

I appreciate the authors' efforts in addressing the important issue of aerosol pH variability in an ammonia-rich urban atmosphere. The manuscript combines field observations, thermodynamic modeling, and an explainable machine-learning framework to examine how meteorological and chemical factors are associated with aerosol pH. The topic is relevant, and the regime-dependent RH effect is potentially interesting. However, I think the manuscript requires substantial revision before the conclusions can be fully supported.

Overall comment

The central concern is that the scientific role of the XGBoost-SHAP framework is not sufficiently justified. In this study, aerosol pH is not directly measured but calculated using ISORROPIA II, and the ML model is subsequently trained to reproduce this ISORROPIA-derived pH using variables that are either direct inputs to the thermodynamic calculation or closely related derived indicators, including Excess NH_x , N/S ratio, PM_{10} mass loading, temperature, and RH. As a result, the SHAP analysis mainly characterizes how the thermodynamic-model-derived pH responds to these variables under the observed covariance structure of the dataset. This can still be useful, but its added value beyond conventional thermodynamic sensitivity analysis needs to be clearly articulated. In particular, the authors should clarify whether the main contribution of the XGBoost-SHAP approach is to diagnose nonlinear pH responses under real-world co-varying conditions, rather than to identify independent causal mechanisms. The manuscript should therefore more carefully distinguish thermodynamic-model behavior, statistical associations in the observational dataset, and actual atmospheric chemical processes. Some current interpretations appear to overstate SHAP-based associations as mechanistic evidence, and these should be revised accordingly.

Specific comments

1. The model validation strategy should be strengthened. The current random training/testing split may overestimate model performance for continuous time-series observations, because temporally adjacent samples, similar pollution episodes, and data from the same season may be included in both training and testing sets. I suggest adding blocked or time-aware validation, such as leave-one-month-out cross-validation, season-based cross-validation, or training in one season and testing in another. This would better show whether the model captures generalizable relationships rather than temporally correlated patterns.
2. The very high model performance should be interpreted cautiously. Since the target pH is calculated by ISORROPIA II and the predictors are closely related to the thermodynamic inputs, a high R^2 is expected. The manuscript should avoid presenting the high R^2 as strong evidence that the ML framework independently captures atmospheric processes.
3. The dependence among input features should be evaluated, and the robustness of the SHAP results to feature selection should be tested. Excess NH_x , N/S ratio, and PM_{10} mass loading are not independent variables. Excess NH_x is derived from ammonia/ammonium and acidic species; N/S ratio is derived from nitrate and sulfate; and PM_{10} mass loading is partly composed of the same inorganic species involved in the pH calculation. These correlations may influence SHAP importance rankings and interaction values. The authors should provide a correlation matrix and test whether the main conclusions remain robust when alternative feature sets are used, for example by removing PM_{10} mass loading, replacing N/S ratio with nitrate and sulfate separately, or excluding strongly derived variables.
4. The regime thresholds should be supported more quantitatively. The manuscript identifies transition values around $15\text{ }^\circ\text{C}$, $\text{N/S} \approx 1.25$, and $\text{PM}_{10} \approx 50\text{ }\mu\text{g m}^{-3}$. These thresholds appear to be inferred from SHAP interaction plots, but their statistical robustness is unclear. The authors should test whether these values remain stable under different random seeds, train/test splits, bootstrapping, or smoothing methods. If not, they should be described as empirical transition ranges rather than fixed thresholds.

5. The uncertainty in the pH calculation should be expanded. The assumptions of $\gamma_{H^+} = 1$, neglect of organic aerosol liquid water, and exclusion of alkaline cations should be discussed more carefully. Since all ML results depend on the ISORROPIA-derived pH, these assumptions may affect the SHAP interpretation.
6. The upper end of the calculated pH range needs further explanation. The manuscript reports pH values up to 6.7, which is close to neutral and unusually high for fine-mode urban aerosols. The authors should examine the high-pH cases, for example $pH > 5$ or $pH > 6$, and clarify under what conditions they occur. In particular, they should show whether these values are associated with very low ALWC, low inorganic mass, extremely ammonia-rich conditions, or other situations where thermodynamic pH predictions may be uncertain. It would also be useful to evaluate whether the NH_3/NH_4^+ partitioning remains reliable for these high-pH samples, rather than only reporting overall model performance.
7. The causal language in the discussion should be moderated. The RH-related pH reversal is interesting, but interpretations involving secondary inorganic aerosol formation, heterogeneous reactions, aqueous-phase chemistry, and H^+ accumulation are not directly resolved by the XGBoost-SHAP model. These explanations should be framed as plausible mechanisms or patterns consistent with known chemistry, rather than as direct evidence from the ML analysis.
8. The manuscript should be more cautious in generalizing the results. The analysis is based on one urban site in Beijing during 2014-2015. The identified thresholds and interaction patterns may reflect local meteorology, emissions, and aerosol composition. Statements about broader predictive capability or future scenarios should therefore be toned down unless supported by additional datasets or validation.

9. Minor issues:

- (1) The sentence in the Abstract, “their behavior differs across environmental regimes and remain incompletely understood,” is grammatically inconsistent. It should be revised to “their behavior differs across environmental regimes and remains incompletely understood,” or rewritten.
- (2) The word “systematical” should be changed to “systematic.”
- (3) The phrase “That means” is too informal for a scientific manuscript and should be replaced with “This indicates that” or “This suggests that.”
- (4) Equation (2) should be checked carefully. The notation for the sample index and feature index is confusing, and the summation should clearly refer to feature contributions.
- (5) The terminology for the nitrate-to-sulfate ratio should be standardized throughout the manuscript. For example, “NtoSratio” in the figures should be changed to “N/S ratio” to match the text.
- (6) Figures 2 and 3 are difficult to read because the fonts and data points are too small. The c-ICE panels in Fig. 2 are particularly dense. The authors should enlarge the figures, improve labels, or move some panels to the Supplement.