This manuscript presents an innovative and practical approach to identifying foehn events using automatic weather station data, offering valuable insights into their climatology and impact on air pollution in Beijing. The strengths include the novel methodology, comprehensive spatial-temporal analysis, and the exploration of foehn-pollution interactions, which address a critical gap in the literature. However, the study could be enhanced by justifying key methodological choices, validating findings with additional data or techniques, and expanding the discussion of physical mechanisms. I have a few points that could be addressed to strengthen the manuscript and some minor comments.

General Comments:

1. The paper proposes a foehn identification method based solely on AWS data. While this approach is innovative, the authors should further clarify and justify the selection of criteria—for example, the choice of thresholds for temperature change, wind direction, and relative humidity. In several places, the rationale for setting specific ranges (e.g., wind directions "250–405°") is not fully explained. The upper bound exceeding 360° raises concerns about potential typographical or conceptual errors that need clarification. The authors should provide a detailed rationale explaining why this specific criterion was selected over alternatives (e.g., a fixed temperature threshold or a percentile-based approach). Additionally, a sensitivity analysis assessing how variations in this criterion affect foehn identification would strengthen the method's credibility and robustness.

The method uses temperature increases (>1°C per hour), humidity decreases, and specific wind directions to identify foehn events. However, these characteristics could also result from non-foehn processes, such as diurnal heating or synoptic-scale warm fronts, especially at downstream plain stations. The authors acknowledge potential overestimation at these sites (Section 6, lines 424–425), but more discussion or quantitative validation is needed to assess the method's specificity. How well does it distinguish foehn events from similar meteorological phenomena?

Besides, the use of the Self-Organizing Maps (SOM) technique is described, yet details on how the optimal number of nodes (six for Type I and four for Type II events) were determined are somewhat brief. A sensitivity analysis or additional justification would strengthen confidence in the clustering results.

3. I am not following the logic of relating the identified foehn events to air pollution. The authors filtered out high PM2.5 episodes first and then explored how the foehn events behave during these episodes. I think the proper way is the opposite by looking into how PM2.5 changes with and without foehn events. In addition, the study categorizes foehn impacts on PM2.5 into Type I (rapid decrease) and Type II (slight decrease followed by rapid increase), based on manual classification of 204 pollution episodes (Section 5, lines 386–389). This manual process may introduce subjectivity. Could an automated approach (e.g., clustering based on PM2.5 change rates and temperature increases) or statistical validation (e.g., significance testing of differences

between types) be applied to ensure objectivity and reproducibility? This would bolster confidence in the findings. The direct (strong pressure gradients) and indirect (weak gradients and boundary-layer changes) mechanisms linking foehn to pollution are well-described (Section 6, lines 429–451). However, the indirect mechanism—where weak foehn winds lead to pollutant accumulation via local circulations and inversions—could be elaborated with more evidence or modeling support. For instance, how does the "seesaw-like exchange" (line 446) manifest in wind fields or boundary-layer height data? Referencing specific observations from the case study (Figure 11) or prior work (e.g., Li et al., 2020) could clarify this process.

Minor Comments:

Lin 131-132: Pollution episodes are defined as periods with city-wide average PM2.5 >35 μg m^{-3} and a mean >75 μg m^{-3}. This threshold aligns with practical air quality concerns but a brief justification or comparison with established criteria in the literature would contextualize this choice and enhance its applicability.

Line 136: I think it is ERA5. Better to include the version here. The NCEP data mentioned in Line 159 is a typo? Also, The reanalysis data at 0.25° × 0.25° resolution for weather pattern classification is appropriate for large-scale patterns but may miss fine-scale topographic effects critical to foehn dynamics in Beijing's complex terrain. Have the authors considered higher-resolution datasets (e.g., regional models) to verify small-scale features? Discussing this limitation would strengthen the methodology.

Figure 1: White triangles are hard to see. You may need to make them thicker.

Line 193: How were these 22 cases identified?

Figure 2: Is CE-PAWS included in Non-NM-PAW? The national AWS is used to identify the events first. Is it possible that the regional stations have foehn events while national stations do not so you are missing some event days?

Figure 3: The color scale should be the same to have an easy comparison. The band-like distribution of foehn days from the northwestern mountains to the southeastern plains is visually compelling, but the paper lacks a detailed physical explanation. Is this distribution driven by the Taihang Mountains' topography, prevailing northwest winds, or a combination of factors? Linking the spatial pattern to specific atmospheric or topographic mechanisms would enhance the analysis.

Line 242: Any explanations about what drives this narrowing trend?

Table 2: While this seasonal variation is well-documented, the underlying reasons are not explored. Are winter maxima linked to stronger pressure gradients, more frequent cold fronts, or topographic amplification of downslope winds? Including a brief discussion of potential meteorological drivers would provide deeper insight into the climatology of foehn events in Beijing.

Table 3: Can relative humidity and wind conditions be discussed in addition to temperature?

Figure 8: It is confusing how the figure is made and more explanations in the texts would aid comprehension. It says “each episode’s initiation is marked by its Local Standard Time (LST)”, but the axis has larger numbers than 24. Also, the subtitle of this figure is not very clear caused by the semi-colons. The axis labels are too small to read, which also applies to some other figures (such as Figure 9).

Line 490. The AWS-based foehn identification method is a key strength, enabling long-term analysis with widely available data. However, the paper does not discuss its potential application to other regions with similar topography (e.g., other parts of the North China Plain or mountain-adjacent cities globally). Addressing generalizability would increase the study’s impact and relevance to the broader atmospheric science community.