Response to the reviewer of:

An intercomparison of satellite, airborne, and ground-level observations with WRF-CAMx simulations of NO_2 columns over Houston, TX during the September 2021 TRACER-AQ campaign

We thank the reviewer for their additional round of comments on the manuscript titled above. In this document we respond to their comments using the following notation: plain text indicates the reviewers' comments, **bolded text** indicates our responses, and *italicized text* indicates quotations from the original and updated manuscripts. For quotations, we will indicate whether they come from the original manuscript or the updated manuscript using "before" and "after", respectively; new text is indicated with red coloring. We have endeavoured to address the concerns of the reviewers and believe that these changes have improved the quality of the work.

Anonymous Referee #2:

Major Comments:

The authors say there is good agreement also against the surface observations, but how do I know that their agreement is good? Table S6 gives an R2 for wind direction of 0.26 which is not very good.

The authors should better discuss the meaning of their meteorological comparison statistics and how any biases could impact their results. This could be a very positive discussion that could inform future studies about what model biases are impacting emissions assessments. Currently the manuscript is a little confusing about what is 'good agreement' and what is a meaningful bias that needs to be discussed.

Thank you for this comment. You are right that "good" agreement is not sufficiently objective. We have rephrased the description to highlight the performance metrics. The model is being used to represent the dispersion of pollution. We note that the GCAS retrievals show that the plumes are indeed being transported in the same direction in the model as in reality. Because the wind speed bias is so minimal, the speed of advection and associated dispersion should be accurate despite the low R^2. We revised the presentation of Table S6 to clarify that it refers to *wind speed* not *wind direction*. We find a correlation in simulated wind direction (Table S6) compared to observations of R^2=0.76 and a minimal bias of MBE=-8.0°. There is a mean absolute error of MAE=26° across all the ground-monitors, hours, and days. We agree that wind speed performance (Table S5) is indeed poorer (R^2=0.26) and that there is some unsystematic error (MAE=1.2 m/s) but almost no systematic error across all days (MBE=-0.02 m/s). We agree with your larger point – that we could more clearly discuss what we mean by *good performance* and refer to specific meteorological conditions – and have modified the text to address this:

Before:

Generally, meteorological conditions simulated by WRF agree with ground-level observations especially on the more data rich non-cloudy days that are the most important for our intercomparison. Across all days, the WRF wind direction MBE was -8°, wind speed MBE was -0.02 m/s, temperature MBE was 0.39 K, and water vapor mixing ratio MBE was -1.45 g/kg.

After:

Generally, meteorological conditions simulated by WRF agree with ground-level observations especially on the more data rich non-cloudy days that are the most important for our intercomparison; ***however, performance depends on the specific measure of meteorology considered.*** *Across all days, the WRF wind direction was well correlated* ***(R^2=0.76) and had minimal bias (MBE=8°) but some unsystematic errors (MAE=26°) compared to observations. This indicates that the model generally captures variability in wind direction without a notable bias; however, considering any individual observation the simulated direction may differ by 20°-30°. For non-cloudy days – that are more relevant for our intercomparisons due to more***

data – correlation for wind direction was similar (R^2=0.73) and the bias and error were reduced (MBE=-5° and MAE=21°). Simulations of wind speed were more poorly correlated (R^2=0.26) and had some unsystematic error (MAE=1.20 m/s); however, there was very little systematic bias in the wind speed simulation (MBE=-0.02 m/s). Correlation and unsystematic errors improve on the non-cloudy days (R^2=0.37 and MAE=1.08) while there is still no notable systematic bias (MBE=-0.13 m/s). Considering wind speeds at 9am and 1pm (Fig. S2-11), it appears that observations in the afternoon degrade correlation compared to the morning and that, generally, simulated wind speeds are better correlated with observations in downtown Houston than in the south-eastern part of the domain near the Galveston Bay. Comparisons between GCAS observations and WRF-CAMx simulations show that the model represents the dominant direction and dispersion of identifiable plumes from known sources. The wind speed bias is sufficiently low that model uncertainty will not lead to systematic errors in plume advection.

Before:

Lastly, given the relatively minimal biases in WRF simulated meteorology compared to observations, low NO_2 biases in the simulated CAMx column concentrations imply that current TCEQ NO_X emissions inventories in the Houston area used to drive the CAMx simulation may be underestimated, and that this underestimation is likely attributable to a source with weekday-weekend differences and correlated with roadways and/or population density.

After:

While there are some errors in the meteorology – notably only a modest correlation between simulated and observed wind speed, albeit with little systematic bias, and mixed capturing of vertical structure compared to ozonesondes observations – these errors are unlikely to fully explain the low bias in simulated NO_2. *Given the relatively minimal biases in WRF simulated* ***wind speed and direction at the surface*** *compared to observations, low NO_2 biases in the simulated CAMx column concentrations imply that current TCEQ NO_X emissions inventories in the Houston area used to drive the CAMx simulation may be underestimated, and that this underestimation is likely attributable to a source with weekday-weekend differences and correlated with roadways and/or population density.*

Before:

This overall underestimate in the CAMx simulations is potentially attributable to a number of confounding factors including an inability of the WRF simulation to capture local meteorology and an underestimate of emissions in sectors that are more spatially located in downtown and western Houston like on-road mobile emissions.

After:

This overall underestimate in the CAMx simulations is potentially attributable to a number of confounding factors including an inability of the WRF simulation to capture local meteorology ***–***

WRF simulated wind speeds had only modest correlation with observations (R^2=0.26) although there was little systematic bias (MBE = -0.02) *– and an underestimate of emissions in sectors that are more spatially located in downtown and western Houston like on-road mobile emissions.*

Before

We generally find good agreement in the WRF simulated meteorology (Table S5-S8 and Fig. S2-11); however, the non-systematic differences in wind direction on the order of 20° would likely degrade correlation between observed and simulated NO_2 columns. Given that there is no apparent systematic bias in the meteorology, the negative bias in NO_2 columns is likely attributable to an underestimate of NO_X emissions.

After

We ***find that the WRF simulated wind direction (R^2=0.76 and MBE=8°), temperature (R^2=0.71 and MBE=0.39K), and water vapor mixing ratio (R^2=0.86 and MBE=-1.45 g/kg)*** *(Table S5-S8 and Fig. S2-11****) are generally well correlated and minimally-biased compared to observations****; however,* ***there are some unsystematic*** *errors* ***in wind direction (MAE=26°) and poor correlation in wind speed (R^2=0.26) that*** *would likely degrade correlation between observed and simulated NO_2 columns.* ***While there are errors in the meteorological conditions, the biases at the surface are all small – including minimal bias in the wind speed (MBE=-0.02 m/s) – indicating that*** *the negative biases in NO_2 columns are likely attributable to an underestimate of NO_x emissions;* ***however, the WRF meteorological performance could partially explain the poor correlation and absolute errors in simulated NO_2 columns.***

I would have preferred to see scatterplots rather than (or in addition to) tables to determine whether the R2 values were being driven by outliers or overall model performance.

We agree with you that whether R^2 values are being driven by outliers is an important consideration and a feature that should be discussed in the main text. While we do not include scatterplots to address this, the influence of outliers can be inferred from Figures S2-11. These figures reveal that poor performance in wind speed is possibly driven by values in the afternoon and the emergence of the Bay/Gulf breeze as there is generally improved agreement in wind-speed in the morning. Additionally, these figures reveal that simulated wind speeds at the downtown sites – that are most relevant to our conclusions regarding bias in the emissions inventory – generally perform better than the south-east sites near the Galveston Bay. We have now modified the text to address this:

New Text:

Considering wind speeds at 9am and 1pm (Fig. S2-11), it appears that observations in the afternoon degrade correlation compared to the morning and that, generally, simulated wind speeds are better correlated with observations in downtown Houston than in the south-eastern part of the domain near the Galveston Bay.

I am still concerned by the lack of evaluation of model vertical structure. As the model has a strong negative bias in NO2, it is possible this is due to errors in wind direction (as the authors suggest) but also to potential excessive vertical mixing. As vertical structure is essential to successful satellite interpretation, I do not agree that comparisons between vertical profiles of temperature, RH etc is out of the scope of this project. I do not see how this is different than the comparison done with the surface stations. I would again strongly suggest the authors add an evaluation against the data from the TRACER-AQ ozonesondes (https://www-air.larc.nasa.gov/cgi-bin/ArcView/traceraq.2021?SONDE=1) that include temperature, pressure, windspeed, wind direction, RH, and of course ozone. Possibly the model output was not saved for this which then certainly would be burdensome to produce? If so, then a statement about how this could be done in the future (or with other datasets like HSRL-2 or TolNet) would be helpful.

We thank you for this comment, and we agree that model vertical structure is an important consideration when evaluating model performance for column quantities, although we note that columns are likely less sensitive to vertical structure than ground-level measurements. To address this comment, we have briefly compared model vertical structure against the ozonesondes at five different sites representing different days and times that we have now added to the supplement as Figures S12-S16 (included in this document for convenience). We note that we do not include RH as this model output was not saved out except at the surface. We find that the model simulates temperature and pressure well across all five ozonesondes; however, vertical profiles in O_3 mixing ratio, wind speed, and wind direction are more mixed. For some sondes (Figure S16-S17) we find great agreement, especially below 1km; however, for others there does appear to be poorer correlations and error (Figure S18). The caveat is that O_3 has a much longer atmospheric lifetime than NO_2 so nothing definitive can be concluded by this analysis.

To make more conclusive statements about the performance of the model's vertical structure we would need to consider observations from all the ozonesondes (around 100); however, this would require substantial analysis on our end to process these data and accurately represent performance in summary statistics that account for the heterogeneous distribution of observations from different areas, days, and time periods. We agree that this analysis could further support our conclusions; however, this could be a separate study in its own right so we believe that this more detailed investigation is beyond the scope of this study but is worth investigating in the future. We have now modified the text:

Before:

Although we evaluate the performance of WRF at the surface, we do not consider vertical advection and the vertical mixing scheme in WRF merits further investigation.

After:

*Although we **primarily** evaluate the performance of WRF **meteorology** at the surface, **we also briefly investigate model vertical structure for five ozonesondes from different locations and***

days (Fig. S14-S18) and find great agreement in temperature and pressure; however, there is more mixed agreement in the ozone mixing ratio, wind speed, and wind direction. Future evaluation of 3D model simulated vertical structure for NO_2 using observations from NASA – such as measurements from the High Spectral Resolution Lidar 2 (HSRL-2) instrument, the Tropospheric Ozone Lidar Network (TolNet), or TRACER-AQ – may be helpful for diagnosing the distinct influences of emissions, meteorology, and chemistry on column NO_2.

Before:

Additionally, there can be substantial differences in vertical mixing coefficients in different schemes in the models, and these can impact the biases in column concentrations (de Foy et al., 2007; Riess et al., 2023).

After:

Additionally, there can be substantial differences in vertical mixing coefficients in different schemes in the models, and these can impact the biases in column concentrations (de Foy et al., 2007; Riess et al., 2023). ***We briefly compare meteorology and the ozone mixing ratio in the WRF-CAMx simulation to ozonesondes data (https://www-air.larc.nasa.gov/cgi-bin/ArcView/traceraq.2021) and find that while temperature and pressure are captured well, there is variable performance in the vertical structure for ozone mixing ratio, wind speed, and wind direction (Fig. S14-S18).***

Before:

Additionally, we compare simulated hourly NO_2 (Fig. S12) and maximum daily eight-hour average or "MDA8" O_3 (Fig. S13) to observations from seventeen TCEQ continuous air monitoring stations (CAMS) operating in Houston. We find poor performance and a strong negative bias in the simulated surface-level NO_2 (NMB=-59%) and generally good performance in the simulated surface-level MDA8 O_3 (NMB=-2.5%) compared to observations

After:

Additionally, we compare simulated hourly NO_2 (Fig. S12) and maximum daily eight-hour average or "MDA8" O_3 (Fig. S13) to observations from seventeen TCEQ continuous air monitoring stations (CAMS) operating in Houston. We find poor performance and a strong negative bias in the simulated surface-level NO_2 (NMB=-59%) ***while*** *simulated surface-level MDA8 O_3* ***has a much weaker bias*** *(NMB=-2.5%) compared to observations.* ***Comparisons to ozonesondes (Fig. S14-S18) suggest that WRF simulates more aggressive vertical mixing than what is observed; this is consistent with our findings of a stronger negative bias at the surface-level than for the columns as emitted NO_2 at the surface is advected vertically quicker in WRF-CAMx than in reality.***

Minor Comments:

Line 170 – Please add the swath size for GCAS.

We thank you for this comment. We have already mentioned the swath size but not the average pixel size which we have now added to the text:

Before:

The flight strategy of the aircraft included flying the plane in a 'lawnmower' fashion with flight lines spaced 6.3 km apart, ensuring overlap at flight altitude (FL280) with the instrument field of view of 45 degrees creating one gapless map of NO2 up to three times per flight day.

After:

The flight strategy of the aircraft included flying the plane in a 'lawnmower' fashion with flight lines spaced 6.3 km apart, ensuring overlap at flight altitude (FL280) with the instrument field of view of 45 degrees creating one gapless map of NO2 up to three times per flight day ***with an average differential slant column pixel size of 250 m × 250 m.***

Line 237 – Previously I commented "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." The response is as follows:

"In section 3.4 we have done this already and found generally comparable performance albeit improved correlation when comparing results at a coarser resolution."

Please add a reference to Section 3.4 about this then on line 237, otherwise as a reader I would be thrown off.

We agree that it would be good to reference Section 3.4 here so that the reader is not thrown off and have now updated the text to reflect this:

Before:

Spatially, we identify the CAMx grid cell in which each Pandora instrument is located and only consider TROPOMI measurements that were regridded to these grid cells

Before:

Spatially, we identify the CAMx grid cell in which each Pandora instrument is located and only consider TROPOMI measurements that were regridded to these grid cells. ***We intercompare GCAS, TROPOMI, and CAMx at this resolution but also compare the three datasets at a coarser resolution (Section 3.4) to account for resolution-dependent errors.***

Line 446 - The authors state that there is a non-systematic difference in wind direction on the order of 20°, but then state two sentences later that there is no apparent systematic bias in meteorology. This is contradictory.

We apologize for the confusion. In the original text what we were trying to convey is that the model does have some random error (i.e., unsystematic; MAE) but it is not biased (MBE). The original text was confusingly written and did not properly emphasize this point, so we have now modified it to reflect this:

Before

We generally find good agreement in the WRF simulated meteorology (Table S5-S8 and Fig. S2-11); however, the non-systematic differences in wind direction on the order of 20° would likely degrade correlation between observed and simulated NO_2 columns. Given that there is no apparent systematic bias in the meteorology, the negative bias in NO_2 columns is likely attributable to an underestimate of NO_X emissions.

After

We ***find that the WRF simulated wind direction (R^2=0.76 and MBE=8°), temperature (R^2=0.71 and MBE=0.39K), and water vapor mixing ratio (R^2=0.86 and MBE=-1.45 g/kg)*** *(Table S5-S8 and Fig. S2-11*___) are generally well correlated and minimally-biased compared to observations___*; however,* ***there are some unsystematic*** *errors* ***in wind direction (MAE=26°) and poor correlation in wind speed (R^2=0.26) that*** *would likely degrade correlation between observed and simulated NO_2 columns.* ***While there are errors in the meteorological conditions, the biases at the surface are all small – including minimal bias in the wind speed (MBE=-0.02 m/s) – indicating that*** *the negative biases in NO_2 columns are likely attributable to an underestimate of NO_x emissions;* ***however, the WRF meteorological performance could partially explain the poor correlation and absolute errors in simulated NO_2 columns.***

Also, Tables S5-S6 give statistics for all days and non-cloudy days, but not windy vs. calm.

We thank you for this comment. While we do not specifically include average statistics for windy vs. calm days this can be easily inferred by looking at the wind speeds in Table S6. We find that on the days with calmer conditions (< 3 m/s) there is usually poorer corelation (R^2=0.07, 0.1, and 0.25) while on windier days (> 4 m/s) there is better correlation (R^2=0.5 and 0.32). Model performance is well established as being poorer when simulating calmer conditions than windier conditions (e.g., Yu et al., 2022) which is consistent with our results. We have now added a sentence to specifically discuss windy vs. calm conditions:

New Text:

We also note that generally, the model performance is stronger on windier days – when speeds exceed 4 m/s (R^2=0.5 and 0.32) – than on calmer days – when speeds are below 3 m/s (R^2=0.07, 0.1, and 0.25).

Line 550 – Can you give us some more information about the performance of the scheme used here (YSU?) compared to others?

We thank you for this comment, we have indicated all the schemes used in Table S1. You are correct that we use YSU. A previous study evaluated the performance of YSU for the same period and domain (Liu et al., 2023). We have now mentioned this in the text:

New Text:

We note that the YSU scheme used in the WRF-CAMx simulation (Table S1) has been shown to underestimate PBL height in the Houston area during the TRACER-AQ campaign (Liu et al., 2023) which would likely impact the vertical distribution of NO_2.

Line 990 – The authors have discussed errors in meteorology throughout the paper including a bias of 20° in wind direction. I am not convinced errors are "minimal". It would be better to say that: with caveats that there are some errors in meteorology, they are unlikely to fully explain the low NO2 bias in the CAMx column and some of this bias may be attributable to underestimated emissions.

We are sorry for the confusion; we do want to note that there is not a bias (MBE) of 20° in wind direction but there is an unsystematic error (MAE). We believe that we have now responded to this comment in response to your first major comment and have also included your suggested caveat of our conclusions in the main text:

New Text:

While there are some errors in the meteorology – notably only a modest correlation between simulated and observed wind speed, albeit with little systematic bias, and mixed capturing of vertical structure compared to ozonesondes observations – these errors are unlikely to fully explain the low bias in simulated NO_2.

Figures and tables referenced in responses:

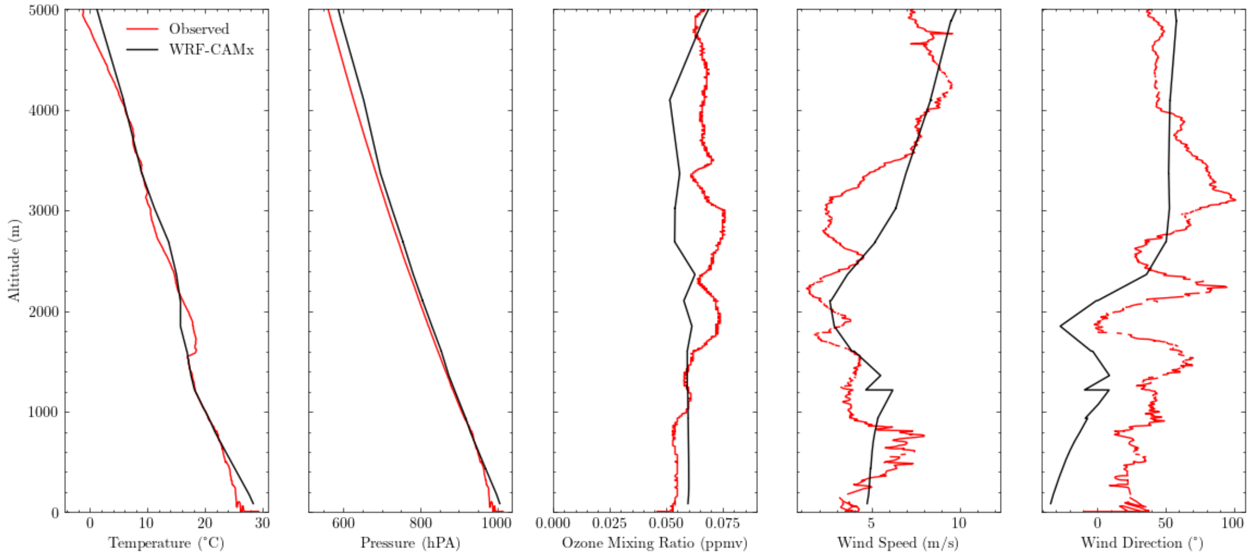

Figure S14: Comparison between WRF-CAMx simulated meteorology and ozone mixing ratios and ozonesondes observations at 11 am on 9/8 at 29.324° N and 94.552° W (Gulf)

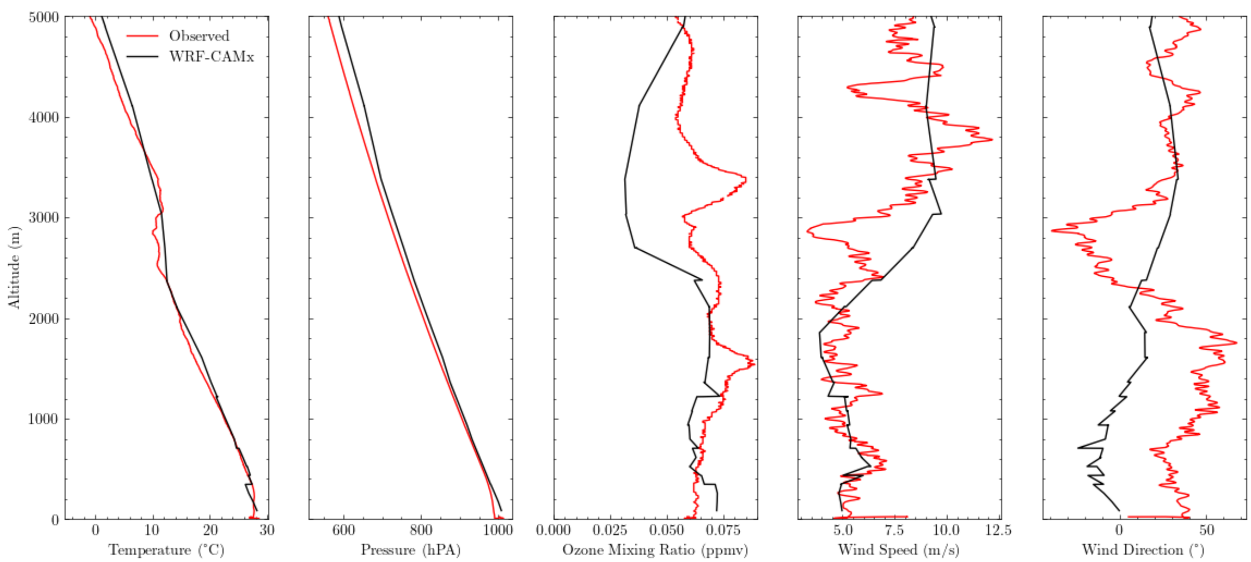

Figure S15: Comparison between WRF-CAMx simulated meteorology and ozone mixing ratios and ozonesondes observations at 10am on 9/9 at 29.383° N and 94.831° W (Galveston Bay)

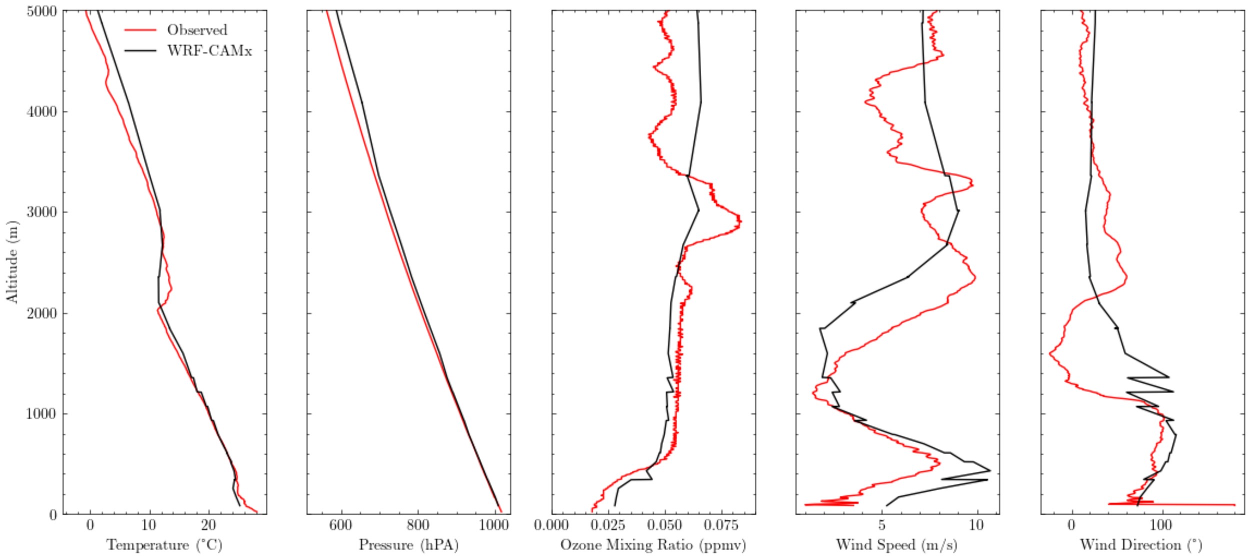

Figure S16: Comparison between WRF-CAMx simulated meteorology and ozone mixing ratios and ozonesondes observations at 8am on 9/10 at 29.724° N and 95.339° W (University of Houston)

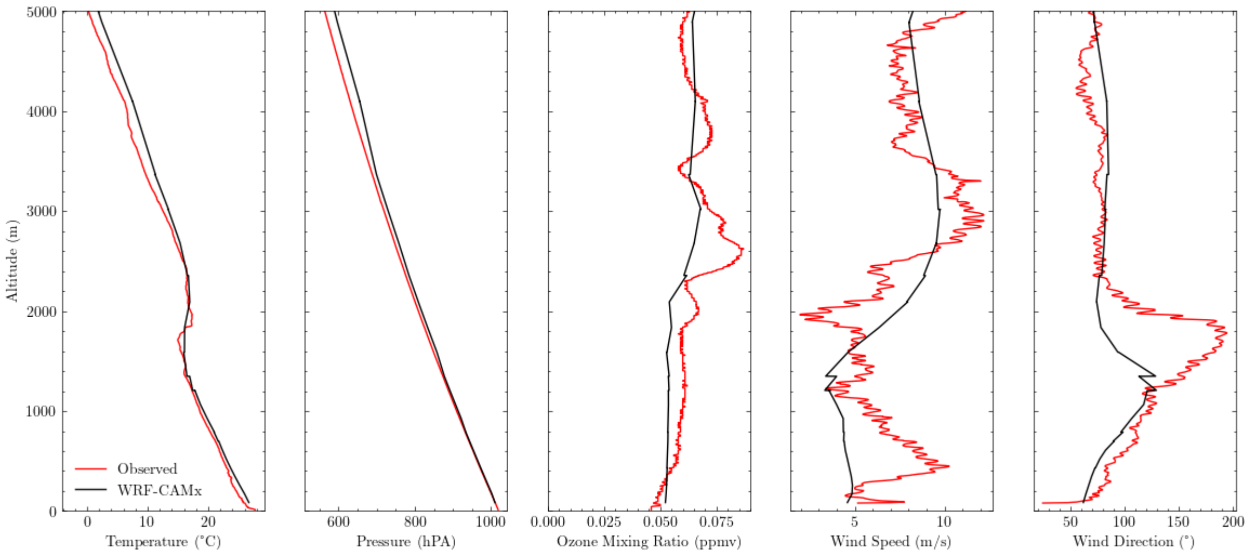

Figure S17: Comparison between WRF-CAMx simulated meteorology and ozone mixing ratios and ozonesondes observations at 9am on 9/11 at 29.67° N and 95.06° W (LaPorte)

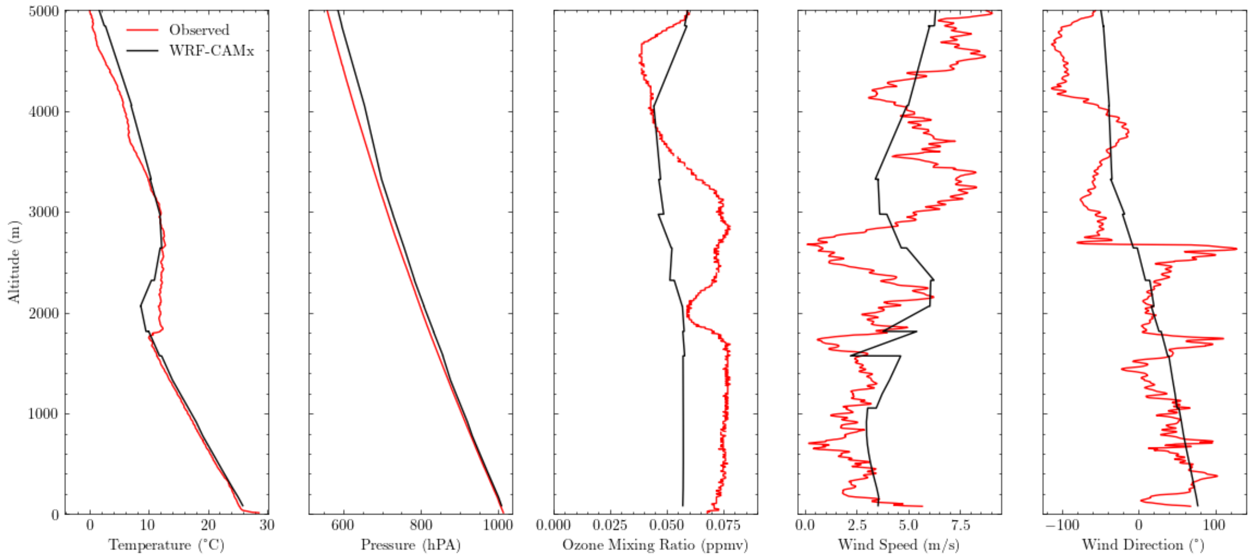

Figure S18: Comparison between WRF-CAMx simulated meteorology and ozone mixing ratios and ozonesondes observations at 1pm on 9/23 at 29.546° N and 95.53° W (Houston SW Airport)

Table S1: WRF physics options and data sources

WRF Option	Option Selected
Analysis Data	0.25° GDAS (IC/BCs and analysis nudging on the 36 and 12 km domains)
Microphysics	Thompson
Longwave Radiation	Rapid Radiative Transfer Model (RRTMG)
Shortwave Radiation	RRTMG
Surface Layer Physics	Revised MM5 surface layer scheme
LSM	Noah
PBL scheme	Yonsei University (YSU)
Cumulus scheme	Multi-Scale Kain-Fritsch (MSKF) on 36/12 km; none for 4/1.333/0.444 km

Table S5: Comparison of observed (OBS) and simulated (WRF) wind direction and associated statistics. Red shading indicates days with limited GCAS observations due to cloud coverage. The statistical measures of mean bias error (MBE), mean absolute error (MAE), and Pearson-R squared (R^2) are computed using the astropy circular statistics python module (https://docs.astropy.org/en/stable/stats/circ.html).

Date	WRF Dir (°)	OBS Dir (°)	R^2	MBE	MAE	N
09/01	184.0	180.0	0.04	-1.0	36.0	130
09/03	136.0	185.0	-0.04	-49.0	60.0	137
09/08	13.0	17.0	0.57	-10.0	29.0	137
09/09	20.0	37.0	0.78	-11.0	24.0	149
09/10	81.0	75.0	0.38	6.0	17.0	149
09/11	89.0	92.0	0.68	-3.0	14.0	151
09/23	63.0	88.0	0.3	-24.0	28.0	118
09/24	87.0	97.0	0.47	-8.0	20.0	137
09/25	70.0	59.0	0.57	9.0	24.0	114
09/26	97.0	96.0	0.8	0.0	19.0	117
All Days	82.0	87.0	0.76	-8.0	26.0	1339
Not Cloudy Days	67.0	73.0	0.73	-5.0	21.0	1072

Table S6: Comparison of observed (OBS) and simulated (WRF) wind speed and associated statistics. Red shading indicates days with limited GCAS observations due to cloud coverage. The statistical measures of mean bias error (MBE), mean absolute error (MAE), and Pearson-R squared (R^2) are defined at the end of this supplement in section S2.

Date	WRF Spd (m/s)	OBS Spd (m/s)	R^2	MBE	MAE	N
09/01	3.47	2.93	0.07	0.55	1.51	145
09/03	3.46	3.17	0.0	0.29	1.8	151
09/08	3.46	2.79	0.1	0.68	1.25	150
09/09	3.7	3.76	0.23	-0.06	0.94	152
09/10	4.6	4.69	0.5	-0.09	0.98	149
09/11	4.11	5.04	0.32	-0.93	1.35	151
09/23	3.14	3.2	0.2	-0.06	1.17	130
09/24	3.32	3.77	0.41	-0.45	0.96	142
09/25	2.73	2.66	0.25	0.07	0.94	135
09/26	3.1	3.29	0.36	-0.2	0.99	131
All Days	3.53	3.55	0.26	-0.02	1.2	1436
Not Cloudy Days	3.55	3.68	0.37	-0.13	1.08	1140

Table S7: Comparison of observed (OBS) and simulated (WRF) temperature and associated statistics. Red shading indicates days with limited GCAS observations due to cloud coverage. The statistical measures of mean bias error (MBE), mean absolute error (MAE), and Pearson-R squared (R^2) are defined at the end of this supplement in section S2.

Date	WRF Temp (K)	OBS Temp (K)	R^2	MBE	MAE	N
09/01	305.5	304.2	0.0	1.3	2.28	159
09/03	305.0	303.0	0.0	1.99	2.61	158
09/08	304.5	304.1	0.75	0.42	1.16	160
09/09	305.1	304.4	0.78	0.66	1.15	160
09/10	303.1	303.2	0.76	-0.1	0.93	158
09/11	302.5	302.3	0.74	0.18	0.99	157
09/23	298.4	298.6	0.69	-0.26	1.14	147
09/24	299.2	299.3	0.72	-0.11	1.02	150
09/25	299.5	299.9	0.81	-0.33	1.02	148
09/26	300.0	300.1	0.69	-0.02	1.18	150
All Days	302.4	302.0	0.71	0.39	1.35	1547
Not Cloudy Days	301.6	301.5	0.86	0.06	1.07	1230

Table S8: Comparison of observed (OBS) and simulated (WRF) water vapor mixing ratio (WVMR) and associated statistics. Red shading indicates days with limited GCAS observations due to cloud coverage. The statistical measures of mean bias error (MBE), mean absolute error (MAE), and Pearson-R squared (R^2) are defined at the end of this supplement in section S2.

Date	WRF WVMR (g/kg)	OBS WVMR (g/kg)	R^2	MBE	MAE	N
09/01	17.5	19.4	0.26	-1.92	1.97	99
09/03	17.7	18.6	0.21	-0.88	1.22	98
09/08	12.3	12.2	0.78	0.09	1.31	100
09/09	10.7	13.8	0.28	-3.07	3.11	100
09/10	9.1	10.7	0.67	-1.51	1.6	98
09/11	10.6	12.6	0.21	-2.03	2.28	97
09/23	6.0	7.6	0.54	-1.55	1.58	97
09/24	7.4	8.8	0.32	-1.34	1.46	100
09/25	8.2	9.7	0.54	-1.51	1.53	98
09/26	10.8	11.6	0.48	-0.76	0.87	100
All Days	11.1	12.5	0.85	-1.45	1.69	987
Not Cloudy Days	9.4	10.9	0.64	-1.46	1.72	790

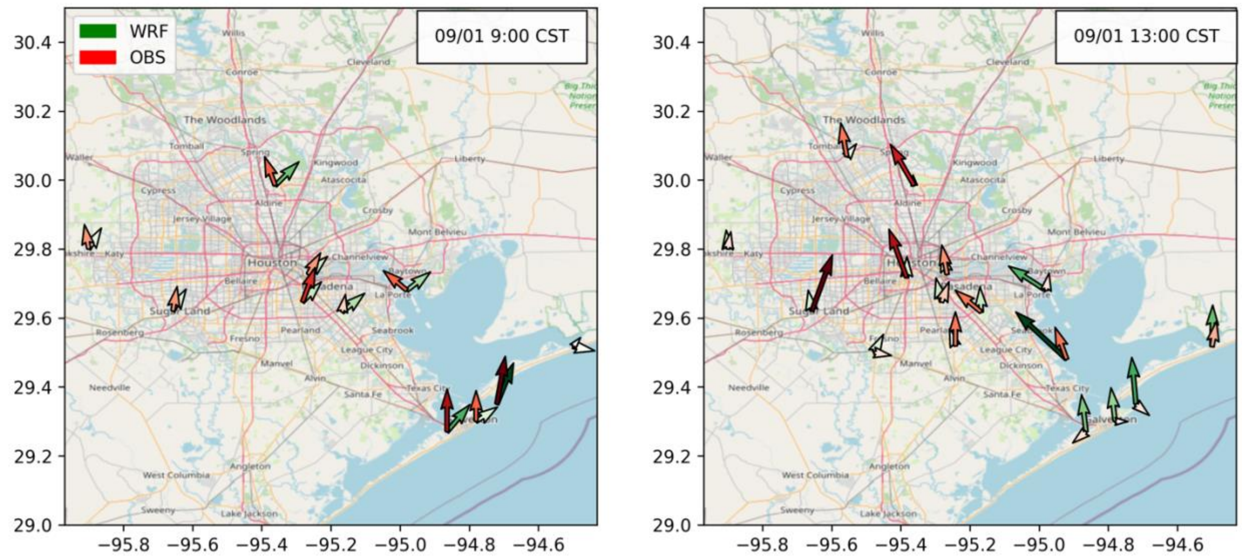

Figure S2: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 1, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

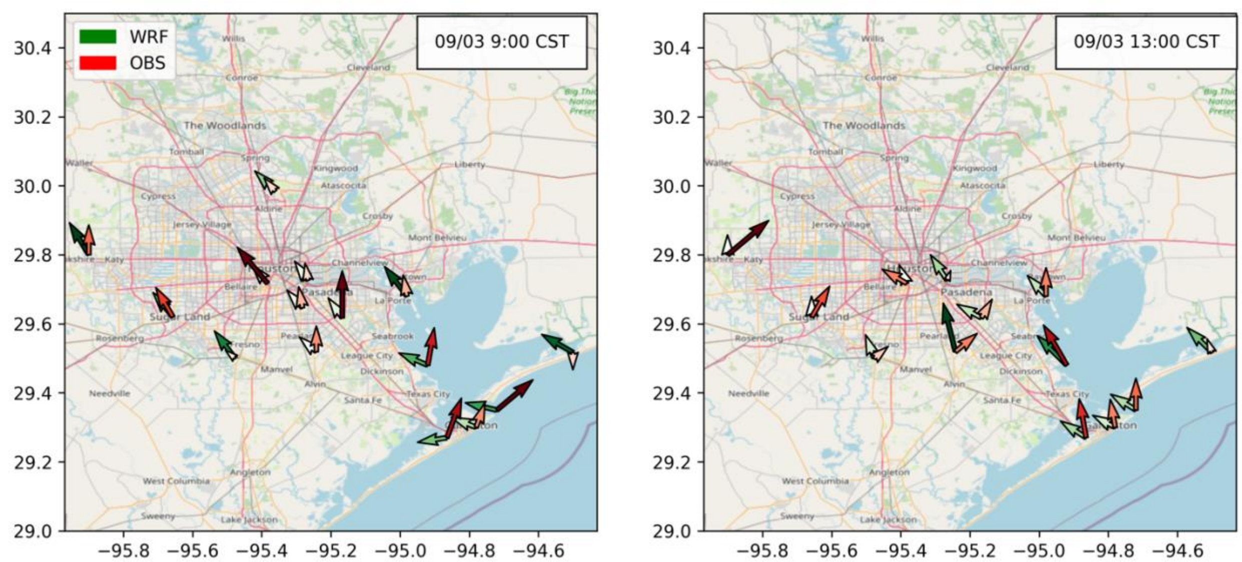

Figure S3: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 3, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

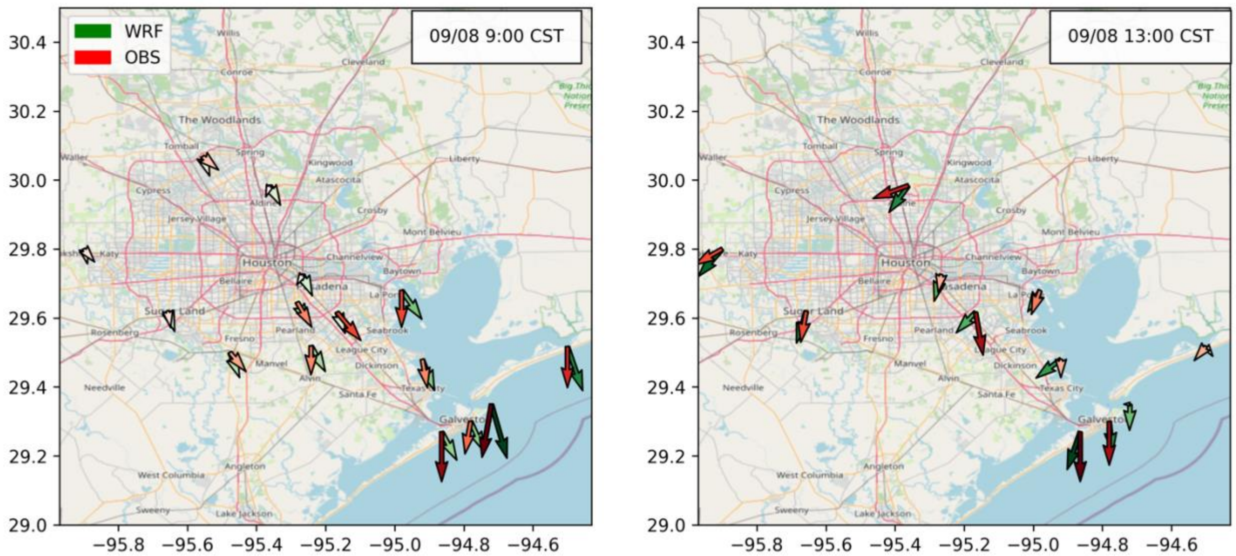

Figure S4: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 8, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

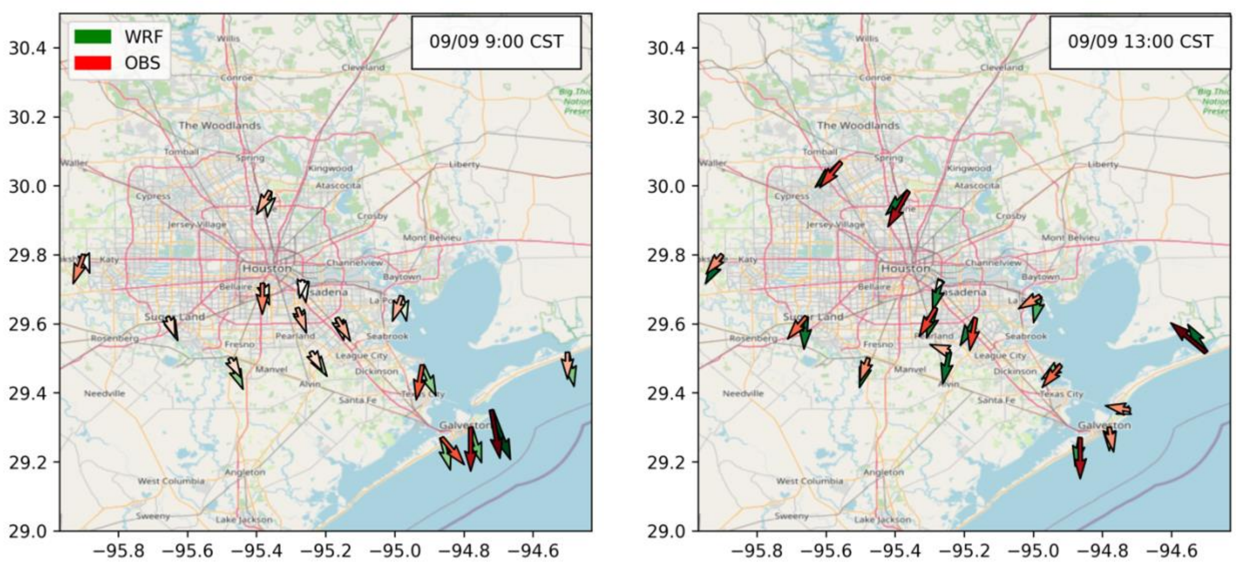

Figure S5: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 9, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

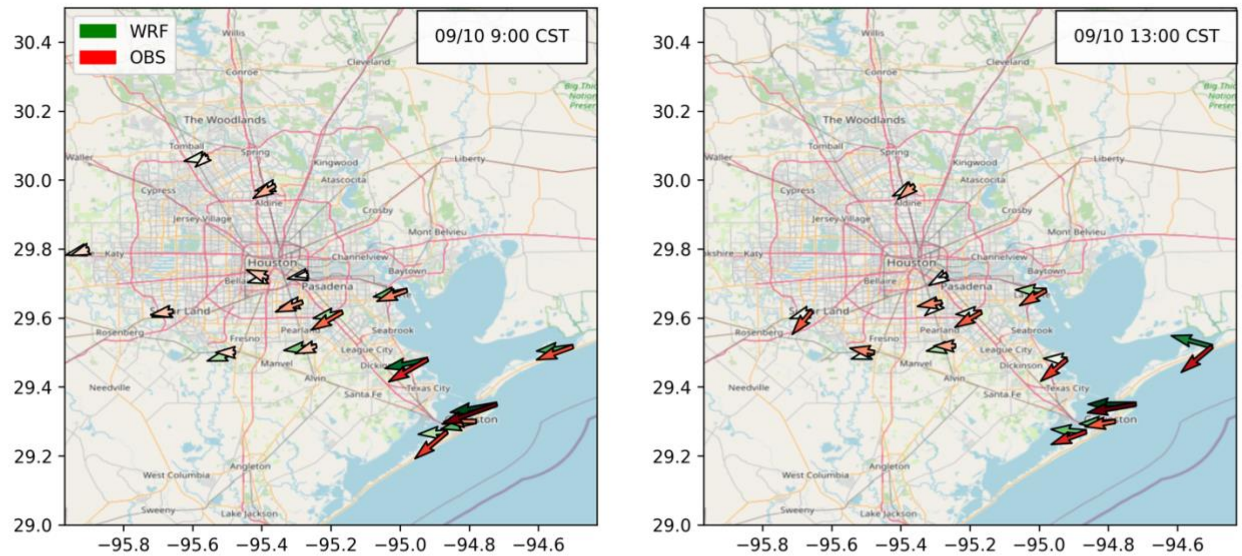

Figure S6: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 10, 2021.

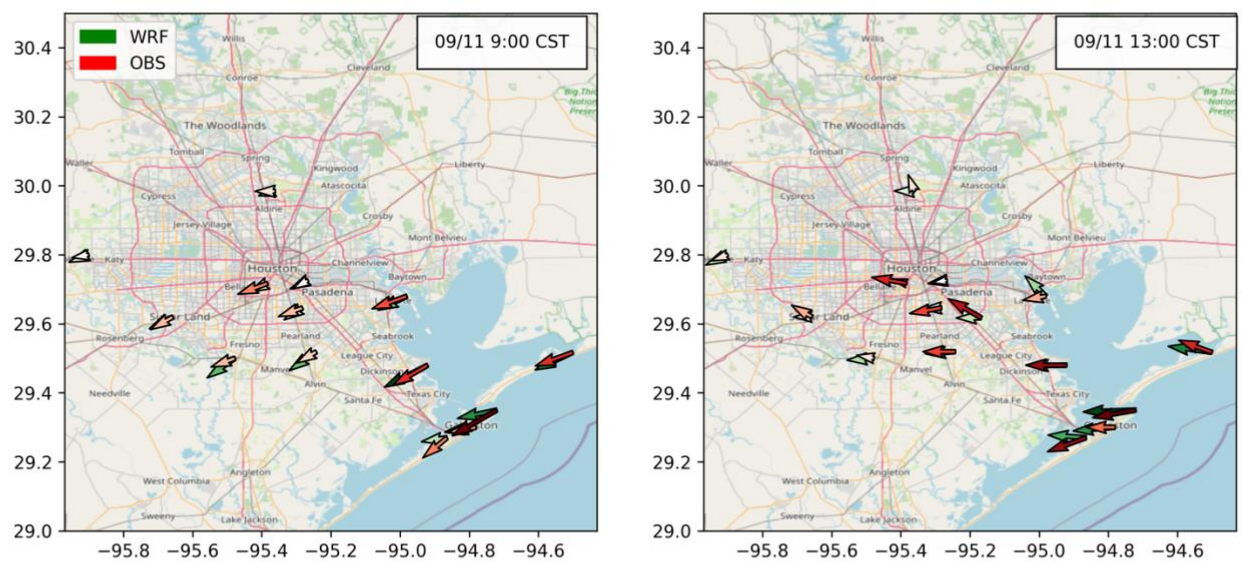

Figure S7: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 11, 2021.

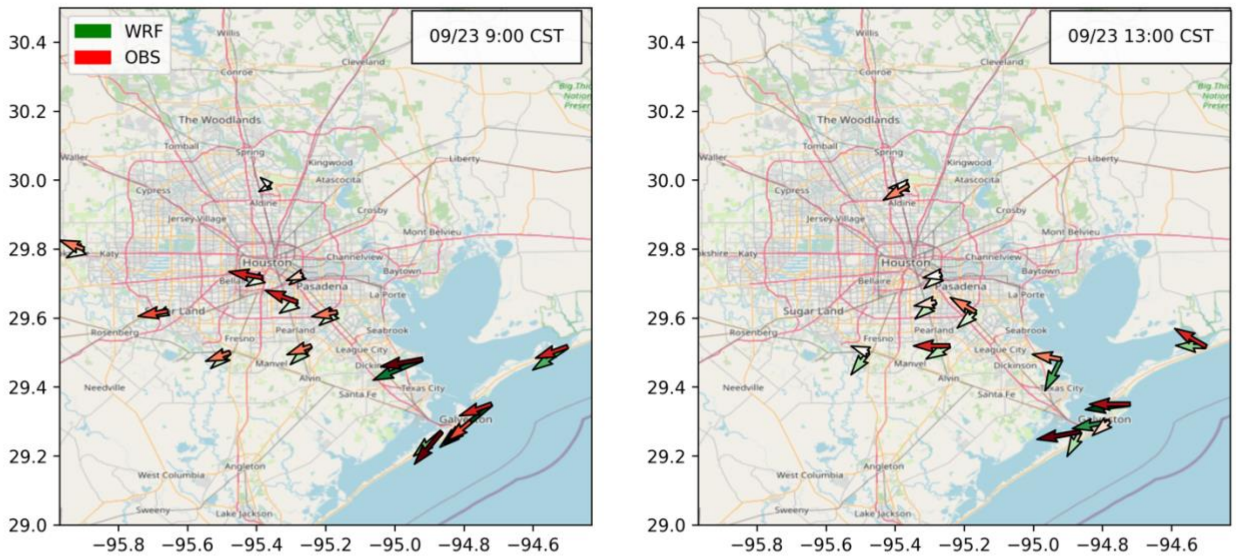

Figure S8: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 23, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

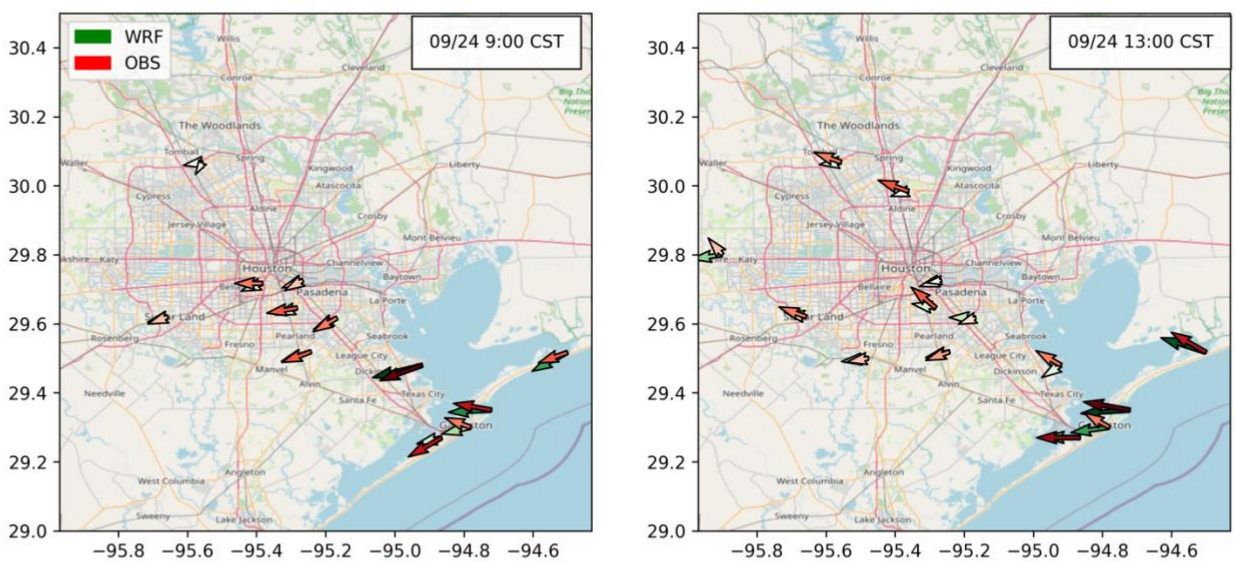

Figure S9: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 24, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

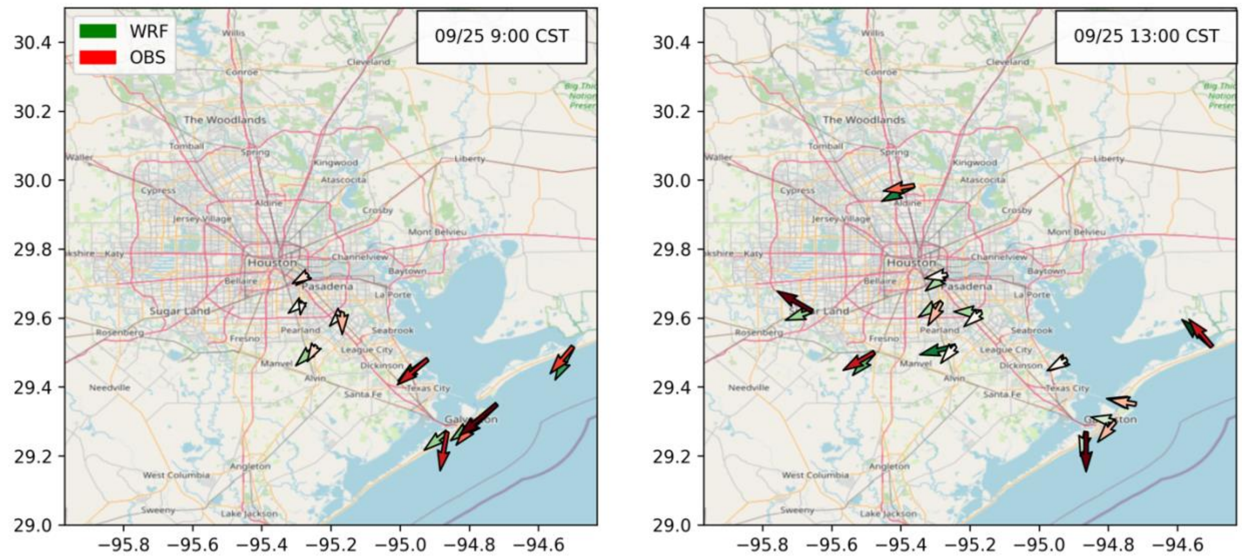

Figure S10: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 25, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

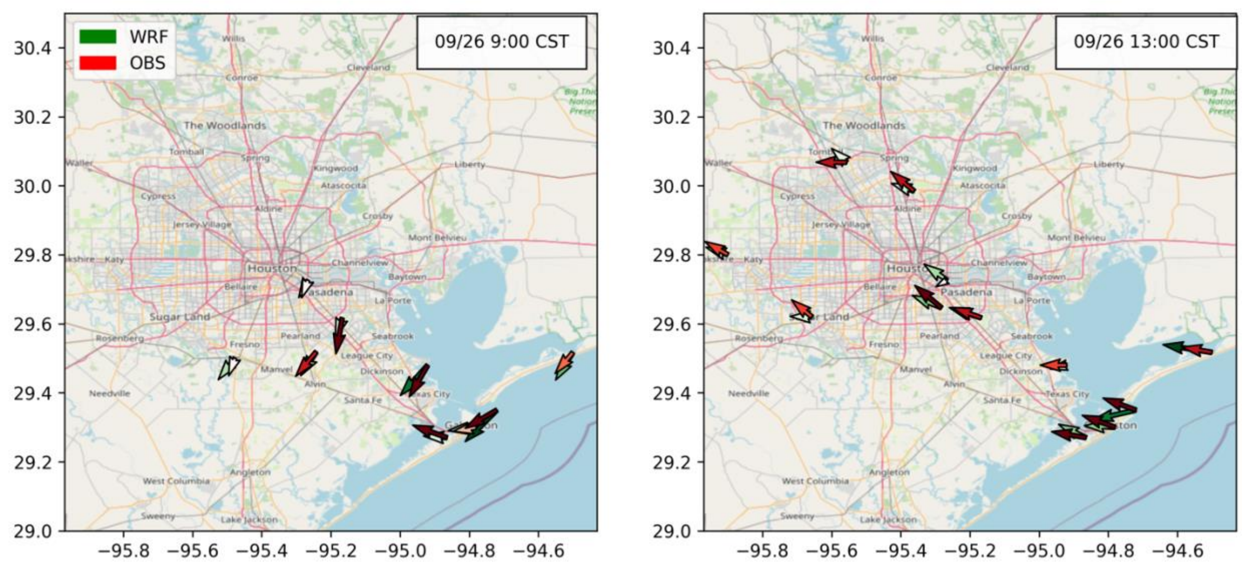

Figure S11: Comparison of observed (OBS) and simulated (WRF) wind at 9am and 1pm CST across sixteen ground-level monitors on September 26, 2021. © OpenStreetMap contributors *2023*. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

References

de Foy, B., Lei, W., Zavala, M., Volkamer, R., Samuelsson, J., Mellqvist, J., Galle, B., Martínez, A.-P., Grutter, M., Retama, A., & Molina, L. T. (2007). Modelling constraints on the emission inventory and on vertical dispersion for CO and SO_2 in the Mexico City Metropolitan Area using Solar FTIR and zenith sky UV spectroscopy. *Atmospheric Chemistry and Physics*, *7*(3), 781–801. https://doi.org/10.5194/acp-7-781-2007

Liu, X., Wang, Y., Wasti, S., Li, W., Soleimanian, E., Flynn, J., Griggs, T., Alvarez, S., Sullivan, J. T., Roots, M., Twigg, L., Gronoff, G., Berkoff, T., Walter, P., Estes, M., Hair, J. W., Shingler, T., Scarino, A. J., Fenn, M., & Judd, L. (2023). Evaluating WRF-GC v2.0 predictions of boundary layer height and vertical ozone profile during the 2021 TRACER-AQ campaign in Houston, Texas. *Geoscientific Model Development*, *16*(18), 5493–5514. https://doi.org/10.5194/gmd-16-5493-2023

Riess, T. C. V. W., Boersma, K. F., Van Roy, W., de Laat, J., Dammers, E., & van Vliet, J. (2023). To new heights by flying low: Comparison of aircraft vertical NO_2 profiles to model simulations and implications for TROPOMI NO_2 retrievals. *Atmospheric Measurement Techniques*, *16*(21), 5287–5304. https://doi.org/10.5194/amt-16-5287-2023

Yu, E., Bai, R., Chen, X., & Shao, L. (2022). Impact of physical parameterizations on wind simulation with WRF V3.9.1.1 under stable conditions at planetary boundary layer gray-zone resolution: A case study over the coastal regions of North China. *Geoscientific Model Development*, *15*(21), 8111–8134. https://doi.org/10.5194/gmd-15-8111-2022