

General Comments

The authors compare WRF-CAMx simulated column NO₂ at 444 m resolution with Pandora, GCAS, and TROPOMI observations. The model is biased low against Pandora and has poor agreement with GCAS NO₂ in downtown Houston, suggesting a potential underestimate of vehicle NO_x. The main general comment is that the authors should provide more support for their model performance before making conclusions about emissions. Could the authors compare the model vertical profiles of temperature and relative humidity against weather soundings, and/or the model ozone profiles against TRACER-AQ ozonesondes? This would better support that the difference between GCAS and the model is due to emissions. In addition, it would be helpful to show the model comparison against any surface meteorology or air quality data (NO₂, ozone?) available during the campaign. With this major revision and the comments below, this paper provides a useful study of the pitfalls of using TROPOMI for constraining emissions and the benefit of high resolution modeling and aircraft data.

Specific Comments

Line 43 – Does NO₂ exceed the NAAQS in Houston? Or do you mean that elevated NO₂ is indicative of higher emissions from combustion sources generally and is a precursor for ozone and PM_{2.5}? Where does Houston fall in compliance for the NAAQS for these species?

Line 84 – What do you mean by “remote-sensing columns”? Satellite?

Line 159 – This is confusing. Why would you compare Pandora to re-gridded GCAS and not the native resolution GCAS?

Line 160 – 30 minutes seems like a long time given that direct sun Pandora data are more frequent and could be matched more closely. What is the variability in the data over that window and is it really reasonable to use such a long time period?

Line 190 – For comparison to TROPOMI, you need to regrid the model to the coarser TROPOMI resolution of 3.5x5.5 km², otherwise the comparison will certainly look poor.

Line 242 – Can you explain what you mean by “Third, we reprocess link-based on-road mobile emissions for the higher resolution domains.”? What is the native resolution of the emissions? Is this some sort of down-scaling procedure?

Line 249 – High degree of uncertainty does not seem like a good reason to remove fire or lightning emissions. Does the model suggest that they are unimportant in Houston and thus you remove them to save computational expense?

Line 256 – Why are you including GEOS-CF in this comparison? You haven’t given us a motivation to do this.

Line 332 – Do you think the model successfully captures horizontal advection and vertical mixing in Houston? How could errors in transport relate to the model underestimate in NO₂ columns? Does the model perform significantly differently on days with slower or faster wind speeds?

Line 348 – Again, any windspeed or direction dependence? What about hour of day dependence?