Author response to Referee 3 (RC3)

All line numbers in this document refer to the revised manuscript.

RC3

Review preamble: "the manuscript already appears publishable with minor revisions" — acknowledge the repost note ("I initially misunderstood the review process…") graciously.

RC3-1

Comment: A fundamental question is the lack of a clear comparison baseline. Additional justification may be beneficial regarding why the proposed framework is particularly necessary under smoke conditions, unless there are cases where conventional approaches perform well under normal air-quality conditions but fail during smoke events.

Author response: This was a flaw of our study, so the revision adds in a comparison, while attempting to not further bloat the paper. New Section 3.6 (L418–440) defines three baseline neighbour-comparison rules. The absolute neighbour deviation, the neighbour z-score, and the median/MAD robust z-score. Each was computed on the same networks and hourly measurements as our methods, evaluated at their conventional cutoffs and additionally at the cutoffs that maximized their performance on our benchmark, against a common, method-independent outlier definition. Section 4.1 reports the outcome under both objectives (L514–537): under the precision-weighted objective the tuned equation leads ($F_{\beta=0.5}$ = 0.797 versus 0.778 for absolute neighbour deviation), while under F_1 the ranking inverts and absolute deviation wins (0.743) at three- to four-fold higher false-positive rates. To the reviewer's specific question, as to whether there are cases where conventional approaches perform well under normal air-quality conditions but fail during smoke, that case which prompted this work is now shown directly. Absolute neighbour deviation is capable in ambient air (44 false positives) but fails when smoke arrives (1,102), while the tuned equation's false-positive rate moves only modestly across the boundary (0.57% to 1.30%) (Section 4.1.3, Table S4).

Changes in manuscript: New Section 3.6 (L418–440); baseline-comparison paragraphs in Section 4.1 (L514–537); smoke-stratified false-positive comparison in Section 4.1.3; supplement Tables S3 and S4.

RC3-2

Comment: A stochastic spatiotemporal modeling framework may also be considered as a comparison approach… Including a brief discussion of physically informed spatiotemporal transport formulations in the Discussion section may help further contextualize the limitations and assumptions of the current network-based approach. For example: Choi, B., and Hummel, M. A. (2025)…

Author response: We appreciate the suggestion, and had we been aware of this work, it may have led to a different methodological approach. Given time and energy constraints, this had to be confined to an acknowledgement in the discussion. Section 6.4 now contains a physically-informed spatiotemporal paragraph (L908–912). Our networks encode proximity alone, whereas formulations such as the stochastic advection–diffusion model of Choi and Hummel (2025) represent pollutant transport explicitly, and could supply physically grounded expectations of inter-device agreement as a plume moves through a network. This helps the discussion contextualizes the assumptions and limitations of distance-based network generation as the reviewer suggested.

Changes in manuscript: Physically-informed future-work paragraph added to Section 6.4 (L908–912) with Choi and Hummel (2025) cited.

RC3-3

Comment: (Suggestion) Reducing redundant explanations in several sections may improve the flow of the manuscript.

Author response: Agreed. The reorganization consolidated all methods into Section 3 and removed redundant explanations across sections, the original Figure 9 was deleted, and reference material moved to the new supplement, reducing the main-text figure count from 17 to 12. We note the tension between the requests we received: shorten the manuscript while also providing additional information, comparisons, and clarification. This resolved by reducing figures, moving reference material to the supplement, and condensing the text where we were able.

Changes in manuscript: Reorganization and redundancy reduction as described in the response to Reviewer 2's first comment; figure count 17 → 12; new supplement.

Additional changes (author-initiated)

While revising Section 4.2, we identified an error in the originally reported XGBoost false-positive rate of 15.2%. That value was computed from the wrong column of an internal results table and does not describe the classifier's performance. Recomputed over the full evaluation dataset against the model's stated training target (every device-hour within a perturbation window), the false-positive rate is 0.26% at the 0.95 probability threshold. The corrected value appears in Table 3 (L571–572), which replaces the affected prose in Section 4.2; no other reported quantity was affected.

Additional changes in the final revision round, following co-author review: the device descriptions of Sect. 2.2 were condensed to a summary with full detail moved to the supplement (Sect. S1 and Table S1); the standalone false-positive-rate figure was consolidated into supplement Table S4, with the key values retained in Sect. 4.1.3; the duration-stratified validation table moved to the supplement as Table S6; and the local-smoke case study moved to the supplement as Fig. S5, reducing the main text to 10 figures and 3 tables. Two corrections were also made: the neighbour-density values in Sect. 5.3, previously mislabelled as percentages, are mean neighbour counts per device (12.8 versus 23.6 at 7,500 m).