

Summary and overall assessment

The authors compare two winter PM_{2.5} episodes in Henan (E1: 22 Dec 2023 to 3 Jan 2024; E2: 8 to 14 Jan 2024) under different emergency-control intensities. They build a daily emission inventory from CEMS, traffic, non-road machinery and satellite data. They find NO_x and primary PM emissions rose about 13% in E2, while PM_{2.5}, nitrate and OM rose about 36%, 44% and 25%. Partial correlations and WRF-CAMx simulations are used to attribute the increase to emissions.

The study has several strong points. The back-to-back episodes are a reasonable way of looking at this issue. Usage of CEMS data is rare and it adds value to the article; the same observation can be made about the satellite-based proxy for residential-combustion emissions. I find it also useful that the authors have decided to subset the meteorological screening into “per episode” parts. The authors are also open about the winter time nitrate underestimation in the model.

The main concerns I have with the study are: attribution (meteorology, chemistry and calendar effects are not separated from emissions), an incomplete and thinly analyzed modeling component, possible circularity in the residential proxy, and limited transparency on data and methods. Numerous numerical inconsistencies across sections also need fixing.

Recommendation: major revisions. The study has a strong design and rare activity data. It can be a great addition to the literature if its main problems are solved. However, the central causal claim, that weaker controls and the resulting emissions were the key driver of the E2 increase, is not yet supported by the analysis.

Major comments

- The model simulation cannot attribute the increase to emissions

The WRF-CAMx run changes meteorology and emissions together. Reproducing an E1 to E2 increase therefore does not show which factor caused it. A sensitivity run is needed, for example E2 with E1-level emissions, or fixed emissions throughout. Without it, the “closed-loop” and “mutually verify” claims are not supported. The model also reproduces only part of the observed change: PM_{2.5} +15.5% vs +33.9%, nitrate +16.2% vs +48.3%. Calling this “comparable within calculation uncertainties” is not justified, especially since the Fig. 7 error bars are spatial standard deviations across cities, not model or observational uncertainty. A fair reading is that the simulated emission changes explain roughly half the PM_{2.5} increase and a third of the nitrate increase. The authors need to explain this discrepancy which could be related to the next point.

- The modeling is under-documented and thinly analyzed

While an effort is made to document the modeling parameters (which is appreciated), it does ignore to talk about some important modelling choices. Table S5 gives the physics and grid configuration, but key inputs are missing (WRF input data sources, anthropogenic emissions outside of Henan, biogenic emissions, map of the domains, etc.)

The aerosol treatment is not described at all, although this is a particle-focused study. It would be important to add information such as (but not limited to) the size representation (two-mode vs sectional), the inorganic thermodynamics module and version, the SOA scheme (given that OM is a headline species), whether primary OA is treated as volatile, heterogeneous chemistry (in particular N₂O₅ uptake), which model species are summed into PM_{2.5}, nitrate and OM for the comparison with observations. The authors mention that they use CAMx ICBCPREP for boundary and initial conditions; would that mean they use the constant concentrations or time varying ones?

The paper attributes the nitrate bias to thermodynamic partitioning and heterogeneous chemistry without saying how the model treats them or why its attributed to the thermodynamic partitioning specifically.

Evaluation is limited to hourly PM2.5 ($R^2 = 0.35$, NME 47.9%) while the PM2.5. There is no meteorological evaluation, which could give some indication about model performance. No statistics for nitrate or OM are given either, which carry the conclusions. Fig. S6 plots simulated values of nitrates on a separate right-hand axis, which hides a roughly a factor of 4 underestimation for the model. Please use a common axis and report statistics per species and per site. Please try to explain why this is happening as well; it would seem the authors have data for all inorganic particles, it would be interesting to know if those show the same bias or not.

Some unusual choices need a brief justification: urban canopy on for the 36 km domain but off for the 12 km domain, WSM3 microphysics. Why these choices?

- The residential-combustion proxy may be partly circular

Residential emissions are scaled by a PM2.5/CO ratio from ML-based concentration products that are trained in part on ground observations. These emissions are then correlated with observed OM ($r = 0.848$). Some of that correlation may be built into the method. It would be interesting to know what the machine learning training sample included.

The PM2.5/CO ratio also rises when secondary aerosol forms, so part of the 17.5% residential increase in E2 may reflect the episode itself. Note also that E2 was about 3 C warmer than E1, which would usually reduce heating demand. Please discuss these points and, ideally, test the proxy against an independent indicator.

There is also a discrepancy of what the emission database represents vertically. In one section it is mentioned that emissions are "ground-level" while in another one it is mentioned that emissions are "tropospheric column". Please clarify which is used.

- The role of meteorology is dismissed too quickly

The EMI worsened by 24% between E1 and E2. Wind speed increased in E2, which normally favors dispersion, yet it is presented as part of the deterioration. The partial correlations control only for temperature, RH and wind speed. Boundary-layer height, likely the most important factor in winter haze, is not included. Please add it, or explain why it was omitted. Could the authors try to explain this?

- The nitrate formation analysis partly contradicts the main conclusion

Section 3.1 shows that E2 had markedly more favorable conditions for nitrate formation, both day and night. This is a non-emission explanation for a nitrate increase (+44%) that far exceeds the NO_x increase (+13%). The conclusions should weigh this more explicitly rather than presenting emissions as the key driver.

The formation indices include NO₂ (SSRD x NO₂ by day, NO₂ x O₃ at night), so they mix emission and condition effects. Please discuss this.

The method is non-standard and heavily approximated (SSRD as an OH proxy, steady-state N₂O₅), yet it appears only in Text S3 while its results are in the main text and conclusions. The core equations and assumptions should be moved to Section 2. Day and night anomalies are presumably each normalized to their own means (please confirm, its unclear), so plotting them as one hourly series is hard to interpret.

- Calendar effects

Are E1 and E2 both in non special non particular parts of the year? Are there any major holidays that could/would affect these episodes? The new year's holiday is in the middle of the first episode, would that affect the measurements/emissions and the differences between the two episodes? The calendar situation of the two episodes should be discussed, at least in a small paragraph.

- Statistical support for the causal claims is limited

The section titled "causal analysis" uses correlation. The analysis rests on about 20 to 25 daily points, and both emissions and concentrations are smooth, autocorrelated and trending over the window. The raw NO_x-nitrate correlation is 0.053. The nitrate argument rests on a partial r of 0.426 (p about 0.04). Please address autocorrelation and multiple testing, and soften the causal language.

The methods say x is an hourly concentration and y a daily emission. Please clarify how these were paired. If emissions are distributed on a daily profile it should be mentioned.

- Data, measurements and site representativeness

The data are not publicly available; the statement offers them on request and after acceptance. Please deposit the processed data and code, in line with the journal's data policy, or explain which components (for example traffic and manufacturer data) cannot be shared. I am not sure "data can be provided when paper gets accepted" is intended in the journal's policy.

The composition measurements are not described: instruments, time resolution, the OC-to-OM conversion factor. The OM results depend directly on that factor.

The five composition sites are justified qualitatively in Table S1 (one per sub-region), with no quantitative criteria. Please state How many composition sites exist in Henan and why these five were chosen. The data completeness for each site should also be noted. It would help the argument of the representative of these 5 stations if the five-site mean PM_{2.5} tracks the all-station provincial in the supplementary information.

It is unclear whether the PM_{2.5} in Fig. 2 and the mass fractions comes from the same five sites or from all cities. Fig. 2 also shows an unexplained gap between the stacked components and the PM_{2.5} line, "Others" is not defined and the gap seems substantial in the second episode. Please discuss mass closure and explain the gap.

Fig. 7 uses four cities instead of five; please explain the exclusion of Sanmenxia.

- References

There is a distinct lack of referencing. None of the tools (CAMx, WRF) are referenced, some of the mentioned data are not referenced (ERA5 for example).

Minor comments

Several headline values differ between sections. Please reconcile them and state which periods and sites each value refers to. For example, the PM_{2.5} increase from E1 to E2 is described as 42% (abstract), 36.5% (conclusions), 33.9% (Sec. 3.3.2); 104 to 142 $\mu\text{g}/\text{m}^3$ gives 36.5%. This occurs for other values and includes the values in SI (nitrates, OL, day/night formation anomalies, peak formation anomaly).

While it is somewhat obvious sometimes, each figure caption should mention if they're showing measurements or simulations (Fig. 1 and 2 are good examples). In Fig. 7, the error bars are spatial

standard deviations and should not be described as uncertainties (or at least made clear what they are). Explain why Sanmenxia is excluded.

In Eq. 1, the denominator sums over sources (i); it should presumably sum over days in the month. In Eq. S8, $SSRD < 0$ should presumably be $SSRD = 0$.