General comments:
The study by Wang et al. provides a PM2.5 data assimilation technique in a regional air quality model (AQMv7) and apply the modelling system to September 2020 western US wildfires. The authors conducted several experiments including a control run without any assimilation and runs with AirNow and PurpleAir data assimilated to demonstrate the improved forecast performance with data assimilation. The manuscript fits in the scope of the journal. However, there are several unaddressed issues that limits the strength of the manuscript for the reader in its current form. Hence I strongly suggest that the following comments are addressed before consideration of the manuscript for publication.

We thank the reviewer for the thorough evaluation of our manuscript and for the constructive comments and suggestions.

In response to the reviewer's comments, we have revised the manuscript accordingly and addressed the issues raised to strengthen the clarity and robustness of the study. Below, we provide point-by-point responses to the reviewer's comments, with **our replies shown in blue font. The changes in the revised manuscript are also shown in the blue font.**

Major comments:
(1) How is this study different than previous similar studies conducted over US (for example, compared to US based studies mentioned in lines 79-80). Novelty part should be articulated clearly.

Thank you for pointing this out. We have revised and expanded the manuscript to more clearly describe the novelty of this study relative to previous work. A new paragraph has been added to summarize the unique aspects of this research (see also the second-to-last paragraph of the Introduction), and the revised text is provided below.

> "This study aims to develop and evaluate an initial aerosol analysis capability for the AQMv7 system by assimilating $PM_{2.5}$ observations using the JEDI 3D-Var framework. Compared to previous $PM_{2.5}$ data assimilation studies, this research adopts the NOAA's regional operational AQMv7 system and incorporates a new $PM_{2.5}$ transform in JEDI for assimilating $PM_{2.5}$ observations. In addition to evaluating the impact of assimilating AirNow $PM_{2.5}$ measurements on air quality prediction, this study also examines the impact of assimilating low-cost PurpleAir observations. Although PurpleAir data are valuable for real-time air quality monitoring, their impact on numerical air quality prediction has not been thoroughly investigated. To the authors' best knowledge, this is the first study to demonstrate the value of PurpleAir observations for air quality prediction during a major fire event using the AQMv7 system."

(2) Usually for studies utilizing the air quality models, physics and chemistry options selected are usually reported. Does this model use a fixed suite of options or there are several options available for physical and chemical processes. In any case, the options used should be reported somewhere. Also, how are meteorological initial and boundary conditions prepared?

The model configuration of the operational NOAA regional AQMv7 system is described in detail by Huang et al. (2025). In that Bulletin of the American Meteorological Society paper (open access; DOI: 10.1175/BAMS-D-23-0053.1), the authors provide a comprehensive description of the AQMv7 configuration, including the physical and chemical parameterizations, emissions, and meteorological initial and boundary conditions.

In the revised manuscript, a new paragraph is added in Section 3.1 to summarize the physics and chemistry options, as well as the meteorological initial and boundary conditions. Here is the new paragraph:

> "In this research, the model configuration is almost the same as the operational AQMv7 setup except for running over the CONUS domain with a 3 hourly cycling interval. The AQMv7 system is configured over the CONUS domain with a grid-spacing of 13 km and 65 vertical levels, extending up to 0.2 hPa. The system uses the Global Forecast System version 16 (GFSv16) physics package within the Common Community Physics Package (CCPP) framework to generate the meteorological fields driving air quality predictions. Meteorological initial conditions and lateral boundary conditions are generated using GFS forecast outputs with lead times up to 30 hours at 3-hour intervals from the previous GFS cycle. Fire-related emissions are represented using real-time Regional hourly Advanced Baseline Imager (ABI) and Visible Infrared Imaging Radiometer Suite (VIIRS) Emissions (RAVE) data at 0.03° spatial resolution. Anthropogenic emissions are based on the 2016 U.S. EPA NEI Collaborative (NEIC2016v1) modeling platform. Gas-phase chemistry is simulated using the Carbon Bond Mechanism version 6 (CB6r3) with updated isoprene chemistry and revised photolysis rates. More detailed information including physics, chemistry options, anthropogenic emissions, and fire emissions about the model configuration can be found in Huang et al. (2025). "

Readers are referred to Huang et al. (2025) for a complete description of the model setup.

(3) There is no clear information on emission inputs? Which anthropogenic and natural emissions are included in the model.

Please see the reply to the previous comment. The emission info can be found in a new paragraph in Section 3.1.

(4) It would help the reader if there is a flow chart to explain the complete system used in this study.
A schematic of the data assimilation and forecasting cycles was added into Section 3.5 Experiments.

(5) Lines 245-246 - Why are the error standard deviations different for AirNow and PupleAir and how the results would change if error standard deviations are kept same for both AirNow and PupleAir (to say 10%). In other words, would it make more sense to compare DA_PA and DA_AN if error standard deviations are kept same?

The quality of AirNow data is superior to the low-cost PurpleAir PM data.

Below are reasons that PurpleAir PM was assigned larger errors than AirNow PM. PurpleAir sensors are designed for affordable, wide deployment by individuals or community groups. They use low-cost optical particle counters to count particles and estimate PM2.5 mass concentrations. These data require calibration to improve accuracy because the sensors measure particle counts and infer mass concentration using assumptions that do not always match real-world aerosol properties. In addition, the sensors are installed and maintained by individuals, either outdoors or indoors. For our application, PurpleAir data are reported by outdoor sensors, and must undergo quality control (QC) and apply appropriate corrections before being assimilated.

In contrast, AirNow/EPA data are collected using regulatory-grade instruments that physically collect and weigh particles. These instruments are installed, maintained, calibrated, and quality-checked by trained agencies, ensuring high accuracy and comparability. As a result, no additional QC or correction is required prior to data assimilation. The current observational errors setup might not be an optional one, but it is reasonable to initiate data assimilation cycles and evaluate data impact.

There are other objective methods to estimating observation errors in a data assimilation system. This will be investigated in the near future.

(6) Line 255-256 - Explain more clearly about these different thresholds for large PM2.5 observations mentioned for AirNow and PurpleAir?
The above lines were removed in the updated manuscript.

(7) Equation 9 is difficult to understand. Also, what is T in superscript?
In Equation (9), the superscript T denotes the transpose of a vector, and $E(\cdot)$ denotes the mathematical expectation operator. Equation 9 has been revised and its description extended in the revised manuscript. The notation has also been updated to ensure consistency with that used in the previous sections.

(8) Is there any possible explanation for the hourly changes (for example - first increase and then decrease in bias) in figure 3 and other similar figures like figure 10. Also, put axis title on the x-axis in figure 3?

We did discuss the temporal bias evolution and explained the bias at forecasts 1 from control and data assimilation experiments in the revised manuscript.

In general, the data assimilation experiments exhibit a similar temporal trend to the corresponding forecasts without data assimilation as data assimilation primarily corrects the model state and does not resolve inherent model bias.

This paper primarily focuses on the development of the data assimilation scheme and the evaluation of data impacts. We assume that the observed temporal bias evolution is mainly controlled by diurnal cycles/variations but dominated by certain forecast cycles and their interactions between fire emissions and meteorological processes, particularly boundary layer physics (e.g., the diurnal evolution of PBL height).

We have also added an explanation of the x-axis in the caption of Figure 4 as requested. Specifically, the x-axis represents forecast lead times from 1 to 24 hours.

(9) Line 418-421 - Can you briefly explain about performance diagram with some references? How are these thresholds chosen?

We have restructured this section to focus more closely on the main results and conclusions of the study. As part of this revision, the two figures showing the performance diagrams have been removed from the revised manuscript. Consequently, a detailed explanation of the performance diagram and the associated threshold selection is no longer included.

(10) Line 455-459 and figure 7 – I think it will be better if bias is shown for DA_AN, DA_PA and DA_ANPA with respect to AirNow (as a reference set of observations) for ease of comparison between figures 7 (b,d,f,h) for the forecast performance.
AirNow observations are already overlaid on the 1-h forecasts from each experiment in Figures 7a, c, and e. The purpose of these panels is to illustrate how the 1-h forecasts from the data assimilation experiments are improved relative to the control (CTR) run. For this reason, the CTR experiment is used as the reference in Figures 7d and f, which highlight the relative impact of data assimilation. By comparing these panels with Figure 7b, the improvements in the 1-h forecasts from the data assimilation experiments can be more clearly identified.

(11) Line 482-486 – Can the authors also discuss the performance EPA region-wise (as in lines 517-522) for 1h and 24h forecast. This may help to understand the performance in wildfires impacted regions more clearly.
We appreciate the reviewer's suggestion.
Following another reviewer's recommendation, we restructured the results section. A figure shows the statistical performance metrics for averaged 1–24 h $PM_{2.5}$ forecasts from control and data assimilation experiments over the 10 EPA regions are discussed in the revised section.

(12) Lines 531-532 – Mention some statistics to demonstrate improvement.
Following another reviewer's recommendation, this section has been removed in the revised manuscript to improve clarity and conciseness. The discussion of data assimilation performance has instead been incorporated into an expanded Section 4 (formerly Section 4.1), which now provides a more comprehensive description of the results.

Minor comments:
(1) Line 22 – particles with aerodynamic diameters
Done.

(2) Line 40-41 – "PM2.5 is a major contributor to poor air quality in US" - Provide some references to this line.
Added references. Also replaced 'major' with 'key' in the revised manuscript.

(3) Line 44 – Kindly avoid the usage of term 'primary pollutants' in this context.
We replaced 'primary' with 'important' in the revised manuscript.

(4) Line 76 - replacing the previous offline-coupled the Global Forecast System
Done.

(5) Line 77-78 – "modeling system" repeated twice.

Fixed.

(6) Line 88 – Put 2.5 here (and in other places in manuscript) as subscript
Fixed.

(7) Line 170 – Check the sentence. Should it be: It is noted that a recipe that uses temperature…
The sentence is rephased to:
It is noted that a recipe to derive dry air density from temperature, surface pressure, and delta pressure has been added to VADER for cases where the dry air density is not provided.

(8) Line 203 - Mmountain Rregion?
Done.

(9) Line 213 – Add the last access date.
Done.

(10) Line 242 – and Texas.
Done.

(11) Lines 309-312 – Reduce the sentence length. Break into two sentences for ease of read.

We reduced the sentence length by breaking it into two sentences for easier reading. The sentence has been rephased as fellows:

"Using the innovations and observation errors from subsection 2.2 and 3.2.3 as inputs to Equation 9, the background error variance of $PM_{2.5}$ was first estimated. This error, along with the background values, was then used in Equation 8 to estimate the scaling factor s. "

(12) Lines 345-346 – Is it "the ratio of the particle number concentration to total particulate volume"? Particle number concentration is already number per unit volume.

We agree that particle number concentration is already expressed per unit volume. However, we are not introducing a new variable; this simply states that the ratio of these two variables in the analysis is assumed to be the same as in the background. This allows the particle number concentration to be updated in the analysis according to the updated total particulate volume after data assimilation.

(13) Line 422 - Figure 6 presents shows time series of
Fixed.

(14) Line 580 – September 2025 or 2020?
Sorry for the typo. It is September 2020.

(15) Line 593 – There are two full stops before and after the reference.
Fixed.