J. W., and Coheur, P. F.: Global distributions, time series and error characterization of atmospheric ammonia (NH_3) from IASI satellite observations, Atmos. Chem. Phys., 14, 2905-2922, https://doi.org/10.5194/acp-14-2905-2014, 2014.

Wang, J., Fei, J., Li, M., Cang, H., Hua, L., and Li, H.: Rapid switching of reagent ions in dopant-assisted positive photoionization ion mobility spectrometry for highly selective and sensitive ammonia detection in atmospheric environments, Anal. Chem., 97, 21518-21529, https://doi.org/10.1021/acs.analchem.5c03686, 2025.

Wentworth, G. R., Murphy, J. G., Benedict, K. B., Bangs, E. J., and Collett Jr., J. L.: The role of dew as

https://doi.org/10.5194/egusphere-2026-2742
Preprint. Discussion started: 28 July 2026

a night-time reservoir and morning source for atmospheric ammonia, Atmos. Chem. Phys., 16, 7435-7449, https://doi.org/10.5194/acp-16-7435-2016, 2016.

Yang, L. and Shami, A.: On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing, 415, 295-316, https://doi.org/10.1016/j.neucom.2020.07.061, 2020.

Yuan, Y., Shen, F. Z., Sheng, C. Y., Zhang, Z. M., Guo, W. H., Zhu, W. G., and Zhu, H.: Evaluating the impact of anthropogenic drivers and meteorological factors on air pollutants by explainable machine learning in Shandong Province, China, Atmos. Pollut. Res., 16, 102694, https://doi.org/10.1016/j.apr.2025.102694, 2025.

Zhai, S., Jacob, D. J., Wang, X., Liu, Z., Wen, T., Shah, V., Li, K., Moch, J. M., Bates, K. H., Song, S., Shen, L., Zhang, Y., Luo, G., Yu, F., Sun, Y., Wang, L., Qi, M., Tao, J., Gui, K., and Liao, H.: Control of particulate nitrate air pollution in China, Nat. Geosci., 14, 389-395, https://doi.org/10.1038/s41561-021-00726-z, 2021.

Zhang, L., Chen, Y., Zhao, Y., Henze, D. K., Zhu, L., Song, Y., and Huang, B.: Agricultural ammonia emissions in China: reconciling bottom-up and top-down estimates, Atmos. Chem. Phys., 18, 339-355, https://doi.org/10.5194/acp-18-339-2018, 2018.

Zhang, X., Wu, Y., Liu, X., Reis, S., Jin, J., Dragosits, U., Van Damme, M., Clarisse, L., Whitburn, S., Coheur, P.-F., and Gu, B.: Ammonia emissions may be substantially underestimated in China, Environ. Sci. Technol., 51, 12089-12096, https://doi.org/10.1021/acs.est.7b02171, 2017.

Zhang, Y., Li, B., Nie, Y., Luan, Q., and Gu, P.: Study of multi-source river discharge data assimilation in collaboration with EKF and MIKE11HD, Hydrol. Sci. J., 70, 1554–1569, https://doi.org/10.1080/02626667.2025.2492884, 2025.

Zhang, Z., Li, P., and Wang, J.: A meteorological data quality control framework for tea plantations using association rules mined from ERA5 reanalysis data, Agriculture, 16, 226, https://doi.org/10.3390/agriculture16020226, 2026.