General comments:

This study applies XGBoost and deep neural network models to reproduce aerosol Fe solubility and dissolved Fe concentrations, and further combines SHAP with independent component analysis (ICA) to interpret model-derived feature contributions. The current discussions require more rigorous justification. For example, several interpretations of the SHAP-ICA components are not sufficiently supported, and statistical components are sometimes directly linked to specific sources or atmospheric processes without adequate evidence. In addition, the manuscript contains many technical problems, including incorrect figure references, inconsistent or wrong section numbering, and an extra Japanese annotation. These issues should be carefully checked and corrected throughout the manuscript. The current version cannot meet the standards required for ACP.

Major Comments:

1. The statements about missing values appear inconsistent. At Line 108, the author states that both XGBoost and DNN cannot handle samples with missing values, whereas Line 142 states that XGBoost can handle missing values but DNNs require removal or imputation.
2. Line 110: The use of [d-Fe]/[d-Al] as an input feature may introduce mathematical coupling with the target variables. Fesol% is defined from d-Fe, and d-Fe is also the direct target variable in the model. How did the authors consider this potential mathematical coupling when evaluating model performance and interpreting the SHAP results?
3. Line 250: Why was the number of ICA components fixed at four? Since the number of input features does not necessarily correspond to the number of underlying physical or chemical processes, the authors should briefly justify this choice.
4. Line 263: The authors conclude that the DNN model outperforms XGBoost because it yields lower MAE and RMSE. However, this statement should be carefully checked against the values shown in the figure, as the improvement does not appear equally clear for all target variables and metrics. The authors should discuss model performance separately for Fesol% and d-Fe concentration, rather than making a general statement that DNN performs better in all cases.
5. Line 270: The authors attribute the poorer performance of XGBoost at low Fesol% and low d-Fe concentrations to data sparsity, and argue that DNN can better interpolate in sparse regions by learning a continuous nonlinear function. This explanation is plausible but remains largely qualitative. The authors should provide error statistics stratified by Fesol% and d-Fe concentration ranges, especially for the low-value samples, to demonstrate that the DNN advantage indeed arises from improved performance in these regions.
6. Line 431: In Section 3.4.2, The authors first argue that the negative SHAP response of T-Fe/T-Al indicates that anthropogenic Fe in fine particles is mainly emitted as insoluble Fe, rather than as high-solubility fly ash from heavy-oil or gasoline combustion. However, the subsequent interpretation of IC2 attributes this component to fly ash emissions from heavy-oil and gasoline combustion (Line 458), in which both T-Fe and d-Fe are already high at emission. This apparent inconsistency requires further explanation.
7. Line 472: the author attributes the poorly reproduced high-Fesol% samples to possible “marine chemical alteration.” What specific processes are meant by this term? Does it refer to cloud/fog processing,

acidification during marine transport or interactions with marine biogenic species? The authors should define this process more explicitly and provide relevant literature support.

8. Line 509: The author first emphasizes the poorer reproducibility of d-Fe concentrations, especially for low-d-Fe samples. However, the subsequent G1–G3 classification is based on the relative error of Fesol% rather than d-Fe. Why is a Fesol%-based error metric used to classify samples when the preceding discussion focuses on deviations in d-Fe concentration?
9. Line 522: The source attribution for the G3 samples requires further clarification. First, are there any evidences supporting the statement that volcanic rocks from Big Ben are characterized by high T-Fe/T-Al? Second, why would high T-Fe/T-Al in volcanic materials necessarily lead to elevated [d-Fe]/[d-Al]? A high total Fe/Al ratio does not directly imply a high dissolved Fe/Al ratio unless the volcanic Fe is also more soluble relative to Al. Finally, the author suggests that the impact of anthro-Fe on Fesol% may differ between East Asia and coastal Australia. Since both are described as anthropogenic Fe influences, what specific differences in source type, Fe mineralogy, emission form, or atmospheric processing would cause anthro-Fe to affect Fesol% differently in these two regions?
10. At Lines 567–570, the SHAP discussion states that d-Fe is mainly controlled by d-Al and [d-Fe]/[d-Al], while T-Fe contributes little. However, at Lines 573–578, the IC1 interpretation states that T-Fe also constrains the magnitude of d-Fe supply to the North Pacific. Could the author clarify these seemingly inconsistent statements?
11. Lines 652–656: IC1 is described as reflecting both [d-Fe]/[d-Al] and Alsol%, but the conclusion states that South Pacific Fesol% is primarily controlled by chemical alteration of mineral dust. Why is the contribution from [d-Fe]/[d-Al], which is used elsewhere as an indicator of anthro-Fe contribution and processing, not reflected in the interpretation?
12. Lines 657–665: IC2 is reported to have only a weak relationship with observed Fesol%, and the scatter plot shows no clear correlation. However, the text then interprets high IC2 scores as episodic influences from volcanogenic particles and coastal anthro-Fe that can elevate Fesol%. If IC2 has no clear correlation with Fesol%, on what basis can it be interpreted as a process that elevates Fesol%? The distinction between a weak statistical association and a process-level interpretation should be made clearer.
13. The Introduction frames the study partly around the “snapshot problem” of marine aerosol observations and as overarching objectives of this study (Line 95). But the later Results and Discussion do not clearly demonstrate how this problem is overcome. The manuscript shows that some marine aerosol samples can be reproduced by the EA-DNN model, but model reproducibility does not necessarily prove that these short-term cruise samples are regionally or temporally representative. The authors should clarify whether model reproducibility is being used only to assess whether samples follow similar Fe-solubility relationships, or whether it is intended to address the broader representativeness problem of marine aerosol observations.

Minor Comments:

1. Line 48: A comma appears to be missing in this sentence.
2. Line 235: The section and subsection numbering should be carefully checked throughout the manuscript. For example, the Methods section appears to skip Section 2.7, and the subsection numbering in Sections 3.2 and 3.5 also appears inconsistent with the manuscript structure.
3. Line 125: Table 1 appears to contain an extra horizontal line at the top. Please check the table formatting.
4. Line 159: The bottom part of Figure 1 appears to be incomplete or cut off. Please revise the figure layout.

5. Line 170: The definitions of RMSE should be checked. In Eq. (3), RMSE appears to be calculated using the sample mean $\overline{y_i}$ rather than the observed value y_i, n is 'the number of samples' not the 'numbers of sample'
6. Line 198: The DNN was trained for 2000 epochs, and the model at the epoch with the minimum MAE was selected, rather than using early stopping. Could the authors clarify why early stopping was not used for the DNN?
7. Line 257: At the beginning of Section 3.1, the manuscript appears to refer to Figure 3 when discussing model performance, whereas the observed-versus-modeled comparison seems to be shown in Figure 2. The figure numbering should be checked and corrected. Similar figure-reference errors also appear at Lines 297 and 654, and possibly in other parts of the manuscript. Please carefully check all figure references throughout the manuscript.
8. Line 355: Figure 4 contains panels (a–h), including ICA results for both Fesol% and d-Fe concentration, but the caption only describes panels (a–d) and refers only to IC scores versus Fesol%. The caption should be revised to include panels (e–h) for d-Fe concentration.
9. Line 373: The title of Section 3.4 contains "ここから見直す", which appears to be an internal editing note. Besides, too many figure references appear incorrect in this section. For example, at around Line 394, "Figures 5 and 5c" likely should be "Figures 5b and 5c". Around Line 397, IC1 is referred to as "Figure 6b", but the relevant panel appears to be Figure 5b. Around Line 430, the reference to "Figure 6d" also appears inconsistent and likely should be Figure 5d.
10. Line 481: In the Atlantic aerosol subsection, the text states that the model reproduces dissolved Fe concentrations in "North Pacific aerosols." This appears to be a typo and should be corrected to "Atlantic aerosols."
11. Line 720: The caption or text for Figure 12 should clarify what the yellow shaded region in panel (b) represents.
12. Section 3.7: The classification and interpretation of low-reproducibility samples appear internally inconsistent between the North Pacific and Atlantic sections. In the North Pacific section, Group I is defined as samples for which the model overestimates Fesol%, and these samples are characterized by Alsol% > Fesol% and [d-Fe]/[d-Al] < 0.1. However, in the Atlantic section, samples with similar characteristics are described as cases where the model underestimates Fesol%. Why does the same feature pattern correspond to model overestimation in the North Pacific but underestimation in the Atlantic? The authors should clarify whether the group definitions are consistent across basins or whether the Atlantic group descriptions contain an error.