Title

Improving the Representation of Fresh Wildfire Smoke Plumes in Air Quality Forecasts

Author

Thapa, Laura Hughes

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Improving the Representation of Fresh Wildfire Smoke Plumes in Air Quality Forecasts

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Atmospheric and Oceanic Sciences

by

Laura Hughes Thapa

2024

ABSTRACT OF THE DISSERTATION

Improving the Representation of Fresh Wildfire Smoke Plumes in Air Quality Forecasts

by

Laura Hughes Thapa

Doctor of Philosophy in Atmospheric and Oceanic Sciences

University of California, Los Angeles, 2024

Professor Pablo Saide Peralta, Chair

Wildfires are increasing in size and frequency in the Western US due to a complex interplay between climate change and landscape-scale fire exclusion practices. The smoke from these fires is degrading air quality across much of the Continental US. Chemical transport models are vital for warning the public about smoky periods, but uncertainties related to fresh smoke plumes can propagate through these models and cause errors in the resulting air quality forecasts.

We address model uncertainty related to smoke plume vertical extent and total emissions. First, we use aircraft observations obtained during the 2019 Western US wildfires (FIREX-AQ) to evaluate and constrain a commonly used smoke plume rise parameterization in two smoke models (WRF-Chem and HRRR-Smoke). Observations show that free tropospheric smoke layers occur in 35% of observed plumes and up to 95% of modeled plumes. False free tropospheric smoke injections were primarily associated with models overestimating fire heat flux by up to a factor of 25. Next, we present data-driven methods for predicting day-to-day changes in smoke emissions. Our top-performing model (random forest) explains 48% of the variance in observed

daily emissions and outperforms the current operational assumption that emissions will remain constant over a forecast period (persistence, R^2=0.02). This model primarily relies on fire weather data to inform its predictions. Finally, we show preliminary results from WRF-Chem simulations which include random forest-derived emissions and updated heat flux values. We find that in the vicinity of large wildfires in 2020 under less severe fire weather, the random forest-derived emissions can produce better predictions of aerosol optical depth (AOD) and fine particulate matter (PM2.5) than the persistence fire emissions. However, in most cases, persistence and random forest-derived emissions yield very similar AOD and PM2.5 predictions, and that the random forest-derived emissions can both improve and degrade AOD and PM2.5 forecasts. Overall, this work demonstrates the utility of incorporating fire observations to quantify and address uncertainties in our state-of-the-art air quality modeling systems.

The dissertation of Laura Hughes Thapa is approved.

Alex Dean Hall

Arlindo M. da Silva

Jacob Bortnik

James Randerson

Pablo Saide Peralta, Committee Chair

University of California, Los Angeles

2024

To Mom
(if that's even your real name)

"What is fire? It's a mystery.
Scientists give us gobbledegook
about friction and molecules.
But they don't really know."
-Ray Bradbury (Fahrenheit 451)

TABLE OF CONTENTS

Chapter 1 ..... **1**

1.1 Model Uncertainty Due to Plume Rise 2

1.2 Model Uncertainty Due to Emissions 3

1.3 Outline of the Thesis 5

Chapter 2 ..... **7**

2.1 Introduction 8

2.2 Data and Methods 10

2.2.1 Campaign data 10

2.2.2 MASTER algorithm and pixel saturation 11

2.2.3 Modeling systems 12

2.2.4 Sampling method 14

2.2.5 Classifying injection by visual inspection 18

2.2.6 Bootstrapping 20

2.2.7 FRE flux computation 21

2.3 Results 22

2.3.1 Model performance on smoke injection 22

2.3.2 Assessing modeled planetary boundary layer height 23

2.3.3 Associating model performance with additional fire characteristics 26

2.3.4 Assessing heat flux assumptions 27

2.3.5 Sensitivity analysis 31

2.4. Discussion.....33

Chapter 3.....35

3.1 Introduction.....36

3.2. Data.....40

3.3 Methods.....43

3.3.1 Building Polygons.....43

3.3.2 Extracting Data Using Polygons.....44

3.3.3 Predicting FRE with Scaled Fire Indices and Random Forest.....47

3.3.4 Evaluation metrics for predicting categories, scaling factors, and FRE.....50

3.4. Results.....51

3.4.1 Comparison of Two Modeling Approaches with Persistence.....51

3.4.2 Evaluation of the random forest in space, across fire severity levels, and in time...54

3.4.3 Sensitivity Results.....58

3.4.4 Selecting Optimal Subsets of Features.....60

3.4.5 Performance of Optimal Subsets of Features.....62

3.5. Discussion.....65

3.5.1 Comparison with previous studies.....65

3.5.2 Uncertainties and Limitations.....66

3.6. Conclusions.....68

Chapter 4.....70

4.1. Introduction.....71

4.2. Data..... 74

4.2.1 Input emissions..... 74

4.2.2 Evaluation Data: AOD, PM2.5, and polygons..... 74

4.3. Methods..... 75

4.3.1 Chemical transport models..... 76

4.3.2 Emissions scenarios..... 77

4.3.3 Forecast Restarts..... 79

4.3.4 Matching models and observations..... 79

4.3.5 Per-polygon model evaluation..... 80

4.4. Results..... 85

4.4.1 Comparison of daily average black carbon emissions rates..... 85

4.4.2 Bulk evaluation of AOD and PM2.5..... 86

4.3 Per-fire model evaluation..... 88

4.5. Conclusions..... 92

Chapter 5..... 94

5.1 Research Summary..... 94

5.2 What does it all mean?..... 96

5.3 Closing Thoughts..... 98

Appendix A: Supplementary Materials for Chapter 2..... 99

Appendix B: Supplementary Materials for Chapter 3.....164

Appendix C: Supplementary Materials for Chapter 4.....185

References.....0

LIST OF FIGURES

Figure 2.1: Selected transects with their corresponding injection behavior identified, and decision tree for classifying injections. Vertical profiles of a DIAL-HSRL backscatter, b HRRR-Smoke PM2.5, and c WRF-Chem PM2.5 for the Shady Fire on 2019-07-26 01:25-01:35 UTC. Vertical profiles of d DIAL-HSRL backscatter, e HRRR-Smoke PM2.5, and f WRF-Chem PM2.5 for the North Hills Fire on 2019-07-29 22:40-22:55 UTC. g The decision tree that is used to classify each case. In the terminal nodes (red and green boxes) representative cases for each path ending at that terminal node are given in parentheses. Open circles in (a) and (d) denote out of plume boundary layer height, and filled circles denote plume top heights. Light gray solid lines and short dashed gray lines in (b, c) and (e, f) denote the modeled PBLH and mixed layer heights derived from vertical $PM_{2.5}$ gradients, respectively. For WRF-Chem, the mixed layer height at 4PM LT, the time when the mixed layer is thickest, is shown the long dashed dark gray line. 5 min of flight time corresponds to 50–70 km.

Figure 2.2: Box and whisker plots of model parameters and observations grouped by WRF-Chem and HRRR-Smoke injection behavior. a, e Comparison of modeled transect median PBLH with observed transect median PBLH, b, f average FRE flux, c, g savanna and grassland fraction, and d, h terrain height. Red horizontal line denotes medians, blue boxes denote the interquartile ranges, and whiskers denote the 1.5x interquartile range. TP, FP, and FN are defined as in Table 2 caption. a–d Refer to the WRF-Chem model and e–h refer to the HRRR-Smoke model.

Figure 2.3: Comparison of observed FRE flux and assumed FRE fluxes for WRF-Chem and HRRR-Smoke. a Scatter plot of total FRP and flaming and saturated fire area from

MASTER for each fire overpass (colored diamonds, stars, triangles). Black line is a linear fit to the MASTER points of the form $y = mx$. Colored bars represent the ranges of FRE flux based on models' assumptions for convective heat flux and uncertainty in converting between convective and radiant heat flux (see "Methods"). FRE flux increases toward the upper left. Shapes represent the campaign fuel category. Colors of diamonds, stars, and triangles, represent the fractional contribution of flaming combustion (including saturated pixels) to total FRP. Other panels show box and whisker plots of b observed FRP and c observed fraction of FRP due to flaming combustion, grouped by observed injection behavior. Box and whisker elements are as in Fig. 2 caption

Figure 3.1. Workflow example for the 2020 August Complex Fire. a) the process used to build two days of fire polygons for August Complex and b) the final August Complex polygon at daily resolution (light orange for mid-August 2020 through dark red for mid-October 2020). Time series of c) FRE (black) and near surface fire weather indices (HWP—blue, HDW—orange, HRRR VPD—green) extracted using the daily polygons, and d) observed (black) and forecasted FRE using persistence (red) scaled weather (HWP—blue, HDW—orange, HRRR VPD—green) and random forest model trained on all predictors (magenta dashed).

Figure 3.2. Comparison of persistence, scaled, and random forest methods. Performance of the a, f) persistence, b,g) VPD (HRRR), c,h) HD0W0, d, i)HWP, e,j) random forest method on their ability to forecast fire growth category, F=0.67-1.5 is slight change and greater (less) than that value is an increase (decrease). R^2 and SS_{RMSE} and SS_{MAD} are shown for FRE for all 1-day k) and 2-day m) predictions in blue, yellow, and green dots respectively. F1 for all 1-day l) and 2-day n) predictions with increases in red, decreases

in blue and slight changes in tan. Rows are labeled according to the forecasting method, with scaled methods labeled with a "" preceding the name of the scaled variable. Vertical dashed lines denote performance of persistence.

Figure 3.3. Random forest evaluation metrics by state (a-f) and by fire severity (g-j) for the 1-day RF and persistence predictions. a) SS_{MAD} and SS_{RMSE} for the RF model; b) R^2 for the persistence and RF models; c) MAD for the persistence and RF models; d) $F1_{slight\_change}$ for the persistence and RF models; e) RMSE for the persistence and RF models; f) $F1_{increase}$ and $F1_{decrease}$ for the RF model; g) RMSE (yellow) and MAD (teal) for the persistence (dotted) and RF (solid) models split by 20th-percentiles of FWI; h) as in f but split by 20th-percentiles of FRE. i) SS_{MAD} (teal), SS_{RMSE} (yellow), $F1_{increase}$ (red), $F1_{decrease}$ (blue), and $F1_{slight\_change}$ (tan) for the RF models with $F1_{slight\_change}$ (tan dotted) for the persistence model split by 20th percentiles of FWI. j) as in i but split by 20th percentiles of FRE.

Figure 3.4. Performance of the random forest model for different sensitivity tests (row names). a) shows SS_{RMSE} in yellow SS_{MAD} in green, and R^2 in blue, and b) shows $F1_{decrease}$ in light blue, $F1_{increase}$ in red, and $F1_{slight\_change}$ in tan. Black horizontal lines delineate different groups of features for which sensitivity is tested, and these categories are also labeled in text to the left. Feature subsets for which no variables of the given category are used have their points outlined in black. The variable following "only" in the variable name is the sole variable of the given category used in training.

Figure 3.5. Performance of the random forest trained on subsets of the data. a) shows subsets of features selected to optimize performance. R^2 and SS_{RMSE} and SS_{MAD} are shown for FRE for all 1-day b) and 2-day d) predictions in blue, yellow, and green dots respectively. F1

scores for all for all 1-day c) and 2-day e) predictions with increases in red, decreases in blue and slight changes in tan.

Figure 4.1. Black carbon (BC) emissions from August Complex Fire from the a) RAVE emissions inventory for September 2nd-September 22nd, 2020. Reference emissions are plotted in blue, and 1-day persistence and random forest emissions are plotted in green and purple respectively. Note that the persistence assumption dictates that the amount of emissions one day (green bar) is the same as the emissions amount the day before (previous day blue bar).

Figure 4.2. An example of per-polygon model evaluation. a)-f) show AOD from a) MAIAC, and WRF-Chem b) reference, c) 1-day persistence, d) 1-day random forest, e) 2-day persistence, and f) 2-day random forest emissions for September 2nd, 2020 in the vicinity of the August Complex fire (red polygon encircled by black dashed line). Also shown are the Red Salmon Complex (N of August Complex), North Complex (E of August Complex), and the W-5 Cold Springs fires (NE of August Complex on CA/NV border) (red polygons). 50 km (black dotted), 150km, and 300km (gray dotted) buffers around the August Complex fire are shown. g) shows a time series of AOD selected over the 50km buffer for the MAIAC observations (red), and the reference (blue), 1-day persistence (green) and the 1-day random forest (purple). Figures h-j) show model mean bias, normalized mean bias, and pearson correlation over the August complex fire. Colors are as in g), with dotted lines representing the 2-day persistence and random forest forecasts. The green box highlights September 2nd in the time series and the purple box highlights a day where FRP (black and gray lines in g)-j)) is high and model performance is good.

Figure 4.3. Filtering AOD and PM25 measurements over merged polygons. a-b),d-e) show the 150km merged polygons plotted on top of AOD (a,b) and PM2.5 (d,e) for the August Complex Fire on September 2nd (a,d) and the Marion-MRO fire (b,e). Surrounding fires are plotted in red. Unfiltered data is shown in even-numeral plots, and filtered data is shown in the odd-numeral plots. c) and f) show time series of the 10th (blue), 20th (orange) and 30th (green) percentiles of the observed AOD (c) and PM2.5 (f), and threshold percentiles are shown in the black dotted lines.

Figure 4.4. Comparison of RAVEblack carbon (kg) with one day (a-b) and two day (c0d) persistence (a,c, green) and random forest (b,d purple) predictions. The black dotted line is the 1:1 line and points in the red shaded region are predictions within a factor of 2 of the observed value.

Figure 4.5. Time series and comparison of daily, domain-averaged AOD (a-d) and PM2.5 (e-h). Red refers to observations, blue to the reference WRF-Chem model, green to the persistence emissions WRF-Chem model, and purple with the random forest emissions WRF-Chem model.

Figure 4.6. Comparison of the distribution of the median (stars* and triangles^) and spread (bars) of AOD error metrics for a) all fire-days, b) days with low-mod-high FWI, and c) days with very high-extreme FWI. Error metrics plotted include i) mean bias, ii) root-mean-squared error, iii) normalized mean bias, iv) normalized mean error, v) pearson correlation, and vi) ratio. Shaded boxes show the ratio between the change in the persistence (*) and random forest (^) medians when RF is better than P and the change in the persistence (*) and random forest (^) medians when RF is worse than P. Blue means

RF increases overall forecast skill in terms of the metric, and red means RF decreases skill in terms of the metric.

Figure 4.7. As in Figure 6 for PM2.5

Figure A1. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheep Fire on 2019-07-24 21:35-21:45 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

Figure A2-A51. As in Figure A1 for the **A2** Sheep Fire on 2019-07-24 21:50-22:05 UTC, **A3** Sheep Fire on 2019-07-24 22:40-23:00 UTC, **A4** Shady Fire 2019-07-24 23:40-23:53 UTC, **A5** Shady Fire 2019-07-24 23:55 - 2019-07-25 00:07 UTC, **A6** Shady Fire 2019-07-25 00:03-00:10 UTC, **A7** Shady Fire 2019-07-25 00:10-00:20 UTC, **A8** Shady Fire on 2019-07-25 22:30-22:45 UTC, **A9** Shady Fire on 2019-07-26 01:25-01:35 UTC, **A10** North Hills Fire on 2019-07-29 22:40-22:55 UTC, **A11** North Hills Fire on 2019-07-29 23:00-23:15 UTC, **A12** Tucker Fire on 2019-07-30 02:10-02:35 UTC, **A13** Tucker Fire on 2019-07-30 04:00-04:20 UTC, **A14** Tucker Fire on 2019-07-30 04:15-04:40 UTC, **A15** Left Hand Fire on 2019-07-31 00:35-00:55 UTC, **A16** Left Hand Fire on 2019-07-31 00:55-01:15 UTC, **A17** Ridge Top Fire 2019-08-02 22:10-22:25 UTC, **A18** Ridge Top Fire on 2019-08-02 23:40-23:50 UTC, **A19** Mica/Lick Creek Fire 2019-08-03 00:35-00:55 UTC, **A20** Mica/Lick Creek Fire 2019-08-03 02:20-02:40 UTC, **A21** Mica Creek Fire 2019-08-03 21:30-21:40 UTC, **A22.** As in Figure A1 for the **A22**

Williams Flats Fire on 2019-08-03 21:40-22:00 UTC, **A23** Williams Flats Fire on 2019-08-04 00:05-00:25 UTC, **A24** Williams Flats Fire on 2019-08-04 02:30-02:55 UTC**, A25** Williams Flats Fire 2019-08-06 18:40-19:00 UTC, **A26** Williams Flats Fire 2019-08-06 20:30-20:40 UTC**, A27** Williams Flats Fire 2019-08-06 21:45-22:00 UTC, **A28** Snow Creek Fire 2019-08-06 22:25-22:35 UTC, **A29** Horsefly Fire 2019-08-06 22:45-23:00 UTC**, A30** Horsefly Fire 2019-08-07 00:25-00:55 UTC, **A31** Sheep Fire on 2019-08-07 23:00-23:20 UTC, **A32** Williams Flats Fire on 2019-08-08 00:45-01:05 UTC, **A33** Williams Flats Fire on 2019-08-08 02:20-02:40 UTC, **A34** Williams Flats Fire on 2019-08-09 02:00-02:15 UTC, **A35** Williams Flats Fire on 2019-08-09 02:15-02:30 UTC, **A36** Springs Fire 2019-08-12 22:10-22:20 UTC, **A37** Springs Fire 2019-08-12 22:25-22:30 UTC, **A38** Castle Fire 2019-08-12 23:19-23:30 UTC, **A39** Castle Fire 2019-08-12 23:30-23:50 UTC, **A40** Castle Fire 2019-08-13 22:30-22:45 UTC, **A41** Castle Fire on 2019-08-13 22:40-22:55 UTC, **A42** Sheridan Fire on 2019-08-16 00:25-00:40 UTC, **A43** Sheridan Fire on 2019-08-16 00:40-00:50 UTC, **A44** Sheridan Fire on 2019-08-16 01:00-01:15 UTC, **A45** Sheridan Fire on 2019-08-16 01:15-01:30 UTC, **A46** Saber Fire 2019-08-16 01:45-01:55 UTC, **A47** Boulin Fire 2019-08-16 01:55-02:00 UTC, **A48** Sheridan Fire 2019-08-16 02:05-02:20 UTC, **A49** Ikes Fire 2019-08-16 04:10-04:25 UTC, **A50** Sheridan Fire 2019-08-17 00:00-00:15 UTC, **A51** Sheridan Fire 2019-08-17 00:15-00:30 UTC

Figure A52. Comparison of observed (green) and modeled (red) planetary boundary layer height (in m above sea level) distributions for the WRF-Chem (a) and HRRR-Smoke (b) models. Cases for which these plots were generated are the WRF-Chem false positives

Figure A53. Plume top heights from the 1D Freitas forced by 1D WRF-Chem meteorology (blue bars), and 1D Freitas forced by 1D HRRR-Smoke meteorology (red bars) for cases where WRF-Chem underpredicted the PBLH and HRRR-Smoke accurately captured the PBLH. PBLHs associated with each meteorology configuration are overlaid as box and whisker plots over their corresponding bar, and the PBLH derived from the DIAL-HSRL is given as a dotted line. Selected fires include a) Shady, b) Mica Creek, c) Williams Flats, d-e) Spring, f) Castle, g-h) Sheridan

Figure A54. Box and whisker plots of (a) HRRR-Smoke time of day, (b) HRRR-Smoke terrain variability (standard deviation in 3x3 grid box around fire), (c) FT Brunt-Vaisala Frequency, (d) PBL Brunt-Vaisala Frequency (e) FT bulk Richardson number, (f) PBL bulk Richardson number. Red line denotes medians, blue box denotes the interquartile range, and whiskers denote the 1.5x interquartile range. TP, FP, and FN are defined as in Table 2.2 caption.

Figure A55. Box and whisker plots of (a) WRF-Chem time of day, (b) WRF-Chem terrain variability (standard deviation in 3x3 grid box around fire), (c) FT Brunt-Vaisala Frequency, (d) PBL Brunt-Vaisala Frequency (e) FT bulk Richardson number, (f) PBL bulk Richardson number. Red line denotes medians, blue box denotes the interquartile range, and whiskers denote the 1.5x interquartile range. TP, FP, and FN are defined as in Table 2.2 caption

Figure A56. Analogous to Fig. 2.3, but for different combustion phases: (a) total FRP and total area, (b) flaming/saturated FRP and total area, and (c) flaming/saturated FRP and flaming/saturated area

Figure A57. MASTER FRP vs (a) VIIRS FRP and (b) MODIS FRP determined using VIIRS and MODIS overpasses within +/- 2hrs from MASTER overpasses. MASTER values tend to be higher than MODIS values and lower than VIIRS values. Red dashed lines are 95% confidence bounds on the best fit line.

Figure A58. Comparison of campaign fuels and model fuels covering three broad fuel categories for the 51 test cases in this study: grassland, savanna, and forest. Entries along the main diagonal represent cases where the model and the campaign fuel type agreed, and off diagonals represent cases where the model and campaign fuel type disagreed.

Figure A59. Illustration of the spatial shifting algorithm for the Castle fire, observed on 8/12/2019 23:20 20-23:30 UTC. (a) Observed backscatter profiles from the DIAL -HSRL. (b) Modeled PM2.5 curtain plot without shifting applied. (c) Modeled PM2.5 curtain plot with shifting applied. (d) Spatial aerosol optical depth (AOD) plot overlaid with unshifted (black) and shifted (red) flight flight track.

Figure A60. Illustration of time shift for the Shady Fire, observed 7/29/2019 00:00-00:38 UTC. (a) spatial emissions plot and (b) vertical curtain of PM2.5 sampled at the model time matching the observed time. Note how emissions are not present in the spatial or the vertical plot. (c) spatial emissions plot and (d) vertical curtain plot sampled at the model time shifted to one hour after the observed time. Emissions are now present in the spatial plot and vertical plume structure is developed.

Figure B1. Time series of FRE (black), hourly wildfire potential (blue), hot-dry-windy index (orange), Canadian Fire Weather Index (green), US Burning Index (red) for the August Complex fire.

Figure B2. Example of the impact of grid resolution on polygon-averaged fuel density (a-b) and slope (c-d). a) and c) show a time series of FRP (black) and fuel density/slope at different grid resolutions (colored lines) for the Bobcat Fire. b) and d) show pearson correlation for fuel density/slope at different grid resolutions

Figure B3. Distributions of a) total daily FRE, b) log(FRE), and c) δFRE, and d) log(δFRE) for the training years, 2019 and 2021. a) and c) use a log scale for the y axis, c) and d) use a linear scale. Outliers which lie more than 1.5 interquartile ranges from the median of the log of δFRE are removed.

Figure B4. Density plots of observed and modeled FRE as predicted by a,f) persistence, b,g) scaled VPD from HRRR, c,h) scaled HD0W0, d,i) scaled HWP, e,j) the random forest trained on all variables. The red line is the best fit to the log of the data and the black line is the 1:1 line. The best fit equation, R^2, RMSE, and MAD are also provided for each model. The top row is for the 1-day forecast and the bottom row is for the 2-day forecast.

Figure B5. Density plots of observed and modeled scaling factors as predicted by a,f) persistence, b,g) scaled VPD from HRRR, c,h) scaled HD0W0, d,i) scaled HWP, e,j) the random forest trained on all variables. The red line is the best fit to the log of the data and the black line is the 1:1 line. The best fit equation, R^2, RMSE, and MAD are also provided for each model. The top row is for the 1-day forecast and the bottom row is for the 2-day forecast.

Figure B6. Random forest evaluation metrics by state (a-f) and by fire severity (g-j) for the 2-day RF and persistence predictions. a) SS_{MAD} and SS_{RMSE} for the RF model; b) R^2 for the persistence and RF models; c) MAD for the persistence and RF models; d) $F1_{slight\_change}$ for the persistence and RF models; e) RMSE for the persistence and RF

models; f) $F1_{increase}$ and $F1_{decrease}$ for the RF model; g) RMSE (yellow) and MAD (teal) for the persistence (dotted) and RF (solid) models split by 20th-percentiles of FWI; h) as in f but split by 20th-percentiles of FRE. i) SS_{MAD} (teal), SS_{RMSE} (yellow), $F1_{increase}$ (red), $F1_{decrease}$ (blue), and $F1_{slight\_change}$ (tan) for the RF models with $F1_{slight\_change}$ (tan dotted) for the persistence model split by 20th percentiles of FWI. j) as in i but split by 20th percentiles of FRE.

Figure B7. Scatter plots of average error vs average values of predictor by state for the 1-day random forest predictions. Shown are the top 3 predictors whose state-averaged values had the highest person correlation (r^2) with each error type. The metrics presented (y-axes) include a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{no\ change}$, g) $F1_{increase}$, and h) $F1_{decrease.}$ The x-axes are labeled with the predictor whose average values are plotted.

Figure B8. Pearson correlation (r^2) and correlation strength (absolute value of r^2, shading, annotated numbers) between error metrics and state-average random forest input variables (row labels) for the 1-day forecasts. Pearson correlations are shown for a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{decrease}$, g) $F1_{increase}$, h) $F1_{no\ change}$

Figure B9. Scatter plots of average error vs average values of predictor by state for the 2-day random forest predictions. Shown are the top 3 predictors whose state-averaged values had the highest person correlation (r^2) with each error type.The metrics presented (y-axes) include a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{no\ change}$, g) $F1_{increase}$, and h) $F1_{decrease.}$ The x-axes are labeled with the predictor whose average values are plotted.

Figure B10. Pearson correlation (r^2) and correlation strength (absolute value of r^2, shading, annotated numbers) between error metrics and state-average random forest input

variables (row labels) for the 2-day forecasts. Pearson correlations are shown for a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{decrease}$, g) $F1_{increase}$, h) $F1_{no\ change}$

Figure B11. Time series of total FRE (black line, left axis) and model 1-day forecast error (teal and gold lines) for the 2020 fire season. a) shows daily total FRE over the western US (black) along with RMSE (gold, right axis) and MAD (green, right axis) for the random forest (solid) and persistence (dashed) forecast methods. b) shows daily total FRE over the western US along with the SS_{RMSE}(gold) and SS_{MAD} (teal) compared with persistence. The dashed black line is at 0 on the skill score axis; points above this line are days where the random forest improves on persistence on average and points below this line are days where the random forest is worse than persistence on average. In a) and b) the size of the stars is proportional to the number of fires burning in the Western US domain on that day.

Figure B12. Time series of total FRE (black line, left axis) and model 2-day forecast error (teal and gold lines) for the 2020 fire season. a) shows daily total FRE over the western US (black) along with RMSE (gold, right axis) and MAD (green, right axis) for the random forest (solid) and persistence (dashed) forecast methods. b) shows daily total FRE over the western US along with the SS_{RMSE}(gold) and SS_{MAD} (teal) compared with persistence. The dashed black line is at 0 on the skill score axis; points above this line are days where the random forest improves on persistence on average and points below this line are days where the random forest is worse than persistence on average. In a) and b) the size of the stars is proportional to the number of fires burning in the Western US domain on that day.

Figure B13. Correlation heatmap for near-surface weather variables. Scaled weather variables are not shown.

Figure B14 Correlation heatmap for a) stability, b) fuel loading, and c) terrain variables.

Figure B15. Correlation heatmap for moisture variables.

Figure B16. Correlation heatmap for human variables.

Figure C1. Black carbon (BC) emissions from August Complex Fire from the a) RAVE and b) QFED emissions inventories for September 2nd-September 22nd, 2020. Reference emissions are plotted in blue, and *2-day* persistence and random forest emissions are plotted in green and purple respectively. Note that the persistence assumption dictates that the amount of emissions one day (green bar) is the same as the emissions amount the day before (previous day blue bar).

Figure C2. Map of fires burning in September 2020, with colored shading representing the day of burning, defined as 12Z on the labeled calendar day to 12Z on the following calendar day. Certain fires are noted by name if mentioned in the text or figures.

Figure C3. Time series of the number of AOD grid cells which are counted once (blue line) and more than once (other colored lines) for 50 km (a-b) and 150 km (c-d) buffers. a,c) show the number of multi-counts for polygons based on individual fires and b,d) show the number of multi counts for merged polygons.

Figure C4. Time series of the number of PM2.5 stations which are counted once (blue line) and more than once (other colored lines) for 150 km (a-b) and 300 km (c-d) buffers. a,c) show the number of multi-counts for polygons based on individual fires and b,d) show the number of multi counts for merged polygons.

Figure C5. Time series of the number of AOD grid cells (a,b) and PM2.5 stations (c,d) which are counted once (blue line) and more than once (other colored lines) for merged polygons. a,c) show the number of multi-counts for unfiltered data and b,d) show the number of multi counts for filtered data.

Figure C6. August Complex Fire time series of AOD (a,c,e,g) and PM2.5 (b,d,f,h) predictions, observations and error metrics. a,b) are a time series of observed AOD (a)/PM2.5 (b) (red), reference model (blue), 1-day persistence (green), and 1-day random forest (purple) distributions. The other time series show mean bias (c,d), normalized mean bias (e,f) and pearson correlation (g,h) for one day (line with filled circle markers) and two day (line with star markers) predictions. (a,c,e,g) is a repeat of Figure 2 g-j)

Figure C7-C15. As in S6 for the **C7** SCU Lightning Complex fire in CA, **C8** Creek fire in CA, **C9** North Complex fire in CA, **C10** Pearl Hill fire in WA, **C11** Cold Springs fire in WA, **C12** Lionshead fire in OR, **C13** Beachie Creek fire in OR, **C14** Cameron Peak fire in CO, **C15** Mullen fire in WY

Figure C16. Distributions of the value which compares the distances of the persistence and random forest of AOD (e.g.) from its ideal value. If the random forest gets the statistic closer to its ideal value, this number is positive, and if the random forest gets the statistic further from its ideal value this number is negative. This value was computed for a) mean bias, b) root-mean-squared error, c) normalized mean bias, d) normalized mean error, e) pearson correlation, and f) ratio. The black dotted lines are a +/- 0.025, representing the cutoffs for the 3 categories of model performance: RF better than P, RF about the same as P, and RF worse than P. The values plotted here are an example computed from the filtered AOD over merged 50 km buffer polygons.

Figure C17. Sample maps of r^2 and mean bias in AOD over the WRF Chem model domain. a) and b) show the two metrics for the whole reference model period, with the comparison done at daily resolution. In a) red means a positive correlation in time and blue means a negative correlation in time. In b) red corresponds to overestimated AOD blue

corresponds to underestimated AOD. c-f) show the difference in mean bias between the reference simulation and the c) 1-day persistence, d) 1-day random forest, e) 2-day persistence, and f) 2-day random forest predictions. In c-f) red means that higher AOD than the reference simulation is predicted, and blue means that lower AOD than the reference simulation is predicted.

Figure C18. Comparison of median error metrics for AOD in smoky regions downstream of fires for a) all fire days, b) the subset of fire days showing low, moderate, or high fire weather index (FWI) values, and c) the subset of fire days showing very high or high extreme FWI values. Numbers in the grid cells are the distance between the median of the error distribution and the ideal value for the metric (1 for rat and pc, 0 for all others). The boxes are shaded blue if the random forest emissions get the median AOD much closer to the ideal value than persistence, and red if vice versa.

Figure C19. As in Figure C18 for filtered PM2.5 data in smoky regions downstream of fires.

Figure C20. Comparison of the distribution of the median (stars* and triangles^) and spread (bars) of AOD error metrics for a) all fire-days, b) days with low-mod-high FWI, and c) days with very high-extreme FWI. AOD statistics were computed over merged polygons and included smoke and non-smoke values. Error metrics plotted include i) mean bias, ii) root-mean-squared error, iii) normalized mean bias, iv) normalized mean error, v) pearson correlation, and vi) ratio. Shaded boxes show the ratio between the change in the persistence (*) and random forest (^) medians when RF is better than P and the change in the persistence (*) and random forest (^) medians when RF is worse than P. Blue means RF increases overall forecast skill in terms of the metric, and red means RF decreases skill in terms of the metric.

Figure C21. As in Figure C18 for unfiltered AOD statistics computed over merged polygons

Figure C22. As in Figure C20 for unfiltered PM2.5 computed over merged polygons

Figure C23. As in Figure C18 for unfiltered PM2.5 statistics computed over merged polygons

Figure C24. As in C20 for unfiltered AOD computed over single polygons

Figure C25. As in Figure C18 for unfiltered AOD statistics computed over single polygons

Figure C26. As in C20 for unfiltered PM2.5 computed over single polygons

Figure C27. As in Figure C18 for unfiltered PM2.5 statistics computed over single polygons

LIST OF TABLES

Table 2.1. Summary of model details**.** The first column lists model details for WRF-Chem, the second feature lists model details for HRRR-Smoke, and the third column lists model features for the 1D Freitas plume rise model. References point to those in (Thapa et al., 2022).

Table 2.2 Confusion matrices for comparing observed and modeled injection behavior counts. Counts of injection behavior are shown for HRRR-Smoke using the HRRR-Smoke PBL height to evaluate injection, HRRR-Smoke using the DIAL-HSRL PBL height to evaluate injection, WRF-Chem using the WRF-Chem PBL height to evaluate injection, and WRF-Chem when the DIAL-HSRL PBL height is used to evaluate injection. Columns are labeled with forecasted plume behavior and rows are labeled with observed plume behavior. TN stands for true negative, where non-injection was forecasted and observed. Similarly, TP stands for true positive. FN stands for false negative, where injection was not forecast but did occur. FP stands for false positive, where injection was forecast but did not occur. HRRR-Smoke has fewer total cases than WRF-Chem due to certain fires missing in the model (see "Methods", Sampling Method).

Table 3.1. Summary of data used in this work, including temporal coverage, temporal and spatial resolution, aggregation method, and primary reference.

Table 4.1. Summary of important WRF-Chem model meteorological parameterizations. In the WRF-Chem references column, numbers refer to the WRF-Chem namelist option.

Table A1. Details for each case, including the figure number, fire name, date, and time of the measurements, WRF-Chem modeled and observed injection behavior (0=non-injection, 1=injection), and terminal node of the decision tree that results in each classification.

Not-a-Number (Nan) means that the model missed the fire and therefore the injection behavior could not be evaluated.

Table A2. Details for each case, including the figure number, fire name, date, and time of the measurements, HRRR-Smoke modeled and observed injection behavior (0=non-injection, 1=injection), and terminal node of the decision tree that results in each classification. Not-a-Number (Nan) means that the model missed the fire and therefore the injection behavior could not be evaluated.

Table A3. Assumptions for selecting the out-of-plume observed PBLH.

Table B1. δ**BI Outliers.** Values of δBI used to generate the 1-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Table B2. δ**BI Outliers.** Values of δBI used to generate the 2-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Table B3. δ**HWP Outliers.** Values of δHWP used to generate the 1-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Table B4. δ**HWP Outliers.** Values of δHWP used to generate the 2-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Table B5. Comparison of R^2 and R^2_{adj} for a) sensitivity analysis, b) 1-day optimal subsets, c) 2-day optimal subsets, d) 1-day RF (all features) and e) 2-day RF (all features)

ACKNOWLEDGEMENTS

Chapter Two is a version of the following published work:

Thapa, L. H., Ye, X., Hair, J. W., Fenn, M. A., Shingler, T., Kondragunta, S., Ichoku, C., Dominguez, R., Ellison, L., Soja, A. J., Gargulinski, E., Ahmadov, R., James, E., Grell, G. A., Freitas, S. R., Pereira, G., & Saide, P. E. (2022). Heat flux assumptions contribute to overestimation of wildfire smoke injection into the free troposphere. *Communications Earth & Environment*, *3*(1), Article 1. https://doi.org/10.1038/s43247-022-00563-x

Each author contributed as follows to Chapter Two: L.H.T. designed the study, wrote MATLAB scripts to generate all figures, analyzed the data, and wrote the paper with contributions from all co-authors. X.Y. provided Matlab scripts to generate curtain plots of DIAL-HSRL, WRF-Chem, and HRRR-Smoke and to extract relevant variables from the model. J.W.H., M.A.F., and T.S. conducted DIAL-HSRL observations. S.K. processed the MASTER L1B data and contributed MASTER and satellite comparison, including data used to generate Fig. S55. C.I. and L.E. processed MASTER observations and generated the MASTER FRP L2 data files using their original algorithm. R.D. conducted and provided MASTER observations. A.J.S. and E.G. assisted with interpretation of results and provided the fuels information used to generate Fig. 2. R.A., E.J., G.A.G., and G.P. provided HRRR-Smoke forecast data and emissions. S.R.F. provided the 1D standalone plume rise code and aided in the interpretation of results. P.E.S. designed the study, provided MATLAB scripts to generate curtain plots, edited the manuscript, and advised L.H.T. during the project.

Chapter Three is a version of the following accepted work:

Thapa, L. H., Saide, P.E., Bortnik, J., Berman, M.T., da Silva, A., Peterson, D.A., Li, F., Kondragunta, S., Ahmadov, R., James, E., Ye, X., Soja, A.J., Wiggins, E.B., Gargulinski,

E. (Accepted). Forecasting daily fire radiative energy using data driven methods and machine learning techniques. *Journal of Geophysical Research-Atmospheres*

In Chapter Three, Pablo Saide served as PI and provided feedback throughout the study design and analysis. Advice related to the study design was provided by Jacob Bortnik and David Peterson. Xinxin Ye, Eric James, Johana Romero-Alvarez, and provided scripts to process data. Fangjun Li, Shobha Kondragunta, Eric James, Johana Romero-Alvarez, Ravan Ahmadov, Amber Soja, Elizabeth Wiggins, and Emily Gargulinski provided data used in the analysis.

In Chapter Four, Pablo Saide served as PI and provided feedback throughout the study design and analysis. Calvin Howes provided WRF-Chem data preprocessing and insight on how to run and configure WRF-Chem. Arlindo da Silva, Ravi Govindaraju, Meng Zhou, and Allie Collow provided GEOS forecasts and insight on how GEOS is configured. Fangjun Li and Shobha Kondragunta provided RAVE data and insight into its behavior in WRF-Chem. Xinxin Ye configured WRF-Chem to run on the UCLA Hoffman2 cluster. Ravan Ahmadov, Eric James, Johana Romero-Alvarez provided the data that was used to generate the random forest emissions. Alexi Lyapustin and Yujie Wang provided the MAIAC data that was used to benchmark model aerosol optical depth.

This work has been supported by the following grants: NASA 80NSSC18K0629, NASA 80NSSC20K1650, NASA 80NSSC18K0685, NASA 80NSSC21K1456, NOAA NA18OAR4310107, NSF 2013461,, NSF 2238338. This work was supported by funding from the Anthony and Jeanne Pritzker Family Foundation and the Eugene V. Cota-Robles Fellowship.

VITA

2022 NOAA Global Systems Laboratory/Cooperative Institute for Research in Environmental Sciences Summer Internship Program, Boulder, CO

2021 Master of Science in Atmospheric and Oceanic Sciences, University of California, Los Angeles (UCLA)

2020 Naval Research Enterprise Internship Program, Monterey, CA

2018 Bachelor of Science in Applied Mathematics, University of California, San Diego (UCSD)

PUBLICATIONS

Thapa, L. H., Saide, P.E., Bortnik, J., Berman, M.T., da Silva, A., Peterson, D.A., Li, F., Kondragunta, S., Ahmadov, R., James, E., Ye, X., Soja, A.J., Wiggins, E.B., Gargulinski, E. Forecasting daily fire radiative energy using data driven methods for predicting smoke emissions (accepted, JGR-Atmospheres).

Thapa, L. H., Ye, X., Hair, J. W., Fenn, M. A., Shingler, T., Kondragunta, S., Ichoku, C., Dominguez, R., Ellison, L., Soja, A. J., Gargulinski, E., Ahmadov, R., James, E., Grell, G. A., Freitas, S. R., Pereira, G., & Saide, P. E. (2022). Heat flux assumptions contribute to overestimation of wildfire smoke injection into the free troposphere. In Communications Earth & Environment (Vol. 3, Issue 1). Springer Science and Business Media LLC. 10.1038/s43247-022-00563-x

Saide, P.E., Krishna, M., et al (**Thapa** is 4th author) (2023). Estimating fire radiative power using weather radar products for wildfires. In Geophyscial Research Letters. 10.1029/2023GL104824.

Turney, F. A., Saide, P. E., et al (**Thapa** is 10th author) (2023). Sensitivity of Burned Area and Fire Radiative Power Predictions to Containment Efforts, Fuel Density, and Fuel Moisture Using WRF-Fire. Journal of Geophysical Research: Atmospheres, 128(18), e2023JD038873. 10.1029/2023JD038873

Warneke, C., et al (**Thapa** is 36th author) (2022) Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ), Journal of Geophysical Research: Atmospheres, 128, e2022JD037758, 10.1029/2022JD037758, 2023.

Pagonis, Demetrios et al., (**Thapa** is 22nd author) (2023). Impact of Biomass Burning Organic Aerosol Volatility on Smoke Concentrations Downwind of Fires, Environmental Science & Technology. In Environmental Science and Technology. 10.1021/acs.est.3c05017.

Berman, M. T., Ye, X., **Thapa, L. H.,** et al., (2023). Quantifying burned area of wildfires in the western United States from polar-orbiting and geostationary satellite active-fire detections. In International Journal of Wildland Fire (Vol. 32, Issue 5, pp. 665–678). CSIRO Publishing. 10.1071/wf22022

Peterson, D. A., **Thapa, L. H.,** Saide, P. E., et al., (2022). Measurements from inside a Thunderstorm Driven by Wildfire: The 2019 FIREX-AQ Field Experiment. In Bulletin of the American Meteorological Society. American Meteorological Society. 10.1175/bams-d-21-0049.1

Saide, P. E., **Thapa, L. H.,** Ye, X., et al., (2022). Understanding the Evolution of Smoke Mass Extinction Efficiency Using Field Campaign Measurements. In Geophysical Research Letters (Vol. 49, Issue 18). American Geophysical Union (AGU). 10.1029/2022gl099175

Chapter 1

Introduction

Intensifying wildfires in the Western US due to anthropogenic climate change (Westerling et al., 2006; Zhuang et al., 2021) are undoing decades of regulatory progress on harmful pollutants such as surface level fine particulate matter (PM2.5) and ozone (Burke et al., 2021, 2023; Larsen et al., 2018). Wildfire plume rise has also become more aggressive, with smoke plumes being injected higher in the atmosphere due to increased fire energy and plume buoyancy (Wilmot et al., 2022). In a recent extreme example, the summer of 2020 was a period in which 1 in 7 Americans were exposed to at least one day of unhealthy air due to smoke by the end of September (Carlsen, 2020). During this period, the National Ambient Air Quality Standard (NAAQS) for PM2.5 was exceeded over 3000 times, and smoke contributed 40-80% of observed PM2.5 in the Western US on September 14th, 2020 (Y. Li et al., 2021). Smoke from these fires contributed to adverse health outcomes over many US counties (Cromar et al., 2024) and reduced solar energy generation over California and Washington (Bertoletti et al., 2022; Juliano et al., 2022).

Wildfire smoke is also a complex climate forcing agent, since smoke particles can scatter and absorb solar radiation and act as cloud condensation nuclei (CCN). In the vicinity of large fires, smoke has been shown to block solar radiation via both scattering and absorption and cool the surface (Yu et al., 2016; Kochanski et al., 2019). However, the total smoke direct effect remains uncertain and dependent on modeling choices like emissions inventory and aerosol-radiation interaction parameterization, resulting in some studies computing a cooling direct effect (e.g. Carter et al., 2020) and other studies computing a warming direct effect (e.g. Jiang et al., 2016; Wilcox 2012) due to smoke. Smoke particles may also act as CCN (Petters et

al., 2019), enhancing the albedo and lifetimes of clouds, and exerting a cooling effect on the climate (e.g. Jiang et al., 2016). Finally, the vertical placement of smoke with respect to clouds may also generate a climate forcing. When smoke is placed within clouds, the clouds may burn off more quickly, offsetting cloud albedo and lifetime effects (Wilcox, 2012). However, smoke above clouds may increase stability and cloudiness, and a more persistent cloud deck may cancel out the direct warming from above-cloud smoke (Wilcox, 2012; Allen et al., 2019). Generally speaking, smoke radiative effects are highly non uniform in space due to the sporadic nature of fires and the short lifetime of smoke aerosols.

Chemical transport models (CTMs) are vital for warning the public about episodes of poor air quality and for quantifying the climate impacts of wildfire smoke, but uncertainty in smoke emissions and vertical structure can propagate through these predictions and cause large errors (Carter et al., 2020; Ye et al., 2021). Including smoke emissions substantially improves model estimates of PM2.5, ozone, and CO compared to no-smoke model simulations (Baker et al., 2016; Roy et al., 2007; Tao et al., 2020; Pavlovic et al., 2016), but CTMs are sensitive to the amount, timing, and vertical distribution of smoke emissions (Garcia-Menendez et al., 2014; Herron-Thorpe et al., 2014; Rooney et al., 2020; Shi et al., 2019). These sources of uncertainty affect air quality-relevant variables, such as PM2.5 and aerosol optical depth (AOD). PM2.5 may be underestimated or overestimated (Herron-Thorpe et al., 2014; Rooney et al., 2020), both in bulk model evaluation and on a station or regional basis. AOD tends to be underestimated (Herron-Thorpe et al., 2014; Shi et al., 2019).

1.1 Model Uncertainty Due to Plume Rise

Plume rise refers to the vertical transport of hot smoke gasses and particles through the atmosphere. This process may result in plume injections, where smoke reaches the free

troposphere or lower stratosphere, or in plume non-injections, where smoke remains confined within the planetary boundary layer (PBL). Plume rise occurs on scales much smaller than most CTMs are able to simulate. Therefore, a range of methods exists for predicting plume heights, with some models assuming even dispersion of smoke throughout the boundary layer (e.g. Collow et al., 2024) and other models deriving plume height through physics based parameterizations (e.g. (Ahmadov et al., 2017; Thapa et al., 2022). When comparing multiple models, as done in Ye et al., (2021), there is a large spread in plume heights when multiple models simulate the same fire event. The widely used physics-based plume rise model described in Freitas et al., (2007) tends to outperform models that assign emissions to a single level or a fixed vertical distribution (Mallia et al., 2018; Roy et al., 2018; Sessions et al., 2011). However, a major challenge in deploying the Freitas model is obtaining accurate estimates of fire size and heat flux to use as inputs to the model (Freitas et al., 2007). Previous studies have highlighted the Freitas model's sensitivity to these inputs and have also pointed to issues with the Freitas model itself, such as the representation of entrainment, as causes for uncertainties like the overprediction of free tropospheric injection and the limited range of predicted plume heights (Val Martin et al., 2012; Ye et al., 2021). Therefore, even the best available plume rise models remain somewhat unconstrained.

1.2 Model Uncertainty Due to Emissions

In CTMs, fire emissions are an essential model input, but there remains uncertainty regarding the best way to compute and predict these emissions. Emissions are commonly estimated using the burned area-based conventional approach (Seiler & Crutzen, 1980) or the FRP-based method (Ichoku et al., 2008; Wooster et al., 2005). Computing emissions starts with a satellite detection of burned area or fire radiative power (FRP), which is then converted to an

amount of fuel burned using an empirically derived constant called the emission coefficient. A second conversion factor, called an emission factor, is then used to compute the emissions of a given species. Historically, emissions were computed using observations from polar orbiting satellites (Darmenov and da Silva, 2015; Kaiser et al., 2012; Wiedinmyer et al., 2011), but in recent years geostationary measurements have been used to create hourly emissions (Darmenov and da Silva, 2015; Li et al., 2022), an improvement on the daily/twice-daily resolution of older inventories. Emissions inventories have their own ways of dealing with obscured fires, usually assuming that the fire continues in some form even while covered with heavy smoke or clouds (Darmenov and da Silva, 2015; Li et al., 2022). Generally speaking, FRP-based methods tend to predict higher emissions values than burned area-based methods (Pan et al., 2020; Ye et al., 2021).

In forecasting mode, many air quality models persist the most recent daily satellite-derived emissions throughout the forecast period. Sub daily emissions are forecasted using a prescribed climatological diurnal cycle, which assumes a peak in fire intensity in the local afternoon (e.g., Ahmadov et al., 2017). A recent study (Ye et al., 2021) has shown that persistence can result in over- or underpredicted emissions in wildfires when daily changes in fire activity are large using twelve forecast systems deployed in the Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) field campaign. Persistence may delay the onset or extend the lifetime of fires within the forecast window (Di Giuseppe et al., 2018; Thapa et al., 2022), which is useful for regions of the world like Southern Africa or the Southeast US where intentional burning is a regular human practice (Earl and Simmonds 2018). In spite of advances such as inverse modeling of emissions (e.g., Saide et al., 2015) and statistical methods to forecast fire evolution (Di Giuseppe et al., 2018; Graff et al., 2020; D. Peterson et al., 2013;

Preisler & Westerling, 2007; Huot et al., 2022; Wang et al., 2022) operational smoke forecasts rely primarily on persistence to forecast daily emissions totals. Precipitation is simplistically included in some of the models (Ahmadov et al., 2017).

1.3 Outline of the Thesis

This thesis focuses on how we can address issues with plume rise and the persistence assumption in the WRF-Chem modeling framework for Western US wildfires.

In chapter 2 we evaluate the Freitas plume rise parameterization using field data and suggest observationally-constrained improvements to Freitas model inputs (Thapa et al., 2022; Warneke et al., 2023). We compare observed and modeled plume profiles from two regional CTMs and determine that the frequency of injection into the free troposphere is overpredicted in the models. This is in a large part due to modeled plumes being artificially buoyant as a result of heat flux assumptions made when implementing the Freitas model. To a second order, the accuracy of predicting injections into the free troposphere is also influenced by the model representation of the planetary boundary layer (PBL), as the height of the PBL is the boundary we use to determine if an injection occurs. Finally, we show that when we place an observational constraint on the heat flux, the frequency of injected plumes is much more accurately predicted.

In chapter 3, we address the persistence assumption using weather indices alone and random forest (RF) machine learning trained on weather indices, land surface, and firefighting data to forecast daily changes in fire behavior (Thapa et al., accepted). In particular, we are interested in forecasting the daily change in fire radiative energy rather than the ignition of new fires (FRE, the time-integrated FRP), so this result is of particular importance to emissions inventory which are based on FRP. We show that the RF method leads to the biggest improvement over persistence, but that using weather indices alone also leads to substantial

improvement. The RF gains the most forecast skill by incorporating weather and FRE information, and can be trained on a subset of the full training set and achieve comparable performance.

In chapter 4, we discuss a WRF-Chem modeling study to test the impact of analysis, persistence, and RF-derived wildfire emissions on air quality relevant variables like AOD and PM2.5. We find that WRF-Chem mainly underpredicts AOD and PM2.5, and that the model is not sensitive in bulk to the emissions scheme used. When we zoom in on smoke plumes directly downwind of fires or groups of fires, we find that model error is a little more sensitive to emissions schemes and on some days for some fires, the RF derived fire emissions lead to much lower model error. However, the results are not enough to determine definitively if using RF emissions in the current configuration of WRF-Chem represents a significant advantage over persistence. Therefore, the main advancement of this chapter is a flexible pipeline to evaluate models downstream of fires.

Chapter 2

Heat flux assumptions contribute to overestimation of wildfire smoke injection into the free troposphere

An edited version of this paper was published in Nature Communications Earth and Environment.

Abstract

Injections of wildfire smoke plumes into the free troposphere impact air quality, yet model forecasts of injections are poor. Here, we use aircraft observations obtained during the 2019 western US wildfires (FIREX-AQ) to evaluate a commonly used smoke plume rise parameterization in two atmospheric chemistry-transport models (WRF-Chem and HRRR-Smoke). Observations show that smoke injections into the free troposphere occur in 35% of plumes, whereas the models forecast 59–95% indicating false injections in the simulations. False injections were associated with both models overestimating fire heat flux and terrain height, and with WRF-Chem underestimating planetary boundary layer height. We estimate that the radiant fraction of heat flux is 0.5 to 25 times larger in models than in observations, depending on fuel type. Model performance was substantially improved by using observed heat flux and boundary layer heights, confirming that models need accurate heat fluxes and boundary layer heights to correctly forecast plume injections.

2.1 Introduction

Previous studies have shown that smoke plumes injected above the planetary boundary layer (PBL) and into the free troposphere (FT) tend to occupy layers of ambient stability and that plume injections are associated with high fire radiative power (FRP), and favorable fire weather (Kahn et al., 2007; Soja et al., 2021; Val Martin et al., 2010). Observations from the Multi-angle Imaging Spectroradiometer (MISR) indicate that injection into the FT occurs in 4–12% of North American smoke plume tops (Kahn et al., 2007, 2008; Val Martin et al., 2010, 2018), and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) data indicate that injection of the total smoke column into the free troposphere occurs in 78% of North American smoke plumes (A. Soja et al., 2021). However, the MISR percentage may be biased low due to satellite overpasses occurring in the morning (~10:30 AM LT for MISR), when plumes have not fully developed (Kahn et al., 2008). Injected plumes can be advected downwind with little dilution (Schum et al., 2018) and thus tend to have more regional impacts. In very extreme cases, stratospheric smoke injections may self-loft due to the absorbing component of smoke, further enhancing smoke lifetimes (Das et al., 2021). On the other hand, non-injected plumes tend to have more local impacts due to more efficient downward mixing by ambient PBL turbulence.

The Freitas plume rise parameterization is a 1-dimensional cloud-resolving model. It is typically embedded in a 3-dimensional host model that specifies the ambient environment (Freitas et al., 2007). This model represents fires as surface buoyancy fluxes that depend on instantaneous fire size, and convective heat flux, and fuel type. Proposed causes of the Freitas model's shortcomings include uncertainties in lateral entrainment and in input parameters. Entrainment is dependent on fire size, so versions of the Freitas model allowing variable fire sizes can improve the modeled range of plume heights (Sessions et al., 2011; Val Martin et al.,

2012). In addition, explicitly adding entrainment to a later version of the model improved performance (Freitas et al., 2010; Walter et al., 2016). Uncertainty in the input parameters, notably fire size and FRP, may be caused by the plume masking FRP retrievals, incorrect fire shape or size, variability of burning with vegetation type, or uncertainty in converting between radiant and convective heat fluxes (Paugam et al., 2016; Val Martin et al., 2012; Freeborn et al., 2008).

Here, we use aircraft and model data from the NASA-NOAA Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) field campaign to evaluate the Freitas model in developed plumes. We show that the Freitas model, implemented in the Weather Research and Forecasting Coupled with Chemistry (WRF-Chem) model and the High Resolution Rapid Refresh Smoke (HRRR-Smoke) model, overpredicts injection in comparison with airborne lidar measurements. Injection occurs in 35% of observed plumes and in 80% and 95% of WRF-Chem and HRRR-Smoke plumes, respectively. When the observed boundary layer height is used to evaluate injection in the both models, injection occurs in 59% of the WRF-Chem plumes and 72% of the HRRR-Smoke plumes. Comparing the fire radiant energy flux (FRE flux, 2.2.7) observed by the Moderate Resolution Imaging Spectroradiometer (MODIS)-Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) airborne simulator (MASTER) instrument with those assumed by WRF-Chem and HRRR-Smoke shows that the models overestimate FRE fluxes by up to a factor of 25 (0.55 kWm^{-2} in the observations and 0.29–13.86 kWm^{-2} in the models). When WRF-Chem total heat fluxes were reduced so radiant fractions matched observed FRE fluxes, the occurrence of false positives became less common, and further improvement was achieved when the WRF-Chem sensitivity simulations were evaluated using the observed PBL as the injection

boundary. This effect was strongest for forest fires, which are assumed by the WRF-Chem implementation of the Freitas model to have the highest heat fluxes (Freitas et al., 2007).

2.2 Data and Methods

2.2.1 Campaign data

The FIREX-AQ field campaign targeted smoke plumes using the DC-8 airborne laboratory. This dataset contains high-resolution measurements of plume backscatter, FRP, fire area, and fuels burned for each fire. In addition, WRF-Chem and HRRR-Smoke simulations implement slightly different versions of the Freitas parameterization. WRF-Chem was run in near-real time to support flight planning efforts and HRRR-Smoke is a NOAA operational system.

An airborne Differential Absorption Lidar-High Spectral Resolution Lidar (DIAL-HSRL, Hair et al., 2008) provided aerosol backscatter coefficient at 532 nm (vertical resolution = 30 m, horizontal resolution = 10 s) to evaluate the plume injection height. FRP, area burned, and fire phase measurements (flaming, smoldering, or saturated) were taken using the MASTER multispectral imager (Hook et al., 2001). For each fire, the campaign fuel category was determined as the dominant fuel category according to the Fuels Characteristic Classification System (FCCS), a high-resolution dataset built using remote sensing and in-situ measurements of land cover and fuel loading. In this study, FCCS fuel categories were mapped to the fuel categories assumed by the MODIS-International Geosphere Biosphere Programme (IGBP, Prichard et al., 2013; Strahler et al., 1999). A comparison of the WRF-Chem modeled and observed fuels for FIREX-AQ fires (Figure A58) shows that the WRF-Chem model uses the correct fuel type in 47% of the cases. Since the fuel type in a WRF-Chem grid cell prescribes the Freitas model input heat flux, errors in model fuel can lead to the heat flux being over or

underestimated with respect to heat flux that would be assumed had the observed fuel been used to determine the heat flux. A majority of the errors (present in 39% of the cases) lead to a lower prescribed heat flux than would be prescribed if the observed fuels had been used. In spite of this heat flux estimates are still too high.

2.2.2 MASTER algorithm and pixel saturation

In this study, FRP retrievals and combustion phase classes from the MASTER instrument are used to evaluate the models' assumptions of fire heat flux. MASTER pixels can vary in horizontal resolution based on the aircraft distance from the surface, and resolution varied between 10-30m during FIREX-AQ. with Fire detection from MASTER is based on the contextual algorithm concept originally developed for the MODIS instruments aboard the Terra and Aqua satellites (Giglio et al., 2003). Regarding the retrieval of FRP, previous studies indicate that 98% of the variance in fire radiative energy is explained by variations in the emitted mid-infrared (MIR) spectral radiance at ~4 μm (Wooster et al., 2003). Thus, retrieval of FRP from MASTER observations is based on the MIR radiance approach (Wooster et al., 2005) using the MASTER 3.91 μm spectral band (hereafter 4 μm). To classify MASTER-detected fire pixels by combustion phase (i.e., flaming and smoldering), efforts to derive and apply constant thresholds were unsuccessful, as the range of values representing these phases varied widely from scene to scene, partly because of differences in the fire regimes and variations of pixel fractions with different fire phases, and partly because of complex instrument radiometric calibration dynamics beyond the scope of this discussion. Thus, an approach similar to the contextual fire detection algorithm (Giglio et al., 2003) is used to accomplish this classification, with thresholds empirically derived from each scene of the MASTER imagery. Flaming pixels were determined as fire pixels for which the 4 μm brightness temperature is higher than 3

standard deviations above the background mean 4 μm brightness temperature and for which the difference between the 4 and 11 μm brightness temperatures exceeds 100 K. Smoldering pixels were determined as fire pixels for which the 4 μm brightness temperature is higher than 2 standard deviations above the background mean 4 μm brightness temperature and for which the 11 μm brightness temperature is higher than 1 standard deviation above the mean background 11 μm brightness temperature. Since some of the more intense fire pixels were found to be saturated due to the limited dynamic range of the MASTER 4 μm fire channel relative to the high temperatures that can be attained by pixels which are fully covered by flames (up to 1500 K), saturated pixels were determined using the 4 μm radiance values. Given the scene with 20% valid background pixels, these pixels with 4 μm radiance values >99.5% of max value are flagged as saturated. These maximum values may vary from scene to scene and even scan line to scan line, but in a given scan line, they are generally equal (within 0.005%). When saturation occurs, clusters of pixels around saturated pixels tend to have the same 4 μm radiance value. These pixels are also flagged as saturated, which is a limitation of this method.

Preliminary analysis (Figure A57) showed that MASTER FRP is comparable to FRP from MODIS, VIIRS, and the Geostationary Operational Environmental Satellite-Advanced Baseline Imager (GOES-ABI). MASTER tends to underpredict VIIRS FRP measurements (bias = −66.11 MW), and overpredict MODIS (bias = 379.6 MW). In the MODIS plot, the strong association ($R^2 = 0.97$) is driven in large part by the outlier, but removing the outlier still results in MASTER overpredicting MODIS ($R^2 = 0.72$, bias = 190,215.3 MW, slope = 1.39). See Figure A55 for the full comparison. Note that not all cases where we have plumes have associated MASTER data.

2.2.3 Modeling systems

In this study, we compared the Freitas plume rise parameterization in WRF-Chem and HRRR-Smoke, the details of each are listed in Table 2.1. The relevant differences for this research are in the way each model deals with heat flux and fire size. WRF-Chem assumes a constant fire size of 0.25 km^2 per model grid cell and fuel-dependent total heat fluxes that range from 4.4 kWm^{-2} for grass fires to 80 kWm^{-2} for forest fires. HRRR-Smoke, on the other hand, assumes fuel-dependent FRE fluxes which range from 3.44−4.76 kWm^{-2} for all fuel types and uses these constants along with measurements of FRP to estimate fire size. Since HRRR-Smoke assumes FRE fluxes, it does an internal conversion to total heat flux (including radiant, convective, and conductive) using a factor of 1.6 (Paugam et al., 2015, McCarter & Broido, 1965). In both models, convective energy is computed using prescribed or calculated total heat flux based on the fuel category and assuming that convective energy is 55% of the total heat flux (McCarter & Broido, 1965). Both models assume that only flaming emissions undergo plume rise and assume a fixed fraction of emissions are flaming. The remaining fraction of the emissions are released at the surface (Table 2.1, “Flamig Fraction of Emissions Assumption”). The plume rise module is run every 30 min of model time.

Model	WRF-Chem v3.6.1	HRRR-Smoke	1D Freitas
Resolution	4 km horizontal, 53 vertical levels	3 km horizontal, 40 vertical levels	70 vertical levels, 100 m spacing
Initializaiton/Spin-Up	00UTC initialization, 12 h spin up, 76 h analysis, 84 h forecast, model meteorology passed to Freitas model[34]	00, 06, 12, 18 UTC initializations, 48 h forecasts, model meteorology passed to Freitas model[34]	Uses model profles of T, P, geopotential, RH, wind, potential temperature, specific humidity.
FRE Flux Assumptions	2.69–13.86 kWm^{-2} for forest, 0.4–3.10 kWm^{-2} for woody savanna/cerrado, and 0.57 kWm^{-2} for grassland/pasture/cropland.	3.4 kWm^{-2} for grassland to 4.8 kWm^{-2} for forest for 14 vegetation categories as defined by the IGBP[28,43,44]	Choose the Freitas maximum heat flux for the dominant fuel category
Fire Size Assumptions	0.25 km^2 per burning grid cell and per fuel category[14,45]	Converts MODIS/VIIRS FRP into area using fuel-varying heat fluxes	0.25 km^2 per burning grid cell and per fuel category[14,45]
Flaming Fraction of Emissions Assumption	50% flaming, undergo plume rise	75% flaming, undergo plume rise	N/A
Key Parameterizations	Rapid Radiative Transfer Model for GCMs[45], Noah Land Surface Model[46], aerosol-aware microphysics[47], Grell-Freitas cumulus[34], Mellor–Yamada-Janjić PBL[48], MEE = 3.5[49], MOSAIC Aerosols[50]	Aerosol-aware microphysics[47,51] Mellor–Yamada–Nakanishi–Niino PBL scheme[52]	N/A
Data Assimilation	Near-real-time inversion of MAIAC AOD V6[53,54] based on 6 largest fires in the region[19]	None	N/A
Emissions Source	QFED v2.4, 0.1 degrees resolution, regridded[38]	Top-down from MODIS/VIIRS FRP[39]	N/A
Variables Extracted	PM2.5 curtains, model PBL and theta-derived mixed layer heights[19], land use fractions, Richardson number and Brunt-Vaisala frequency via centered-difference at twice and half the PBLH.	PM2.5 curtains, model PBL height, Richardson number and Brunt-Vaisala frequency calculated via centered-difference at twice and half the PBLH.	Maximum plume top height

Table 2.1 Summary of model details. The first column lists model details for WRF-Chem, the second feature lists model details for HRRR-Smoke, and the third column lists model features for the 1D Freitas plume rise model. References point to those in (Thapa et al., 2022).

The Quick Fire Emissions Dataset (QFED) emissions used in WRF-Chem were regridded to the model resolution by assigning emissions to the closest grid cell. The QFED daily emissions were regridded to hourly resolution by assigning a percentage of the daily emissions to each hour, with the peak of the diurnal cycle occurring at 16:00 LT and containing 17% of the daily emissions. In addition, this diurnal cycle assumes that a majority of the emissions are released during daytime hours (around 10:00–19:00 LT) and less than 1% of the daily emissions are released per hour during nighttime hours (Western Regional Air Partnership, 2002).

2.2.4 Sampling method

Figure A1–A51 show the 51 transects that were selected for analysis in this work. Observations were taken during transects wherein the aircraft flew over the fire source and upwind or downwind along the longest axis of the plume. Modeled columns of particulate matter less than 2.5 microns in diameter ($PM_{2.5}$) concentration were extracted along the flight path, considering fires that were slightly shifted in space and time in WRF-Chem relative to observations (Figure A59–A60). All cross sections are shown in Figure A1–A51 and Tables A1–A2 indicate injection behavior of all samples. Terminal nodes of the classification (Figure 2.1g) are used to determine which PBL to evaluate injection against and to compare with the DIAL-HSRL PBLH.

In a few cases, the aircraft observed fires which are not present in the emissions inventory driving the HRRR-Smoke model. HRRR-smoke does not interpolate fires with missing satellite detections, which means that FRP was not ingested into the model and thus fire emissions and plume height are never computed for those fires. This issue impacted the Shady fire (observed July 24, 2019) and the Tucker Fire (observed July 29, 2019). These missing detections are likely due to clouds in the area, which could lead to missed satellite FRP retrievals (Tucker), or persistent clouds dampening fire behavior and leading to missed detections even after the clouds cleared (Shady). These missing HRRR-Smoke cases are present in the WRF-Chem dataset, because the QFED emissions inventory has the capability to estimate emissions for cloud-covered regions (Darmenov and da Silva, 2015). Because these cases have WRF-Chem curtain plots available, they are present in our analysis, but for the HRRR-Smoke samples, the injection behavior has been labeled as not a number (nan, Table A1–A2).

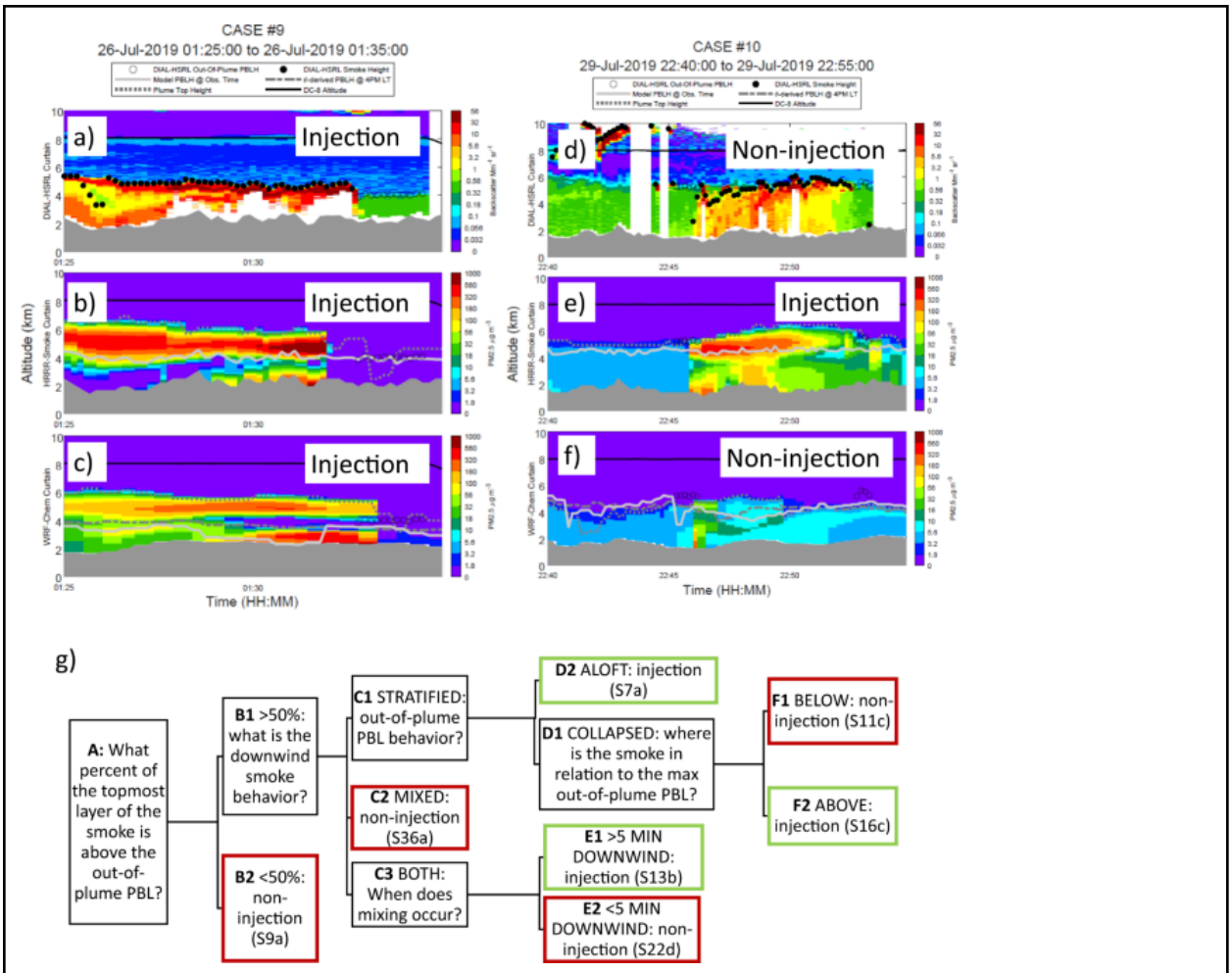

Figure 2.1: Selected transects with their corresponding injection behavior identified, and decision tree for classifying injections. Vertical profiles of **a** DIAL-HSRL backscatter, **b** HRRR-Smoke PM2.5, and **c** WRF-Chem PM2.5 for the Shady Fire on 2019-07-26 01:25-01:35 UTC. Vertical profiles of **d** DIAL-HSRL backscatter, **e** HRRR-Smoke PM2.5, and **f** WRF-Chem PM2.5 for the North Hills Fire on 2019-07-29 22:40-22:55 UTC. **g** The decision tree that is used to classify each case. In the terminal nodes (red and green boxes) representative cases for each path ending at that terminal node are given in parentheses. Open circles in (**a**) and (**d**) denote out of plume boundary layer height, and filled circles denote plume top heights. Light gray solid lines and short dashed gray lines in (**b**, **c**) and (**e**, **f**) denote the modeled PBLH and mixed layer heights derived from vertical $PM_{2.5}$ gradients, respectively. For WRF-Chem, the mixed layer height at 4PM LT, the time when the mixed layer is thickest, is shown the long dashed dark gray line. 5 min of flight time corresponds to 50–70 km.

In about half of the WRF-Chem cases, modeled fires appear shifted with respect to observations. When sampling the model in the location of the flight, this appears as missing portions of the plume (Figure A59a) and only capturing the plume edge (Figure A59d). This

apparent shift is a result of the resolution differences between the WRF-Chem model (4 km) and the QFED emissions inventory (0.1-degree, ~12 km). As each QFED sample is regridded to a single WRF-Chem grid cell, a shift in fire location can result. HRRR-Smoke is not subject to this issue because it is run at 3 km resolution and ingests 375 m VIIRS FRP to calculate and spatially allocate fire emissions, area, and plume rise (Ahmadov et al., 2017).

Rather than exclude these cases and lose half our sample size in the analysis, we implemented spatial shifts in the sampling algorithm to assure the model was sampled over the densest part of the plume. The shift distance is defined as the distance between the flight's closest approach to the modeled fire location and the modeled fire location itself, which is defined as the center of mass of the emissions in a 9 × 9 set of grid boxes around the observed fire location (FIREX-AQ data repository). The shift distance is then used to translate the flight track in space, and the model is sampled at these new locations. This assumes that the shifts are small enough that model meteorology and plume rise drivers do not change over the relatively small spatial range of the shifts (shifts are often a fraction of a degree). This substantially increased the number of comparisons between the model and the observations. However, it is acknowledged that shifting may result in higher error when the fuel type changes over the distance shifted, as this impacts the plume rise calculation (Freitas et al., 2007). Figure A59c, d (red line) shows how shifting allows us to sample the model in the location where plume rise occurs and how this impacts model performance.

In two cases, poor model-observation agreement was found for a different reason. In Figure A60b, the fire plume is not evident in the WRF-Chem vertical profile of PM2.5, and Figure A60a shows that this is due to QFED emissions not being present in the model at the sampled time. As noted above, QFED daily emissions are assigned hourly values based on UTC

time and each day (at 01 UTC) a new set of emissions is imported to QFED which tends to be around 5–6 pm local time for the FIREX-AQ domain. This leads to discontinuities in the emissions diurnal cycle which typically coincided with the DC-8 observing the fire plume. Therefore, a time shift was also implemented for the two cases where this occurred. In these cases, the model was sampled at a timestep one hour later than the time the DC-8 overpassed the plume. Similar to the spatial shift, the time shift assumes no change in meteorology (i.e., wind direction) from one timestep to the next. Although in reality meteorology can change drastically over the course of an hour, the timing of the overpasses (within 20 or 30 min of the next hour) we are able to neglect some of this variability. Figure A60c–d shows how sampling the WRF-Chem model one hour later brings us to a time where model emissions are present, and plume rise behavior can be evaluated.

2.2.5 Classifying injection by visual inspection

We define injection as a stratified layer of smoke above the PBL or residual layer above the fire whose stratified shape is maintained for at least five minutes (50–70 km based on aircraft groundspeed) of downwind transport. This definition is motivated by the differences between injected and non-injected smoke and the tendency of injected smoke to aggregate in free tropospheric stable layers. With this definition, injections are independent of the presence of PBL smoke. Examples of injection are shown in Figure 2.1a–c, e, and examples of non-injection are shown in Figure 2.1d, f.

In the DIAL-HSRL curtains (Figure 2.1a, d), backscatter values of 1–56 Mm^{-1} sr^{-1} (yellow-dark red, black filled circles) represent smoke, and values of 0.18–0.56 Mm^{-1} sr^{-1} (dark green-light green, black open circles) represent the out-of-plume aerosol mixed layer. During the day, this aerosol mixed layer is roughly coincident with the PBL, and after sunset the aerosol

mixed layer retains the day's maximum PBL height. Since the presence of fire influences the mixed layer height observed by the DIAL-HSRL, upwind and downwind locations where the PBLH remained relatively stable were chosen as the out of plume region (black open circles in 1a–3f). In both models (Figure 2.1b–c, e–f), $PM_{2.5}$ values of 5.6–1000 μg m^{-3} (light blue-dark red) are assumed to be smoke. Determining the modeled out-of-plume PBL height is more complex due to the fact that WRF-Chem simulates background aerosols and HRRR-Smoke does not. Therefore, across all model cases, we use the out-of-plume modeled PBL height or the potential temperature-derived mixed layer height (following Ye et al., 2021) at 4PM local time (light gray solid or dark gray dashed lines in Figure 2.1b–c, e–f in the location of black open circles) to classify injection depending on if injection is computed with respect to a daytime PBLH or a nighttime residual layer (Figure 2.1g). This methodology evaluates modeled injection with respect to the observed and modeled PBL and observed injection with respect to the observed PBL. This is different from previous studies that use modeled PBL heights only as the standard to judge injection (Kahn et al., 2008). Comparing lidar backscatter curtains with modeled $PM_{2.5}$ curtains is sufficient for our purposes because smoky regions and non-smoky regions are visually distinct from each other and have a satisfactory degree of spatial coincidence in both the backscatter and the $PM_{2.5}$ curtains.

Figure 2.1g shows the decision tree-style process that was applied to all cases to determine whether or not injection occurred. The first split in the tree (node A) asks what fraction of the topmost layer of smoke is above the out-of-plume PBLH. Node A identifies cases where smoke remains well-mixed throughout the local PBL (Figure 2.1d) and is especially useful for evaluating the models (i.e., Figure 2.1f) as the Freitas model injects flaming emissions at the height where the simulated plume stops rising but for a range of user-prescribed heat fluxes

(Freitas et al., 2007), allowing the possibility for plumes to be injected into multiple vertical levels that could be below the PBLH.

For cases where a majority of the lofted plume resides in the free troposphere, node B1 indicates that downwind behavior will play a role in whether or not plumes will be classified as injections. If plumes maintain their stratified shape (i.e., in all panels of Figure 2.1a), then they are considered injections. This comes with the caveat (node F1) that smoke in the residual layer, although it may appear stratified in shape, is not considered to be an injection in this work. These cases occur when fires were observed after dark, so the daytime convective PBL had been replaced with a less turbulent residual layer, leading to the stratified structure that likely will have more localized air quality impacts the day after if similar PBLH is reached. Cases where the plume mixes with the PBL within 5 min flight time are then considered to be non-injections (nodes C2, E2). If mixing occurs after 5 min of flight time, the case is considered an injection (node E1).

2.2.6 Bootstrapping

Some fires were observed multiple times in a single day, which can bias the data towards those fires. To mitigate such a bias, a bootstrapping approach was used that generates box and whisker plots by subsampling the dataset such that only one aircraft overpass per fire was selected per day. Final boxplots (Figure 2.2 and Figure A50–A52) were generated using the average statistics of all box plots from 2000 such iterations. The false non-injection category did not contain enough cases to include meaningful statistics for either model. HRRR-Smoke did not have enough true non-injection cases to generate statistics, but the true and false injection distributions are similar to the corresponding WRF-Chem distributions (i.e., Figure 2.2).

2.2.7 FRE flux computation

Fire radiative energy (FRE) flux (units of $kJ/m^2/s$) was chosen as the standard quantity with which to compare heat fluxes across the models and observations, and it is derived from the measurements as follows. First, the FRP is summed over all the MASTER pixels present in a single scene and all master pixels are counted, yielding the total FRP and the number of pixels in the scene. Next, the sizes of the MASTER pixels are derived from the altitude of the aircraft above the terrain using a linear relationship (MASTER flight planning), and a fire area is computed as the product of the per-pixel fire area and the total number of pixels. FRE flux is then the slope of the linear fit $FRP = m \cdot AREA$, which in the main text is calculated on the entire dataset as well as subsets. This analysis can be done using smoldering pixels, flaming and saturated pixels, or all pixels.

HRRR-Smoke reports heat fluxes as area-to-FRP ratios which are reported as fuel-dependent climatological values, so we take their inverse to translate them into FRE fluxes. WRF-Chem uses the default convective heat fluxes assumed by the Freitas model, which are assumed to be 0.55 of the total heat fluxes (McCarter and Broido, 1965; Freitas et al., 2007). The resulting convective heat flux is turned into a FRE flux using measured radiant and convective fractions (Freeborn et al., 2008). These radiant and convective fractions were measured at ~3.9 μm, which is similar to wavelengths used by MASTER to measure FRP. The conversion factor is as follows: radiant fraction/convective fraction = 0.12/0.51 = 0.24 (Freeborn et al., 2008). Radiant and convective fraction measurements contain considerable uncertainty, and two instruments were used to measure the radiant fraction (Freeborn et al., 2008). In order to produce the WRF-Chem ranges in Figure 2.3, error propagation calculations were performed and the range including the errors from both radiant fraction measurements are plotted. We can also use

error propagation to get a range of the radiant fraction/convective fraction ratio, which yields 0.239 ± 0.076 and 0.234 ± 0.07 when the different instruments are used to estimate the radiant fraction (Freeborn et al., 2008). The inverse of these calculations can be done to obtain total heat fluxes from measured FRE fluxes. This produces analogous results to the ones described above ("Results", Assessing heat flux assumptions).

2.3 Results

2.3.1 Model performance on smoke injection

Table 2.2 shows the injection behavior of modeled and observed smoke plume profiles using the observed planetary boundary layer height (PBLH) to classify the observations and the modeled PBLH to classify the models (see "Methods", Figure A1–A51, Table A1–A3). Injections occur in 35% of observed cases, 80% of WRF-Chem cases, and 95% of HRRR-Smoke cases. WRF-Chem and HRRR-Smoke have accuracies of 0.51 and 0.44, respectively. Both models tend to capture observed injections well, with WRF-Chem and HRRR-Smoke having true positive rates of 0.94 and 1.0, respectively. However, WRF-Chem and HRRR-Smoke also exhibit false positive rates of 0.73 and 0.92, respectively. For WRF-Chem, in cases where the plumes are within the uncertainty range of the modeled PBLH (i.e., cases which exhibit different injection behavior when the modeled and observed PBLHs are used to evaluate injection, see next section), are excluded from the analysis, then accuracy is 0.67, the true positive rate is 1.0, and the false positive rate is 0.61. This only slightly improves on the error achieved by WRF-Chem when all cases are used. These statistics indicate that overprediction of injection leads to the high true positive rate, low accuracy error type present in both models, consistent with previous work which quantified uncertainty in the Freitas model (Sessions et al., 2011; Val Martin et al., 2012).

HRRR-Smoke against HRRR-Smoke PBL		Non-injection forecasted	Injection forecasted	Total
	Non-injection observed	TN = 2	FP = 24	26
	Injection observed	FN = 0	TP = 17	17
	Total	2	41	43
HRRR-Smoke against DIAL-HSRL PBL		Non-injection forecasted	Injection forecasted	Total
	Non-injection observed	TN = 9	FP = 17	26
	Injection observed	FN = 3	TP = 14	17
	Total	12	31	43
WRF-Chem against WRF-Chem PBL		Non-injection forecasted	Injection forecasted	Total
	Non-injection observed	TN = 9	FP = 24	33
	Injection observed	FN = 1	TP = 17	18
	Total	10	41	51
WRF-Chem against DIAL-HSRL PBL		Non-injection forecasted	Injection forecasted	Total
	Non-injection observed	TN = 20	FP = 13	33
	Injection observed	FN = 2	TP = 16	18
	Total	21	30	51

Table 2.2 Confusion matrices for comparing observed and modeled injection behavior counts. Counts of injection behavior are shown for HRRR-Smoke using the HRRR-Smoke PBL height to evaluate injection, HRRR-Smoke using the DIAL-HSRL PBL height to evaluate injection, WRF-Chem using the WRF-Chem PBL height to evaluate injection, and WRF-Chem when the DIAL-HSRL PBL height is used to evaluate injection. Columns are labeled with forecasted plume behavior and rows are labeled with observed plume behavior. TN stands for true negative, where non-injection was forecasted and observed. Similarly, TP stands for true positive. FN stands for false negative, where injection was not forecast but did occur. FP stands for false positive, where injection was forecast but did not occur. HRRR-Smoke has fewer total cases than WRF-Chem due to certain fires missing in the model (see "Methods", Sampling Method).

2.3.2 Assessing modeled planetary boundary layer height

Underestimation of the PBLH is a known problem with the WRF-Chem model (X.-M. Hu et al., 2010; Shin et al., 2012) that may cause plumes placed at the correct height to be misidentified as injections into the free troposphere. Figure 2.2a and e shows the bias on the modeled PBLH for the WRF-Chem and HRRR-Smoke models, respectively. For the false positive cases, WRF-Chem tends to underpredict the PBLH by 500–1000 m, and HRRR-Smoke tends to underpredict the PBLH by <500 m. Both models show large spread and even contain cases where the models overpredict PBLH and false injection still occurs. In addition, the WRF-Chem model correctly captures non-injections for generally the most underpredicted PBLHs.

Next, we re-classify injection behavior for WRF-Chem and HRRR-Smoke using the observed PBLH (Table 2). There are 12 cases that get reassigned to the true negative category, two that get reassigned to false positive, and one that gets reassigned to false negative. For this evaluation, injection occurs in 59% of cases, resulting in an accuracy of 0.70, a true positive rate of 0.89, and a false positive rate of 0.39, indicating better model performance. When the observed PBLH was used to evaluate the injection behavior of HRRR-Smoke, the injection rate was 59%, the accuracy was 0.53, the true positive rate was 0.82, and the false positive rate was 0.65. Therefore, we conclude that for WRF-Chem, uncertainties in the boundary layer likely do contribute to the overprediction of injection. However, this comes with the caveat that if WRF-Chem could correctly predict PBLH, then it is possible that plume heights might change (Figure A52). We ran a standalone version of the Freitas plume rise scheme based on simulated meteorology for cases where WRF-Chem underpredicted the PBLH and HRRR-Smoke accurately captured it as determined by overlap in the interquartile ranges of the observed and modeled PBLH distributions (Table 1, Figure A52). It was found that plume height and plume

injection behavior have little sensitivity to variations in PBLH present in the WRF-Chem and HRRR-Smoke meteorology used to drive the 1-d model for the same fire cases (Figure A53, note percent error on plume heights do not exceed 8.8%). For HRRR-Smoke, which runs an newer PBL scheme (Table 1), uncertainties in the boundary layer are likely smaller and play less of a role in generating false injections.

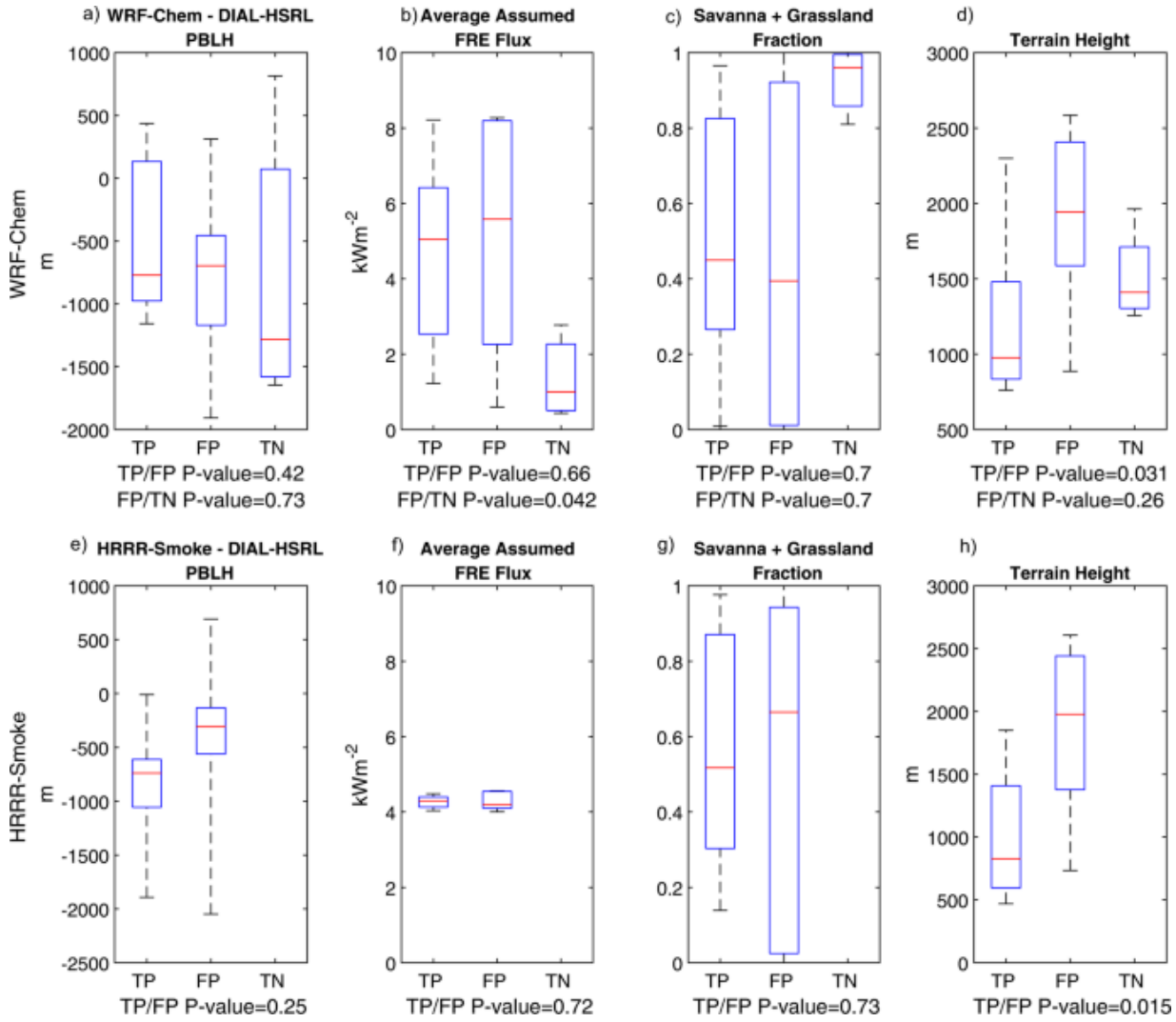

Figure 2.2: Box and whisker plots of model parameters and observations grouped by WRF-Chem and HRRR-Smoke injection behavior. a, **e** Comparison of modeled transect median PBLH with observed transect median PBLH, **b**, **f** average FRE flux, **c**, **g** savanna and grassland fraction, and **d**, **h** terrain height. Red horizontal line denotes medians, blue boxes

denote the interquartile ranges, and whiskers denote the 1.5x interquartile range. TP, FP, and FN are defined as in Table 2.2 caption. **a**–**d** Refer to the WRF-Chem model and **e**–**h** refer to the HRRR-Smoke model.

2.3.3 Associating model performance with additional fire characteristics

Figure 2.2 shows distributions of three additional variables that can help explain the injection behavior of the WRF-Chem and HRRR-Smoke models: average of the model's assumed FRE fluxes (see "Methods", 2b, 2f), model fuel type (2c, 2g), and terrain height (2d, 2h). Since FRE flux is assumed to depend on fuel type in the model, assumed FRE flux and fuel type are strongly correlated with each other. Other variables showing strong correlations with each other include free tropospheric and boundary layer stability, and terrain complexity and heat flux in the HRR-Smoke model (Figure A54–A55). The rest of the variables are at most moderately correlated with each other ($r^2 < 0.46$). Box plots were generated using bootstrapping to avoid sampling bias towards any single fire ("Methods").

Figure 2.2b and f imply that modeled FRE fluxes are likely a strong contributing factor to the overprediction of injection by both models. WRF-Chem captures non-injections correctly when a lower FRE flux is prescribed and either overpredicts or captures injections for larger FRE fluxes. In fact, the difference in FRE flux between correct non-injection and false injection cases (which in the observations are showing the same non-injection behavior) is statistically significant (p-value = 0.042). While HRRR-Smoke did not contain enough correct non-injection cases for a meaningful comparison, HRRR-Smoke assumes similar FRE fluxes for all fuel types, and these values are comparable with what WRF-Chem assumes for cases where false injection is forecasted. Therefore, we may infer that these high heat fluxes contribute to HRRR-Smoke overpredicting injections as well.

Capturing non-injections correctly tends to occur in WRF-Chem for larger fractions of savanna and grasslands (Figure 2.2c), which is consistent as WRF-Chem prescribes lower heat fluxes for savanna and grassland than for forests (Section 2.2.7). Similarly, there is no clear relationship for HRRR-Smoke regarding fuel type as FRE fluxes assumed for all fuels are similar and within the range that WRF-Chem assumes for forest. This indicates that the fuel-dependent heat fluxes assumed by the models (Section 2.2.7) , specifically those corresponding to forest fuels in WRF-Chem and all fuels in HRRR-Smoke, could be high and lead to an overprediction of injection. Other variables such as time of day, static stability, and terrain complexity were investigated on their role on modeled injection behavior but none resulted in being a clear explanatory variable (Figure A54–A55).

Figure 2.2d and h shows that false injections occur at higher terrain elevation than correct injections and that this difference is statistically significant across both models (p-values = 0.031 and 0.015). In addition, Figure 2.2d shows that false injections occur at higher elevations than correct non-injections. Likely, this means that the plume rise scheme's input heat fluxes or meteorology are more uncertain in areas of higher (Figure 2.2d, h) and more complex (Figure A54b, A55b) terrain. The Freitas model is also limited in that it is 1-dimensional and thus cannot simulate interactions between neighboring model columns.

2.3.4 Assessing heat flux assumptions

Figure 2.3a provides an observational constraint on the FRE flux. Scatter points show observations of total FRP versus flaming fire area from the MASTER instrument for each transect and shapes denote the campaign fuel category for each MASTER overpass. Colored regions denote the FRE flux assumptions made by the models. Active fire area is computed for flaming and saturated pixels only, assuming that these fire phases primarily contribute to plume

injection (Freitas et al., 2007). FRP is computed for all categories, as this is what satellites observe. Plots showing fire size and FRP for different combustion phases show similar qualitative results (Figure A56). MASTER, WRF-Chem, and HRRR-Smoke all track heat fluxes slightly differently, so for ease of comparison all have been converted into FRE flux (see "Methods").

Figure 2.3a shows that the best fit for the observed FRE flux (0.55 kWm^{-2}) is within the range of the WRF-Chem savanna (0.40–3.09 kWm^{-2}) and grassland (0.29–0.57 kWm^{-2}) FRE fluxes and is not within the range the WRF-Chem forest FRE flux (2.69–13.86 kWm^{-2}) or any assumed HRRR-Smoke FRE fluxes (3.44–4.76 kWm^{-2}). Put another way, WRF-Chem and HRRR-Smoke assume FRE fluxes equal to 0.5–25 and 6–9 times the average observed FRE flux, respectively. This is consistent with the general lack of overlap between the scatter points and the yellow and orange colored bars in Figure 2.3a. Thus, this provides observational evidence for the hypothesis that the assumed heat fluxes that get used in the Freitas model, specially for forest fuels, are likely too large. Since these heat fluxes are directly used to calculate initial plume buoyancy, this leads to simulated plumes which are more easily lofted into the free troposphere.

A caveat of this finding is that a fraction of the flaming pixels retrieved by MASTER are flagged as saturated, and thus FRP could be underpredicted, affecting our conclusions. Comparing MASTER to satellite observations (which are less prone to saturation due to larger pixel footprint) shows that MASTER tends to provide FRP estimates in between both MODIS (slope = 1.65) and the Visible Infrared Imaging Radiometer Suite (VIIRS, slope = 0.79) (Figure A57). If we assume that VIIRS retrievals are closer to reality, we can use the slope of the scatter plot to correct MASTER retrievals, yielding an average observed FRE flux of 0.7 kWm^{-2} which

does not change our findings. Therefore, we do not expect saturation of MASTER to influence our final result to a large degree.

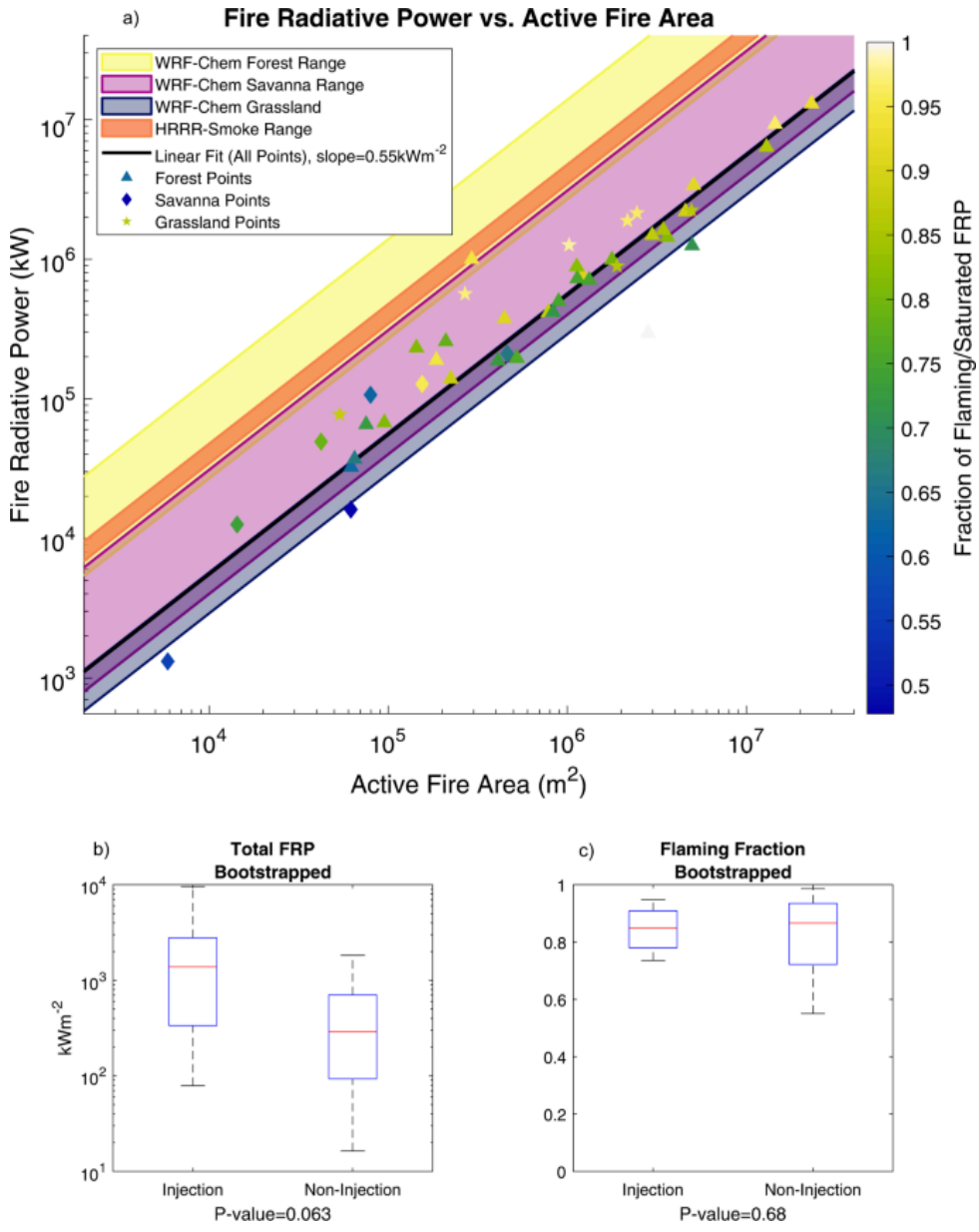

Figure 2.3: Comparison of observed FRE flux and assumed FRE fluxes for WRF-Chem and HRRR-Smoke. a Scatter plot of total FRP and flaming and saturated fire area from MASTER for each fire overpass (colored diamonds, stars, triangles). Black line is a linear fit to the MASTER points of the form $y = mx$. Colored bars represent the ranges of FRE flux

based on models' assumptions for convective heat flux and uncertainty in converting between convective and radiant heat flux (see "Methods"). FRE flux increases toward the upper left. Shapes represent the campaign fuel category. Colors of diamonds, stars, and triangles, represent the fractional contribution of flaming combustion (including saturated pixels) to total FRP. Other panels show box and whisker plots of **b** observed FRP and **c** observed fraction of FRP due to flaming combustion, grouped by observed injection behavior. Box and whisker elements are as in Figure 2.2 caption

To investigate drivers of variability of FRE flux, FRE flux was calculated for MASTER points grouped by the dominant campaign fuel category (see "Methods") and by the flaming fraction of FRP. This resulted in FRE flux of 0.55 kWm^{-2} for forest fires, 0.51 kWm^{-2} for savanna fires, 0.59 kWm^{-2} for grassland fires, 0.56 kWm^{-2} for >85% flaming fraction, and 0.36 kWm^{-2} for <85% flaming fraction. FRE flux is generally constant across fuel type and appears to vary more with fire combustion phase. In other words, given the right spread and fuel consumption conditions, forest and grass fires may burn with the same intensity, which is consistent with previous studies (Stocks & Kauffman, 1997). Thus, heat fluxes used to drive the Fretias model should be small, relatively constant across fuel types, and perhaps more closely associated with fire weather.

Figure 2.3a supports the idea that model heat flux assumptions assign too much buoyant energy per unit area, which is likely the primary contributor to injection overprediction. The overprediction occurs even when a small fire size is assumed in WRF-Chem (0.25 km^2 per grid cell), further supporting the idea that the models assume too much fire energy per unit area. In addition, WRF-Chem tends to correctly identify non-injection events in grass and savanna-dominated fires (Figure 2.2b), as these fires have the lowest modeled heat flux (Freitas et al., 2007). FRE flux used in HRRR-Smoke for all fuel types are generally in the range of FRE fluxes for forest specified in Freitas et al., (2007), consistent with HRRR-Smoke showing a high rate of false injection regardless of fuel type (Table 2). HRRR-Smoke's internal calculation of

total heat flux from FRE flux (see "Methods"), likely keeps plume top heights from being unphysically high. Thus, heat fluxes used to drive the Freitas model should be updated to be consistent with the smaller observed FRE fluxes. The linear relationship between FRP and burned area (black solid line) is consistent across several orders of magnitude which can be used to update the HRRR-Smoke scheme where active fire area is a function of FRP.

Figure 2.3b shows observed injection behavior as a function of observed FRP showing a near-significant trend (p-value = 0.063) of injections happening at higher FRP. Thus, even though the heat fluxes need to be updated in HRRR-Smoke, there is value in driving plume injection with observed FRP, as is already implemented in HRRR-Smoke (Section 2.2.3) and other systems (e.g., Sofiev et al., 2012). MASTER FRP is highly correlated to burned area ($R^2 = 0.96$), and thus approaches using observed burned area to drive plume rise would also be appropriate. On the other hand, Figure 2.3c shows that observed injection behavior has little association with the flaming fraction of MASTER pixels. However, flaming fraction does play an indirect role, as higher flaming fractions are associated with higher FRE fluxes, and thus heat flux could be implemented as a function of flaming fraction if this is available from observations.

2.3.5 Sensitivity analysis

The final step in this analysis was to re-run the WRF-Chem model for selected cases with an adjusted heat flux based on the observed average FRE flux of 0.55 kWm^{-2} (total heat flux of 3.4–5.3 kWm^{-2}). The sensitivity test was run on the WRF-Chem false injections, and a total of 24 of the 51 cases were re-evaluated. These cases are shown as a fourth panel in the relevant transects in Figure A1–A51. WRF-Chem model setup for the sensitivity simulations is as shown in Table 2.1, and HRRR-Smoke is not included in the sensitivity analysis as it was being run in only as a real-time forecast product for the FIREX-AQ period.

Of the 24 cases which were re-run with updated heat fluxes, 18 remained false injections and 6 were re-classified correctly as non-injections. This yields an improved false positive rate of 0.75, given that the initial false positive rate of this subset was 1. If injection behavior is evaluated using the observed PBLH as the injection boundary, then 7 cases remain false positive and 17 cases become correctly re-classified as non-injections for a false positive rate of the subset equal to 0.29. The sensitivity analysis shows that both an accurate boundary layer and an accurate heat flux are needed to correctly model injection behavior.

On average across these 24 cases, in-plume median mixed layer height (a proxy for smoke top height, Ye et al., 2021) decreased by 455 m with the inclusion of more realistic heat fluxes. Grouped by model fuel category, this corresponds to a 687 m decrease for extratropical forest fires (14 cases), a 141 m decrease for savanna fires (7 cases), a 20 m decrease for grass fires (3 cases). Of the six cases that were re-classified as non-injections with respect to the WRF-Chem PBLH, 5 had extratropical forest as their dominant model fuel category and 1 was a savanna fire. All but one saw a reduction in the median in-plume smoke top height (average decrease of 506 m across the 6 cases). Of the 18 cases which remained injections, 9 were extratropical forest-dominated, 6 were savanna-dominated, and 3 were grassland-dominated.

Using heat fluxes and boundary layers that are in line with observed values reduces the occurrence of false positives and reduces mean smoke top height for the WRF-Chem model. This effect is consistent with other studies which have shown that modeled plume top heights decrease in proportion to decreasing heat flux values (Val Martin et al., 2012), with the fact that the Freitas model assumes the highest heat flux values for forest fires (Freitas et al., 2007), and with studies that have looked at the underprediction of the PBLH by WRF-Chem (X.-M. Hu et al., 2010; Shin et al., 2012). While we note that classifying injection behavior with respect to the

observed boundary layer does improve model performance, further improvements on WRF-Chem PBLH are expected by using more recent versions of the model (the one used here dates to 2014) which are more consistent with the HRRR-Smoke model's reduced biases in PBLH.

2.4. Discussion

We compared smoke injection behavior in the WRF-Chem and HRRR-Smoke models against airborne lidar observations and found that both models tend to overpredict injection behavior. Unrealistically high heat fluxes used to drive buoyancy of plumes is identified as a likely contributor to the high rate of false injections in both models, and underprediction of PBLH is identified as a second contributor to the high rate of false injections in the WRF-Chem model. Observations of FRP and fire area from the MASTER instrument provide an observational constraint on heat flux, which we used to determine that the assumed FRE fluxes in WRF-Chem (HRRR-Smoke) are generally higher than observed by a factor of 0.5–25 (6–9). The observations mainly fell within the ranges of FRE flux assumed in WRF-Chem for grassland and savanna fires (average FRE flux = 0.55 kWm^{-2}), consistent with WRF-Chem correctly predicting non-injections for high fractions of grassland and with HRRR-Smoke predicting a high rate of false positives. FRE flux was shown to vary weakly with fuel type and more strongly with fraction of FRP due to flaming combustion. Adjusting the heat fluxes used in WRF-Chem based on field data partially improves the overprediction behavior that dominated the error in this model, and further improvement is achieved when the model injection behavior is evaluated using the observed PBLH as the injection boundary. Sensitivity tests for HRRR-Smoke were not performed, but we hypothesize that reducing heat fluxes will improve injection representation,

although this effect may be somewhat damped by HRRR-Smoke's internal calculation of total heat flux.

The FIREX-AQ dataset we used is limited in that it targets only Western United States wildfires during the relatively normal 2019 fire season. In addition, due to the nature of observing wildfire smoke with an aircraft, more intense and lofted plumes were preferred by the flight planners. In spite of these limitations, the western FIREX-AQ dataset contains a range of fire sizes, injection behaviors, and fuel types (Fig. 2.3a). In addition, these results are consistent with studies which used satellites to evaluate plume rise (Sessions et al., 2011; Val Martin et al., 2012). However, due to the limited spatial domain and its limited ecoregions and meteorological regimes and (e.g. lack of tropical vegetation and deep PBLs), these results are most relevant for North American wildfires.

Our results also show that heat fluxes may not be dependent on fuel in the way the Freitas model assumes. The Freitas plume rise parameterization was originally developed for tropical forest fires (Freitas et al., 2007; Val Martin et al., 2012), so it is possible that these constants are tuned to this region and thus not wholly applicable to Western US wildfires.

Past studies have identified shortcomings with the Freitas model, including underpredicting the injection height range and misidentifying injections (Val Martin et al., 2012; Ye et al., 2021). The data available from FIREX-AQ have allowed us to pinpoint the assumed heat fluxes and boundary layer uncertainty contributors to the known issues with the performance of the Freitas model. We expect our results will improve smoke modeling across a variety of disciplines, including air quality and visibility forecasting, wildfire-climate interactions, and health impact studies.

Chapter 3

Forecasting daily fire radiative energy using data driven methods and machine learning techniques

An edited version of this manuscript has been accepted for publication in the Journal of Geophysical Research-Atmospheres.

Abstract

Increasing impacts of wildfires on Western US air quality highlights the need for forecasts of smoke emissions based on dynamic modeled wildfires. This work utilizes knowledge of weather, fuels, topography, and firefighting, combined with machine learning and other statistical methods, to generate 1- and 2-day forecasts of fire radiative energy (FRE) for existing fires. The models are trained on data covering 2019 and 2021 and evaluated on data for 2020. For the 1-day (2-day) forecasts, the random forest model shows the most skill, explaining 48% (25%) of the variance in observed daily FRE when trained on all available predictors compared to the 2% (<0%) of variance explained by persistence for the extreme fire year of 2020. The random forest model also shows improved skill in forecasting day-to-day increases and decreases in FRE, with 28% (39%) of observed increase (decrease) days predicted, and increase (decrease) days are identified with 62% (60%) accuracy. Error in the random forest increases with FRE, and the random forest tends towards persistence under severe fire weather. Sensitivity analysis shows that near-surface weather and the latest observed FRE contribute the most to the skill of the model. When the random forest model was trained on subsets of the training data produced by agencies (e.g. the Canadian or US Forest Services), comparable if not better performance was achieved (1-day R^2=0.39-0.48, 2-day R^2=0.13-0.34). FRE is used to

compute emissions, so these results demonstrate potential for improved fire emissions forecasts for air quality models.

3.1 Introduction

In this work, we seek to predict fire radiative energy (FRE) for already-burning fires in a more skillful way than assuming persistence. Although predicting ignition is an important piece of this problem, particularly for long-term air quality forecasts, recent work has shown that human related variables such as gross domestic product and the presence of human settlement and cropland are key drivers for fire ignition (Mukunga et al., 2003; De Vasconcelos et al., 2001). However, in the Western US, prescribed and agricultural burning has lagged behind the Southeast (Burke et al., 2020), so predicting fire evolution and predicting ignition are two somewhat decoupled problems.

In order to make FRE predictions that are more skillful than persistence, we must understand the drivers of wildfire spread and intensity. These variables are governed by the interactions between three sides of "the fire behavior triangle": weather, topography, and fuel (Countryman, 1972). First, the atmosphere can drive fire via low humidity, high temperature and wind speed, and vertical instability (Rothermel, 1991; Srock et al., 2018). Second, fires spread more easily in an uphill direction (Rothermel, 1972), as a steep slope allows for preheating of fuels, but complex flows in mountainous regions may confound this relationship (Viegas, 2004). Finally, fuel type, moisture, and loading are key components of determining the intensity of a fire (Rothermel, 1972). The moisture of dead fuels and organic soils respond over the course of a day to changing weather conditions (Wagner, 1987), while the moisture content of living fuels and

organic soils can hold a 1-2 month memory of previous moisture conditions (Krueger et al., 2022). The latter has been shown to be a good indicator of fire intensity (Sazib et al., 2022).

Currently, wildfire indices are used to inform wildfire behavior predictions, plan firefighting strategies, and forecast fire intensity. These indices may represent one aspect of the fire behavior triangle, such as the hot-dry-windy index (HDW) or the continuous Haines index (CHI) which represent the effect of meteorology on wildfire (Srock et al., 2018; Pinto et al., 2020). These indices may represent multiple aspects, such as the Canadian Forest Fire Danger Rating System (CFFDRS), the National Fire Danger Rating System (NFDRS), or the National Oceanic and Atmospheric Administration (NOAA) hourly wildfire potential (HWP) which represents fuel and weather effects on wildfire (Bradshaw et al., 1984; Wagner, 1987; Text B1). To the best of our knowledge, wildfire indices which include all aspects of the fire triangle are rare. However, wildfire indices such as the CFFDRS Fire Weather Index (Di Giuseppe et al., 2018; Di Giuseppe et al., 2017) and vapor pressure deficit, and/or wind speed have been used to successfully modulate persistence-based forecasts (Graff et al., 2020) forecasts or to derive conditional probability estimates of FRP exceedances (Hernandez et al., 2015)**.** Additionally, components of the NFDRS, notably the Energy Release Component, have served as inputs to machine learning frameworks which have predicted monthly PM2.5 and daily fire perimeters (Huot et al., 2022; Wang et al., 2022). Other fire weather indices not covered in this paper, such as the Grassland Fire Danger Index, have shown promise for modulating climatological predictions of fire occurrence in regions such as Sub-Saharan Africa which have a regular burning season (Partanen & Sofiev, 2022).

In addition to the natural fire behavior triangle, firefighting may also shape the evolution of a wildfire. Naturally occurring wildfires may be managed or suppressed to thin fuels or protect

life and property (Young et al., 2019), and this may affect the final burned area (Hu & Ntaimo, 2009). Since burned area is correlated with both FRP and emissions (e.g., Thapa et al., 2022; Wiedinmyer et al., 2011), we would expect containment to impact these quantities as well. However, much of the study of firefighting as it pertains to intensity and burned area has focused on including containment in fire spread modeling (Hu & Ntaimo, 2009), planning the most effective containment strategies (Rodrigues et al., 2020), and understanding containment successes and failures (Arienti et al., 2006). Only recent studies have connected containment efforts, fire spread, and FRP (Jolly et al., 2019; Turney et al., 2023). While these studies underscore the importance of including containment in fire spread forecasts, they also highlight that containment efforts can be hindered by severe fire weather and the inaccessibility of some fires (Arienti et al., 2006; Rodrigues et al., 2020).

Machine learning techniques are a subset of empirical-statistical models which can learn underlying relationships in systems with many predictive variables. The random forest machine algorithm has been applied successfully to multiple areas within the wildfire problem including understanding drivers of fire severity in mediterranean pine forests (García-Llamas et al., 2019), modeling containment probability using land surface variables (Rodrigues et al., 2020), and detecting fires from geostationary satellites over South Korea (Jang et al., 2019). More broadly, statistical models have been applied to predict fire evolution using meteorological quantities, fire weather and moisture indices, and land cover datasets. In Graff et al. (2020), the authors train a poisson regression on reanalysis data from the European Center for Medium-range Weather Forecasts (ERA-Interim) and MODIS land cover and are able to predict fire counts with more skill than persistence out to five days. Peterson et al. (2013) used a maximum likelihood classification method to predict the daily change in fire counts based on CFFDRS fire indices,

convective available potential energy, moisture, and synoptic weather, and achieved better results than persistence. Finally, Pinto et al. (2020) compare the utility of the CFFDRS fire weather index alone and the CFFDRS fire weather index and the continuous Haines index, and they find that using both surface and stability information allows for a more skillful forecast of FRP exceedance probabilities. Generally statistical forecasts which have beaten persistence have tended to focus on two of the three legs of the fire behavior triangle. However, Huot et al., (2022) and Wang et al., (2022) successfully trained various machine learning algorithms on weather, topography, and fuel to generate gridded maps of burned area and monthly PM2.5. In these studies, some human influence variables were included (population density, GDP), but predictors relating to firefighting were not included.

In this work, we use weather indices and random forest machine learning as tools to forecast daily changes in fire radiative energy (FRE, the time-integrated form of FRP) in the Western US with the goal of improving upon the persistence assumption and maintaining computational efficiency. Unlike previous work (e.g., Huot et al., 2022; Wang et al., 2022) we compute changes in FRE on a fire-by-fire basis, making this work directly applicable for emissions inventories which use fire energy to compute emissions. Although we evaluate FRE only in this work, we can assume since emissions are proportional to FRE, any changes in FRE dictated by our models would translate to the same change in emissions. For the machine learning approaches, the training dataset consists of variables describing near-surface weather, atmospheric stability, fuel density and water stress, fuel moisture, and firefighting efforts, and our target variable is based on a blended geostationary/polar orbiting emissions product (Li et al., 2022). All algorithms were trained on data from the 2019 and 2021 fire seasons and tested on comparable data from the 2020 fire season. In this work we look at July-December, as these were

the months which contained significant fire activity from both summertime and wind-driven fire events. The methods tested represent varying levels of computational complexity, from daily changes in weather being assumed to be proportional to daily changes in fire intensity, to random forest machine learning.

3.2. Data

This study focuses on wildfires which occurred between July and December in 2019-2021 in the Western Continental US (Lat: 31N – 49N, Lon: 125W – 101W). Features were extracted (Section 3.3.2) for each day of each fire, totalling 1238 training samples (2019 and 2021) and 1293 1-day (1160 2-day) testing samples (2020). Table 3.1 lists the availability, resolution, aggregation method, and primary reference for each dataset used in this study. GFWED, NCAR Fuel Moisture, and final MTBS perimeters, are only available for portions of the Jan 1st 2019- Dec. 31st 2021 period. RAVE data are not available before Jan. 1st 2019. At the time of analysis, MTBS perimeters for the 2021 fires were considered provisional. For further description of the datasets used, see Text B1.

Dataset ***(Type of data)***	**Available Dates in 2019-2020**	**Resolution**	**Variables Extracted**	**Aggregation Method**	**Reference (date of access)**
VIIRS ***(fire location)***	SNPP: 1/1/2019-12/31/2021 NOAA-20: 1/1/2020-12/31/2021	375m, twice daily	Lat/lon of detection	Aggregate detections into polygons (Section 3.3.1)	Schroeder et al., 2014 (accessed 9/13/2021, 10/20/2022, 4/5/2023)

MTBS ***(fire location)***	All 2019-2020, partial 2021. 2021 filled with MTBS provisional perimeters.	Total burn perimeter, one per fire	Final fire perimeter	N/A	Eidenshink et al., 2007 (MTBS Apr. 2023 and Provisional May 2023)
RAVE ***(fire intensity)***	1/1/2019-12/31/2021	3km hourly	Fire Radiative Energy (FRE)	Summed over day and polygon	Li et al., 2022 (March 2023)
GridMET ***(fire weather, fuel moisture)***	1/1/2019-12/31/2021	4km daily	Vapor pressure deficit (VPD); wind speed (WS); max and min temperature (T_{max}, T_{max}); NFDRS Burning Index (BI) and Energy Release Component (ERC); 100hr and 1000hr fuel moisture	Averaged over day and polygon	Abatzoglou, 2013 (Aug. 2023)
HRRR ***(fire weather, soil moisture)***	1/1/2019-12/31/2021	3km, hourly	VPD, WS, hot-dry-windy index (HDW); hourly wildfire potential (HWP); continuous Haines index (CHI);	Averaged over day and polygon	Dowell et al., 2022 (Mar.-Apr. 2023)

			pyroCb firepower threshold (PFT); surface, 1cm, 4cm, 10cm, 30cm soil moisture		
GFWED *(fire weather)*	1/1/2019-9/30/2021	0.1deg, daily	CFFDRS fire weather index (FWI); buildup index (BUI); fine fuel moisture code (FFMC)	Averaged over day and polygon	Field et al., 2015 (Apr. 2023)
NOAA Soil Moisture *(moisture)*	1/1/2019-12/31/2021	0.25deg, daily	Surface soil moisture (Blended_SM)	Averaged over day and polygon	Liu et al., 2016 (Mar.-Apr. 2023)
NCAR Fuel Moisture *(moisture)*	8/1/2019-11/26/2021	1km, daily	Dead fuel moisture (FMCG2D); live fuel moisture (FMCGLH2D)	Averaged over day and polygon	McCandless et al., 2020 (Sep.-Oct. 2021)
Evaporative Stress Index (ESI) *(moisture)*	1/1/2019-12/31/2021	5km, weekly	ESI	Averaged over day and polygon	Anderson et al., 2011 (Apr. 2023)
Plant Water Sensitivity (PWS) *(moisture)*	Assumed static for 2019-2021 period	4km, static	PWS	Static map copied to daily, averaged over day and polygon	Rao et al., 2022 (Aug 2023)
Fuel Loading	Assumed static for 2019-2021 period	900m, static	Low_N, Moderate_N, High_N,	Static maps copied to daily, averaged	Peterson et al., 2022; Soja et al., 2004 (Feb. 2023)

			VeryHigh_N, Extreme_N	over day and polygon	
Terrain	Assumed static for 2019-2021 period	900m, static	Slope, elevation (mean and stdev)	Static map copied to daily, averaged over day and polygon	Landfire, https://www.landfire.gov/version_download.php (Aug. 2023)
SIT-209 ***(firefighting)***	1/1/2019-12/31/2021	N/A, ~daily	Containment (%), structures, crews, engines, aircraft, construction, overhead	N/A	Jamieson, 2005a,b (2022)
Population Density	Assumed static for 2019-2021 period	2.5 arcmin, static	Population density	Static map copied to daily, averaged over day and polygon	CIESIN (2018)

Table 3.1. Summary of data used in this work, including temporal coverage, temporal and spatial resolution, aggregation method, and primary reference.

3.3 Methods

3.3.1 Building Polygons

Previous work has tracked sub-daily to daily wildfire movement by aggregating VIIRS 375m fire detections into polygons (Berman et al., 2023; McClure et al., 2023). In this work, a similar approach is taken:

1. For each overpass and each active fire, all VIIRS detections potentially corresponding to each fire are selected. Corresponding points are identified as points within a 0.25-degree radius of the previous day's polygon (gray shaded region in Figure 3.1a, Step 2). If it's the first day of the fire, the "previous day's polygon" is the ignition location from the

SIT-209. The radius is increased by 0.25 degrees incrementally and the search repeats recursively until no new points are added, with the 0.25-degree buffer chosen to reduce computing time. This process is shown graphically in Figure 3.1a for the first two days of the 2020 August Complex fire.

2. The per-overpass polygon is built by aggregating all fire detections, each of which have been buffered to a circle with a diameter matching the area of VIIRS pixels at nadir (375m x 375m), using the Python Shapely module's multipolygon datatype (Gillies et al., 2007).
3. Sub-daily polygons are aggregated to daily resolution and clipped using the final burn perimeters from MTBS. Clipping with final MTBS perimeters confirms that the sub-daily polygons generated in Step 2 are associated with the correct fire event. Figure 3.1b shows the final polygon for August Complex.

This method evaluates well against the method developed by McClure et al. (2023) for the August Complex Fire (not shown).

3.3.2 Extracting Data Using Polygons

Due to the variable resolutions of training variable datasets, the polygons discussed above are used to extract only the portions of the training data which directly impact each fire on each day (Figure 3.1c, B1). This method has several advantages. First, it allows us to track the movement of the fire as a whole and the factors affecting the fire (including firefighting, which is reported per-fire) on a daily basis. Second, it is flexible for fires of different sizes. Finally, this method is capable of preserving high resolution information in cases where the feature grid may be finer resolution than the polygon size (Figure B2). A polygon representing a fire on a single day (e.g., Figure 3.1a) may be used to extract data as follows:

1. Grid cells falling in a bounding box defined by the polygon are determined, and each grid cell is transformed into a polygon using the shapely python library (Gillies et al., 2007).
2. The intersection area of each grid cell with the polygon is calculated (Gillies et al., 2007), and the fraction of the fire polygon which intersects each grid cell is saved.
3. Most variables are extracted as a weighted average over all the grid cells which intersect the polygon, with the weights being the results of step 2. FRP is an exception as it is extracted as the unweighted sum of all grid cells which intersect the polygon.
4. Values of the variables are aggregated (averaged or summed, Table 3.1) to daily resolution, with the day beginning at 12Z. 12Z was chosen to preserve the midafternoon peak in fire intensity, which occurs around 0Z.

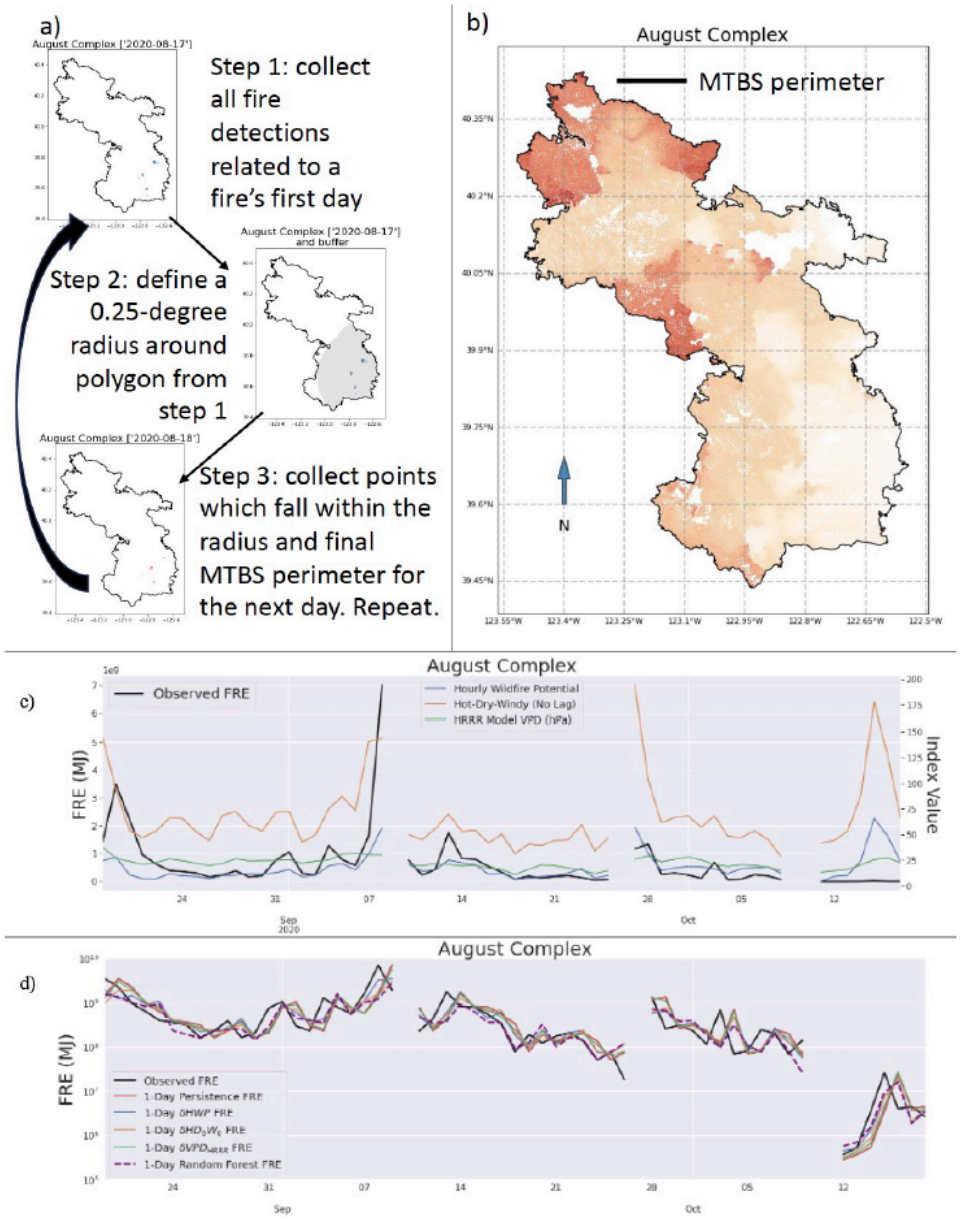

Figure 3.1. Workflow example for the 2020 August Complex Fire. a) the process used to build two days of fire polygons for August Complex and b) the final August Complex polygon at daily resolution (light orange for mid-August 2020 through dark red for mid-October 2020). Time series of c) FRE (black) and near surface fire weather indices (HWP—blue, HDW—orange, HRRR VPD—green) extracted using the daily polygons, and d) observed (black) and forecasted FRE using persistence (red) scaled weather (HWP—blue, HDW—orange, HRRR VPD—green) and random forest model trained on all predictors (magenta dashed).

3.3.3 Predicting FRE with Scaled Fire Indices and Random Forest

Daily FRE values span around six orders of magnitude (Figure B3a-B3b), owing to the extreme nature of the chosen wildfire seasons. Therefore, we introduce the scaling operator (Equation 3.1) which represents the daily change in a variable X over a fire polygon from one day to the next (Figure 3.1d).

(3.1) $\delta X = \frac{X_t}{X_{t-1}}$, where t is the valid time of the forecast, and t-1 is the previous day.

Applying the scaling operator to FRE (δFRE) gives the right-skewed distribution in Figure B3c. To deal with skew, the machine learning models are trained to predict the log of FRE (Figure B3d) so as to give equal weight to large increases (δFRE>>1) and large decreases (δFRE<<1). Applying the scaling operator as well as the log transformation of FRE helps the model to deal with how different aspects of the fire behavior triangle affect the evolution of wildfires at any size. Working with a scaling factor allows us to further simplify the persistence model, as persistence would predict a scaling factor of 1 for all cases. We removed scaling factors which lie more than 1.5 interquartile ranges from the median, since our goal is to predict the evolution of existing wildfires, rather than how fires start or end. To minimize the impact of temporal covariance, 2019 and 2021 were chosen as the training periods and 2020 was chosen as the testing period. Scaled methods are evaluated for 2020 only so that the comparison is consistent.

In a strategy similar to that presented in Graff et al., (2020), Equation 3.1 can be applied to FRE and rearranged to obtain an FRE prediction (Equation 3.2, Figure 3.1d). The latest FRE observation is derived from RAVE as a daily (12Z-12Z) sum of hourly FRE. Equation 3.2 can

also be used to perform multiple steps of integration in time (Equation 3.3). The following paragraphs discuss two methods for how how $\delta_{pred,\,t}$ is derived.

(3.2) $$FRE_{pred,\,t} = \delta_{pred,\,t} * FRE_{obs,\,t-1}$$

(3.3) $$FRE_{pred,\,t+1} = \delta_{pred,\,t+1} * FRE_{pred,\,t}$$

The scaled fire index (SFI) method applies the scaling operator (Equation 3.1) to daily- and polygon- averaged values of fire indices to generate a value with which to modulate persistence. To do this, HRRR wind speed and VPD, GridMET wind speed and VPD, HDW, HWP, FWI, and BI are plugged into Equation 3.1. The lagged version of the HDW (HDmW0, Text B1)was also tested as a scaled variable. In subsequent discussion and figures, SFIs are denoted as Index, for instance HWP for the scaling operator applied to the Hourly Wildfire Potential. SFIs are included to be a simpler alternative to the random forest that does not require training a machine learning algorithm.

We exclude cases where a SFI gives outliers that are further than 1.5 times the interquartile range away from the median. In total 9 (51) 1-day (2-day) forecast fire days contained a scaled index which was considered an outlier. 38 of the 2-day SFIS were zero values, primarily from zeros in the NFDRS BI and CFFDRS FWI. The remaining outliers were from HWP and BI, which contained unphysically large or small values (Table B1-B4). This filtering was done in addition to filtering outliers from the observed FRE. This filtering was applied before the random forest was trained, resulting in 1296 training fire-days for 2019 and 2021 (382 decreases, 435 slight changes, and 479 increases) and 1293 (1160) 1-day (2-day) fire-days to evaluate for 2020.

Random forests (RF) are a machine learning technique that use ensembles of decision trees which split the training dataset into homogeneous subgroups and average predictions across trees in the forest (Breiman, 2001). During exploratory research, we tested the RF, support vector machine, multilayer perceptron neural network, and XGBoost, and the RF model performed the best and had a relatively shorter computing time. In this work, the scikit-learn implementation is used, with the default values used for all but two hyperparameters. We found that min_samples_leaf=20, the parameter controlling the number of sample points needed for a leaf node to be created, and max_depth =15, the parameter for controlling the depth of the trees in the forest, were best for minimizing overfitting. The random_state variable, which controls the bootstrapped subsampling of the feature sets used to train the trees in the forest, was set to 42 to ensure that models were trained on the same data points across feature configurations.

Our RF is trained on the full set of training variables (Table 1), including SFIs (Section 3.3.1) as well as subsets (Section 3.4.2-3.4.5). The RF was trained to predict the log of δFRE to allow the model to learn both the large increases and the large decreases (Figure B3d). The RF was trained on data extracted over observed polygons, assuming that, since these fires were past, conditions on the day of the prediction could be known. On the other hand, the random forest was tested using data extracted over the latest observed polygons to model real case scenarios where the location of the fire into the forecast window is unknown. Forecasted weather variables were used to make the 1- and 2- day predictions of fire growth and decay, and for other variables (fuel, topography, firefighting, moisture, FRE) the day 0 values over the latest observed polygon were persisted.

The 1-day RF prediction was made following Equation 3.2, where $\delta_{pred,\,t}$ was derived from the RF which was given day 1 and day 0 weather inputs and day 0 fuel, topography,

firefighting, moisture, and FRE inputs. The 2-day RF prediction was derived following Equation 3.3 where $\delta_{pred,\,t+1}$ was derived from the RF which was given day 2 and day 1 weather inputs and day 0 fuel, topography, firefighting, moisture inputs, and predicted day 1 FRE values.

3.3.4 Evaluation metrics for predicting categories, scaling factors, and FRE

The SFIs and the RF model were evaluated on their ability to represent the daily change in fire behavior (increase, decrease, slight change) and FRE magnitude. Predicted scaling factors and FRE values (δ_{pred} and FRE_{pred}) were compared with observed scaling factors and FRE values (δ_{obs} and FRE_{obs}) for the testing period.

Whether FRE for a given fire increased, decreased, or remained constant on a day-to-day basis is based on thresholds of the FRE, where FRE=0.67-1.5 (33% decrease-50% increase) represents slight change and anything above (below) that range is an increase (decrease). The ability of the model to forecast the daily change category is assessed using 3-by-3 tables and the associated F1 statistic (Figure 3.2). The F1 statistic (Equation 3.6) is derived from the harmonic mean of the precision (Equation 3.4) and recall (Equation 3.5), and higher values of F1 indicate both a higher proportion of true category instances modeled (recall) and a higher proportion of correct predictions of a given category (precision, Powers, 2008).

(3.4) $P_k = \frac{True\ predictions\ of\ category\ k}{All\ predictions\ of\ category\ k}$ where k = increase, decrease, and slight change

(3.5) $R_k = \frac{True\ predictions\ of\ category\ k}{All\ observed\ occurrences\ of\ category\ k}$ where k = increase, decrease, and slight change

(3.6) $F1_k = 2\frac{P_k * R_k}{P_k + R_k}$ where k = increase, decrease, and slight change

Predicted scaling factor and FRE magnitude are assessed using R^2, mean absolute deviation (MAD), and root-mean-squared error (RMSE). RMSE and MAD are defined as in Willmot and

Matsuura (2005), and R^2 is the coefficient of determination. The adjusted R^2 (Equation 3.7) is also computed and can be interpreted as the percent of the variance explained by the model, adjusted for the number of predictors used to train the model. Adjusted $R^2$and R^2yield very similar results, indicating that more than additional model features are responsible for improved model performance (Table B5).

(3.7) $Adjusted\ R^2 = 1 - \frac{(1-R^2)*(n-1)}{(n-k-1)}$ where R is the coefficient of determination, n is the number of observations, and k is the number of predictors.

We report the unitless skill score (SS), the relative change in performance metric for a model with respect to a reference model. Skill scores are calculated for RMSE and MAD (SS_{RMSE} and SS_{MAD}) using persistence as the reference model (Graff et al., 2020; D. Peterson et al., 2013). R^2 and SS_{RMSE} are sensitive to large values, so these statistics will see larger improvements when large increases are correctly forecasted.

3.4. Results

3.4.1 Comparison of Two Modeling Approaches with Persistence

Figure 3.2 compares the performance of the persistence model, the SFI method (Index, Section 3.3.1), and the RF trained on the entire set of training variables (Table 3.1) over two days for the 2020 fire season. Fig. 3.2a-e (3.2f-j) show modeled and observed growth categories (Section 3.3.4) for selected forecasts of FRE. Fig 3.2k-n summarize error metrics for all methods, with Figures 3.2k, 3.2m showing performance of the FRE forecasts (R^2, SS_{RMSE}, SS_{MAD}) and Figures 3.2l, 3.2n showing performance of the forecasted categories ($F1_{decrease}$ and $F1_{increase}$).

Overall, Figure 3.2 shows that the random forest is the top method tested and also that there is utility in using SFIs to predict FRE. In both the 1- and 2-day forecasts, the RF is the most skillful method (Figure 3.2e, 3.2j, 3.2k-3.2n) except for in R^2 in the 2-day forecast (Figure 3.2m),

where it is beat only by scaled HWP. In the 1-day forecast, skill at predicting FRE or growth category increases as the SFIs get more complex or the RF is used (Figure 3.2k). For the 2-day forecasts (Figure 3.2m, n) the SFIs using simpler fire weather indices or raw meteorological variables (HDW, HWP, VPD) can show comparable FRE prediction skill to the RF. All methods tend to do better than persistence at forecasting FRE, and they all forecast increases and decreases at roughly the same rate.

Certain models (persistence, δVPD (HRRR), δHD0W0, δHWP, and random forest) are highlighted to illustrate common error types (Figures 3.2a-f, B4a-j, B5a-j). Figure 3.2a and 3.2f show, as expected, that the persistence model is only capable of forecasting slight changes in FRE. The "persistence" rows of Figures 3.2l and 3.2n show the F1 score of the increases and decreases is zero, indicating an inability for persistence to capture increases (decreases) of >50% (<33%) over the latest FRE measurement. In spite of this limitation, persistence gives a slightly capable 1-day FRE forecast (Figure 3.2k, R^2=0.02) and a severely limited 2-day FRE forecast (Figure 3.2m, R^2=-0.62) for 2020 fires.

Figures 3.2b-c and 3.2g-h show that assuming that daily changes in FRE are driven by daily changes in VPD (δVPD (HRRR)), HD0W0 (δHD0W0), and HWP (δHWP) adds skill when compared with the persistence forecast. These forecasts are relatively low risk in that they boost performance on FRE (R^2 =0.345,0.437, and 0.418 for the three methods respectively vs. 0.02 for persistence, Figure 3.2k, 3.2l), while not straying too far from the categories persistence would predict (Figure 3.2l, 3.2n). Across both the 1-day and 2-day forecasts, derived indices δ HD0W0 and δHWP are more skilled than raw variables δVPD and δWS (Figure 3.2k-n). However, some of the more complex SFIs (BI and FWI) struggle to reliably capture fire growth

or decay (Figure 3.2l, 3.2n) so perhaps index complexity is not always an indicator of ability to capture fire growth.

Figures 3.2e and 3.2j show that the random forest trained on the entire set of training variables is the top performer of all methods tested. The 1-day (2-day) random forest has R^2 =0.48 (R2 =0.25) and gives SS_{RMSE} =0.27 and SS_{MAD} =0.28 (SS_{RMSE} =0.32 and SS_{MAD} =0.33) with respect to persistence (Figure 3.2k, 3.2m). The random forest also shows the strongest ability to forecast increases and decreases with little bias, with the F1 score being 0.38 and 0.48 (0.50 and 0.57), respectively (Figure 3.2l, 3.2n). The following results sections will focus exclusively on the random forest model, with Section 3.4.2 evaluating the random forest over space, time, and fire severity levels, 3.4.3 determining which predictors add the most skill to the model, and section 3.4.4 and 3.4.5 discussing which subsets of the feature space lead to optimal model performance.

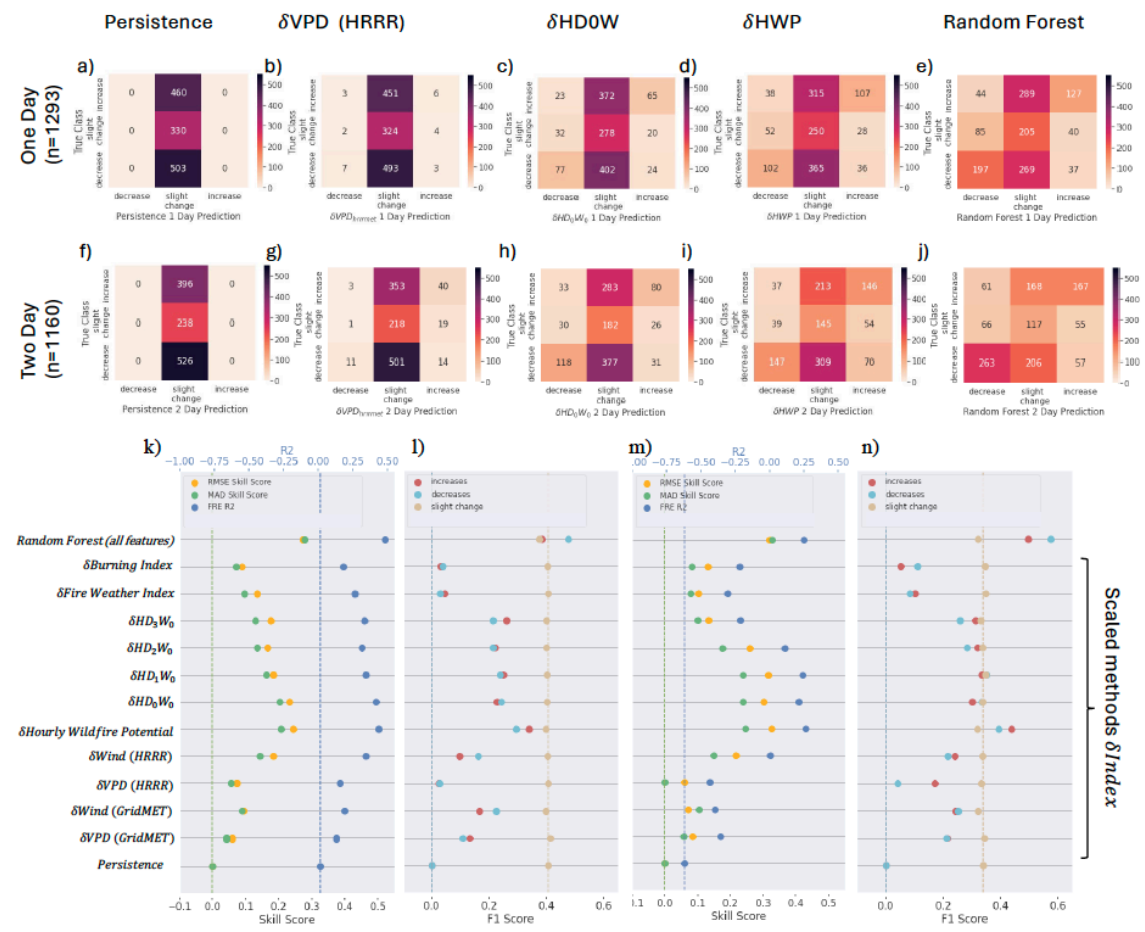

Figure 3.2. Comparison of persistence, scaled, and random forest methods. Performance of the a, f) persistence, b,g) VPD (HRRR), c,h) HD0W0, d, i)HWP, e,j) random forest method on their ability to forecast fire growth category, F=0.67-1.5 is slight change and greater (less) than that value is an increase (decrease). R^2 and SS_{RMSE} and SS_{MAD} are shown for FRE for all 1-day k) and 2-day m) predictions in blue, yellow, and green dots respectively. F1 for all 1-day l) and 2-day n) predictions with increases in red, decreases in blue and slight changes in tan. Rows are labeled according to the forecasting method, with scaled methods labeled with a "" preceding the name of the scaled variable. Vertical dashed lines denote performance of persistence.

3.4.2 Evaluation of the random forest in space, across fire severity levels, and in time

We also evaluate the performance of the random forest by state, by two indicators of fire severity (FRE and FWI), and by day of the fire season. Across all states, using the RF model results in a more skilled or a very slightly less skilled FRE forecast than persistence (Figure 3.3a,

B6a), as shown by the SS_{RMSE} and SS_{MAD} being positive or slightly negative. Generally speaking, the SS_{RMSE} and SS_{MAD} tend to be higher for the 2-day forecast than for the 1-day forecast (Figure 3.3a, B6a), except in UT and CO. More specifically, the RF model gives the largest improvements in CO for the 1-day forecast and the largest improvements in NV for the 2-day forecast. In WY, MAD and RMSE remain high on both days (Figure 3.3c,e, B6c,e). Using the RF also improves R^2 in most states in the 1- and 2-day forecasts. Except for NM (Figure 3.3b), all states show R^2 either changing sign or becoming more positive/less negative (Figure 3.3b, B6b). In the 1- and 2-day forecasts, the pattern of FRE error (RMSE, MAD) is driven mainly by FRE, with higher FRE states having higher error (Figure B7-B10, panels c-d). The remaining FRE error metrics (SS_{RMSE}, SS_{MAD}, and R2) are driven by weather and moisture variables on day 1, with model skill reducing under more extreme fire weather conditions and increasing for more moist fuels (Figure B7-B8 panels a,b,e). On Day 2, the SS_{RMSE}, SS_{MAD}, and R^2 spatial patterns are consistent with patterns in firefighting resources and moisture, with higher skill being achieved for fires in drier fuels with more resources applied (Fig. B9-B10 panels a,b,e).

Across states, the RF maintains the ability to predict days with little change (Figure 3.3d, B6d) while also forecasting increases and decreases with skill (Figure 3.3f, B6f). In the 1- and 2-day forecasts, there is a slight decrease (never more than 0.2) in the $F1_{\text{slight change}}$ score between the persistence and RF forecasts (Figure 3.3d, B6d). In certain states (e.g. MT, ID) increases are forecasted more skillfully than decreases and in other states (e.g. NV) decreases are forecasted more skillfully than increases. Generally speaking, F1 scores for increases and decreases are higher on day 2 than day 1 (Figure 3.3f, B6f). On day 1, the spatial pattern in the F1 scores is driven by fuel moisture, terrain, and weather variables (Figure B7-B8 panels f-h) and on day 2

the spatial pattern in the F1 scores is driven by fuel moisture, firefighting, and weather variables (Figure B9-B10 panels f-h).

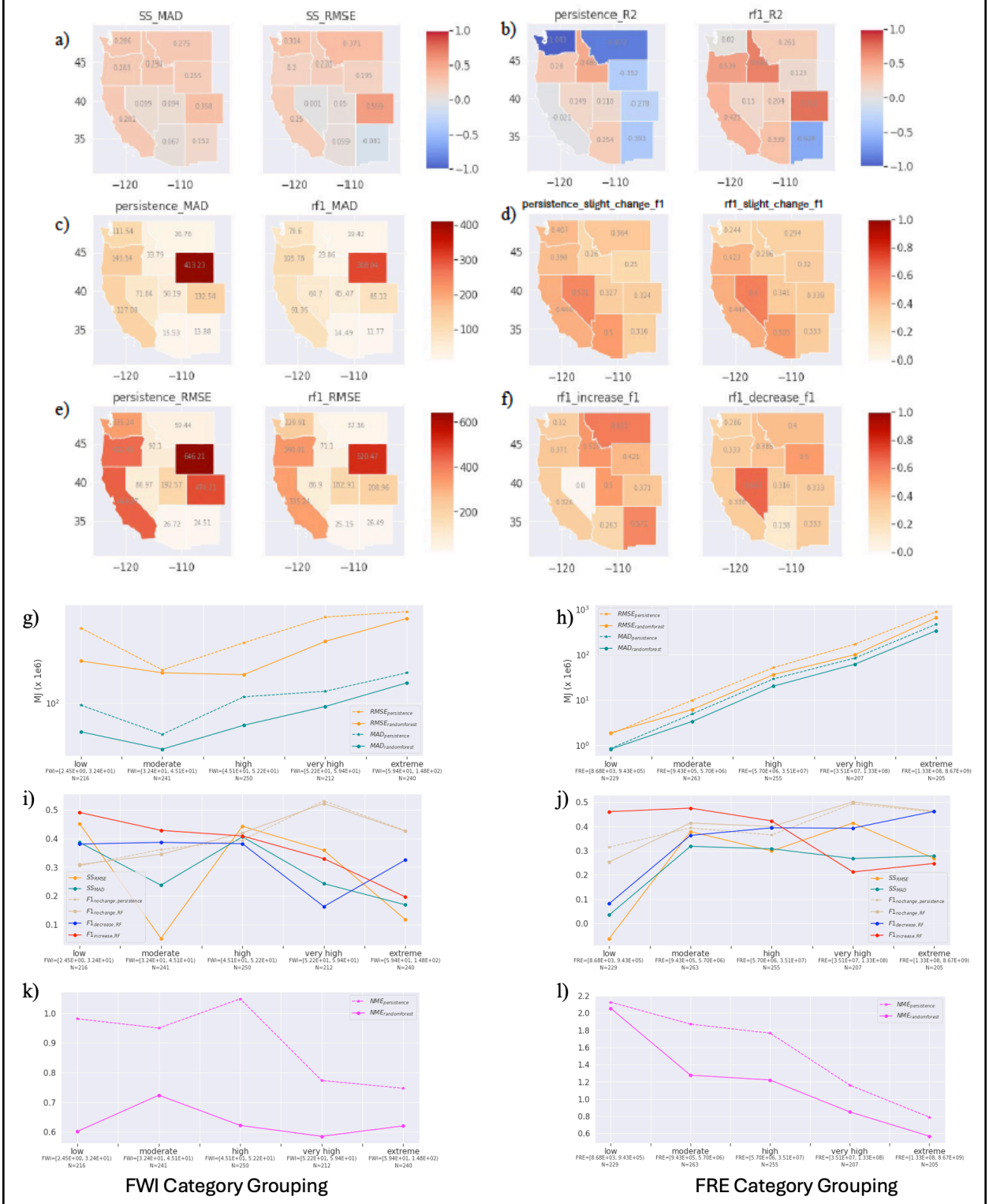

Figure 3.3. Random forest evaluation metrics by state (a-f) and by fire severity (g-j) for the 1-day RF and persistence predictions. a) SS_{MAD} and SS_{RMSE} for the RF model; b) R^2 for the

persistence and RF models; c) MAD for the persistence and RF models; d) $F1_{slight\_change}$ for the persistence and RF models; e) RMSE for the persistence and RF models; f) $F1_{increase}$ and $F1_{decrease}$ for the RF model; g) RMSE (yellow) and MAD (teal) for the persistence (dotted) and RF (solid) models split by 20th-percentiles of FWI; h) as in f but split by 20th-percentiles of FRE. i) SS_{MAD} (teal), SS_{RMSE} (yellow), $F1_{increase}$ (red), $F1_{decrease}$ (blue), and $F1_{slight\_change}$ (tan) for the RF models with $F1_{slight\_change}$ (tan dotted) for the persistence model split by 20th percentiles of FWI. j) as in i but split by 20th percentiles of FRE.

Figure 3.3g-l (B6g-l) shows error for the 1-day forecasts grouped into 20th-percentile bins of FWI (Figure 3.3g,i,k; B6g,i,k; a RF input) and FRE (Figure 3.3h,j,l; B6g,i,l; a RF output). In both the 1- and 2-day forecasts, the RF predictions tend towards persistence at extreme FWI, as shown by the decrease in SS_{MAD} and SS_{RMSE} (Figure 3.3g,i, B6g,i). Extreme FWI may correspond to a large range of fire behaviors (including cases of no fire, e.g. Field et al., 2020a). Also, it appears that persistence is a more capable forecast at higher FWI values than lower FWI values (Figure 3.3k; B6k). Therefore, it is possible that the RF model cannot make the distinctions between persistence and non-persistence fire days at higher values of FWI (see F1 scores in the Extreme class, Figure 3.3i, B6i). We also see RMSE and MAD increasing with FRE (Figure 3.3h, B6h) and normalized mean error (NME, Figure 3.3l; B6l) decreasing with FRE in both the 1- and 2-day forecasts. However, for the 1-day forecast (Figure 3.3j) the SS_{MAD} and SS_{RMSE} tend to increase then plateau after the moderate FRE category, and for the 2-day forecast (Fig. B6j), the SS_{MAD} and SS_{RMSE} tend to increase as the FRE categories intensify. This implies that we are helping more with the upper end of FRE in the 2-day forecast and making improvements across most categories in the 1-day forecast. This is consistent with the spatial analysis, which showed that the model tends to under-perform for more extreme fires and more severe fire weather (Figure B-B10).

Throughout the 2020 fire season, there are some days where the RF performs worse than persistence and some days where it performs better (Figure B11-B12). The days when the RF is

worse than persistence tend to be at the beginning and end of the fire season, but on these days the sample size is very low (<5 active fires per day). In spite of this, the RF beats persistence for a majority of the 1-day (2-day) forecasted days in the 2020 fire season, with SS_{RMSE} exceeding zero in 61% (63%) of days and SS_{MAD} exceeding zero in 67% (69%) of the 114 (113) days containing active fires. If the evaluation is restricted to days where there were at least 10 fires burning, SS_{RMSE} exceeds zero in 75% (73%) of days and SS_{MAD} exceeds zero in 80% (78%) of the 51 (48) days containing active fires.

3.4.3 Sensitivity Results

We performed sensitivity tests to evaluate each training variable's contribution to the overall forecast skill of the random forest (Figure 3.4). Many variables represent similar wildfire drivers (e.g. HWP, HD0W0, FWI, and BI all represent near surface meteorology) and so were grouped together (Figure 3.4, thick black lines and text). To test the random forest sensitivity to different feature groups, all variables in a certain feature group were dropped (Figure 3.4 black outlined points) as done in previous work (Huot et al., 2022). Additional sensitivities are performed where one variable in the group was used at a time, allowing for testing, re-training, and re-testing of the random forest model at each step (Figure 3.4, non-outlined points). This is slightly different from traditional random forest feature importance (Breiman, 2001), which is known to misrepresent the importance of correlated features (Nicodemus & Malley, 2009).

Figure 3.4 shows that groups of features generally fall into two tiers, one that includes feature groups that strongly influence the random forest, and one that includes the rest of the feature groups with less influence. The amount of influence a group of features exerts on the random forest is defined by the change in performance of the random forest when that group of features is excluded from the training dataset (Figure 3.4, black outlined points) with respect to

the performance of the random forest trained on all available features (Figure 3.4, vertical lines). For reference, the metrics for the full random forest are R^2=0.48, SS_{MAD}=0.28, SS_{RMSE}=0.27 $F1_{increase}$=0.38, and $F1_{decrease}$=0.48. The feature groups fall into the tiers as follows:

- Tier 1: This tier of features contributes a large amount of skill to the random forest across metrics. Two feature groups fall into this category, last day FRE and near-surface weather. Without these feature groups we see large drops in performance across all metrics (R^2 =0.30 without surface weather and R^2 =0.45without last day FRE).The model trained including a single one of the simpler fire indices (HDW, HWP) are the top performers, sometimes outperforming the full RF in terms of F1 score and recovering much lost FRE skill when removing surface weatherFinally, the model trained usingHWP as the only weather variable(Figure 3.4 features_only_hwp) indices has comparable R^2, SS_{RMSE}, and SS_{MAD} to the model trained on all variables (Figure 3.4 features_all row).
- Tier 2: This tier of features includes those feature groups that moderately influenced the random forest model compared to Tier 1. Tier 2's feature groups include stability, living and dead fuel moisture, PWS, ESI, terrain, fuel loading, and human variables. When feature groups in this tier are left out, performance metrics change less and may be better or worse with respect to the full random forest's performance metrics (e.g. R^2=0.47-0.48, $F1_{increase}$=0.37-0.40, $F1_{decrease}$=0.41-0.48).

Often, variables within the feature groups of Figure 3.4 are highly correlated. These correlations may be strongly positive, as in the case of different fire weather indices developed by different agencies capturing the same periods of favorable fire weather (Figure B13), fuel loading under different fire weather scenarios which is based on a fraction of the total biomass (Figure B14b) or terrain variables where steeper slopes tend to occur in more variable elevations

(Figure B14c). Sometimes strong negative correlations exist between variables in a feature group. One such case is fuel moisture, which is defined by some methodologies as codes which increase with increasing dryness (CFFDRS) and by others as fractional water content which decreases with increasing dryness (Figure B15). The high degree of correlations between variables within feature groups motivates training different versions of the random forest model subsets of the data to derive simpler models that could behave similarly to the model using all variables.

3.4.4 Selecting Optimal Subsets of Features

Random forest models were trained on subsets of the data which kept agency data (i.e. near-surface weather and dead fuel moisture from HRRR, GridMET, CFFDRS and NFDRS) separate where applicable, considered features which can be operationally forecasted, and chose at most one feature or weather index system from each group to maximize FRE performance and F1 scores (Figure 3.4, non-outlined points). Figure 3.5a summarizes the subset of features that were tested. The subsets are as follows: “features_nfdrs” is based on NFDRS near-surface weather and moisture indices; “features_cffdrs" is based on CFFDRS near-surface weather and moisture indices; “features_gridmet” is based on GridMET VPD and WS and NFDRS moisture indices; “features_hrrrmet” is based on HRRR VPD,WS, and soil moisture predictions; “features_hwp” is based on HRRR HWP and soil moisture predictions; and “features_hd0w0” is based on HRRR HD0W0 and soil moisture predictions. There are two additional subsets “features_hrrrmet_plus_hwp” and “features_hrrrmet_plus_hd0w0”, which are trained on HRRR VPD, WS, and HWP and HD0W0 respectively. These subsets were included to exploit potential nonlinear interactions between the raw (VPD, WS) and derived (HWP, HD0W0) weather variables that are available from HRRR. In this study we used the HRRR model as a reanalysis

product and CFFDRS values from the GFWED version based on MERRA-2 (Text B1), but the HRRR and GEOS-5 models also produce forecasts of these same variables operationally. On the other hand, NFDRS indices, VPD, and WS from GridMET (Text B1) are based on near real time observations only and do not have a forecasting component in GridMET. Therefore, any configuration based on GFWED or HRRR may be used on both forecast- and reanalysis-based meteorology to predict FRE.

Figure 3.5a also shows that the feature categories including population, containment percentage, day of year, NCAR live fuel moisture, and ESI were excluded from the subsets. These categories contained only one feature, and when this feature was dropped from the random forest, FRE scores increased slightly. Alternatively, persistence FRE and PWS are feature categories containing one variable which benefits model performance. When these variables were dropped, FRE and F1 skill decreased from R^2=0.48 in the full model to R^2=0.45-0.47. Therefore, persistence FRE and PWS are important contributors to good model performance, while population, containment percentage, day of year, NCAR live fuel moisture, and ESI have little impact or slightly hinder model skill.

The variables listed for the stability (Figure B14a), dead moisture (Figure B15), fuel loading (Figure B14b), terrain (Figure B14c), and human (Figure B16) categories in Figure 3.5a may co-vary within these categoriesand were all chosen to maximize FRE performance and minimize the impact of correlated features. When certain feature groups (i.e. dead fuel moisture, Figure 3.4, “features_no_dead_moisture”) are left out, FRE performance metrics donot change., We attribute this behavior to the random forest learning compensating errors between less correlated features (e.g. Figure B15, NCAR dead moisture NFDRS 1000hr moisture) within the feature groups or by interaction terms between feature groups.

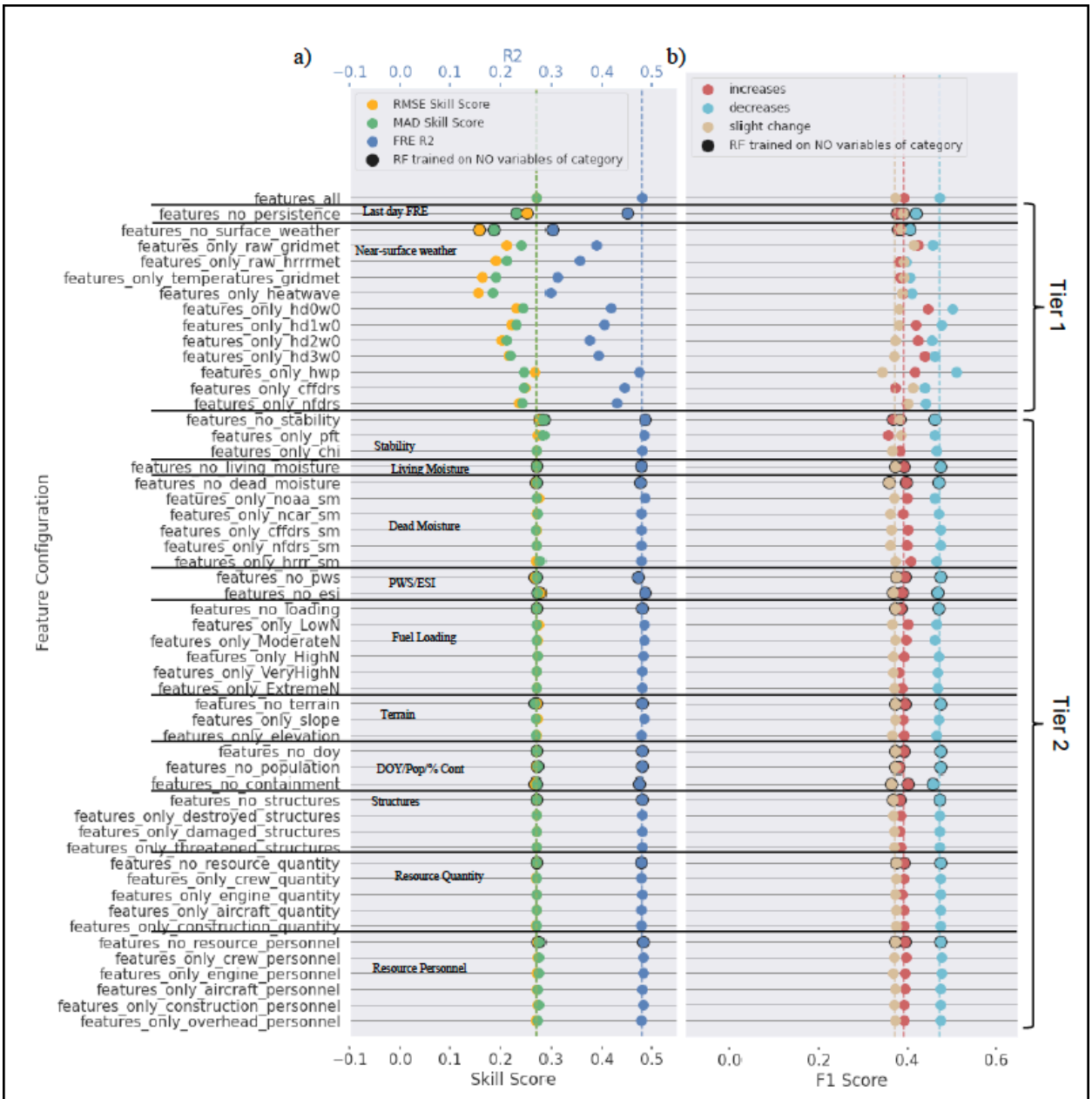

Figure 3.4. Performance of the random forest model for different sensitivity tests (row names). a) shows SS_{RMSE} in yellow SS_{MAD} in green, and R^2 in blue, and b) shows $F1_{decrease}$ in light blue, $F1_{increase}$ in red, and $F1_{slight\_change}$ in tan. Black horizontal lines delineate different groups of features for which sensitivity is tested, and these categories are also labeled in text to the left. Feature subsets for which no variables of the given category are used have their points outlined in black. The variable following "only" in the variable name is the sole variable of the given category used in training.

3.4.5 Performance of Optimal Subsets of Features

Figures 3.5a-d show the performance of random forest models which were trained using the subsets of features in Figure 3.5a. This analysis shows that subsets of the full feature set can be used to train random forest models and obtain performance comparable to or better than the random forest model trained on the full feature set (e.g. for the 1-day forecast R^2=0.39-0.48, SS_{RMSE}=0.21-0.28, SS_{MAD}=0.22-0.28, $F1_{increase}$ =0.39-0.49 and $F1_{decrease}$=0.39-0.49). Both the 1- and 2-day forecasts are similar to the SFI methods, in that RFs trained on more complex fire weather indices (Figure 3.5b, d "features_cffdrs") have better skill at forecasting FRE than RFs trained on raw meteorological inputs (Figure 3.5b,d "features_hrrrmet"). Also like the SFI methods, the RFs trained on the simpler fire indices (Figure 3.5b, d "features_hwp") show better performance than the RFs trained on more complex fire indices (Figure 3.5b,d "features_cffdrs"). Further, the models trained on derived indices (HWP, HD0W0) and raw weather variables (VPD, WS) perform better across most metrics than the models trained on either raw weather variables or derived indices (Figures 3.5b, d "features_hrrrmet", "features_hwp", "features_hrrrmet_plus_hwp"). An ensemble forecasting method, defined as the average of the predictions from the 8 RF models in Figure 3.5a also shows comparable skill. All random forest-based models shown in Figure 3.5 have greater forecast skill than the scaled and persistence methods shown in Figure 3.3.

Overall, these results illustrate that the fire weather and fuel moisture indices developed by each agency independently represent important predictors of wildfire evolution and can be combined using machine learning methods to create forecasts that outperform persistence. Random forests trained on subsets of the predictors show similar behavior to both the ensemble method and the random forest trained on the full set of predictors, further highlighting the

redundancy of including all available fire weather and fuel moisture indices in a single random forest model.

a)

Subset Name (based on source of meteorology)	US Fire Danger Rating System (features_nfdrs)	Canadian Forest Fire Danger Rating System (features_cffdrs)	HRRR-Raw Meteorology + HWP (features_hrrrmet_plus_hwp)	HRRR-Raw Meteorology+ HD0W0 (features_hrrrmet_plus_hd0w0)	HRRR-HDW (features_hd0w0)	HRRR-HWP (features_hwp)	HRRR-Raw Meteorology (features_hrrrmet)	GridMET-Raw Meteorology (features_gridmet)
Persistence	FRE, one day old	FRE, one day old	FRE, one day old	FRE, one day old	FRE, one day old	FRE, one day old	FRE, one day old	FRE, one day old
Near-Surface Weather	NFDRS: BI, ERC	CFFDRS: FWI, BUI	HRRR: VPD, WS, HWP	HRRR: VPD, WS, HD0W0	HRRR: HWP	HRRR: HD0W0	HRRR: VPS, WS	
Stability	PFT	PFT	PFT	PFT	PFT	PFT	PFT	PFT
Dead Moisture	NFDRS: 100-hr, 1000-hr	CFFDRS: DC, DMC, FFMC	HRRR: 30cm moisture	HRRR: 30cm moisture	HRRR: 30cm moisture	HRRR: 30cm moisture	HRRR: 30cm moisture	NFDRS: 100-hr, 1000-hr
PWS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fuel Loading	Extreme	Extreme	Extreme	Extreme	Extreme	Extreme	Extreme	Extreme
Terrain	Elevation	Elevation	Elevation	Elevation	Elevation	Elevation	Elevation	Elevation
Human	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity	Threatened structures, aircraft quantity
Excluded Features	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population	Living fuel moisture (NCAR), ESI, DOY, Containment, population

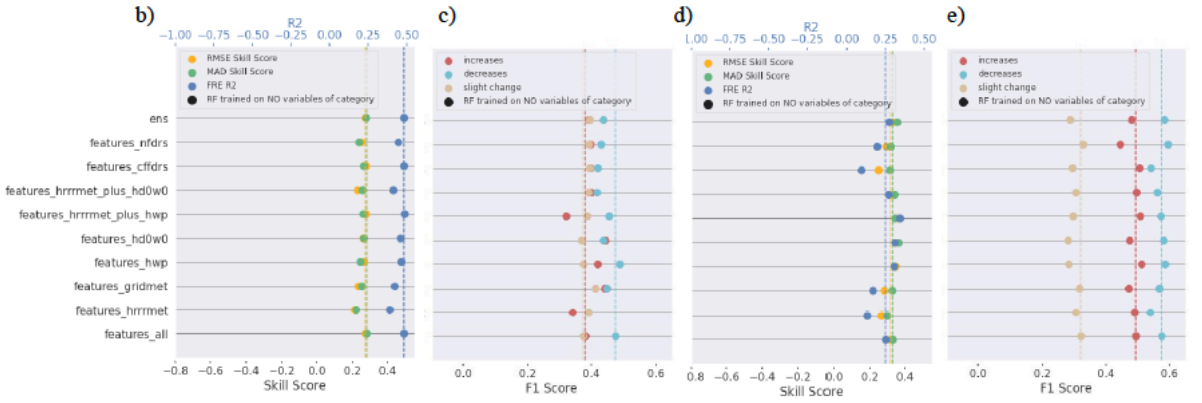

Figure 3.5. Performance of the random forest trained on subsets of the data. a) shows subsets of features selected to optimize performance. R^2 and SS_{RMSE} and SS_{MAD} are shown for FRE for all 1-day b) and 2-day d) predictions in blue, yellow, and green dots respectively. F1 scores for all for all 1-day c) and 2-day e) predictions with increases in red, decreases in blue and slight changes in tan.

3.5. Discussion

3.5.1 Comparison with previous studies

The skill gained with respect to persistence is consistent with previous work which has beaten persistence. Graff et al., (2020) and Peterson et al., (2013) (G20 and P13 respectively) paired weather data with maximum likelihood classification and Poisson regression, respectively, to predict fire counts in Alaska. SS_{RMSE} ranging from 5-25% are reported in both papers, with the best performance occurring for decreases in fire count (P13) and 1- and 2-day forecasts with Poisson regression (G20). It is notable that, like this work, G20 also scales weather indices to forecast fire counts. In the 1-day forecast, the G20 scaled method is a less skillful forecast than persistence and the Poisson hurdle model, and in the 2-day forecast the G20 scaled method is more skillful than persistence but less skillful than the poison hurdle model. The results of this study are also consistent with work by Garcia-Llamas (2019), which explains up to 42% of variance in fire severity using random forest models trained on vegetation indices, rainfall, and time since last burn for Mediterranean pine forests in Spain. The most important variables reported in this study (i.e. near-surface fire weather) are also consistent with those reported in Gray et al., (2018).

This study can also be compared to Huot et al., (2022, H22) and Wang et al., (2022, W22), two studies which used machine learning to predict a next-day fire mask and monthly PM2.5, respectively. H22 uses gridded fuel, weather, topography, and human data to produce gridded next-day predictions of fire location. Similar to our work, H22 finds that the previous day's fire location (persistence information) is key to giving a reliable forecast of next-day fire spread. Unlike our work, they have topography-related variables as another key feature, which

lends credence to the idea that spatially resolved land surface features contain more useful information than area-averaged land surface features. In W22, the authors predict monthly PM2.5 from GFED (Kaiser et al., 2012) using fuel, weather, topography, and human variables. They achieve higher correlation with observations spatially than our work, and we attribute that to our daily data being noisier than their monthly data. In their feature importance ranking, weather is the most important predictor. We may conclude that our work is comparable to the machine learning models present in H22 and W22, and that most differences are attributable to methodological differences and different output variables. Between our work and the work of H22 and W22, a consensus seems to be emerging about the high importance of weather and persistence information for forecasting fire and emissions progression using machine learning.

3.5.2 Uncertainties and Limitations

Since the random forest method is applied as a post-processing step, uncertainties in the input data propagate through to the final model output. These uncertainties can be associated with measurements, such as gap-filling (Li et al., 2022) or missing (Schroeder et al., 2014) smoke- or cloud-obscured fire detections or the best estimates of resources applied to a fire event (Text B1). Additionally, other models were used to generate the training data. For instance, the operational HRRR model was used here, which has a longer record but did not include smoke radiative feedbacks until the end of 2020 (Ahmadov et al., 2017; Dowell et al., 2022). Additionally, model biases such as those for relative humidity in MERRA-2 which are used to calculate FWI and components (Field, 2020), and the outsize impact of temperature on some of the fuel moisture estimates (McCandless et al., 2020) could contribute to uncertainty in these variables and how they behave in the random forest. Finally, population density, PWS, fuel

loading, and terrain slope and elevation were assumed to be static over the 2019-2021 period, which does not consider how either past burns or fuel treatments have altered the landscape. Future work to improve this modeling technique could consider adding georeferenced firefighting information, such as the location of firelines, as such information has proven useful in fire spread modeling (Turney et al., 2023).

Many variables used in this study are not available outside of the 2019-2021 period (Table 1), so the specific trained models evaluated may not be directly applicable to beating persistence for other fire seasons. However, our sensitivity analysis showed that the random forest can utilize wildfire drivers that were identified as far back as the 1970s (Countryman, 1972). We also show that subsets of the predictors can be used to beat persistence. This means that the model is likely learning real relationships between fire weather, fuels, topography, and firefighting, and that this technique is robust to the datasets chosen to represent different sides of the fire behavior triangle. Training the model on subsets of the data could allow future studies to investigate years beyond 2019-2021, because such models would not be limited by datasets which only exist for this limited period. Likely this would lead to improved model performance due to a training dataset which samples the input space more fully. Finally, the configuration here requires a priori knowledge of fire location to extract predictors in the relevant location and generate future forecasts. This work was also limited by the fact that we had no multi-day predictions of fire location over which to extract predictors. This may explain why some features, such as fuel loading, moisture, population, and terrain did not show predictive power, since they may be spatially heterogeneous, and their variability is not captured through persisting polygon-averaged values throughout the forecast window. Therefore, we recommend for future

study coupling the random forest technique with fire spread modeling (Bakhshaii & Johnson, 2019) and extending this methodology to predict hourly evolution of fires.

3.6. Conclusions

We described the development of data-driven fire radiative energy (FRE) forecasting products that can be used as inputs for predicting wildfire emissions. We test the skill of several models at making forecasts of daily wildfire FRE using data extracted over the area affected by the wildfire. Scaled fire indices (SFI), and random forest (RF)-based model predictions were evaluated with respect to the persistence forecast. Overall, the RF showed the largest increase in skill in the 1-day (2-day) forecasts, with an R^2=0.48 (R2=0.25) compared to 0.02 (-0.62) for persistence, SS_{RMSE} =0.27 (SS_{RMSE} =0.32), and SS_{MAD} =0.28 (SS_{MAD} =0.33) when all variables were used to train the model. The random forest was also the most skilled in forecasting whether day-to-day changes in fire behavior would occur, with $F1_{decrease}$=0.26 ($F1_{decrease}$=0.44) and $F1_{increase}$ = 0.48 ($F1_{increase}$ = 0.57). We find that the random forest reduces error compared to persistence across all states, for a majority of the days in Summer 2020, and for all severity categories. Error in both the random forest model and in the persistence model tends to increase as FRE increases. Additionally, the random forest model tends towards a persistence forecast at extreme FWI and for low-FRE fires.

The sensitivity analysis confirmed that most forecast skill came from including reanalysis-derived surface weather variables and persistence FRE as features. Other variables such as vertical atmospheric stability, fuel moisture, human influence, fuel loading, and plant water sensitivity, seem to be most useful for fine tuning the predictions. Comparable skill was achieved by training several random forests on subsets of these variables based on data available to certain agencies. Therefore, there is significant value in using the random forest over a

persistence or scaled fire index approach, and we recommend random forests be applied to scale emissions in future work.

We showed that machine learning plus knowledge of the human and natural drivers of wildfire can dramatically improve upon the current persistence assumption, with the biggest gains made via predicting increases or decreases in FRE. Our work also shows that the daily change in weather indices (HWP, HDW) can be very useful in predicting changes in FRE. FRE is proportional to emissions, so this method may be used to predict daily changes in FRP-based emissions inventories (Darmenov and da Silva, 2015; Kaiser et al., 2012; Li et al., 2022). Many regional models (HRRR-Smoke, Ahmadov et al., 2017) use these FRP-based inventories, and so are poised to take advantage of the methodology presented here. While the methodology outlined here is most suitable for forecasting FRE and computing emissions, coupled chemical transport model-machine learning frameworks have been shown to work (e.g. Kelp et al., 2022) and are a promising future application of this work. Thus, data-driven derived approaches for scaling wildfire emissions should be considered in the next generation of air quality and atmospheric composition forecasting.

Chapter 4

Air quality model sensitivity to machine learning-derived fire emissions

Abstract

Forecasts using chemical transport models are vital for warning the public about poor air quality due to wildfire smoke, but uncertainties related to persistence of smoke emissions can propagate through these models and cause errors in the resulting air quality forecasts. In this work, we present a pipeline for implementing machine learning-derived wildfire emissions in the Weather Research and Forecasting with Coupled Chemistry (WRF-Chem) and test this pipeline for September 2020. We examine the sensitivity of the models to different emissions schemes including analysis, persistence, and machine learning emissions, and we propose an automated way to evaluate forecasts of AOD and PM2.5 in smoke-impacted regions downstream of fires. WRF-Chem is driven by the Regional Advanced Baseline Imager (ABI) and Visible Infrared Imaging Radiometer Suite (VIIRS) Emissions (RAVE).

Preliminary results show that, although the machine learning-derived emissions improve on the persistence emissions, WRF-Chem with analysis RAVE emissions tends to underestimate AOD and both under- and overestimate PM2.5. On average, WRF-Chem is not very sensitive to the chosen emissions. On a per-fire basis, WRF-Chem is more sensitive to the emissions scheme, with random forest emissions measurably impacting some AOD (PM2.5) skill metrics in up to 50% (60%) of fire days. However, persistence and machine learning predictions lead to almost no difference a majority of the fire days in many metrics. In the best case scenario, using random forest emissions improves the median AOD error for some fires and degrades the median AOD

error for other fires, and the improvement is a factor of four higher than the degradation. Therefore, it is likely that the current version of WRF-Chem with RAVE emissions contains too many internal uncertainties to fully benefit from machine learning derived emissions.

4.1. Introduction

When it comes to forecasting air quality in real time, chemical transport models (CTMs) rely on persistence to predict smoke emissions (Jaffe et al., 2020; Ye et al., 2021). Persisting emissions can cause delayed onset or extended burning of fires (Di Giuseppe et al., 2018; Thapa et al., 2022), leading to errors in downstream estimations of aerosol optical depth (AOD) or fine particulate matter (PM2.5) (Ye et al., 2021). Real time air quality forecasts are capable of identifying days of degraded air quality due to smoke (Hung et al., 2020), but prediction skill of AOD and PM2.5 could be improved. For instance, Yao et al., (2013) compared BlueSky PM2.5 forecasts with monitors in British Columbia in the summer of 2010 and found moderate skill in replicating the temporal distribution of surface PM2.5 but underestimation overall across the domain (pearson r=0.4, NMB=-0.45). Similarly, Pavlovic et al., (2016) forecasted air quality over the US and Canada using the smoke-allowing FireWork model for the 2015 fire season and compared it to the non smoke-allowing Regional Air Quality Deterministic Prediction System (RAQDPS). While FireWork improved on RAQDPS, skill for predicting surface PM2.5 was lowest (pearson r=0.42-0.49 and MB ~ -2ug/m3) for the western part of the domain which contained the most fire activity. Finally, Ye et al., (2021) compared PM2.5 and AOD outputs from twelve air quality predictions (all using persistence emissions) for the Williams Flats Fire, an August 2019 Washington state wildfire that was heavily observed during the NASA-NOAA Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ; Warneke et al., 2023) campaign. AOD was underpredicted across all models, and AOD pearson correlation

never exceeded 0.5 across the forecast period. PM2.5 was under- and over-predicted (depending on the model), and PM2.5 pearson correlation never exceeded 0.35.

Persistence of smoke emissions is a dominant source of error in operational air quality forecasting, and several methods including ensemble forecasting, coupled large eddy simulation (LES)-fire spread modeling, and machine learning have been proposed to fill this gap. Recently, ensemble air quality forecasting using models which employ persistence for fire emissions has been used to address forecast error, and across multiple error metrics model ensembles outperformed single models for predicting AOD and PM2.5 in 2020 and PM2.5 in 2022 (Y. Li et al., 2024; Makkaroon et al., 2023). LES-fire spread models fully resolve complex flows and represent the coupling between weather, fuel, topography, and fire spread (Coen et al., 2013; Kochanski et al., 2013; Turney et al., 2023). These LES-fire spread models were coupled to air quality forecasts (Kochanski et al., 2021; Mandel et al., 2019), and a recent study showed that this technique improved forecasts of exceedance for the daily PM2.5 NAAQS against a traditional emissions inventory (Lassman et al., 2023; persistence emissions were not tested). There have been recent papers relating to using statistical and machine learning techniques to predict the evolution of wildfire activity on a daily basis, either in terms of burned area (Huot et al., 2022), FRP (Thapa et al., accepted; Pinto et al., 2020), fire counts (Graff et al., 2020; Peterson et al., 2013), or smoke plume heights (Yao et al., 2018). These models have been shown to be more accurate than the persisting burned area, FRE, or fire counts, but to our knowledge these predictions have not been coupled to an air quality forecast like the LES-fire spread models or evaluated like the LES-fire spread models and ensemble forecasts. Machine learning has also been used to directly predict air quality or interpolate between station measurements, but these studies tend to be retrospective and related to health impact assessment rather than addressing

forecasting problems (e.g., Reid et al., 2015; Watson et al., 2019; Zou et al., 2019; Mirzaei et al., 2018).

In the literature, standard techniques are used to evaluate model performance. One common technique is calculating bulk statistics (correlation, bias, error, spatial alignment of plumes) for quantities like AOD and PM2.5 over a region or for a monitoring station during a given time period (e.g., Makkaroon et al., 2023; Pavlovic et al., 2016; Rooney et al., 2020; Shi et al., 2019; Yao et al., 2013; Ye et al., 2021). Since PM2.5 is a regulated air pollutant, it is also common to classify modeled and observed days as attaining or exceeding the PM2.5 NAAQS. Binary classification metrics can then be applied to determine model performance (e.g., Lassman et al., 2023; Y. Li et al., 2024; Makkaroon et al., 2023; Pavlovic et al., 2016). Spatial variability is evaluated primarily through the use of maps and comparisons across multiple stations, with stations split into near source and downwind groups (Rooney et al., 2020; Shi et al., 2019). Many studies also evaluate specific test case fires using hand selected boxes or stations (Kochanski et al., 2015, 2021; Mallia et al., 2020; Rooney et al., 2020; Shi et al., 2019; Ye et al., 2021), indicating that there is not yet a good automated way to evaluate fires.

In this study, we present a modeling pipeline (recipe) to generate machine learning-derived fire emissions on a per-fire basis based on scaling the emissions input files for the WRF-Chem and GEOS models. The scaling method is based on previous work by Thapa et al., (accepted). We also present an automated method to evaluate air quality (AOD and surface PM2.5) in regions around actively burning fires. This method is flexible; it can be applied in multiple models and for multiple inventories. This method primarily relies on daily wildfire polygons, and so it can be done using the many fire tracking algorithms developed in recent years (Y. Chen et al., 2022; Liu et al., 2023; McClure et al., 2023; Thapa et al., accepted).

4.2. Data

4.2.1 Input emissions

Li et al., (2022) blends FRP measurements from the Geostationary Operational Environmental Satellite-Advanced Baseline Imager (GOES-ABI) FRP measurements from the Joint Polar Satellite System (JPSS) Visible Infrared Imaging Radiometer Suite (VIIRS) to create the Regional ABI and VIIRS Emissions (RAVE) dataset. Within RAVE, FRP measurements are binned into 5 minute intervals and aggregated to 3 km horizontal resolution, with VIIRS measurements used where available and GOES measurements used elsewhere. Gaps in the FRP time series are filled using an interpolation procedure which considers the length of the gap, climatological burning conditions, and fire residence times in different fuel types (F. Li et al., 2022). RAVE uses emission factors from Akagi et al. (2011) and Anderae (Andreae, 2019) and the dry matter conversion factor from Wooster et al. (Wooster et al., 2005). Hourly RAVE emissions are based on hourly fire radiative energy (FRE), the hourly sum of FRP. In order to match observed AOD, RAVE emissions were scaled by fuel-dependent values following the work of (Darmenov and da Silva, 2015).

4.2.2 Evaluation Data: AOD, PM2.5, and polygons

Aerosol optical depth (AOD), computed as the vertical integral of aerosol extinction, is an extensive aerosol property related to total column aerosol loading. Model AOD will be validated using AOD from the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm applied to the Moderate-resolution Imaging Spectroradiometers (MODIS) on board the Terra and Aqua satellites (Lyapustin et al., 2018). We will use version 6.1 of MAIAC for MODIS Aqua, which has specifically adjusted parameters of the retrieval to capture AOD enhancements from smoke, resulting in improved performance compared with AERONET

AOD (Ye et al., 2022a) particularly for thick smoke that other AOD products may exclude. MAIAC AOD values based on Aqua are provided roughly twice daily at 1 km resolution.

Surface PM2.5 data come from the near-real time AirNow reference monitors, the data for which is available from OpenAQ (https://explore.openaq.org/). The PM2.5 monitor data are available at roughly hourly resolution. We assembled data from stations which measured PM2.5 at any point during the simulation period.

Wildfire polygons are necessary to derive the persistence and machine learning emissions estimates and to evaluate the forecasts daily on a fire-by-fire basis. For this study, we use the wildfire polygons developed by Thapa et al., (accepted). These polygons are based on Visible Infrared Imaging Radiometer Suite (VIIRS) 375m fire detections which were aggregated to daily resolution based on S-NPP and NOAA-20 overpasses occurring in 24 hour periods beginning at 12Z. Each polygon is assigned to the day where the 24 hour period begins, e.g. the September 2nd, 2020 polygons cover all detections made between September 2nd, 12Z and September 3rd, 12Z.

4.3. Methods

In this work, we simulate wildfire smoke over the Western US (WUS; Lat: 30N – 51N, Lon: -139W – -97W) during the September 1st-24th 2020 period. We use three fire emissions scenarios, persistence, machine learning, and reference, to drive the regional Weather Research and Forecasting coupled with Chemistry (WRF-Chem) model. In this section, we present details on WRF-Chem, describe the three fire emissions scenarios, and discuss methods with which to evaluate predictions of AOD and PM2.5.

4.3.1 Chemical transport models

The WRF-Chem configuration is based on WRF-Chem v4.4.2 and will produce output at 4km, hourly resolution. The model is run in a "full chemistry" configuration, with the Regional Atmospheric Chemistry Mechanism (RACM) for gas phase chemistry (Stockwell et al., 1997), and the Modal Aerosol Dynamics model for Europe-Volatility Basis Set (MADE-VBS, Ackermann et al., 1998; Ahmadov et al., 2012) for aerosols and heterogeneous reactions. The Kinetic PreProcessor (KPP) tool is used to set up the numerical integration of chemical species to be as fast and efficient as possible. Chemical initial and boundary conditions come from the Copernicus Atmospheric Monitoring Service (CAMS) global chemical reanalysis product (EAC4, Inness et al., 2019), and meteorological initial and boundary conditions come from the National Center for Environmental Prediction's North American Regional Reanalysis (NCEP NARR, Mesinger et al., 2006). WRF-Chem requires biogenic and anthropogenic emissions; biogenic emissions will come from the Model of Emissions of Gases and Aerosols from Nature (MEGAN, Guenther et al., 2006), and anthropogenic emissions will come from the Environmental Protection Agency National Emissions Inventory (EPA NEI, Mason, 2017) estimates. The WRF-Chem configuration will include online calculations of plume rise from the Freitas model with updated heat flux constants (Freitas et al., 2007; Thapa et al., 2022). Table 4.1 summarizes the relevant meteorological parameterizations the WRF-Chem model will use. One thing to note is that this configuration of WRF-Chem updates the PBL scheme based on the results of Thapa et al. (2022) showing that WRF-Chem required a more accurate PBL to show proper plume injection behavior.

Parametrization	WRF-Chem Reference
Planetary Boundary Layer (PBL)	5=Mellor-Yamada-Nakanishi-Niino scheme (Nakanishi & Niino, 2004)
Radiation (longwave and shortwave)	4= RRTMG scheme (Iacono et al., 2008) 2= Goddard Shortwave (Suarez, 1999)
Land Surface Model	2= Noah Land Surface Model (F. Chen & Dudhia, 2001)
Cloud Microphysics	10= Morrison double-moment scheme (Morrison et al., 2009)
Cumulus Parameterization	3= Grell-Freitas Cumulus Parameterization (Grell & Freitas, 2014)
Radiative Feedbacks	Subgrid cloud-radiation interactions (Alapaty et al., 2012) and direct aerosol-radiation interactions

Table 4.1. Summary of important WRF-Chem model meteorological parameterizations. In the WRF-Chem references column, numbers refer to the WRF-Chem namelist option.

4.3.2 Emissions scenarios

We run a reference simulation of WRF-Chem driven by RAVE emissions. Emissions from RAVE are based on FRE observations and so are the closest we can get to a perfect forecast, although not one free from uncertainty (F. Li et al., 2022). Figure 4.1 and C1 (blue bars) show RAVE daily total black carbon (BC) emissions summed over the August Complex fire area. Note that in order to produce realistic AOD with the RAVE emissions inventory, reference RAVE emissions (and thus persistence and machine learning emissions) were scaled by the fuel-dependent scaling factors used in (Darmenov and da Silva, 2015).

In addition to the reference emissions, our study tests model sensitivity to two fire emissions prediction techniques, persistence and machine learning, each of which is applied for each day and for each active fire (on a per fire-day basis) based on fire location polygons (Section 4.2.2). The persistence and machine learning-based methods of emissions calculation

are based on scaling initial emissions values. Initial emissions values are based on gridded RAVE emissions at the most recent locations of daily fire polygons at the time of forecast initializations, based on the intersection method presented in Thapa et al., (accepted). Persistence and machine learning preserve the diurnal cycles from RAVE. Persistence- and machine learning-derived scaling factors are calculated out to two days, as follows:

- *Persistence*. Persistence is a common forecasting assumption wherein emissions values from the initialization day are replicated throughout the forecast period. Therefore, initial emissions values per fire are multiplied by 1 to get the values of emissions on day 1 (Figure 4.1, green bars) and day 2 of the forecast (Figure C1, green bars).
- *Machine learning.* The machine learning method for predicting emissions is based on multiplying persistence emissions for a given fire based on machine learning-derived scaling factors (Thapa et al., accepted). These scaling factors are a nonlinear function of FRE, weather, soil moisture, fuel water stress and loading, structures threatened in the vicinity of the fire, and number of aircraft used to suppress the fire. Previous research (Thapa et al., accepted) determined that this was the set of features with the highest predictive capability for estimating emissions scaling factors. Initial emissions values will be multiplied by the day 1 scaling factor to produce day 1 emissions values (Figure 4.1, purple bars), and day 1 emissions values will be multiplied by the day 2 scaling factors to produce day 2 emissions values (Figure C1, purple values). This is done for every RAVE grid cell which intersects a fire polygon on a given day. The presence of 2-day forecasts leads to the model evaluation time period to be restricted to the September 2nd-22nd time period.

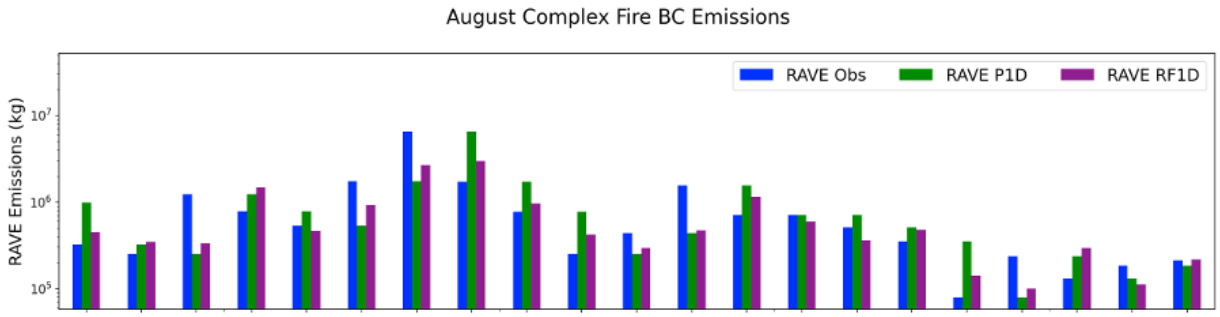

Figure 4.1. Black carbon (BC) emissions from August Complex Fire from the a) RAVE emissions inventory for September 2nd-September 22nd, 2020. Reference emissions are plotted in blue, and 1-day persistence and random forest emissions are plotted in green and purple respectively. Note that the persistence assumption dictates that the amount of emissions one day (green bar) is the same as the emissions amount the day before (previous day blue bar).

4.3.3 Forecast Restarts

The WRF-Chem persistence and machine learning simulations will be initialized daily based on the restart files and will run for two days at a time. Restart files contain the model state at a given timestep. Our reference configuration of WRF-Chem saved restart files daily at the 12Z timestep. In WRF-Chem, the persistence and updated emissions simulations are initialized with the restart files, so we expect in this case that the meteorology will be roughly consistent between the different fire emissions scenarios. Meteorology between the different emissions scenarios may vary slightly due to the presence of cloud and aerosol direct radiative feedbacks (Table 4.1). Some of this variability will be accounted for by the WRF-Chem simulations performing nudging towards the NCEP NARR reanalysis every 6 hours.

4.3.4 Matching models and observations

The MAIAC 1 km granules are regridded to the WRF-Chem WUS domain. The regridding results in MAIAC swaths split by overpass time (1:30 PM local time for Aqua) which must be aggregated to a single map representing AOD at 1:30 PM local time. To do this, the day

is set to cover the period from 12Z on day d to 12Z on day d+1 (12Z-12Z day). All MAIAC and model time slices from each 12Z-12Z day are collected, and the final output contains the map over all points where the model and MAIAC have AOD values that are less than 5. Since model output files are continuous, this step primarily filters the model to the grid cells where MAIAC data is present. MAIAC and model values from the MAIAC overpass time are stored in netCDF files for use in later analysis.

OpenAQ station PM2.5 observations are matched to model values by nearest neighbor interpolation in time and space. In cases where more than one station falls within a model grid cell, the same model value is used for comparison with each station. This is done for each model configuration: reference, persistence, and ML, with 1 and 2 day forecasts selected for the persistence and ML. AOD and PM2.5 performance are evaluated on the station basis, and averaged to the 12Z-12Z day in order to make comparisons (section 4.4).

In this paper we present bulk model evaluations. For the bulk model evaluation, performance statistics including mean bias (MB), root mean squared error (RMSE), normalized mean bias (NMB), normalized mean error (NME), pearson correlation (PC), and ratio (RAT) are computed following Ye et al., (2021) on a daily and per grid cell or station basis. The bulk model evaluations allow us to diagnose uncertainties in the RAVE emissions and in the WRF-Chem model.

4.3.5 Per-polygon model evaluation

To zoom in on the effects of the different fire emissions configurations, we evaluate the models on a fire-by-fire basis. The models are evaluated in the vicinity of the daily burned area (Thapa et al., accepted, Figure 4.2 a-f red polygons), with the vicinity defined as a buffered radius around the burned area polygon (Figure 4.2 a-f gray and black dashed lines). One day

forecasts are evaluated on the near buffer, defined as 50 km for AOD (Figure 4.2a-f black dashed line) and 150 km for PM2.5 (Figure 4.2a-f, inner concentric gray dotted line). Two day forecasts are evaluated on the far buffer, defined as 150 km for AOD (Figure 4.21a-f, inner concentric gray dotted line) and 300 km for PM2.5 (Figure 4.2a-f, outer concentric gray dotted line). Depending on the chosen buffer and forecast timestep, AOD and PM2.5 grid cells or station values falling within that buffer are extracted. Values within the buffer are then considered as a daily distribution (Figure 4.2g), and performance statistics including mean bias (MB), root mean squared error (RMSE), normalized mean bias (NMB), normalized mean error (NME), pearson correlation (PC), and ratio (RAT) are extracted following Ye et al., (2021) (Figure 4.2h-j).

Due to the severe nature of the 2020 wildfire season, multiple fires may occur in the same buffered region. For instance, the 150 km buffer for the August Complex Fire (e.g. Figure 4.2a, inner concentric gray dotted line) contains the polygons for the Red Salmon Complex and North Complex Fires (Figure C2). Therefore, certain points near these fires are likely to be double counted, with the same point(s) included in the statistics for multiple fires. Indeed, for AOD (PM2.5) on the large buffer there are grid cells (stations) which get counted up to 9 (20) times (Figure C3a,c; Ca,c) .

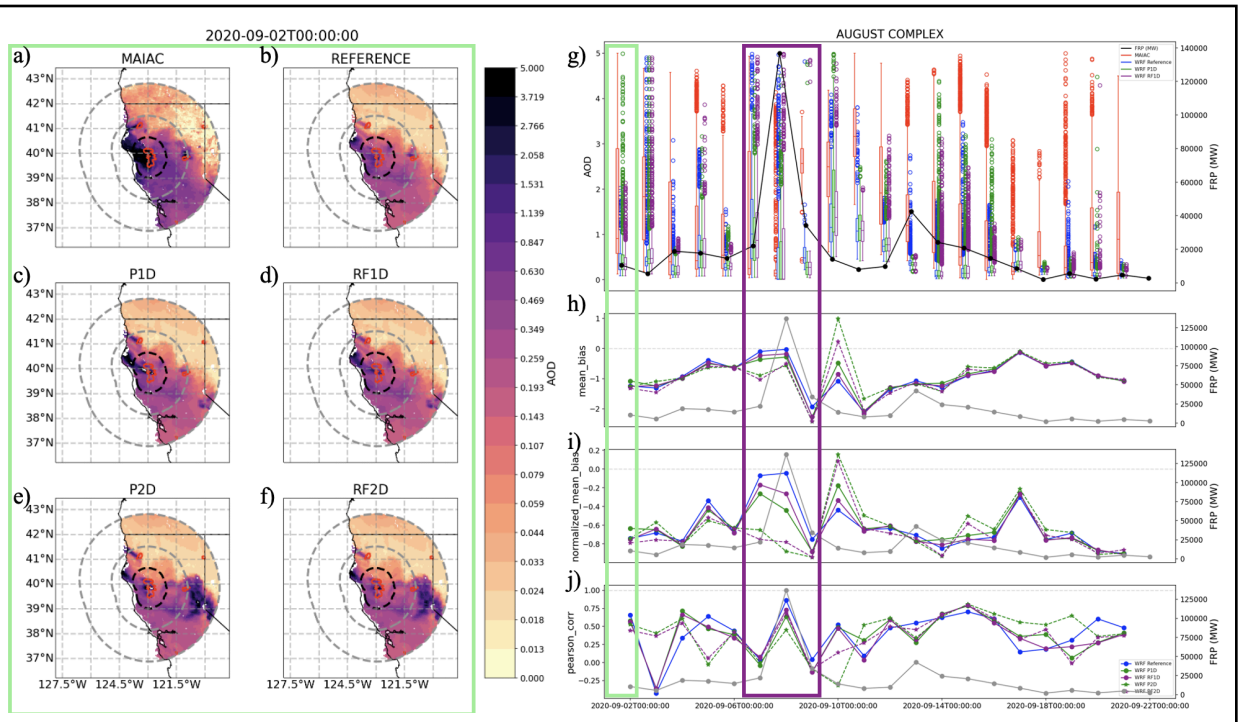

Figure 4.2. An example of per-polygon model evaluation. a)-f) show AOD from a) MAIAC, and WRF-Chem b) reference, c) 1-day persistence, d) 1-day random forest, e) 2-day persistence, and f) 2-day random forest emissions for September 2nd, 2020 in the vicinity of the August Complex fire (red polygon encircled by black dashed line). Also shown are the Red Salmon Complex (N of August Complex), North Complex (E of August Complex), and the W-5 Cold Springs fires (NE of August Complex on CA/NV border) (red polygons). 50 km (black dotted), 150km, and 300km (gray dotted) buffers around the August Complex fire are shown. g) shows a time series of AOD selected over the 50km buffer for the MAIAC observations (red), and the reference (blue), 1-day persistence (green) and the 1-day random forest (purple). Figures h-j) show model mean bias, normalized mean bias, and pearson correlation over the August complex fire. Colors are as in g), with dotted lines representing the 2-day persistence and random forest forecasts. The green box highlights September 2nd in the time series and the purple box highlights a day where FRP (black and gray lines in g)-j)) is high and model performance is good.

In order to address the issue of stations and grid cells being counted more than once, we introduce polygon merging. First, this involves determining which buffered polygons intersect each other, for all buffer sizes. Then, intersecting buffered polygons for which the intersection area makes up more than 30% of the area of either polygon are combined and statistics are extracted as above. Figure 4.3a, and d illustrate merged 150 km buffer polygons built out of the

150 km polygons for the August Complex, Red Salmon, and North Complex fires. This dramatically reduces the number of stations or grid cells being counted more than once. Across all buffer sizes, points are mainly counted once, with a small fraction counted twice and a negligible fraction counted three times (Figure C3b,d; C4b,d). Using the merged buffered polygons reduces the occurrence of double counting, but it can introduce background (non-smoky) AOD or PM2.5 values into computed statistics (e.g Figure 4.3a,b lighter colors in the even numeral plots).

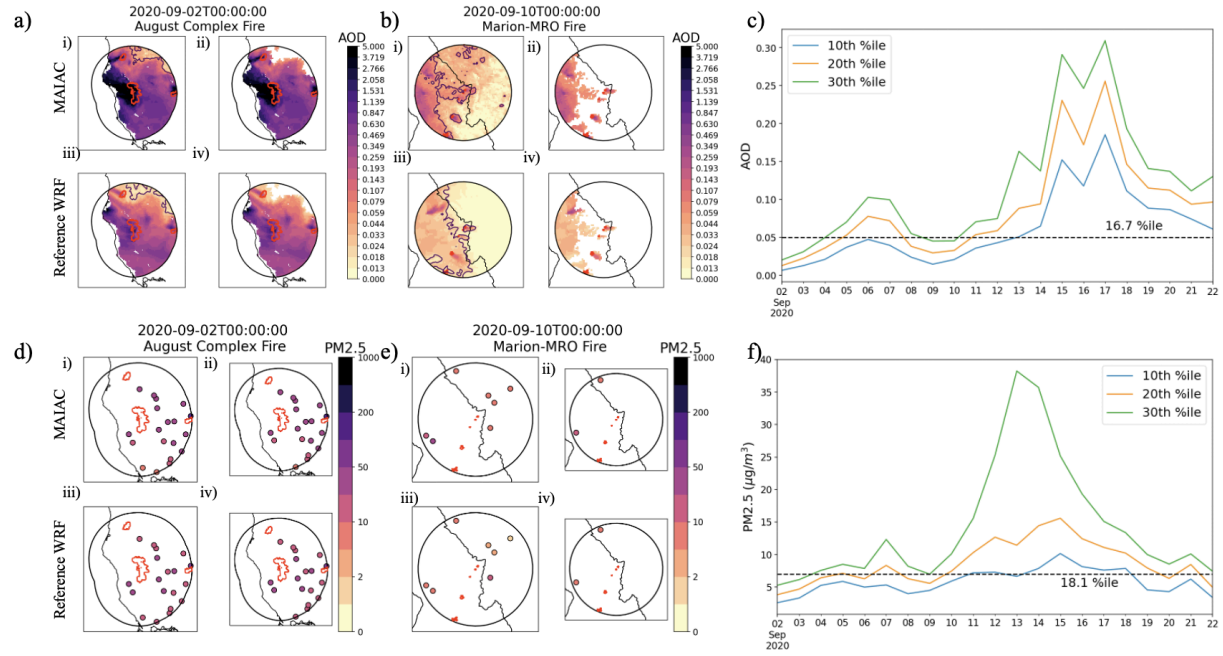

Figure 4.3. Filtering AOD and PM25 measurements over merged polygons. a-b),d-e) show the 150km merged polygons plotted on top of AOD (a,b) and PM2.5 (d,e) for the August Complex Fire on September 2nd (a,d) and the Marion-MRO fire (b,e). Surrounding fires are plotted in red. Unfiltered data is shown in even-numeral plots, and filtered data is shown in the odd-numeral plots. c) and f) show time series of the 10th (blue), 20th (orange) and 30th (green) percentiles of the observed AOD (c) and PM2.5 (f), and threshold percentiles are shown in the black dotted lines.

Therefore, we apply a domain- and time period-wide single threshold to AOD and PM2.5 and only calculate statistics using values above these thresholds (Figure 4.3c, f, dotted line). Note

how 10th and 20th, percentiles (orange and blue lines) are below/near the threshold for the beginning of the time series and above the threshold towards the end of the time series. This threshold was selected to remove background values from the relatively cleaner beginning of September while preserving smoke enhancement throughout the time series. The thresholds shown in Figure 4.3 are based on the observations and correspond to 0.01 for AOD and 7 ug/m3 for PM2.5. The corresponding percentiles were selected for the models, and the values are 0.02 for WRF-Chem AOD and 6 for WRF-Chem PM2.5 Filtering out AOD and PM2.5 values below these thresholds reduces the number of points included in the statistics, and minimally affects the number of points counted more than once (Figure C5). Figure 4.3a,b,d,e, even numeral plots show how smoke enhancements are preserved in the region around the fires.

Per-polygon model evaluations allow us to determine when and where the machine learning derived emissions improve on persistence emissions. For the per-polygon model evaluation performance statistics (see above) are computed on the filtered data for each 1- and 2-day persistence (P) emission- and random forest (RF) emission-derived prediction of AOD and PM2.5. The distance between each performance statistic and its ideal value (1 for PC and RAT, 0 for all other statistics) is then determined for each polygon and each statistic.

Generally speaking, the reference WRF-Chem model tends to underpredict AOD and over and underpredict PM2.5, with the best model performance occurring for the first 10 or so days of the simulation period for the western states and reasonable model performance occurring throughout the simulation period for the intermountain west states (see section 4.4.1). Figures C6-C15 show model evaluation over the top ten fires by size that burned during September 2020. Due to the fact that the reference model has the most reliable emissions of the different emissions configurations, we restrict our comparison of P and RF emissions to days where the reference

model performs well. Good reference performance days first filtered by FRP>0, PC>0 for AOD and PM2.5 to ensure fires are active and there is reasonable spatial alignment of plumes. Then we further restrict the analysis to days where the reference model has |MB|<0.5 (|MB|<50) for AOD (PM2.5) and |NMB|<0.5 for AOD and PM2.5. For an example of a good reference model performance, see Figure 4.2, purple box.

Once good reference model days are selected, statistics are then separated into 3 categories depending on their distances from the ideal values: RF better than P, RF similar to (~) P, and RF worse than P (Figure C16). The distribution of statistics in each category is then compared and conclusions are drawn about the degree to which RF emissions improve on P emissions for predicting air quality.

4.4. Results

4.4.1 Comparison of daily average black carbon emissions rates

Figure 4.4 compares 1- and 2-day persistence- and random forest-based (Thapa et al., accepted) predictions of the daily average emission rate of black carbon (BC) as estimated by the RAVE) emissions inventory respectively. RAVE BC emissions show improvement with the application of the random forest, most notably in terms of R^2 score in both the 1 day (a,b) and 2 day (c,d) forecasts. An improved R^2 score indicates that the rando, forest improves on both the timing and the magnitude of the emissions and thus is a more reliable prediction than persistence.

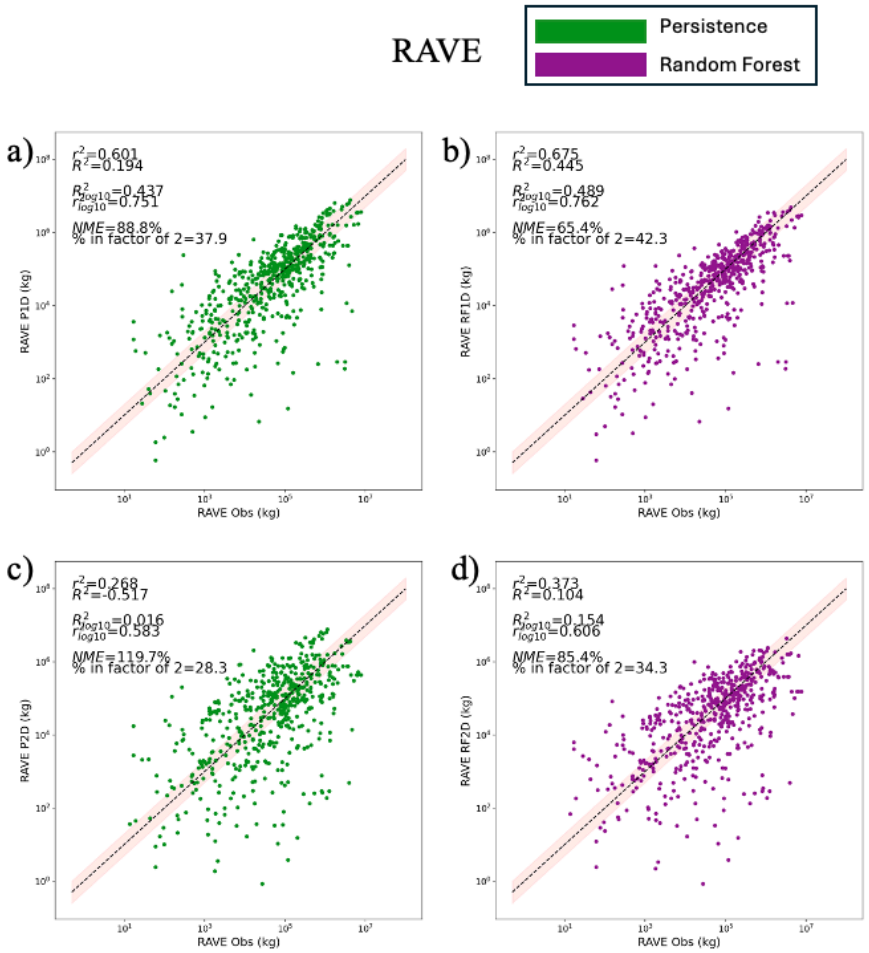

Figure 4.4. Comparison of RAVEblack carbon (kg) with one day (a-b) and two day (c0d) persistence (a,c, green) and random forest (b,d purple) predictions. The black dotted line is the 1:1 line and points in the red shaded region are predictions within a factor of 2 of the observed value.

4.4.2 Bulk evaluation of AOD and PM2.5

Figure 4.5 shows the domain averaged AOD (5a-5d) and PM2.5 (5e-5h) domain average over station locations at times where satellite AOD measurements are available for each day of the simulation period. Model PM2.5 and AOD from the 1-day persistence emissions predictions (green), 1-day random forest emissions predictions (purple), and reference emissions (blue) are compared with observed PM2.5 and AOD (red). All model simulations capture the temporal pattern of AOD and PM2.5 reasonably well (5a,e; r~0.8 for AOD and PM2.5), but all model

simulations tend to underestimate surface and column aerosol loading (NMB=-53.19%- -51.64% for AOD, NMB=-45.65%- -44.66% for PM2.5). Generally speaking, model underestimation is a result of the smokiest days' (e.g., Sept 10th-18th) AOD and PM2.5 being underpredicted. However, missing data also plays a role here, particularly on Sept 9th where many fires peaked in emissions and AOD was not well-retrieved spatially (e.g., Figure C6). There are a few days at the beginning of the simulation period (Sept 5th-10th) where surface PM2.5 is overestimated across the models and one day at the beginning of the simulation period (Sept. 6th) where AOD is overestimated. In the observations, the peak in AOD beginning September 6th is not matched in the PM2.5 indicating a plume which was perhaps more lofted. Here the model overestimates PM2.5 and AOD, indicating perhaps a missing injection or the wrong fraction of the smoke lofted.

In the domain average, it is difficult to see the impact of random forest-derived emissions on PM2.5 and AOD (Figure 4.5c,d, 4.5g,h). Because persistence and random forest emissions are based on the reference emissions, these two emissions scenarios do not seem to address the overall model biases in PM2.5 and AOD (5b, 5f). Maps of evaluation metrics over the whole simulation period, which leverage the complete spatial coverage of the satellite data show that while the temporal evolution of the smoke remains well represented across the domain (Figure C17a), underestimation is most severe in the Western states of CA and OR (Figure C17b) and present across the domain. The differences in AOD between the persistence and random forest simulations are most pronounced in regions in the immediate vicinity of burning fires (Figure C17c-f).This analysis motivates the main methodological advancement of this paper: automated evaluation of models on a fire-by-fire basis.

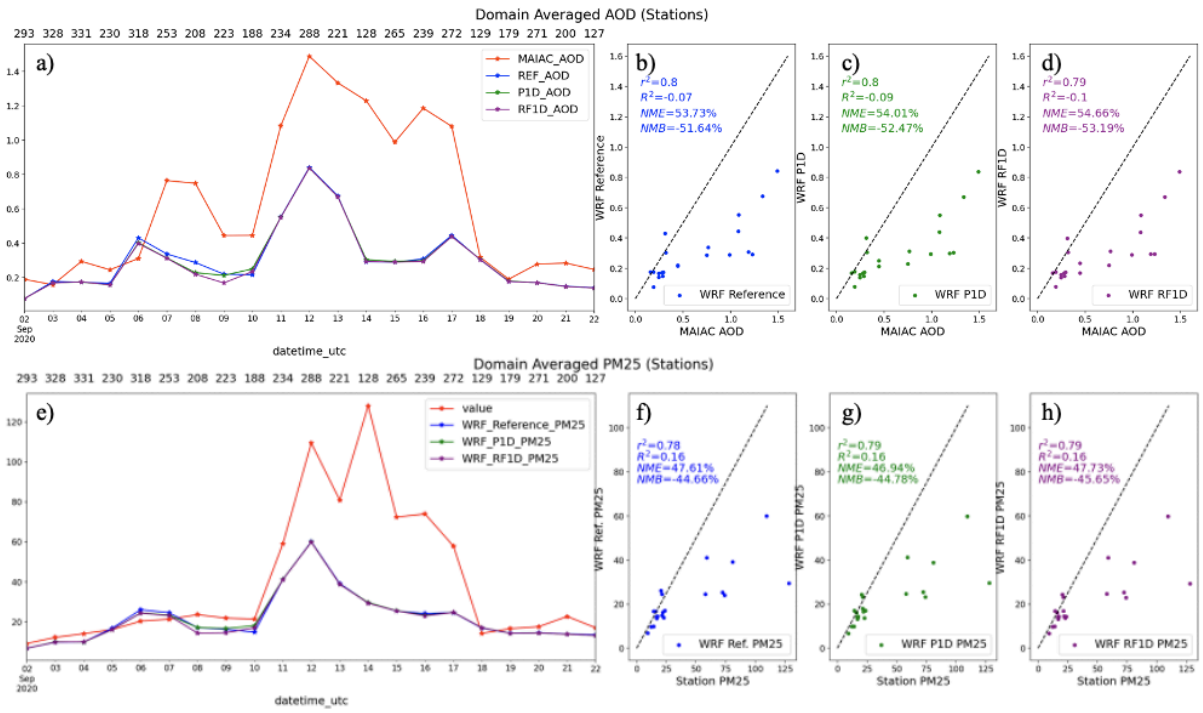

Figure 4.5. Time series and comparison of daily, domain-averaged AOD (a-d) and PM2.5 (e-h). Red refers to observations, blue to the reference WRF-Chem model, green to the persistence emissions WRF-Chem model, and purple with the random forest emissions WRF-Chem model.

4.3 Per-fire model evaluation

In this section we examine how the models represent AOD (Figure 4.6) and PM2.5 (Figure 4.7) on a fire-by-fire basis. These figures compare the distributions of the metrics for P and RF for all fire days (Figure 4.6a,4.7a), days with low, moderate, and high Fire Weather Index (FWI; Wagner, 1987; Figure 6b,7b) and days with very high and extreme FWI (Figure 4.6c,4.7c). Evaluation metrics were calculated over the merged polygons using the filtered data (section 4.3.5), and the day 1 (day 2) forecasts are evaluated on the near (far) buffers. For AOD, a majority of the cases (50-90% depending on the metric), show the random forest and the persistence methods having comparable skill (Figure 4.6, RF~P). For PM2.5, there are a few metrics where RF and P are very similar in less than <50% of the cases (MB, RMSE, Fig

4.7ai,aii), but in most metrics a majority of the cases show RF and P having similar skill. Therefore, we consider the random forest to improve upon persistence in terms of AOD or PM2.5 if the median of the metric distribution (Figure C18,C19) moves further toward the ideal value in the "RF better than P" cases than it does away from the ideal value in the "RF worse than P" cases. Improvement (degradation) to forecast skill is represented by blue (red) shading, with the darkest blue (red) indicating the largest improvements (degradations).

Choosing to use RF emissions instead of P emissions can move us closer to correctly predicting AOD enhancements due to smoke in both 1 and 2 day forecasts (1D, 2D). Across all fire days (Figure 4.6a), the random forest shows mixed success. Generally speaking, 20-50% of the fire days see large changes in error metrics due to the use of RF emissions, and usually a higher percentage of fire-days are classified as "RF worse than P" than "RF better than P". However, RF1D (RF2D) shows improvements on NME, PC, and RAT (RMSE and MB), with these improvements being up to a factor of 4 in the case of RMSE (Figure 4.6aii) and RAT (Figure 4.6avi). It is also notable that even in cases where the median of the distribution is degraded by the implementation of the random forest emissions (e.g., NMB 4.6aiii), the spread of error may still be narrowed, indicating that large, outlier errors have been removed. The RF1D is more skillful than the P1D at predicting the magnitude and timing of AOD enhancements due to smoke, whereas RF2D is more skillful than P1D at predicting magnitude of AOD enhancements due to smoke. For low-moderate-high FWI, RF1D and RF2D tend to be more skillful than P1D and P2D at predicting AOD, with the RMSE, NME, and PC metrics improved by a factor of 4 (Figure 4.6bii, 4.6bvi, 4.6bv). For very high-extreme FWI, RF1D and RF2D are more and less skillful than P1D and P2D respectively, with multiple metrics degraded by a factor of 4 (Figure 4.6ciii, 4.6civ, 4.6cv). This means that AOD is more predictable with RF emissions

than with P emissions under low fire danger conditions, consistent with previous work (Thapa et al., accepted). This trend holds true whether or not merged polygons are used or filtering is applied (Figure C20-C27).

PM2.5 predictions can also see improvement through the implementation of RF emissions. Like AOD, PM2.5 RF can improve or not improve upon P depending on the metric when all fire days are considered (Figure 4.7a). However, for certain metrics 50-60% of the fire days can see large changes in error metrics due to the RF emissions (Figure 4.7ai, 4.7av, 4.7avi). RF1D (RF2D) shows improvements for RMSE, NME, PC, and RAT (RMSE and NME) indicating that 1 day predictions of PM2.5 show improved magnitude and timing and 2 day predictions of PM2.5 show improved magnitude. RF1D-based PM2.5 predictions also tend to be better than P1D-based predictions of PM2.5 for low fire danger than for high fire danger with NME, PC, and RAT (Figure 4.7biv, 4.7bv,4.7bvi) being improved by up to a factor of 4 for low-moderate-high fire days and RMSE being degraded by a factor of 4 across both forecast days (Figure 4.7cii). This is consistent with our AOD predictions and previous work, although PM2.5 forecasts do tend more generally to be improved whereas AOD forecasts are more mixed. Notably, RF2D out-performs P2D for all high fire danger day cases, which is why no data points are present for the RF2D worse than P2D category.

Generally speaking, the RF models tend to improve on error and correlation metrics for AOD and error, correlation, and some bias metrics for PM2.5. We interpret this to mean that the RF emissions help enhance the timing of emissions while not always addressing issues with forecast magnitude and the consistent underprediction of AOD and PM2.5 exhibited by WRF-Chem. It is also important to note that the random forest model is trained to predict changes in the FRE observed by GOES and VIIRS (RAVE input), and so is trained to a dataset

which comes with its own set of uncertainties. Additionally, underlying uncertainties with the WRF-Chem model could also be responsible for the underprediction issues.

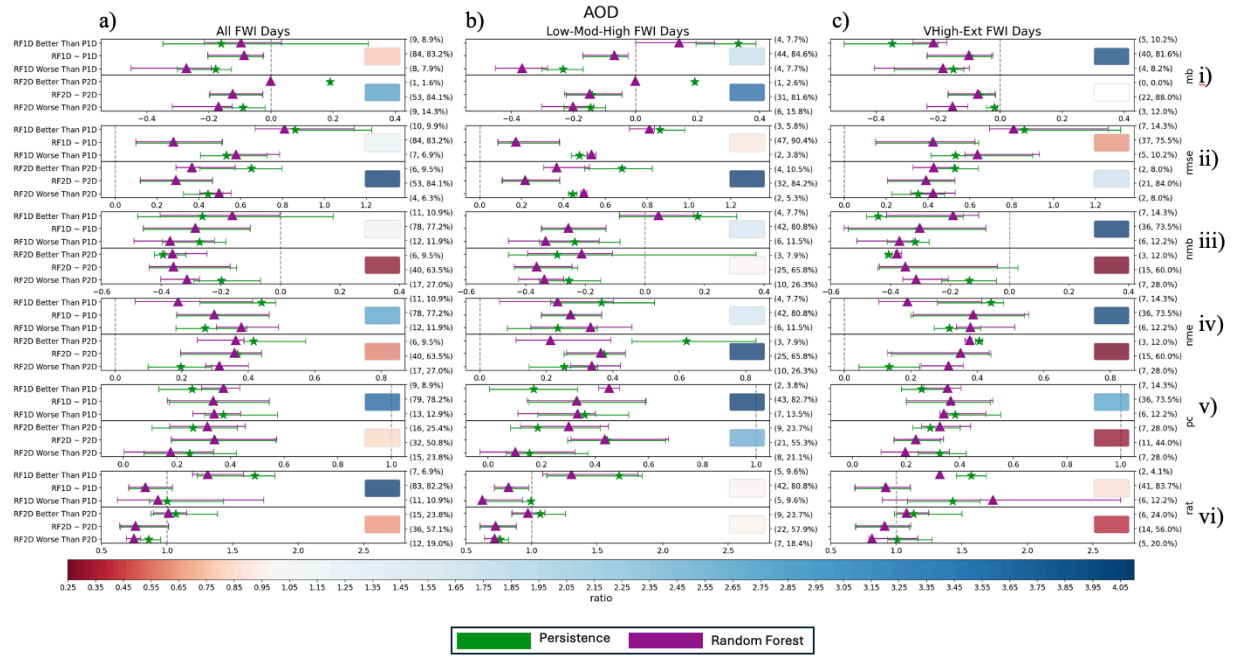

Figure 4.6. Comparison of the distribution of the median (stars* and triangles^) and spread (bars) of AOD error metrics for a) all fire-days, b) days with low-mod-high FWI, and c) days with very high-extreme FWI. Error metrics plotted include i) mean bias, ii) root-mean-squared error, iii) normalized mean bias, iv) normalized mean error, v) pearson correlation, and vi) ratio. Shaded boxes show the ratio between the change in the persistence (*) and random forest (^) medians when RF is better than P and the change in the persistence (*) and random forest (^) medians when RF is worse than P. Blue means RF increases overall forecast skill in terms of the metric, and red means RF decreases skill in terms of the metric.

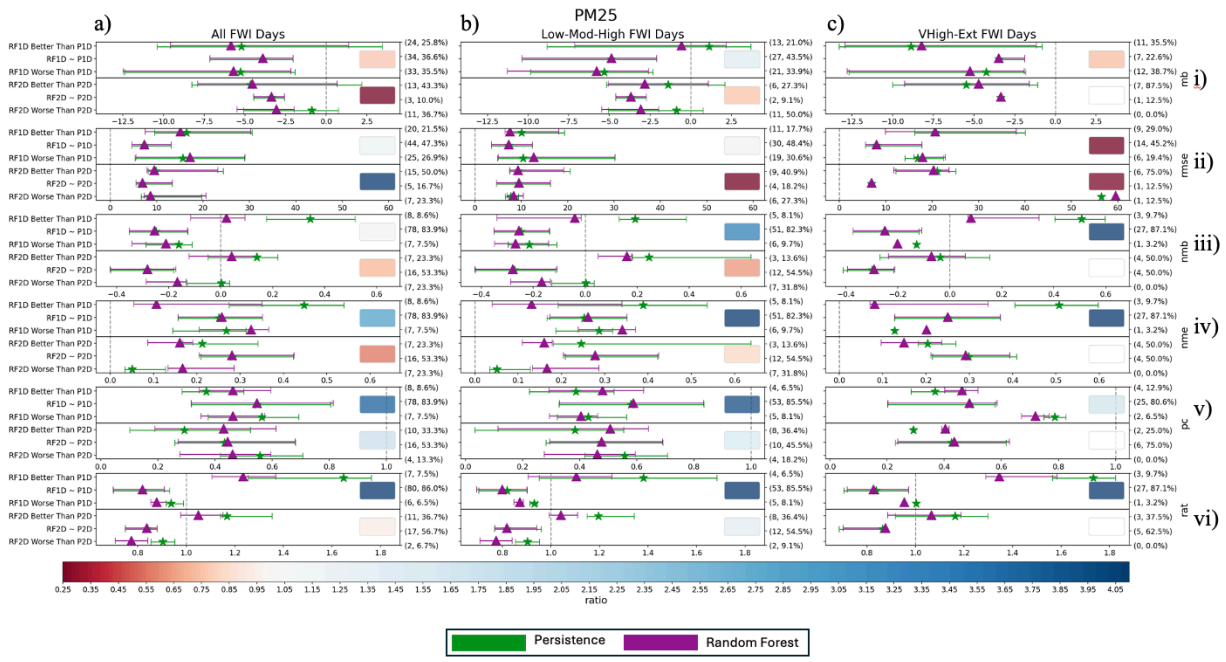

Figure 4.7. As in Figure 4.6 for PM2.5

4.5. Conclusions

Here we investigated a random forest (RF)-based update to the current persistence (P) assumption used for predicting smoke emissions and quantified how the machine learning-based emissions impact air quality predictions. The machine learning model was applied to the Regional ABI and VIIRS Emission (RAVE) dataset and the WRF-Chem model was used to make air quality predictions.

We showed that although the machine learning model substantially improved emissions over persistence, the gains in terms of air quality forecast skill in WRF-Chem were relatively smaller. Indeed, in the domain-average, the WRF-Chem model does not appear to be very sensitive to the emission prediction scheme used. This is likely due to the fact that smoke plumes are relatively small features in most cases and changes in emissions are likely smeared out in the domain average. Therefore, we introduced an automated method to select the region around a given fire or set of nearby fires and evaluate smoke enhancements in AOD or PM2.5 in those

regions. This analysis showed that WRF-Chem is sensitive to the emissions configuration and WRF-Chem AOD and PM2.5 predictions can be improved by using RF emissions instead of P emissions. However, uncertainties in both the emissions and the WRF-Chem model make it difficult to definitively state whether the random forest emissions are an improvement to the current configuration of WRF-Chem with RAVE emissions.

Our results are primarily limited by uncertainties in the baseline emissions and model uncertainty in WRF-Chem. With baseline emissions, WRF-Chem underpredicts AOD, even after RAVE emissions had been scaled following Darmenov (2015). Although RAVE is based on polar orbiting and geostationary satellites and is expected to provide continuous hourly emissions, certain fires, such as those obscured by clouds and thick smoke, could be missing or their emissions underestimated. Internal uncertainty, such as plume rise or the representation of smoke optical properties (Saide et al., 2022) could also impact the way AOD and PM2.5 are represented. Finally, this work carries forward the same errors from the random forest described in Thapa et al., (accepted), most notably is that the random forests tends to default to a persistence forecast during days of severe fire weather as was seen throughout much of September 2020.

Chapter 5

Thesis Conclusions

5.1 Research Summary

This work focused on improving the way fresh smoke plumes are represented in chemical transport models (CTMs). In particular, my work examines uncertainties with the Freitas plume rise model (Freitas et al., 2007) and the persistence assumption for predicting the vertical extent and amount of fresh smoke.

In chapter 2, we learned that both the WRF-Chem and HRRR-Smoke implementations of the Firetas model overestimate the frequency of injection as a result of assumed model heat fluxes being much too high. Indeed, when we compare the model and aircraft-derived heat fluxes, we find that overprediction can be up to a factor of 25 (9) in the worst WRF-Chem (HRRR-Smoke) cases. The observations suggest a heat flux of 0.55 kWm^{-2}, consistent with what WRF-Chem uses as the heat flux for grass and savanna fires and explaining why the baseline WRF-Chem model was a bit more accurate than the baseline HRRR-Smoke model. Finally, we adjusted heat fluxes in WRF-Chem to better match the observed value and found a decrease in the rate of false injections. However, in order to get the best performance out of the WRF-Chem model, the observed planetary boundary layer (PBL) height needed to be used, pointing to the need for using the most recent PBL scheme available in WRF-Chem going forward.

In chapter 3 I describe a method to use machine learning and domain specific knowledge of the drivers of fire to predict the daily changes in fire intensity. I tested scaled weather indices and a random forest (RF) trained on data representing fuel, weather, topography, and firefighting. Although both the scaled and RF methods beat persistence, the RF method showed the most

skill, explaining up to 50% of the variance in observed fire radiative energy (FRE) for the 1 day forecast. The RF maintained its high skill level out to two days of forecast lead time, but RF predictions tended back towards persistence predictions for days of very high and extreme fire weather index (FWI, Van Wagner, 1974). Finally, sensitivity analysis showed that the RF was learning the most from weather and current fire size, and that a pared-down version of the input features performed just as well as the full feature set. Therefore, we recommend further testing machine learning based smoke emissions in air quality prediction frameworks.

In chapter 4 the main takeaway is that uncertainties underlying the Regional Advanced Baseline Imager (ABI) and VIIRS Visible Infrared Imaging Radiometer Suite (VIIRS) Emissions (RAVE) and the Weather Research and Forecasting with coupled Chemistry (WRF-Chem) need to be addressed before we may state with confidence whether random forest (RF) based emissions are an improvement over the current persistence (P) assumption. At this point, the RF emissions do not make much of a difference with respect to P emissions in bulk evaluations of AOD and PM2.5 and over many of the fire days, but it is clear that the RF emissions can improve AOD and PM2.5 in the vicinity of large wildfires on certain days for certain fires. Granted, the improvements on some fire days come at the expense of degradation of forecast skill on other fire days, but for some forecast error metrics RF emissions improve skill much more than they degrade skill. When these improvements do occur, the RF improves the timing and magnitude of AOD and PM2.5 on the 1-day forecasts and improves the magnitude of AOD and PM2.5 on the 2-day forecasts.

Another key result of chapter 4 is the modeling and model evaluation pipeline that I derived. Emissions from any FRP-based inventory can be scaled on a per-fire basis and then those emissions fed into their respective models. I also developed an automated way to find the

regions around fires, combine buffer zones when multiple fires are near each other, and extract the smoke portion of the AOD and PM2.5 signals.

5.2 What does it all mean?

The research and tools presented in this thesis have some important impacts and implications which need to be mentioned. First of all, Chapter 2 indicates that a major change should be made to the Freitas model inputs when modeling Western US fires. The Freitas model is commonly used in regional smoke simulations, including operational models used by NOAA (Ye et al., 2021), so this work has the potential for broad impact in the community.

While it is very clear that machine learning is a promising way to predict the evolution of smoke emissions (Chapter 3), some uncertainty remains as to the best way to use these tools moving forward. Using input data available at daily resolution and hourly emissions summed to daily resolution, we were able to modify some but not all of the persistence predictions to predict increases or decreases instead. Therefore, one way that these machine learning tools could be improved would be to incorporate some idea of sub-daily fire spread into the machine learning models. This could be accomplished by training the model to predict a scaling factor on hourly resolution using hourly polygons (i.e. those developed by Y. Chen et al., 2022), a similar method to what was used here, or by predicting the shape parameters of the diurnal cycle curve. Predicting hourly emissions would be a great place to leverage simple fire spread models like FARSITE (Finney et al., 1998), as these could predict fire growth on a sub-daily basis and provide polygons over which to extract fire driver data. However, fire spread modeling is challenging in that it requires detailed knowledge of the landscape and geolocated firefighting efforts (Turney et al., 2023).

In Chapter 4 we saw that the forecast skill improvements made with machine learning emissions get smudged out in the output air quality forecasts. Since there is a lot of variability between different model setups, the first thing this points to is to repeat the WRF-Chem analysis with a different model and emissions inventory combination. Specifically, with WRF-Chem we noticed that the model tended to underpredict AOD and under and overpredict PM2.5. This points us toward a few places to explore further. AOD being underpredicted indicates that emissions could be too low, and recent work has shown that scaling emissions based on radar data (Saide et al., 2023) can lead to less underestimations of AOD (Howes, 2024). However, the fact that PM2.5 can be overpredicted suggests that simply scaling emissions may not entirely fix the problem. Therefore, it is important to also make sure we are correctly modeling the vertical distribution of smoke as well as the smoke optical properties (Ye et al., 2022b; Saide et al, 2022), as these parameters control the amount of coupling between smoke column loading (AOD) and smoke surface impacts (PM2.5).

The tools developed in my research also have an important role to play in research in the Saide Group and beyond. In Chapter 3 I developed a method to build wildfire polygons out of satellite fire detections. This work was used to select fires for which to build radar-based FRP estimations (Saide et al., 2023). In Chapter 4, I describe a pipeline for scaling emissions based on machine learning derived scaling factors. This tool was used to scale emissions based on radar-derived FRP Howes (2024), and in the future this tool can be used to scale emissions for use in any other model. Chapter 4 also describes using buffered satellite-derived polygons to look at AOD and PM2.5 in the regions around fires. Following the work of Krishna et al., (2024), which used weather radar to create gridded plume heights, the buffered polygons may

also be used to evaluate plume rise as well as AOD and PM2.5 spatially in the vicinity around a fire.

5.3 Closing Thoughts

This work took steps toward addressing two sources of uncertainty in fresh smoke plume modeling: plume rise and the persistence of emissions. Smoke modeling is just one challenge in the broader, climate-driven wildfire crisis (Zhuang et al., 2021), but being able to predict smoky days is key to keeping people safe in a world with more fire. Learning to live with fire will be a group effort, and it has been a pleasure to contribute research to the community.

Appendix A: Heat flux assumptions contribute to overestimation of wildfire smoke injection into the free troposphere

Contents of this file

Figures A1 to A60

Table A1 to A3

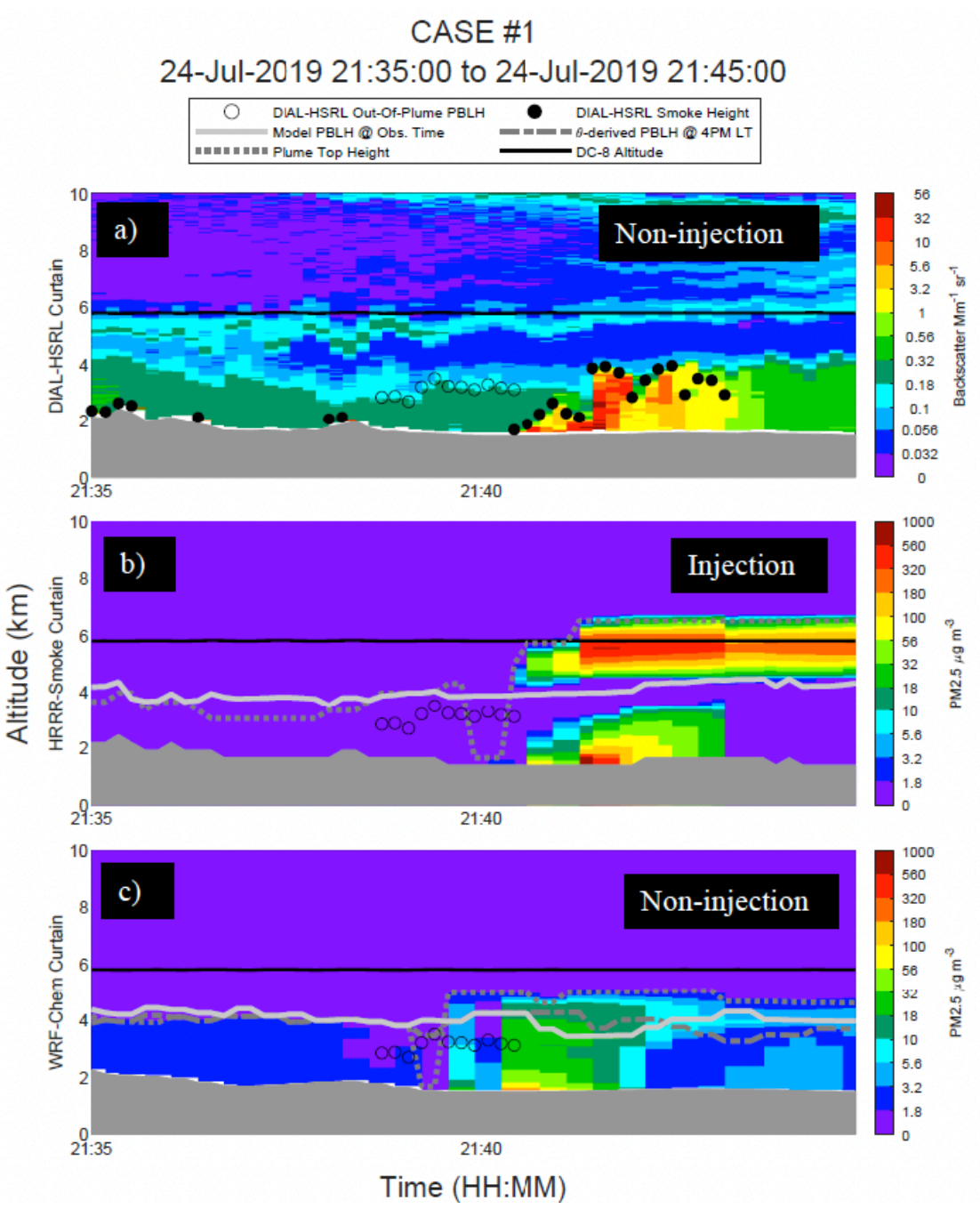

Figure A1. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheep Fire on 2019-07-24 21:35-21:45 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

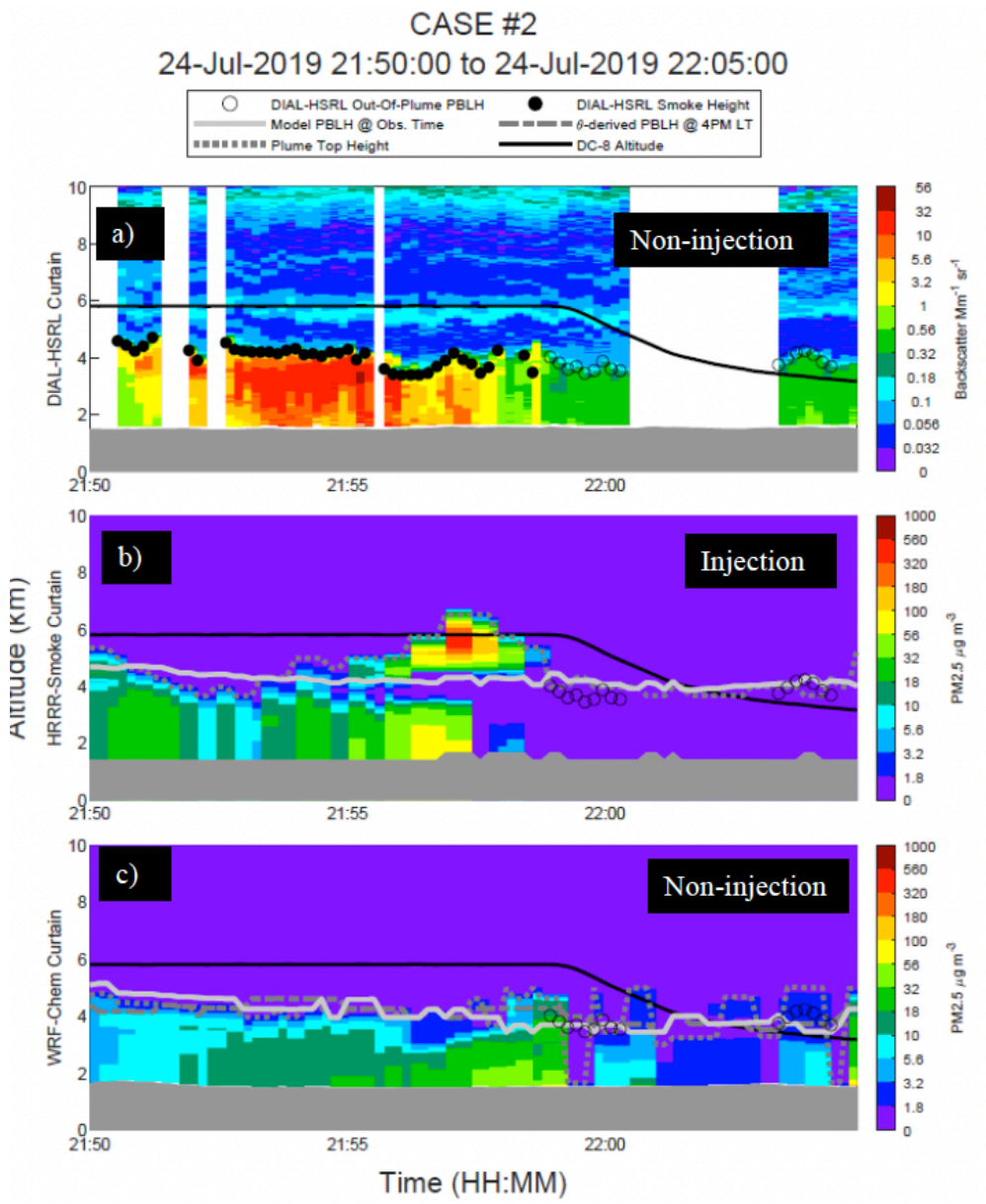

Figure A2. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheep Fire on 2019-07-24 21:50-22:05 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

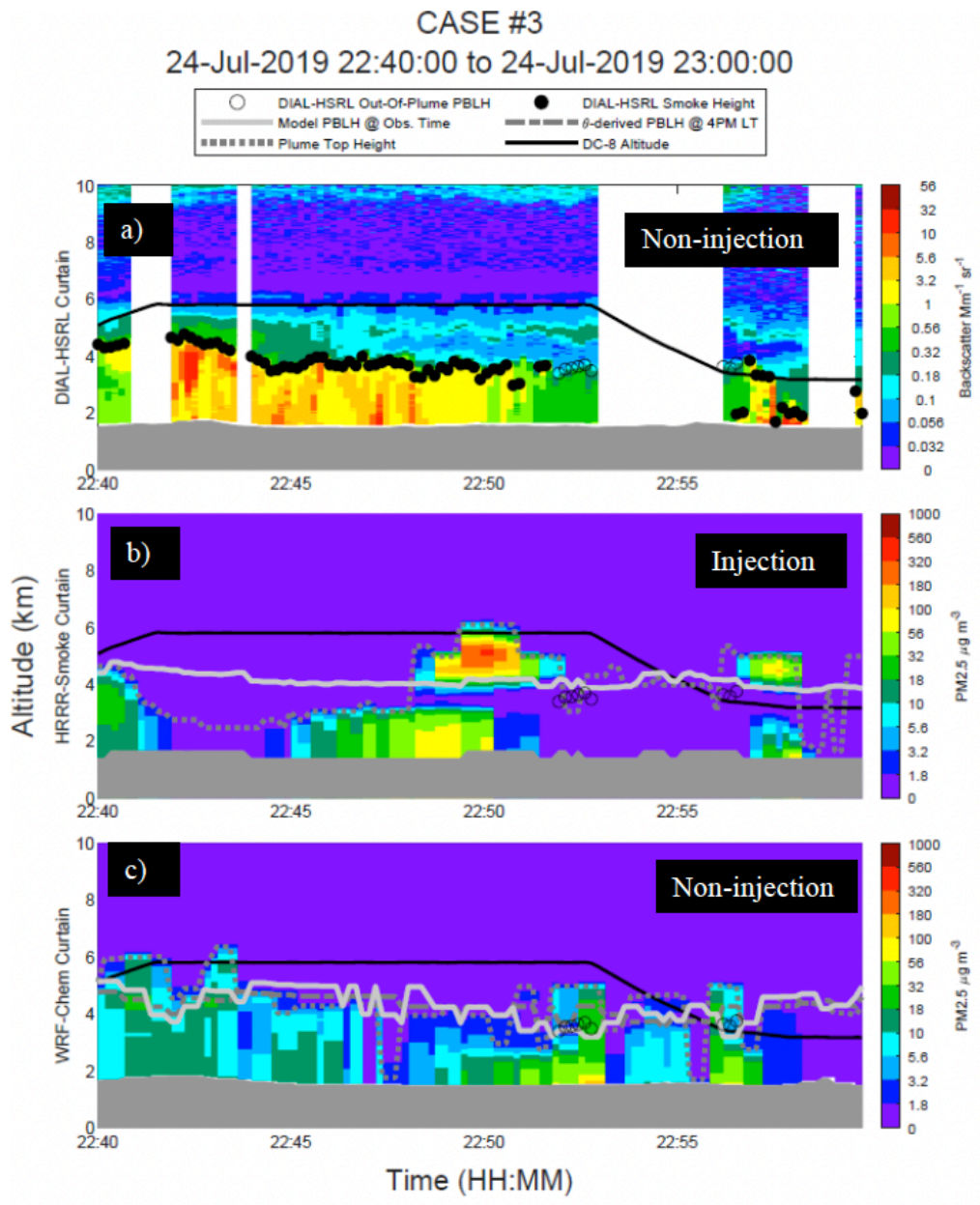

Figure A3. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheep Fire on 2019-07-24 22:40-23:00 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

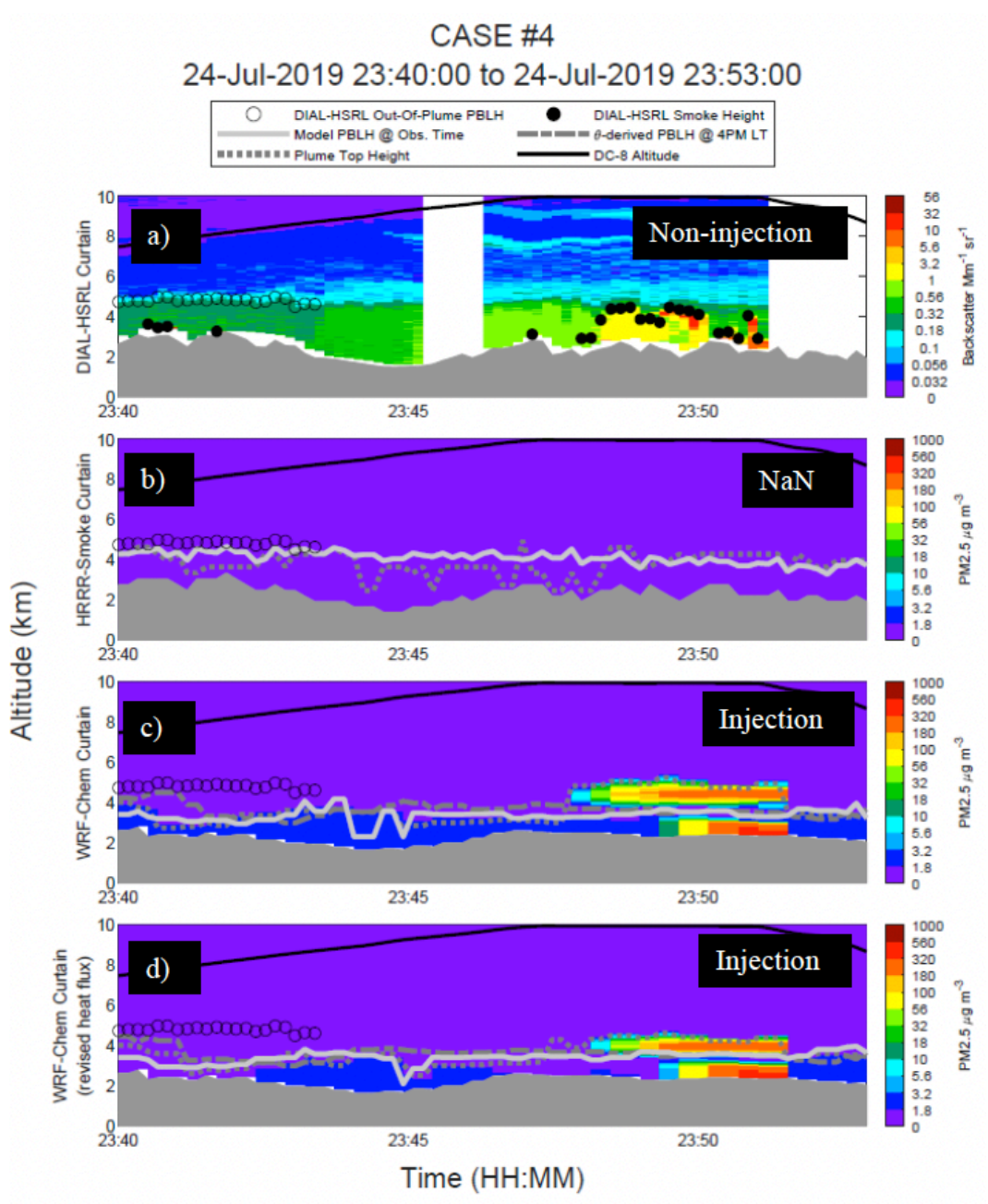

Figure A4. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Shady Fire 2019-07-24 23:40-23:53 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

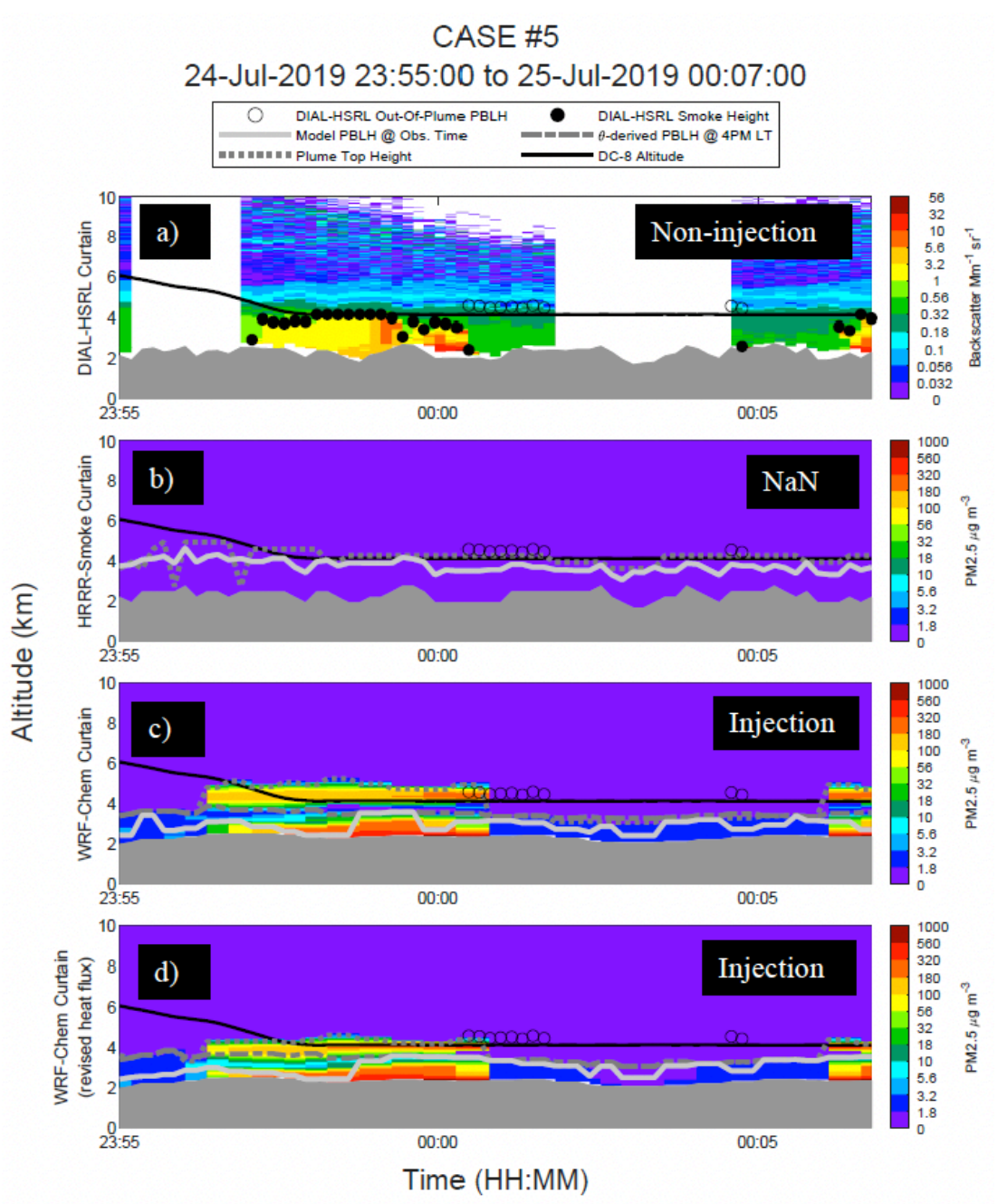

Figure A5. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Shady Fire 2019-07-24 23:55-2019-07-25 00:07 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

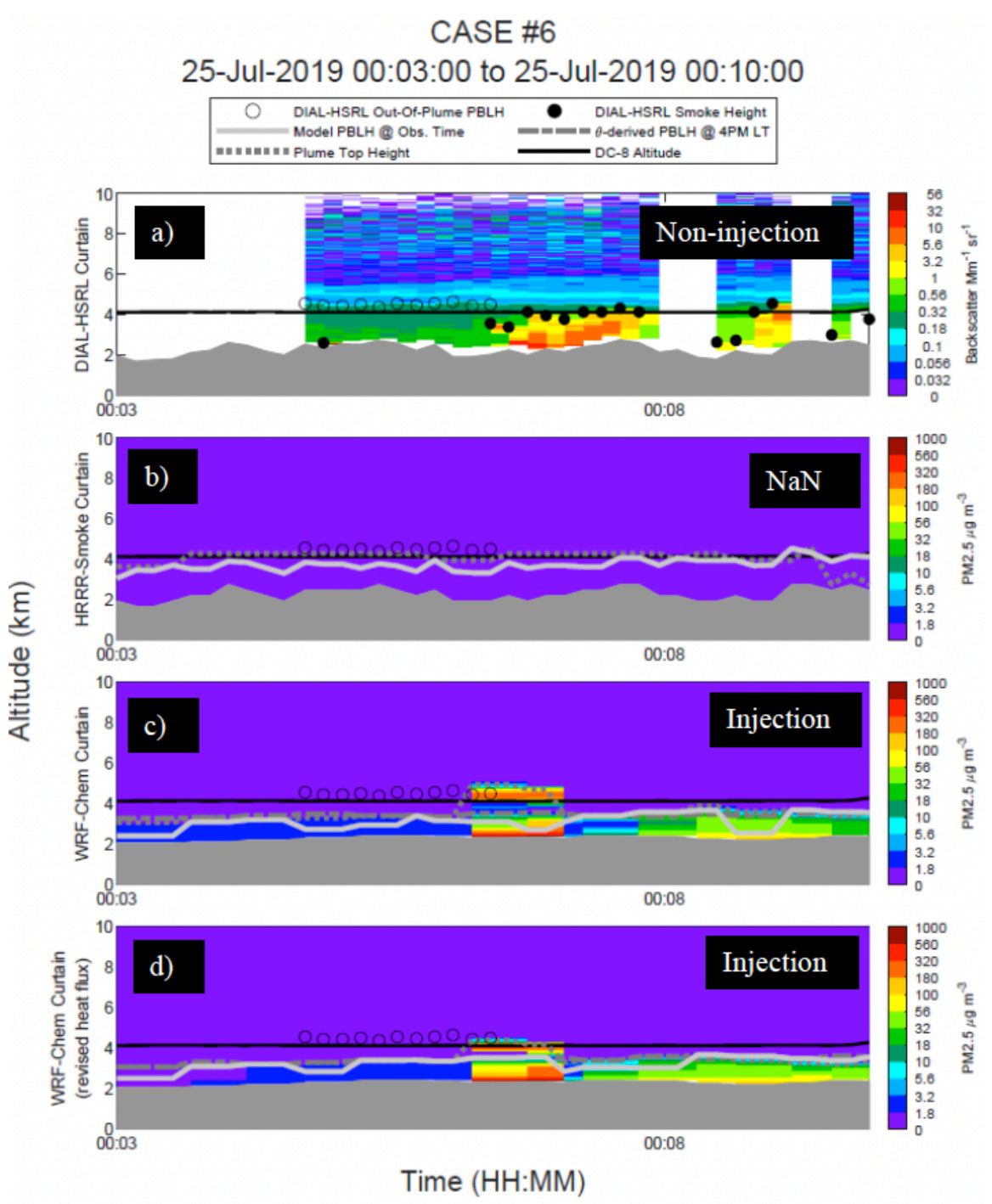

Figure A6. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Shady Fire 2019-07-25 00:03-00:10 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

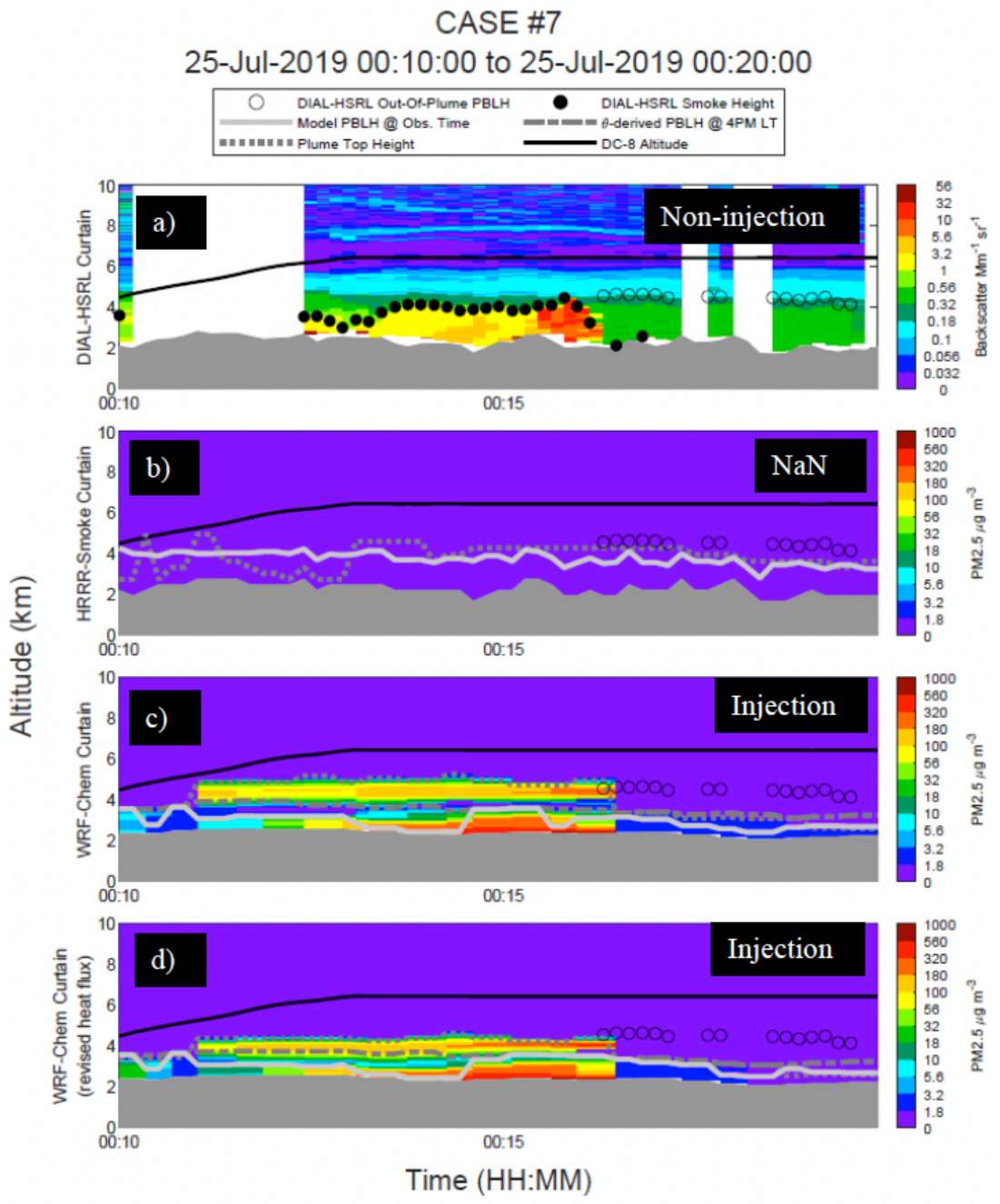

Figure A7. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Shady Fire 2019-07-25 00:10-00:20 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

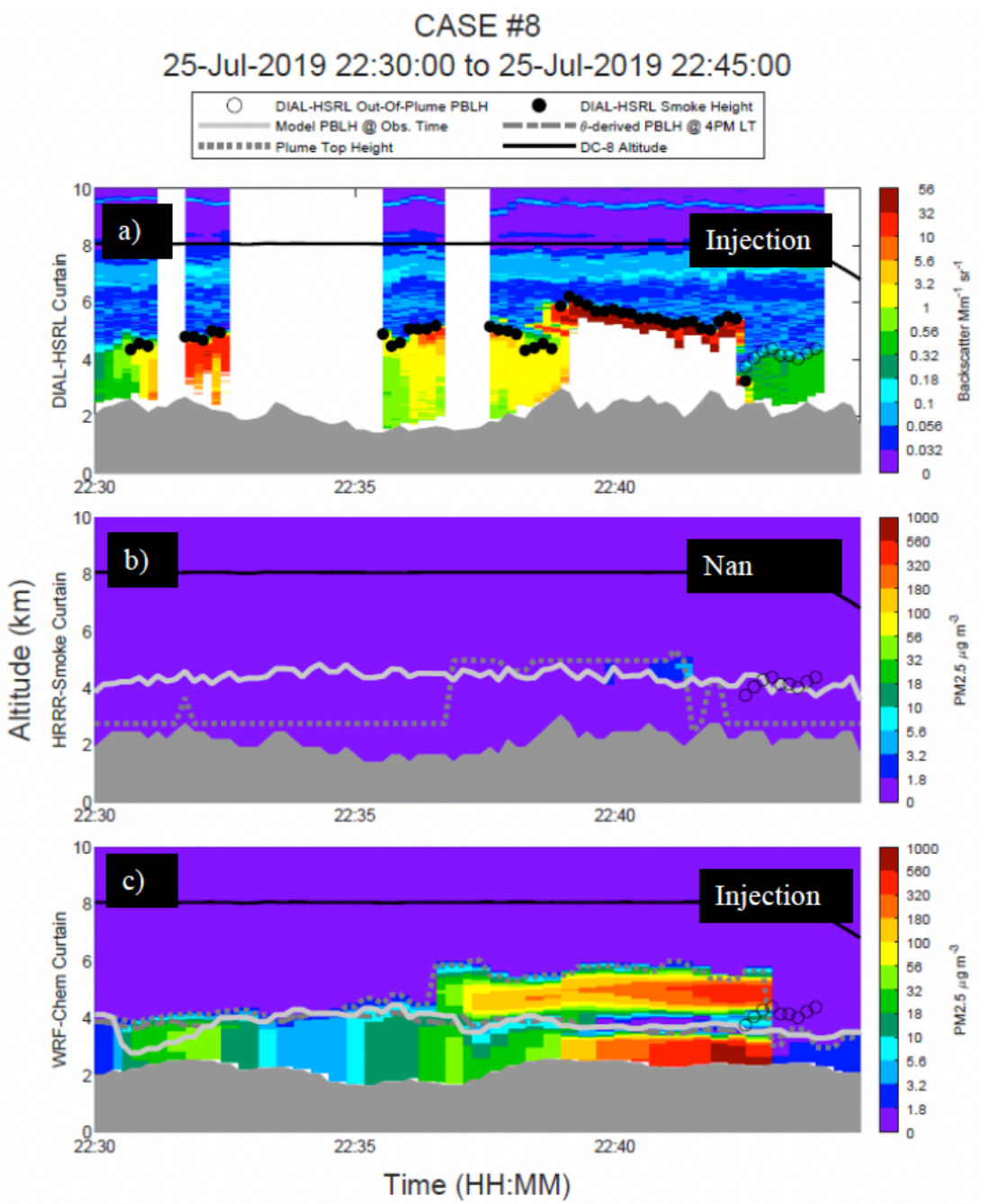

Figure A8. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Shady Fire on 2019-07-25 22:30-22:45 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

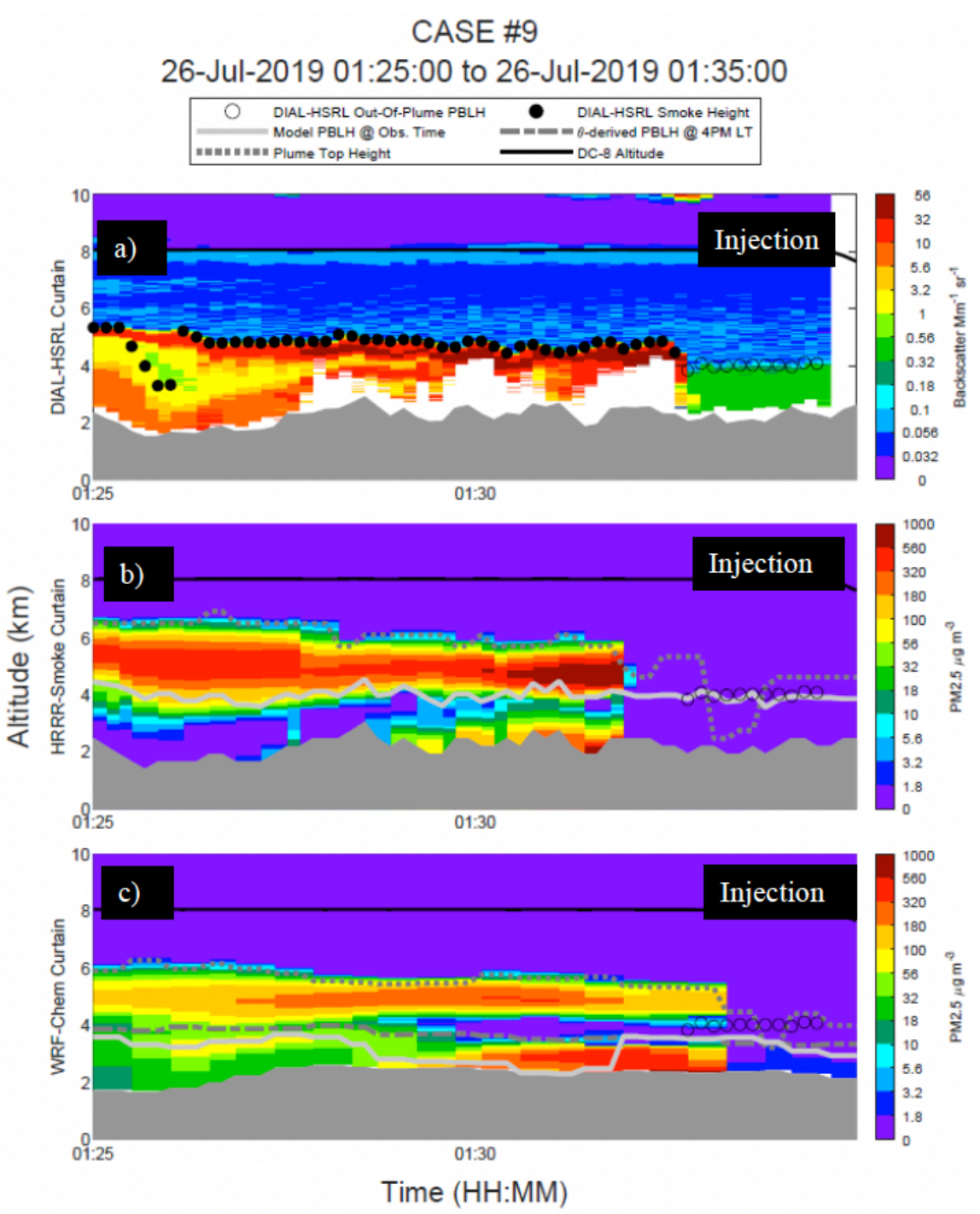

Figure A9. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Shady Fire on 2019-07-26 01:25-01:35 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

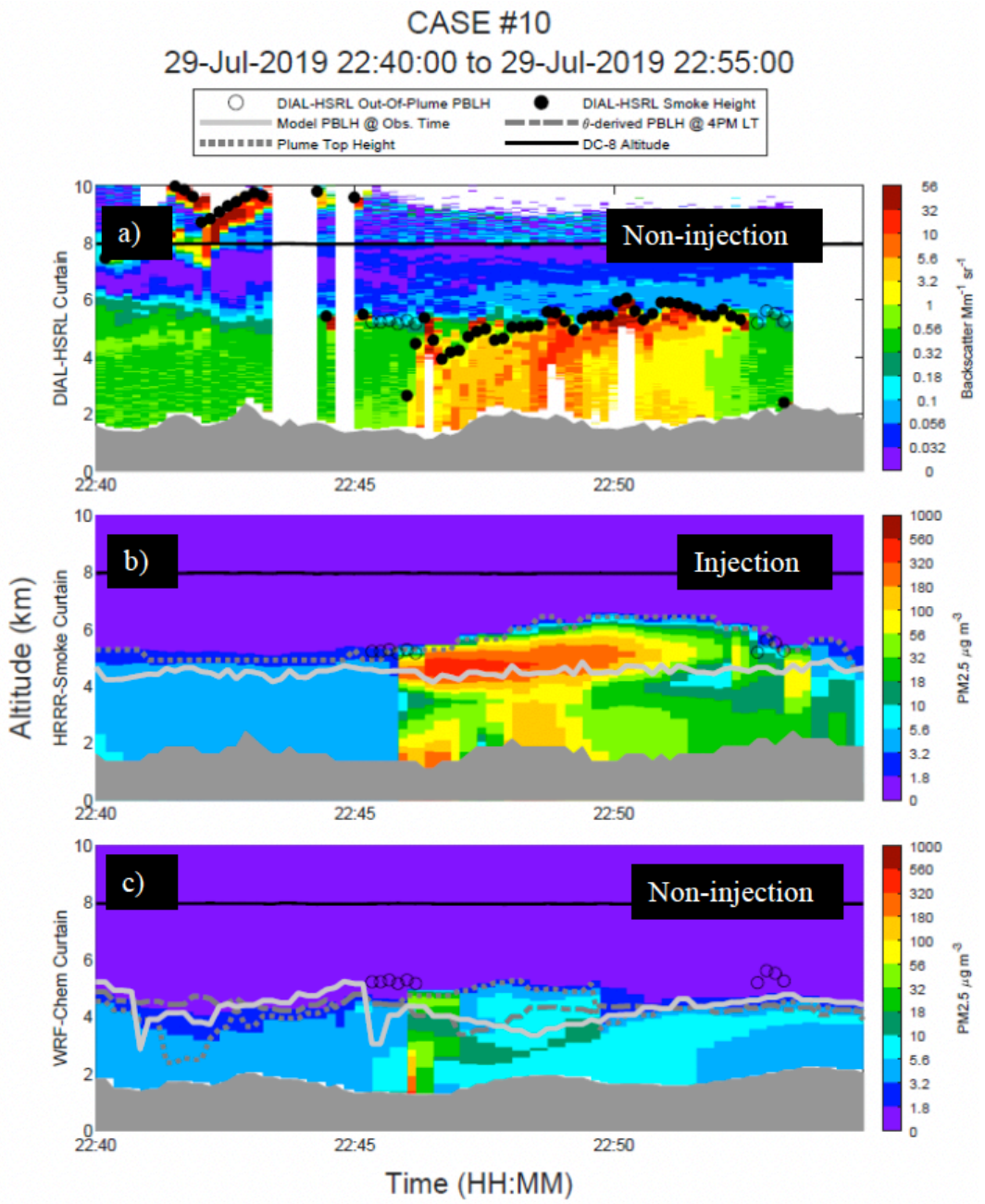

Figure A10. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the North Hills Fire on 2019-07-29 22:40-22:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

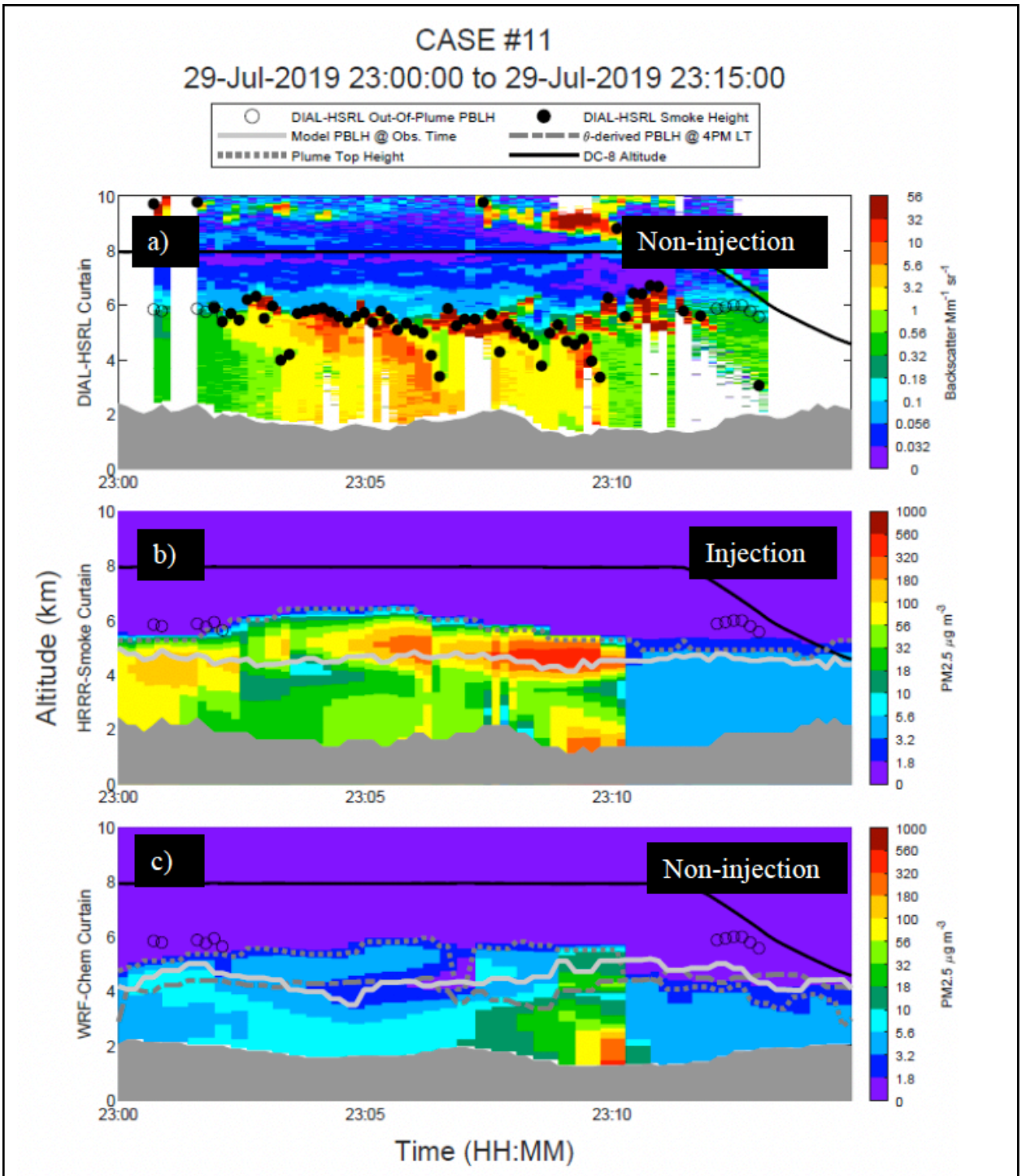

Figure A11. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the North Hills Fire on 2019-07-29 23:00-23:15 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

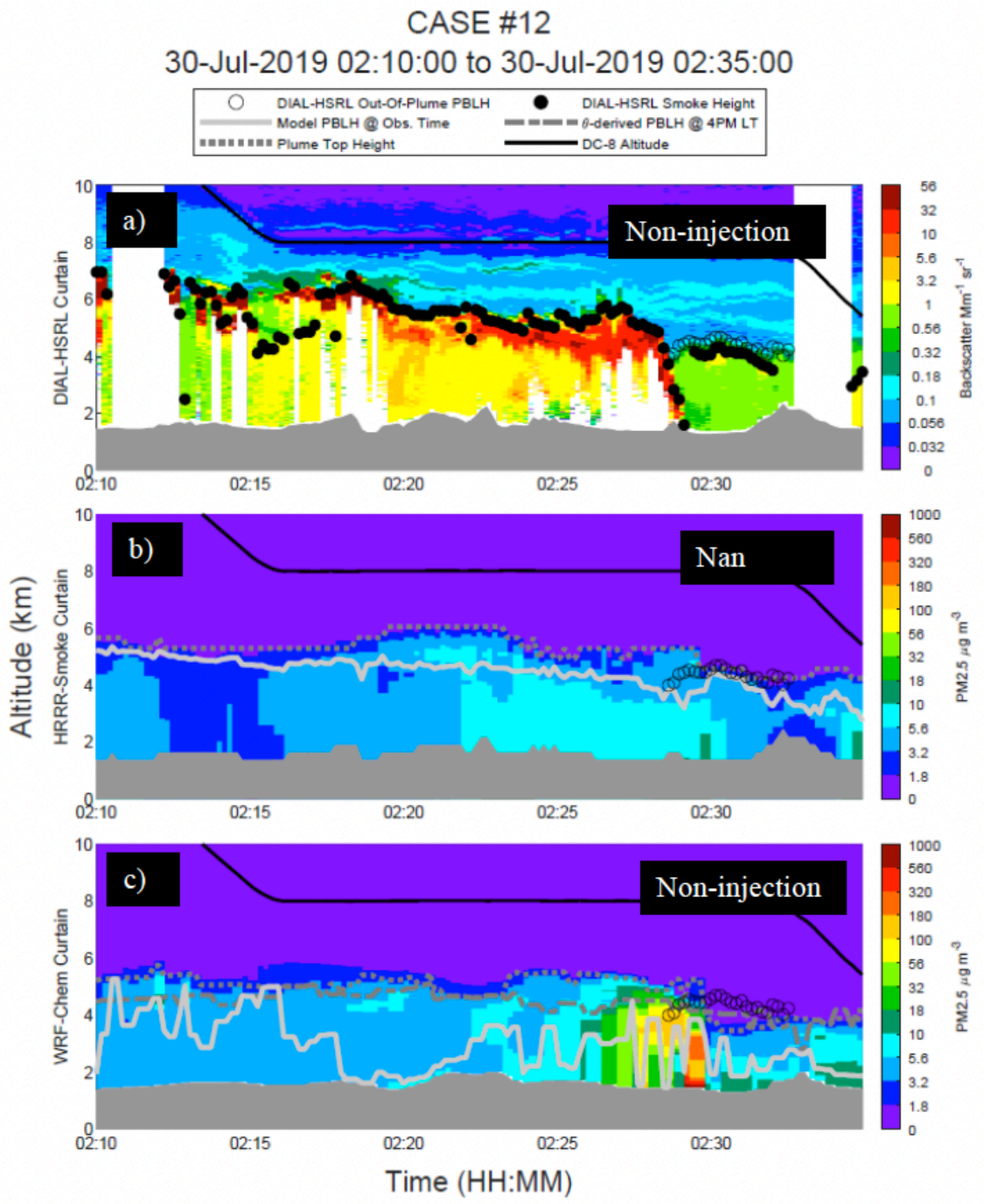

Figure A12. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Tucker Fire on 2019-07-30 02:10-02:35 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

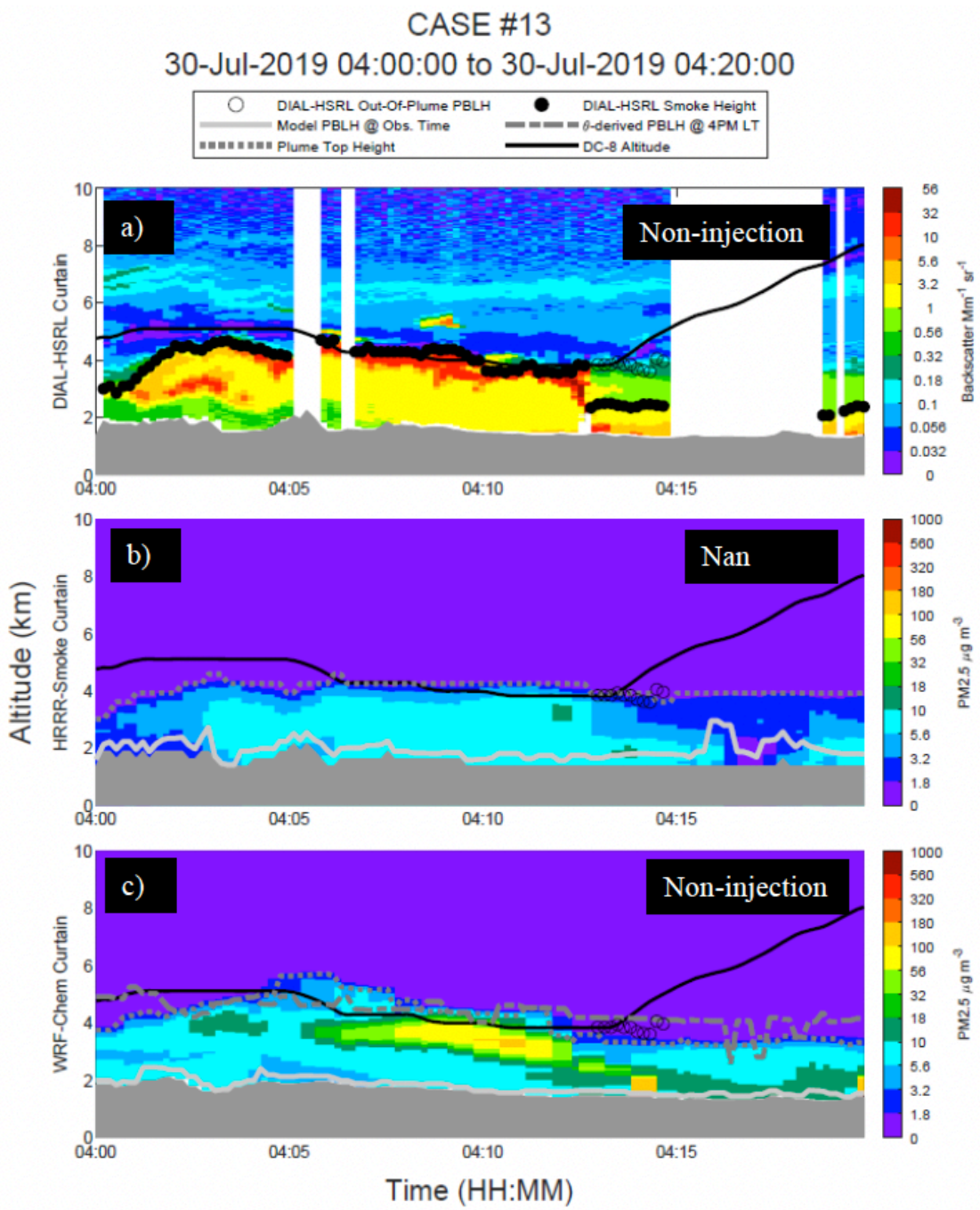

Figure A13. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Tucker Fire on 2019-07-30 04:00-04:20 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

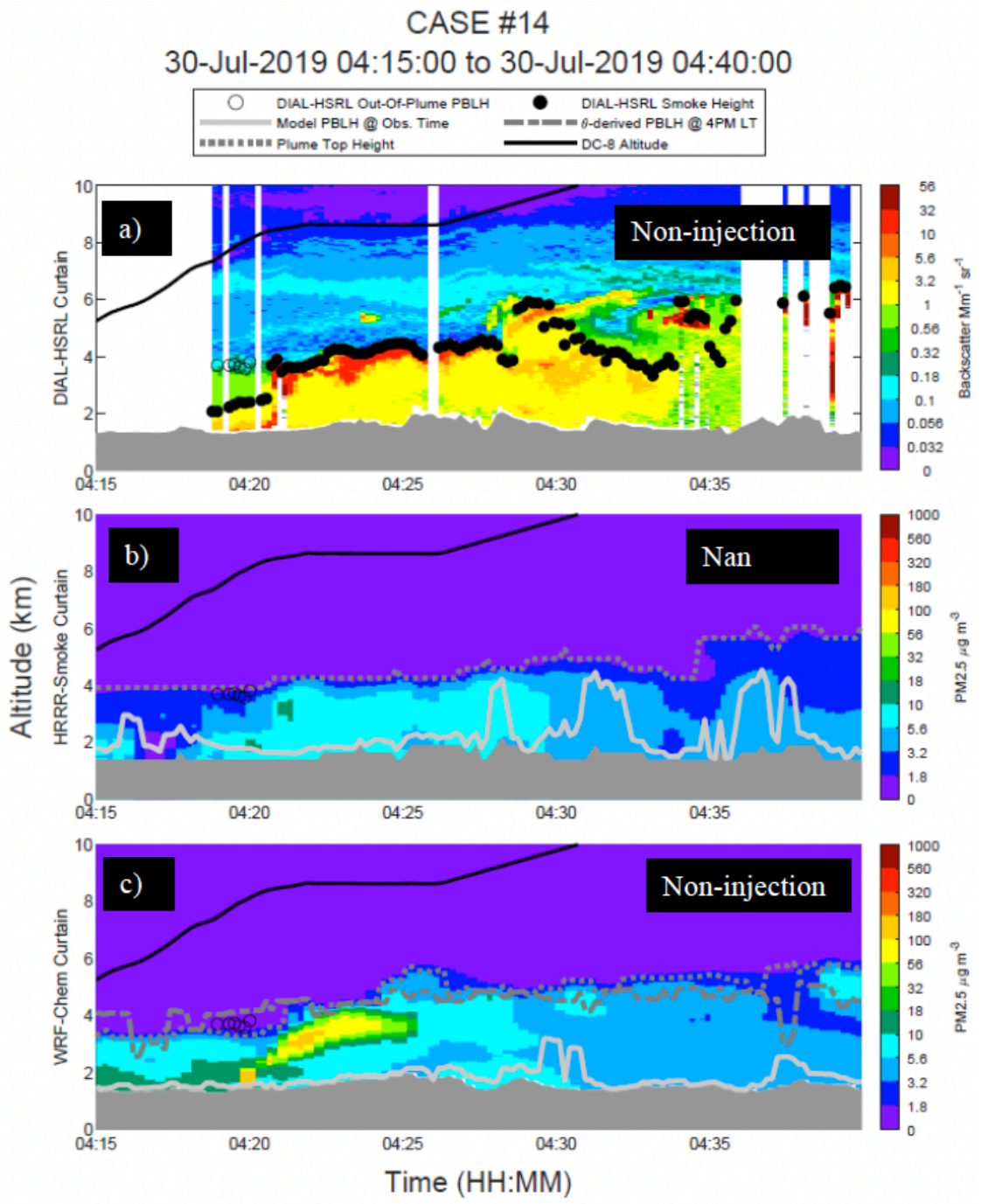

Figure A14. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Tucker Fire on 2019-07-30 04:15-04:40 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

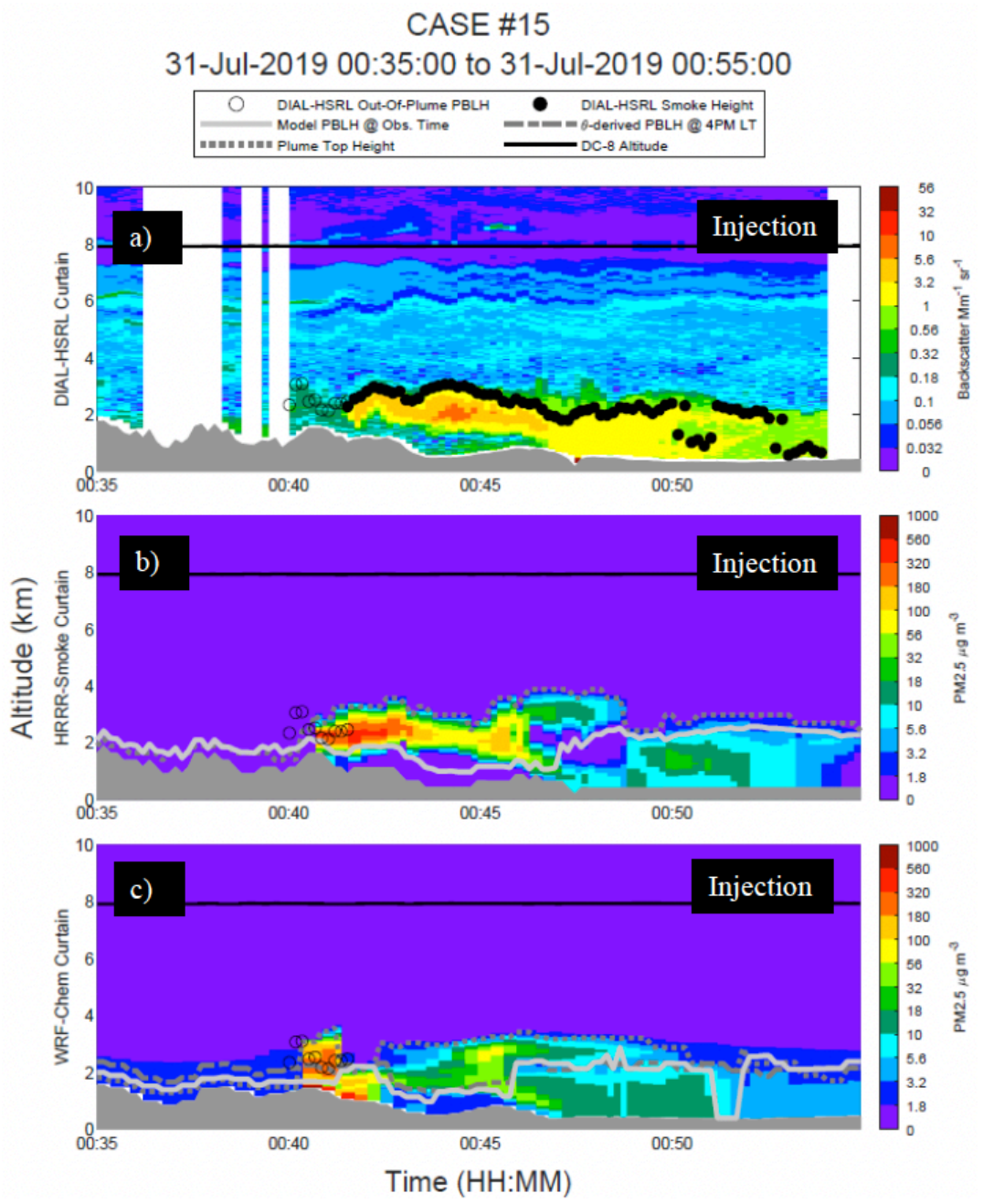

Figure A15. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Left Hand Fire on 2019-07-31 00:35-00:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

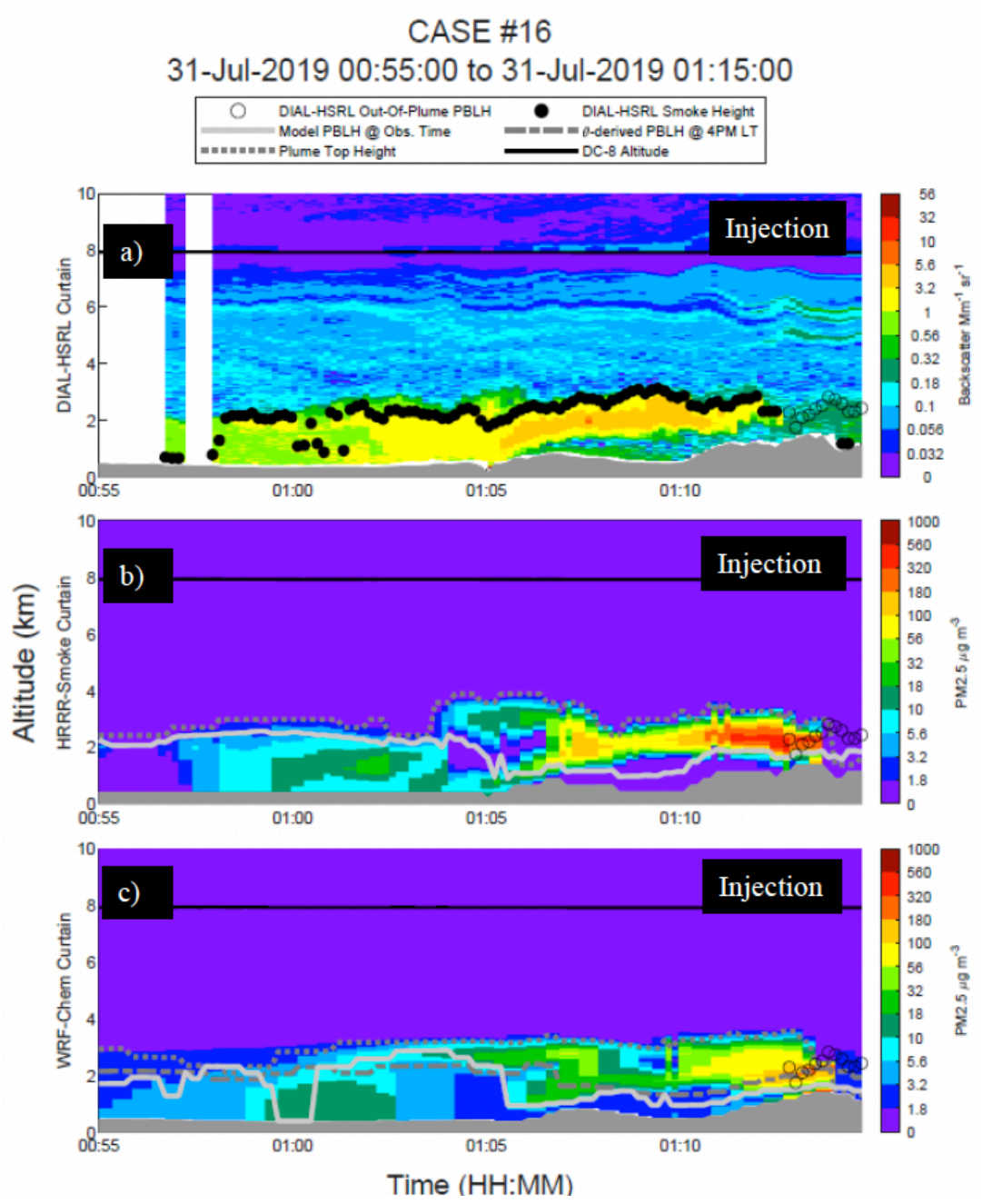

Figure A16. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Left Hand Fire on 2019-07-31 00:55-01:15 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

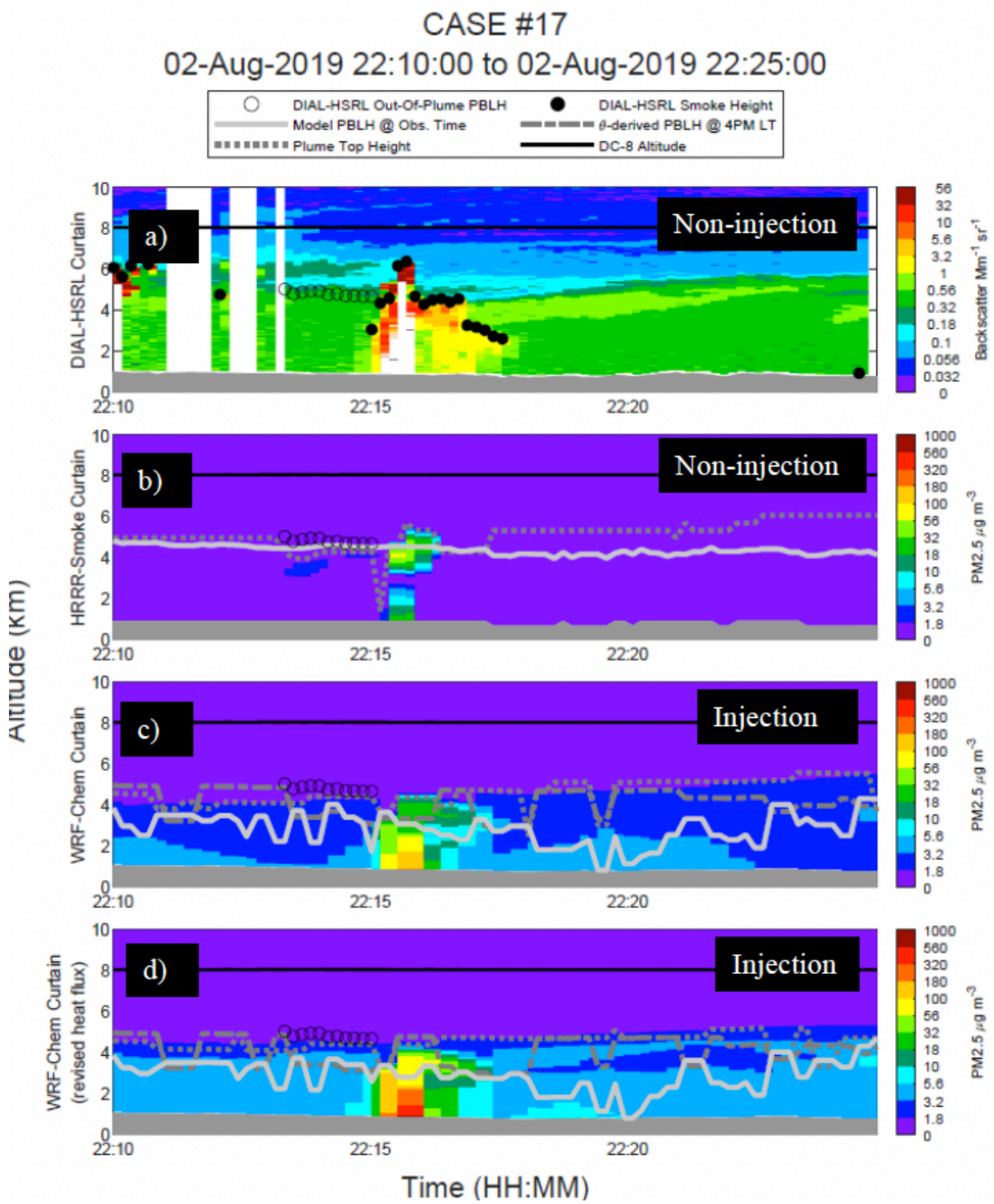

Figure A17. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Ridge Top Fire 2019-08-02 22:10-22:25 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

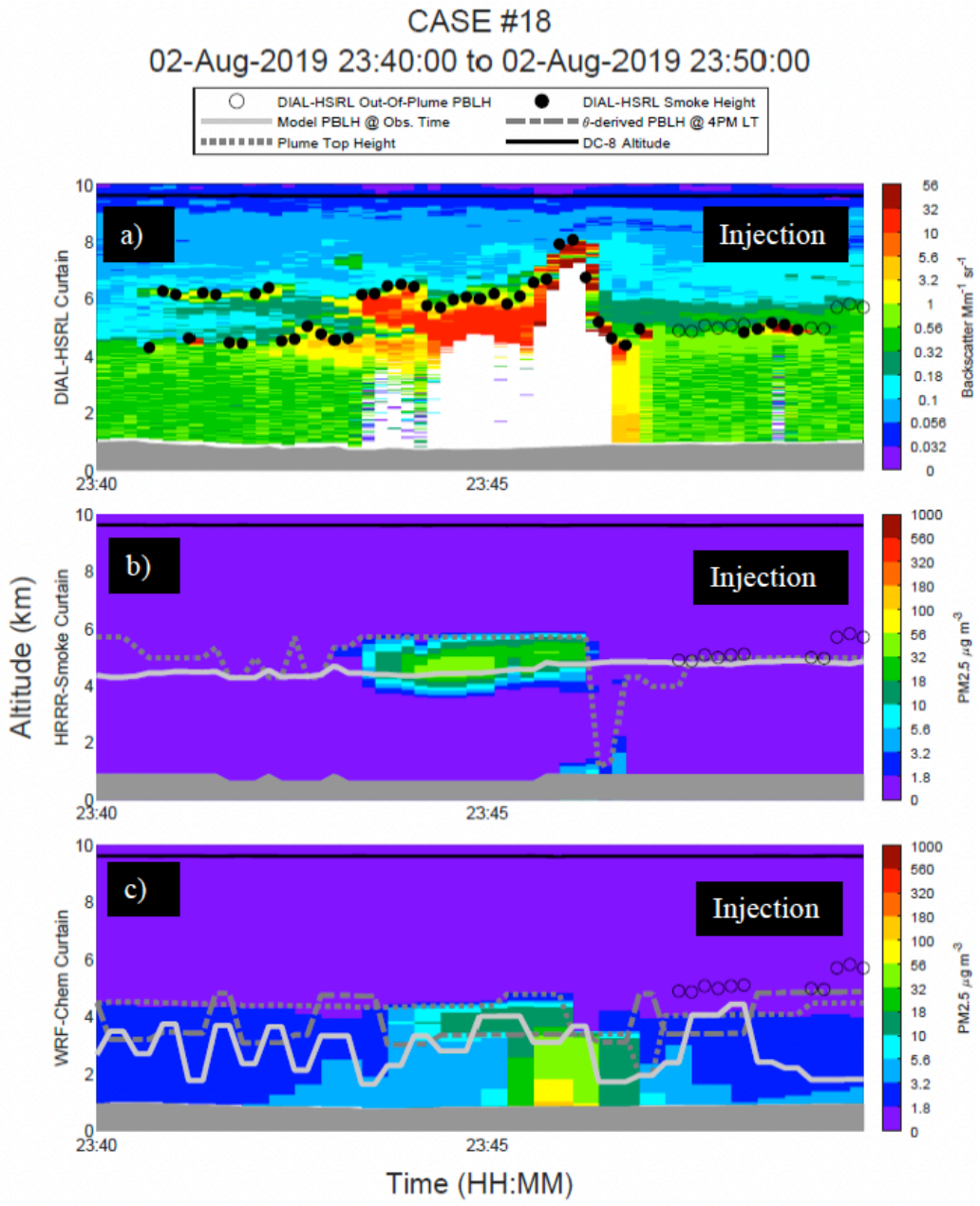

Figure A18. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Ridge Top Fire on 2019-08-02 23:40-23:50 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

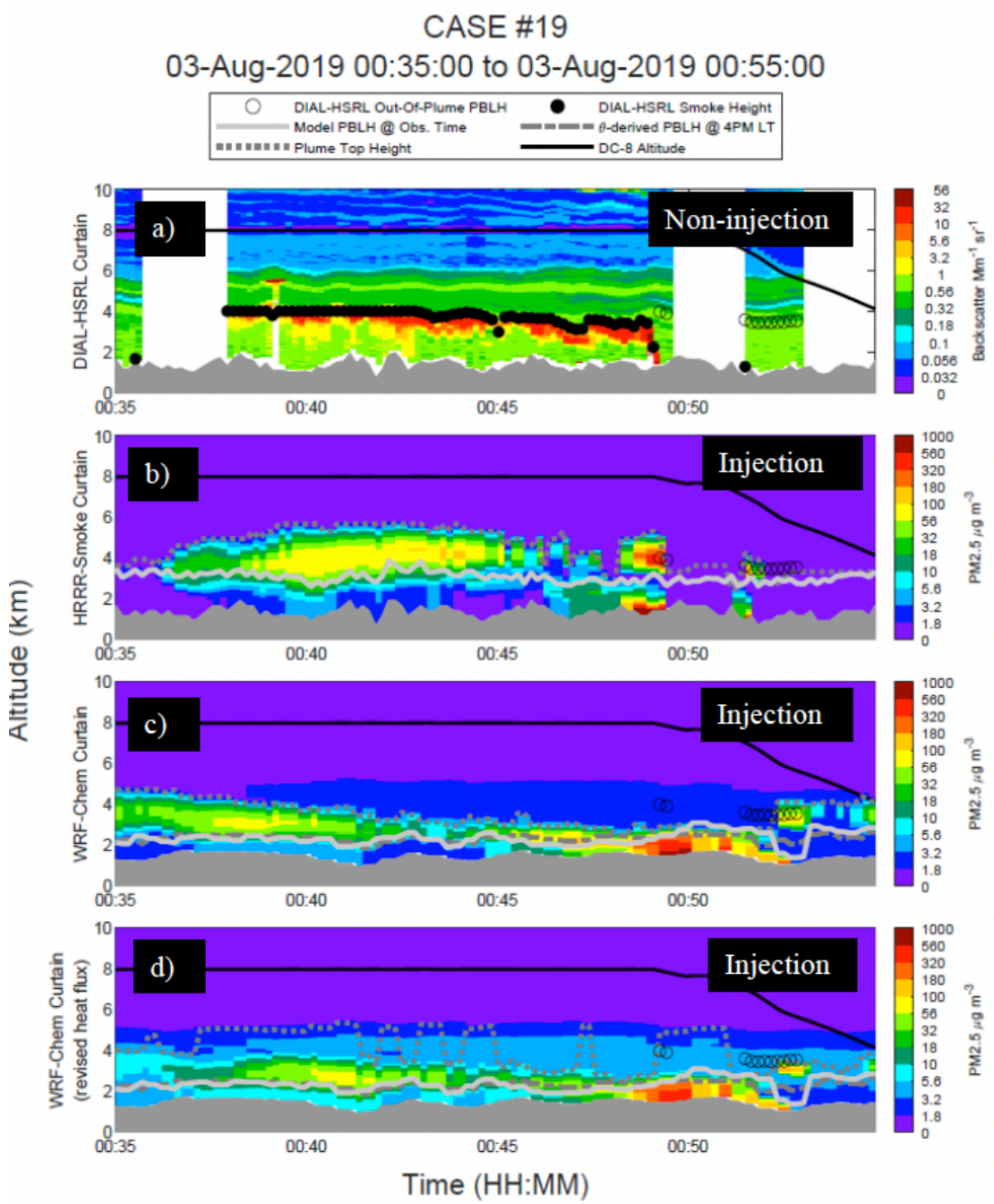

Figure A19. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Mica/Lick Creek Fire 2019-08-03 00:35-00:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

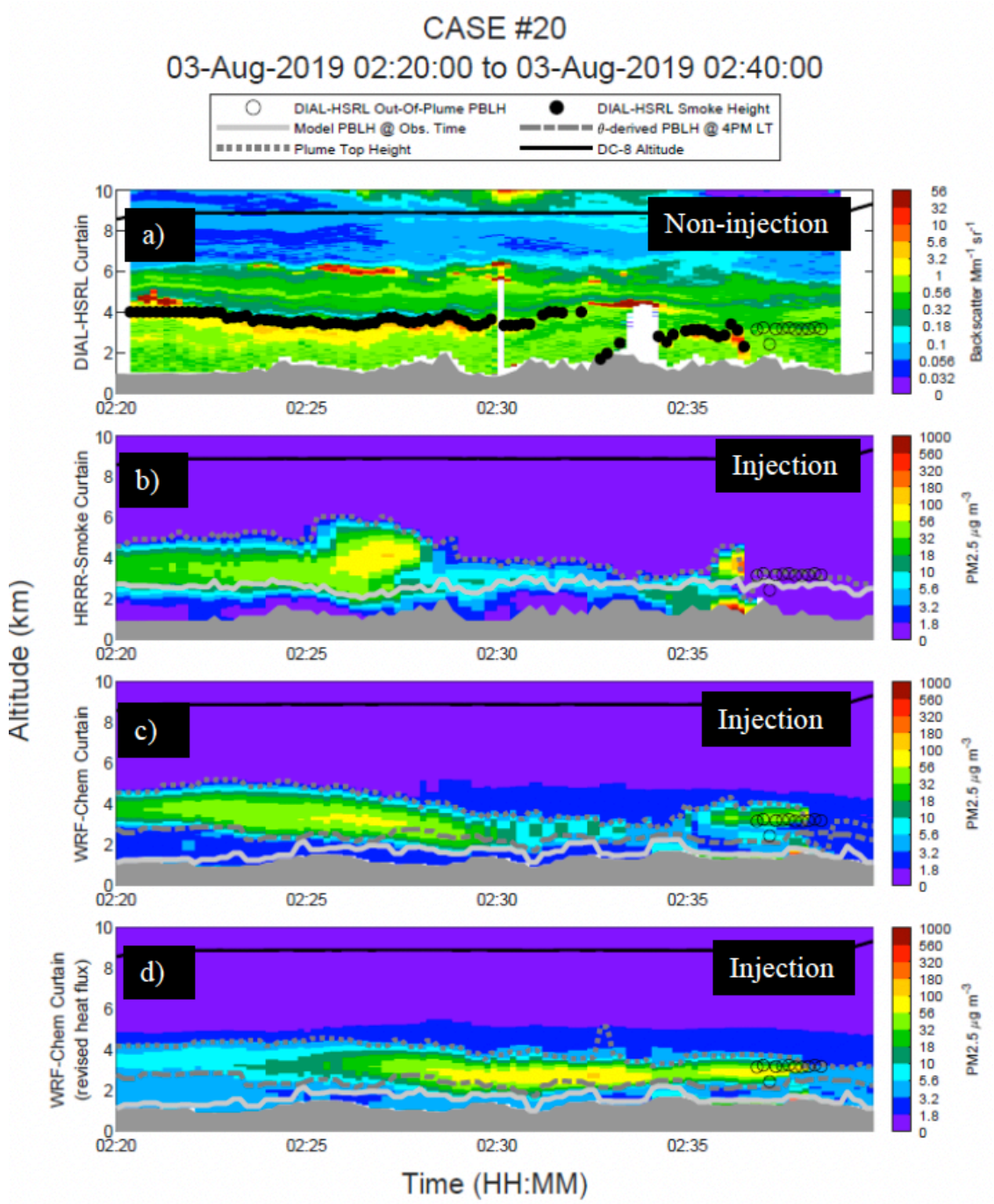

Figure A20. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Mica/Lick Creek Fire 2019-08-03 02:20-02:40 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

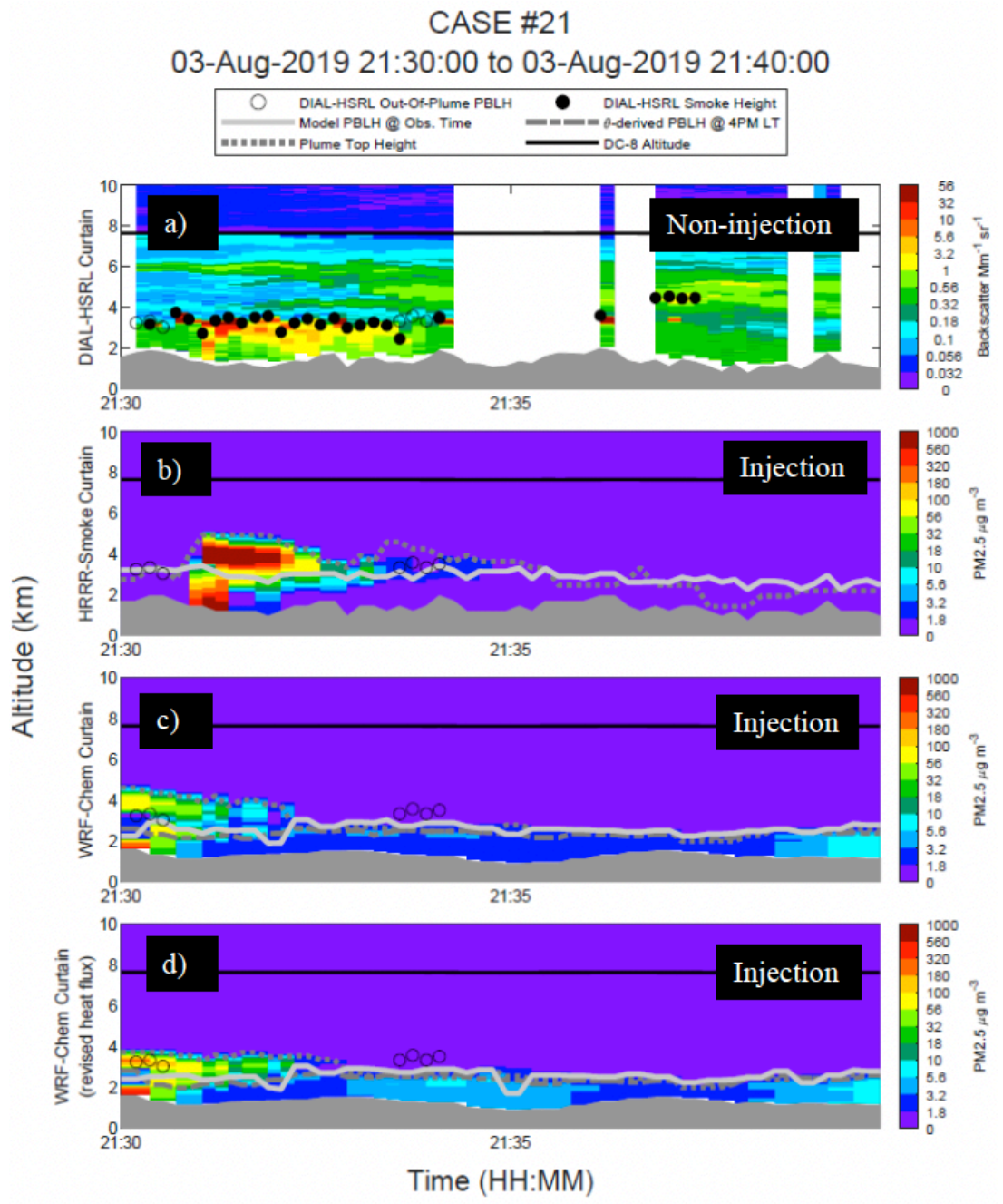

Figure A21. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Mica Creek Fire 2019-08-03 21:30-21:40 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

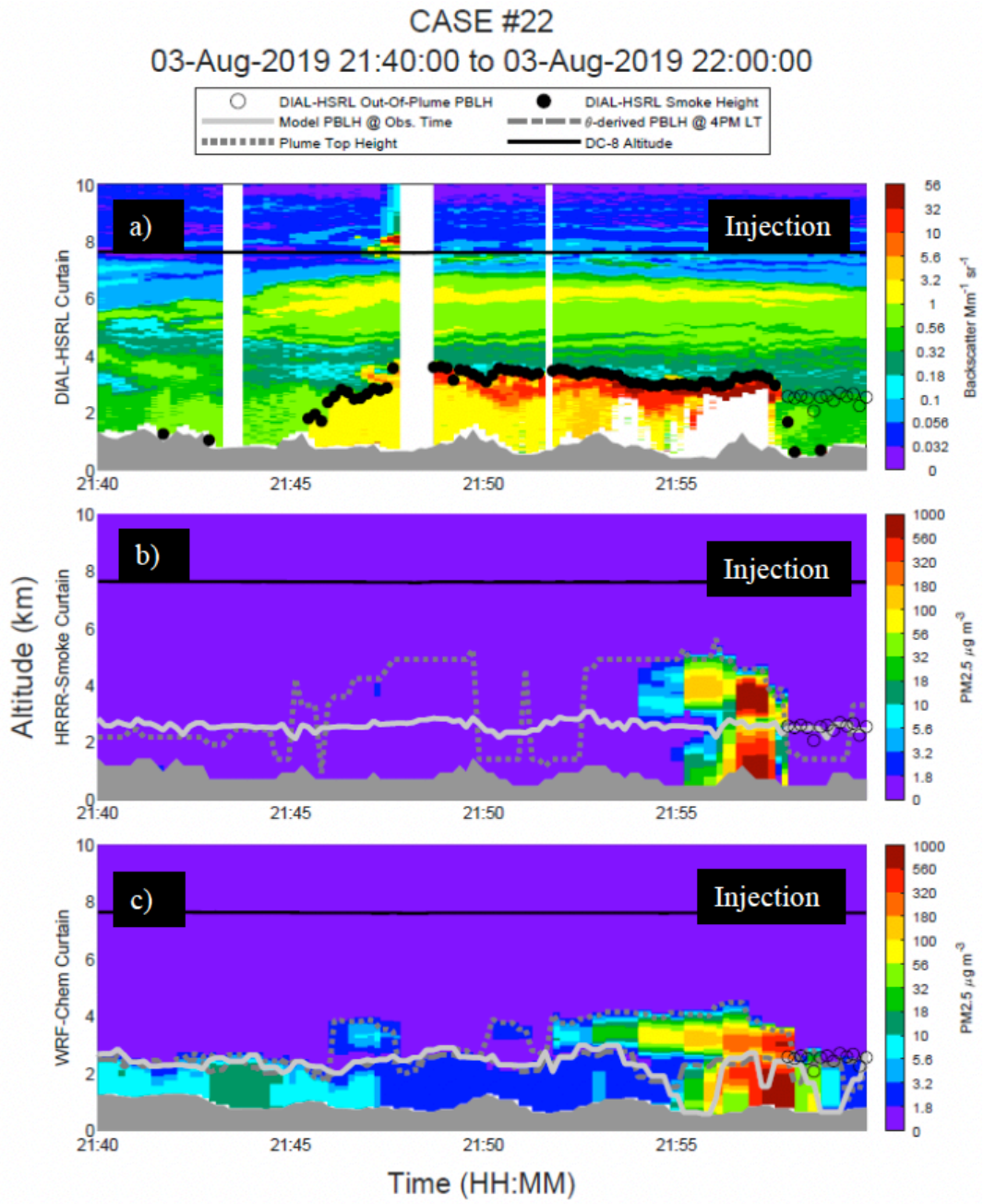

Figure A22. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-03 21:40-22:00 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

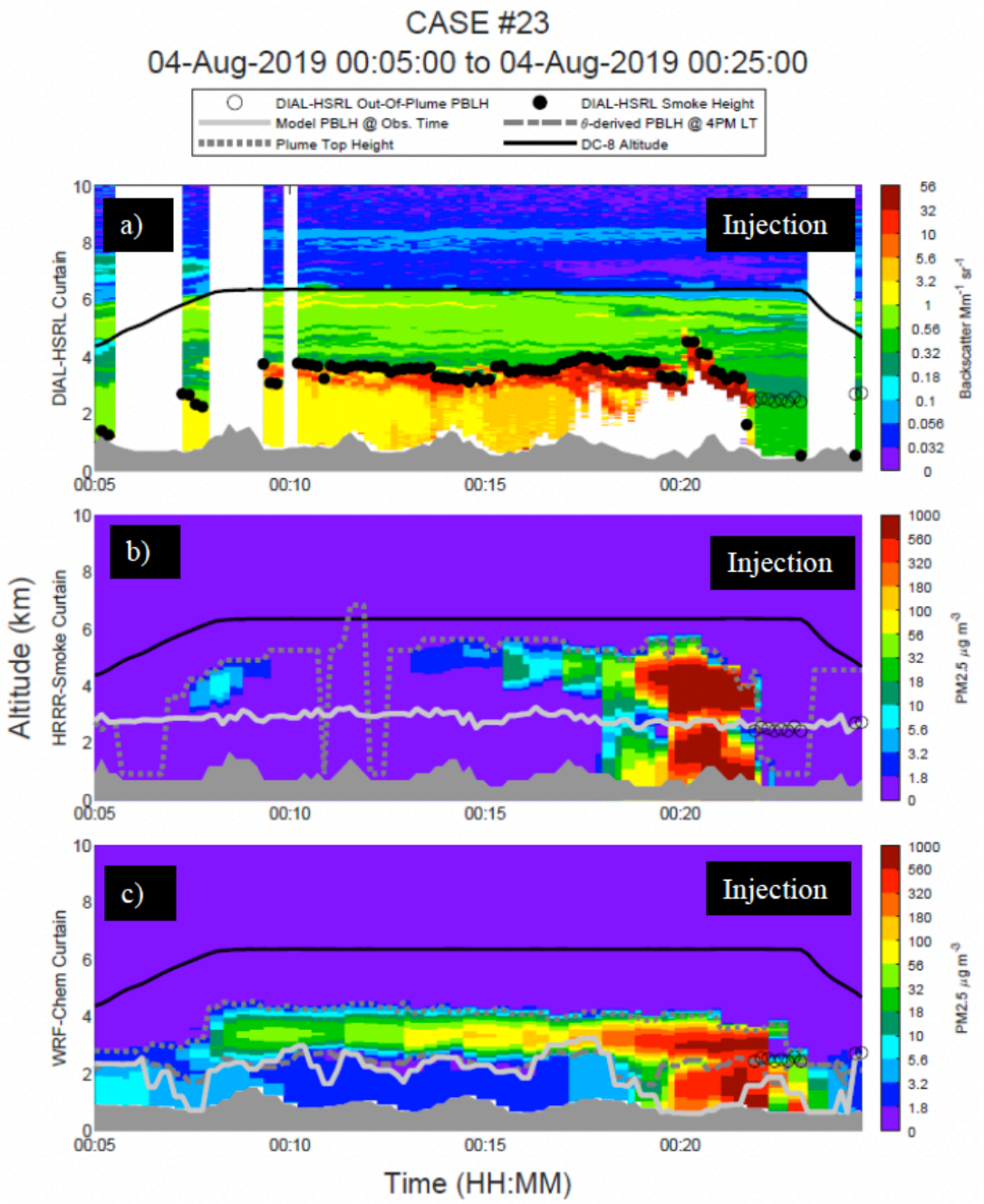

Figure A23. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-04 00:05-00:25 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

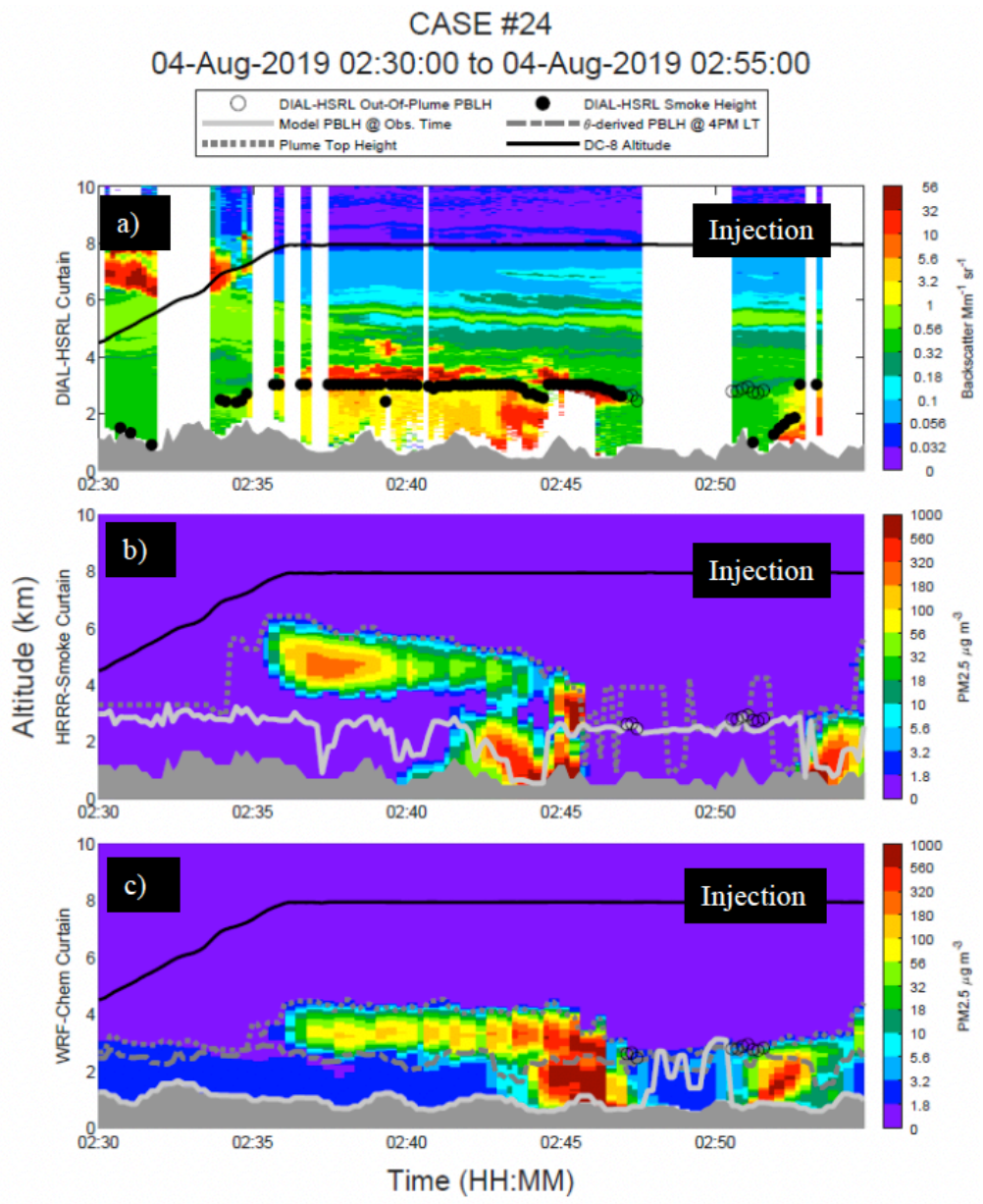

Figure A24. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-04 02:30-02:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

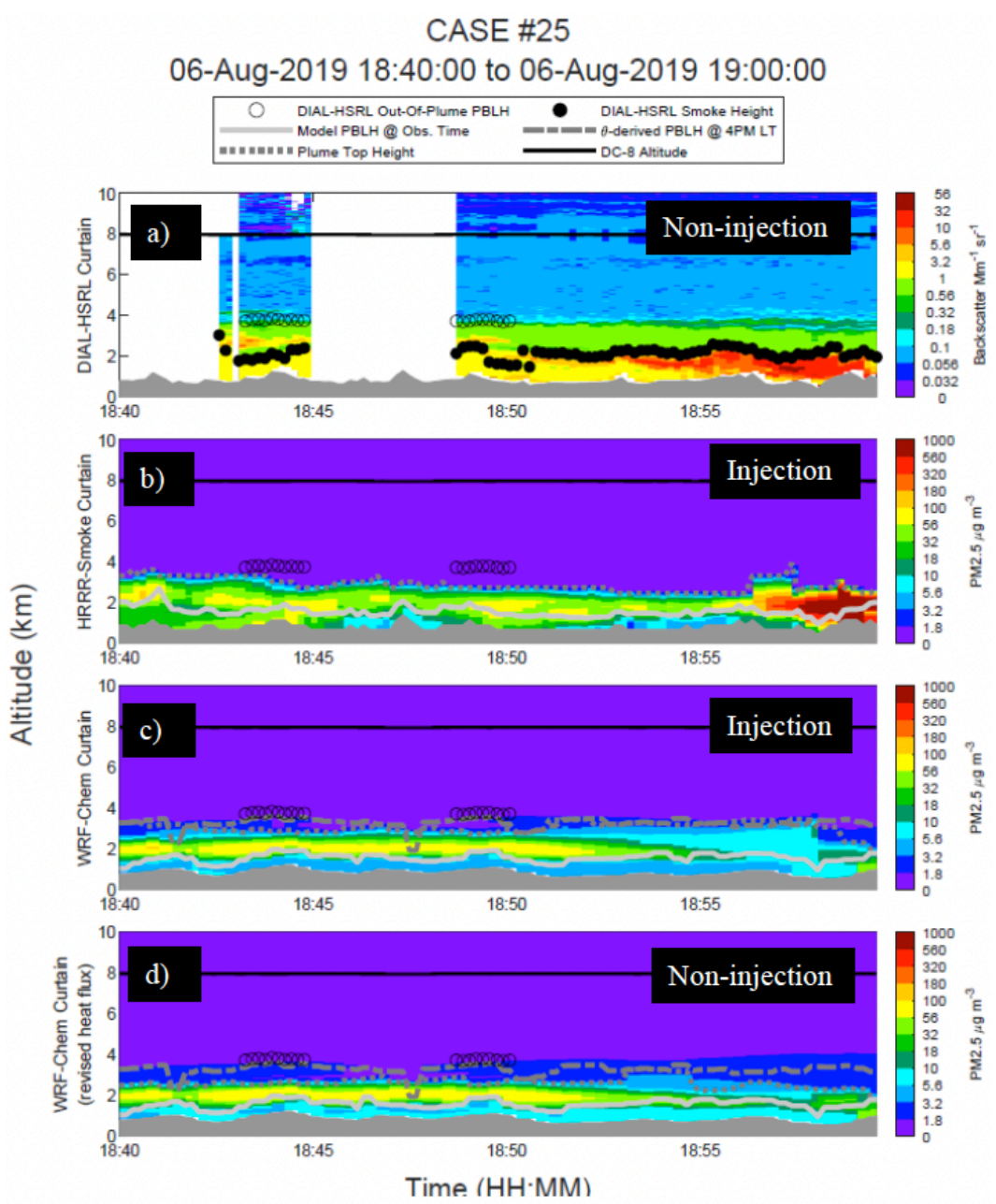

Figure A25. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Williams Flats Fire 2019-08-06 18:40-19:00 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

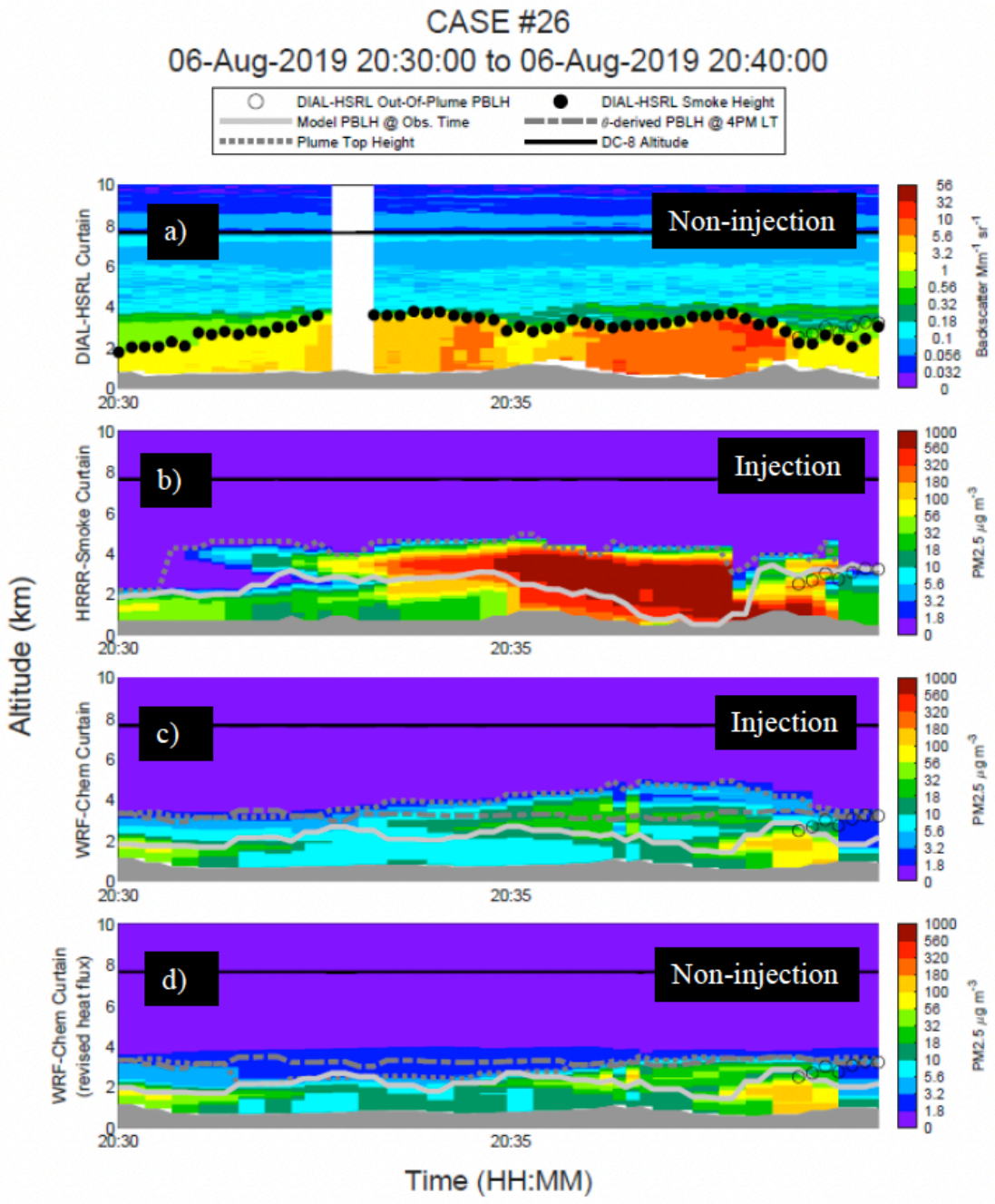

Figure A26. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Williams Flats Fire 2019-08-06 20:30-20:40 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

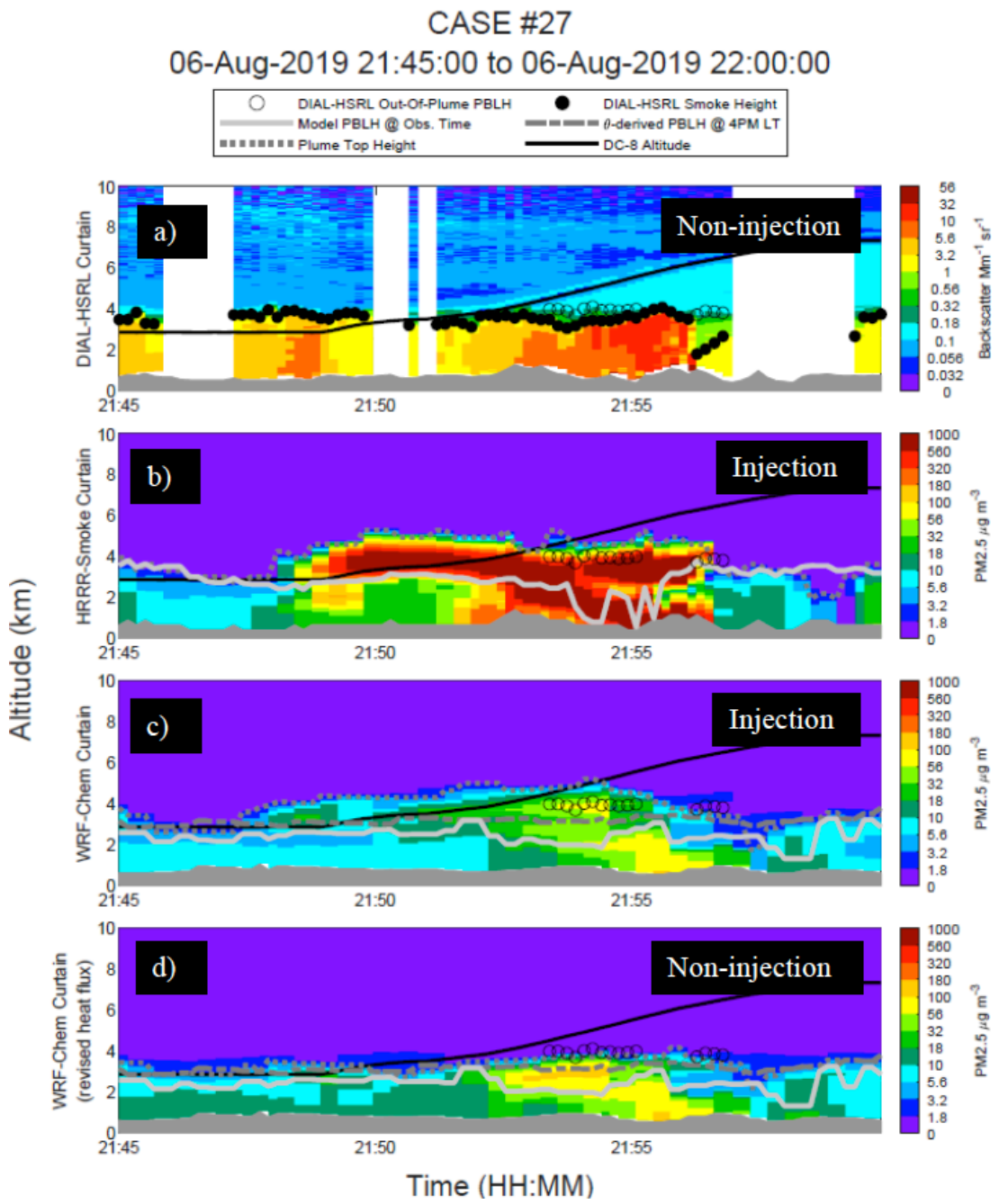

Figure A27. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Williams Flats Fire 2019-08-06 21:45-22:00 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

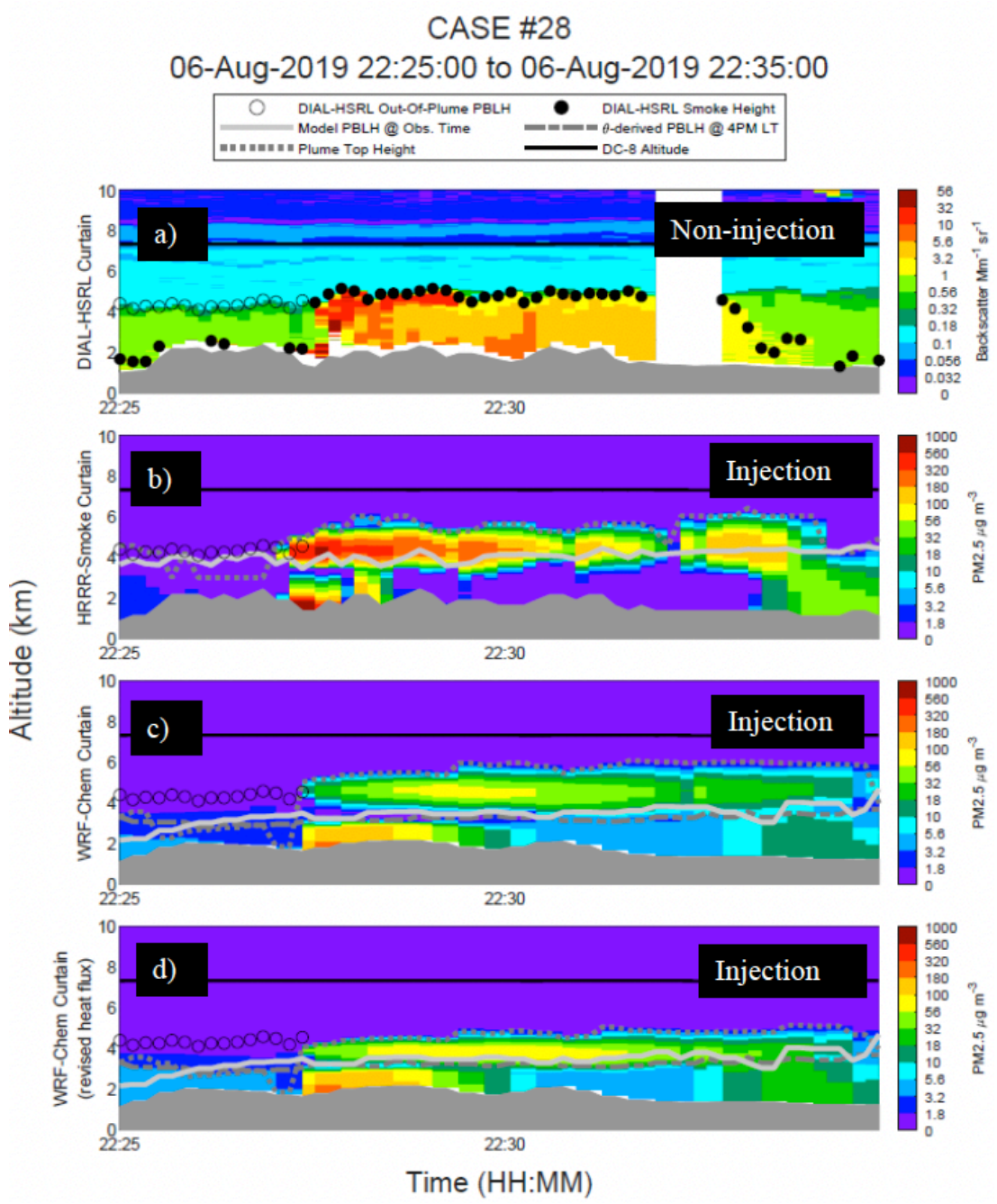

Figure A28. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Snow Creek Fire 2019-08-06 22:25-22:35 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

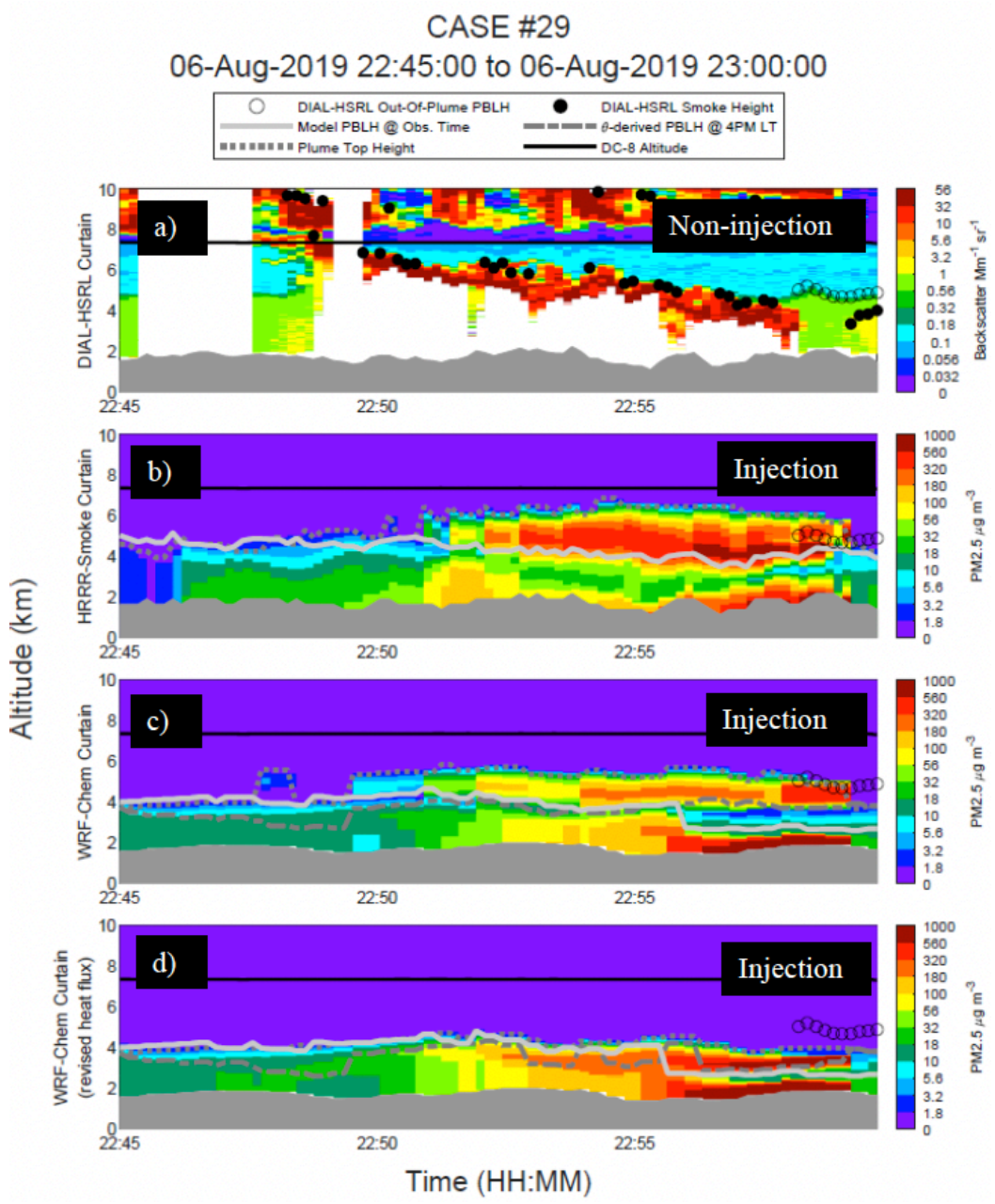

Figure A29. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Horsefly Fire 2019-08-06 22:45-23:00 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

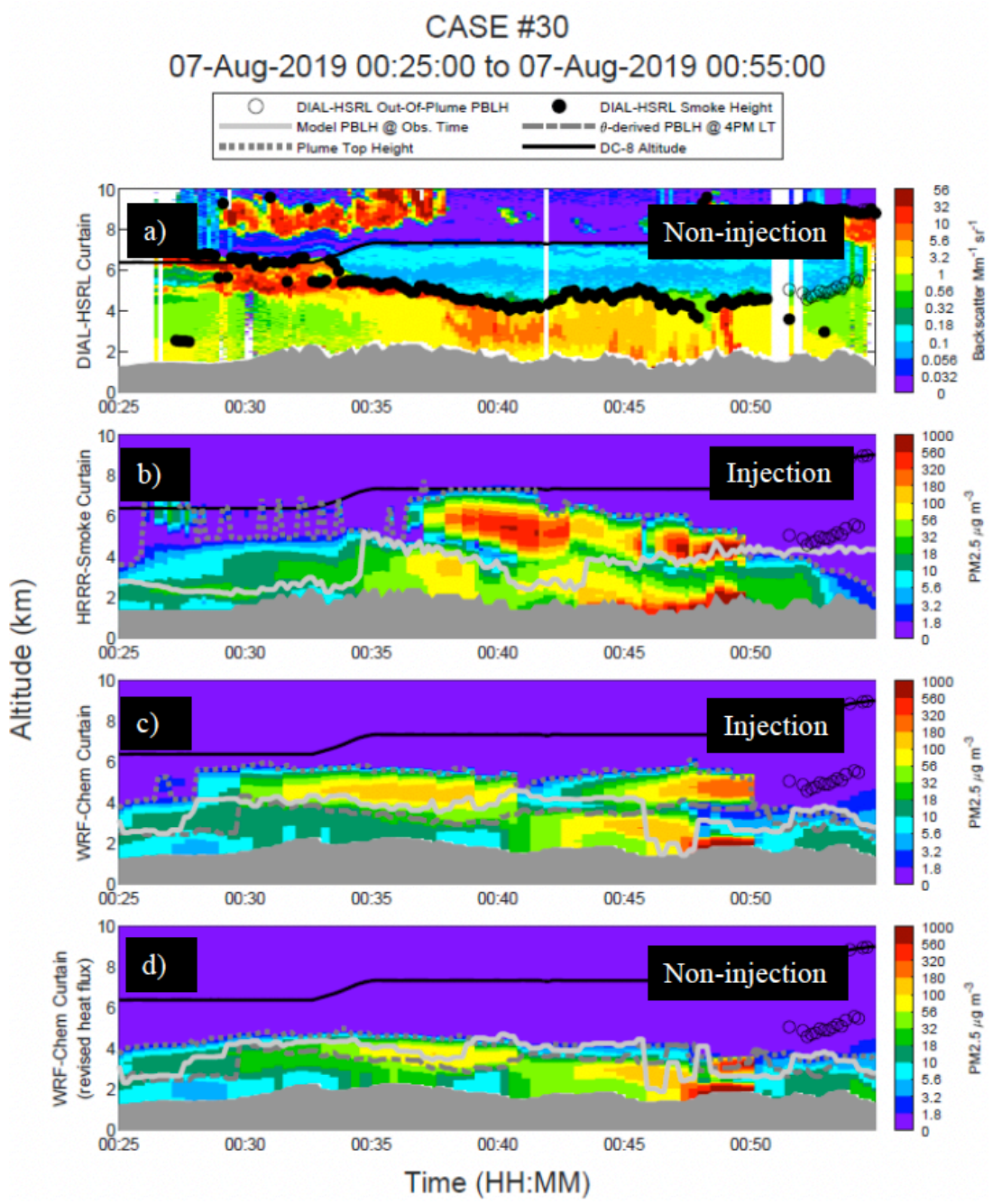

Figure A30. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Horsefly Fire 2019-08-07 00:25-00:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

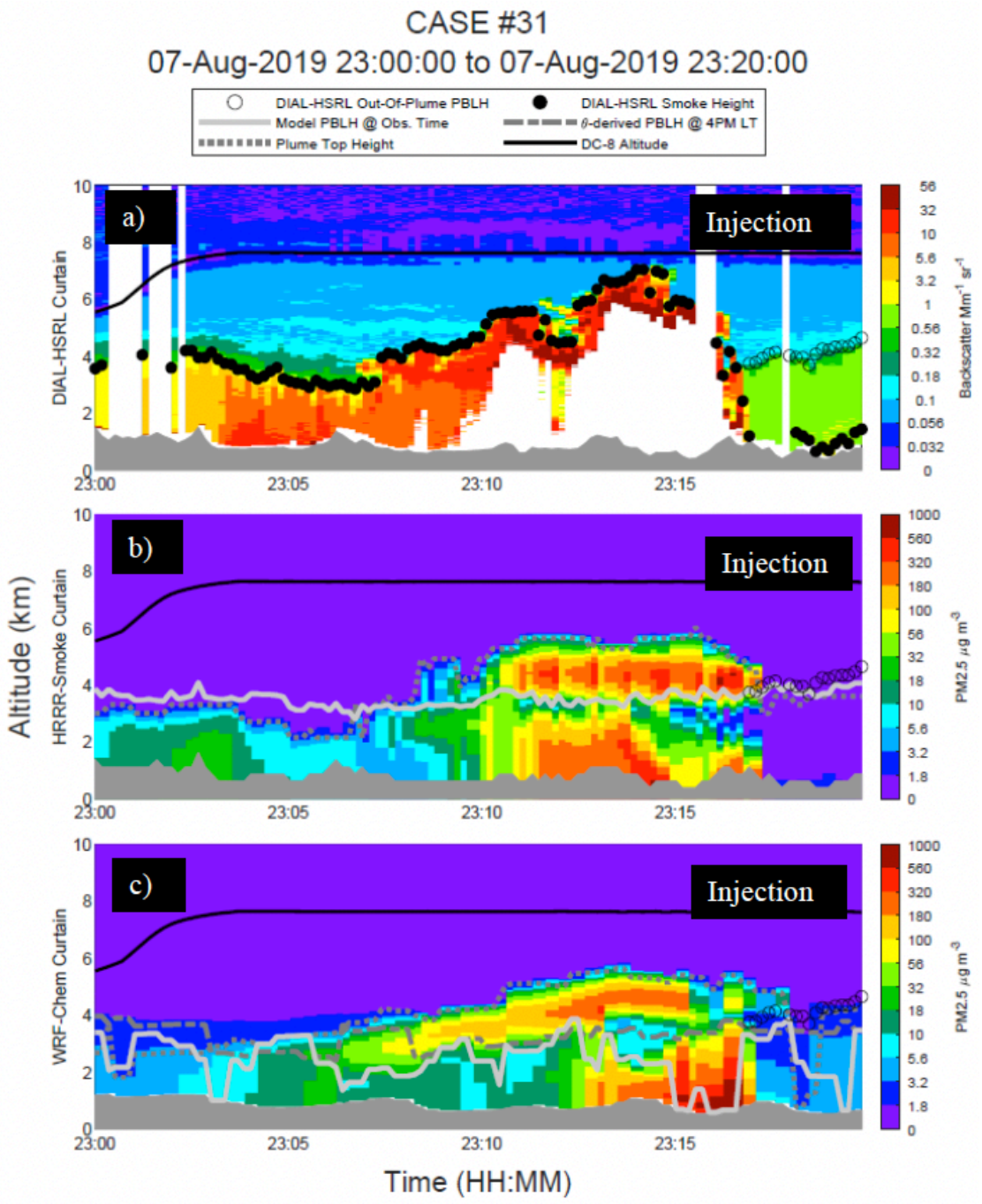

Figure A31. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheep Fire on 2019-08-07 23:00-23:20 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

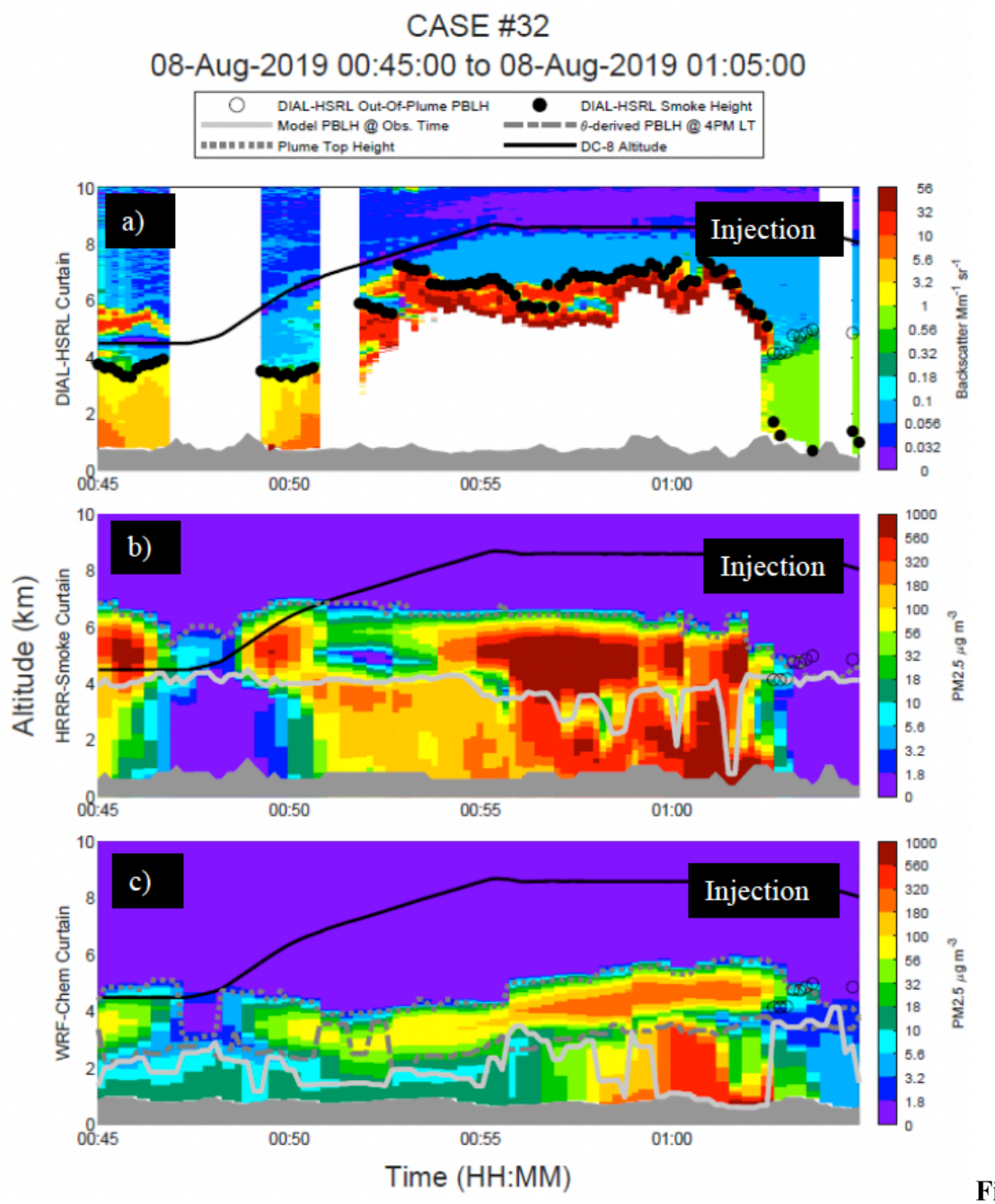

Figure A32. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-08 00:45-01:05 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

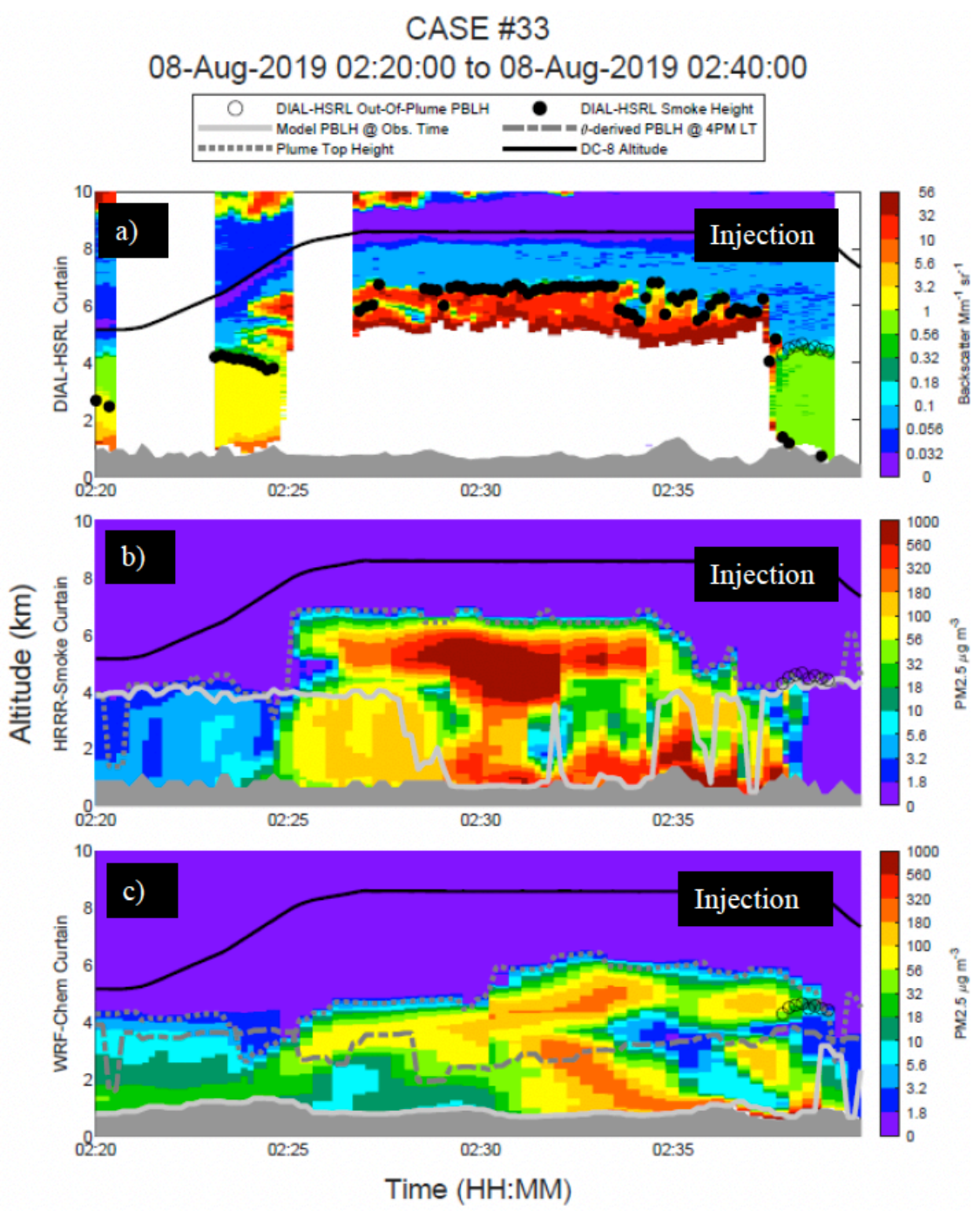

Figure A33. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-08 02:20-02:40 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

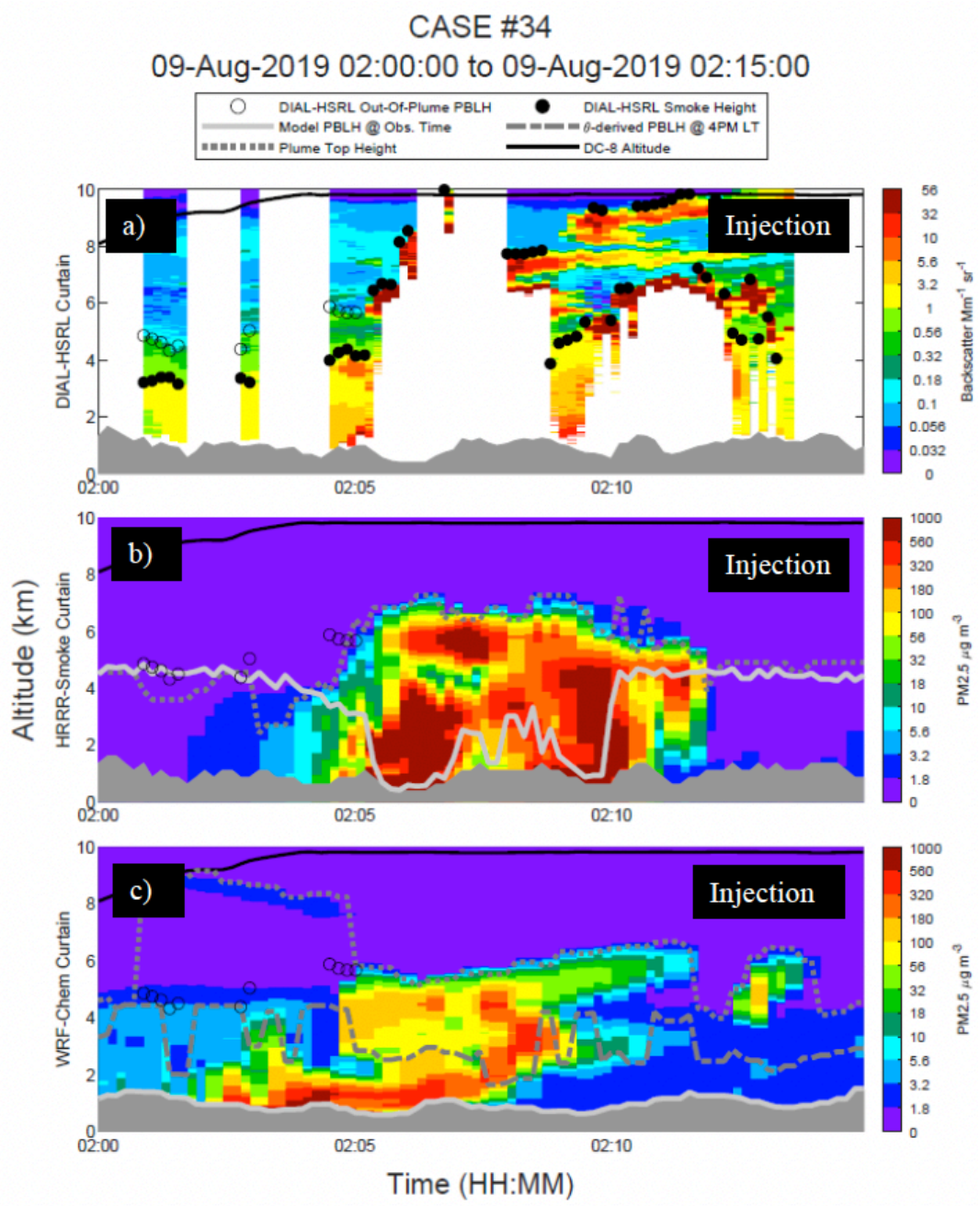

Figure A34. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-09 02:00-02:15 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

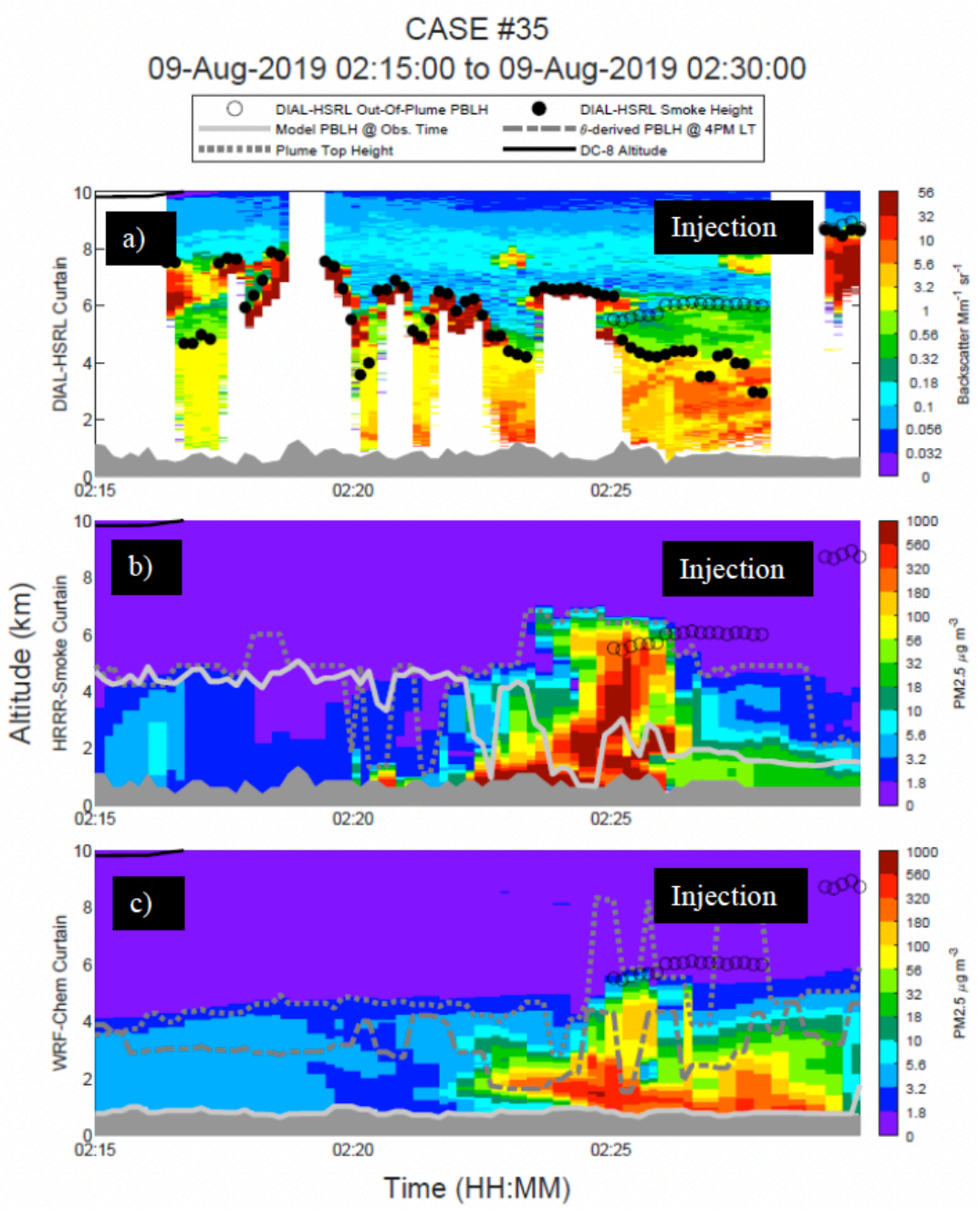

Figure A35. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Williams Flats Fire on 2019-08-09 02:15-02:30 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

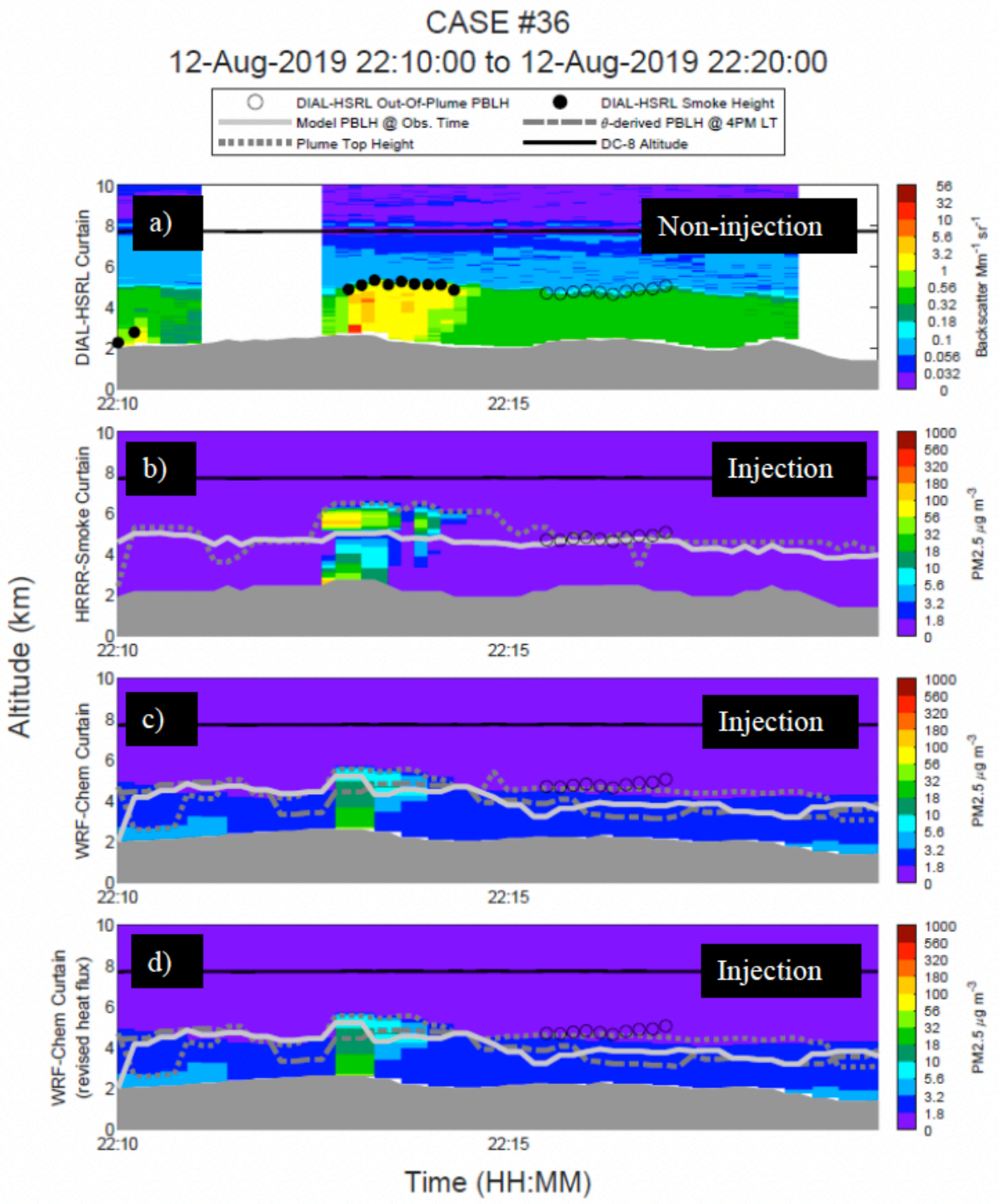

Figure A36. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Springs Fire 2019-08-12 22:10-22:20 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the

modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

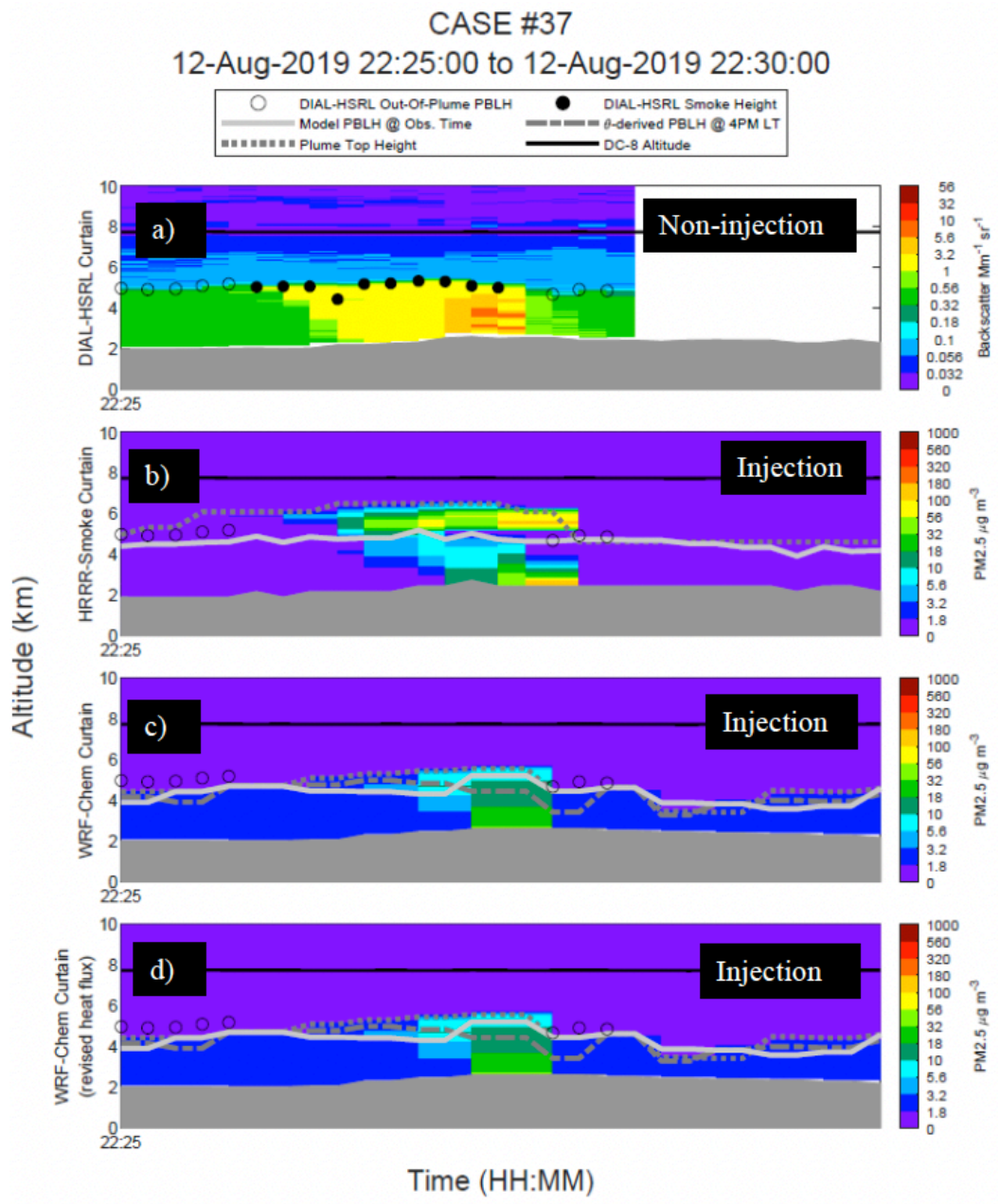

Figure A37. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Springs Fire 2019-08-12 22:25-22:30 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the

modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

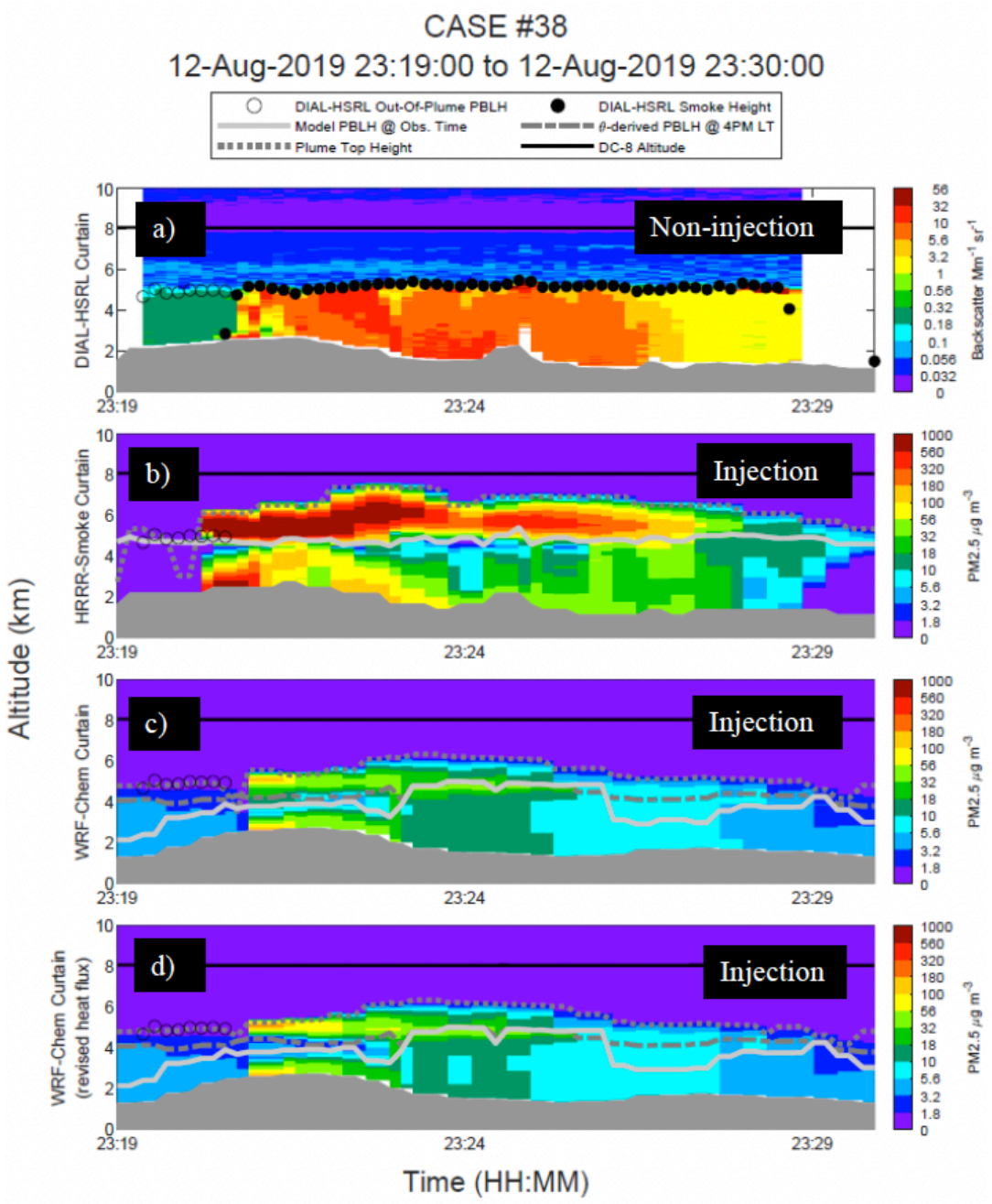

Figure A38. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Castle Fire 2019-08-12 23:19-23:30 UTC. In panel a, open circles show

the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

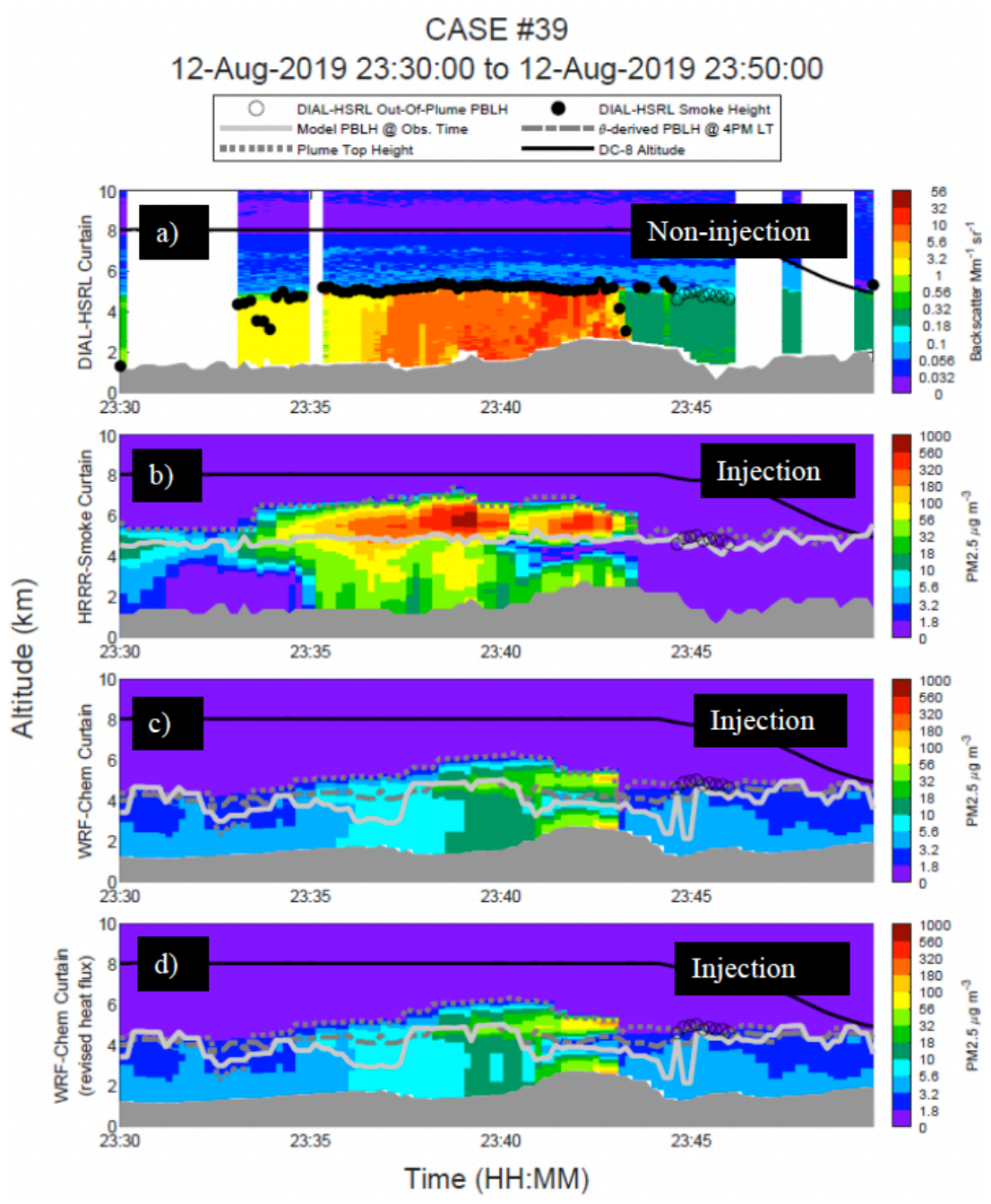

Figure A39. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted

heat fluxes (d) for the Castle Fire 2019-08-12 23:30-23:50 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

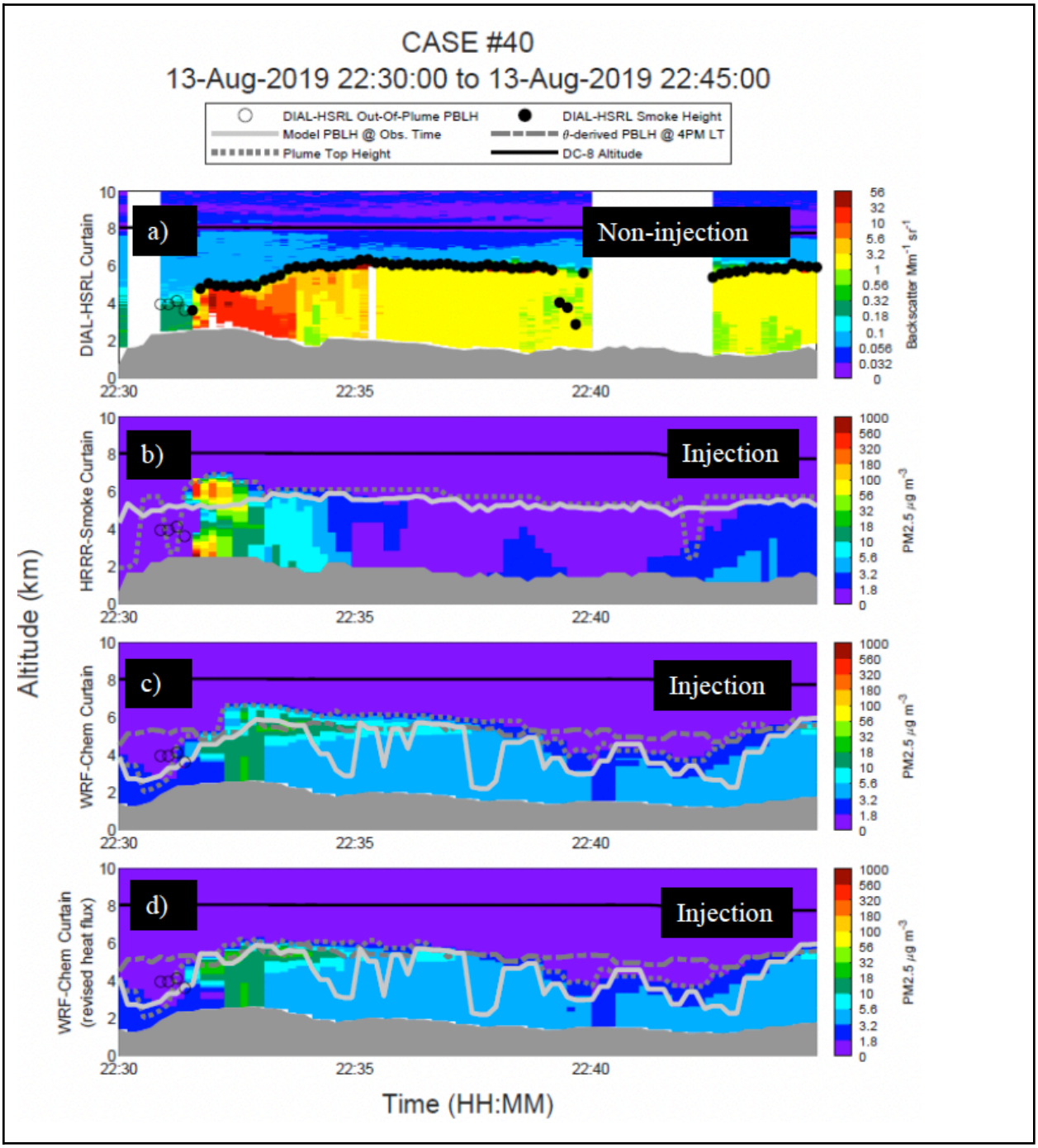

Figure A40. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Castle Fire 2019-08-13 22:30-22:45 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

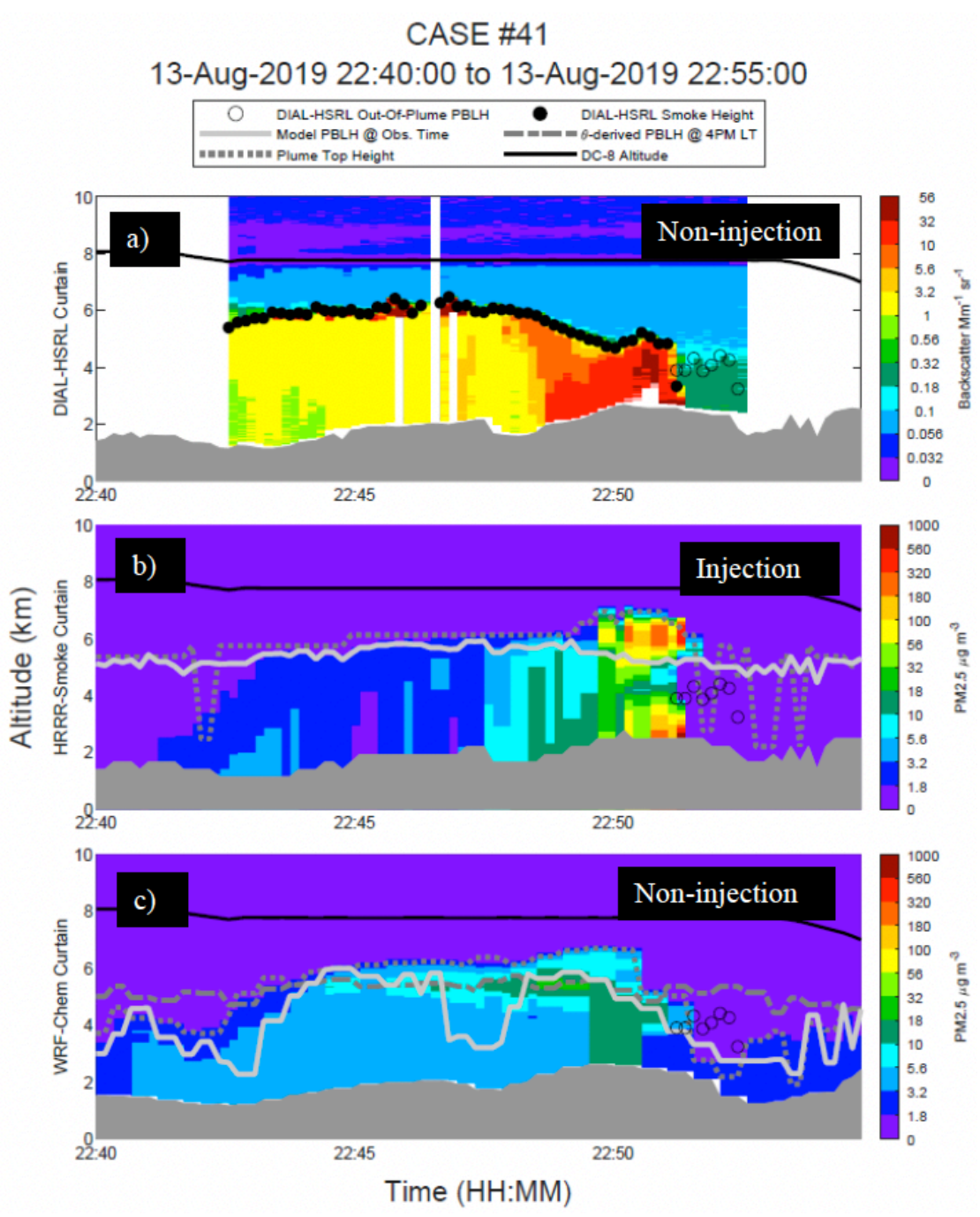

Figure A41. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Castle Fire on 2019-08-13 22:40-22:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

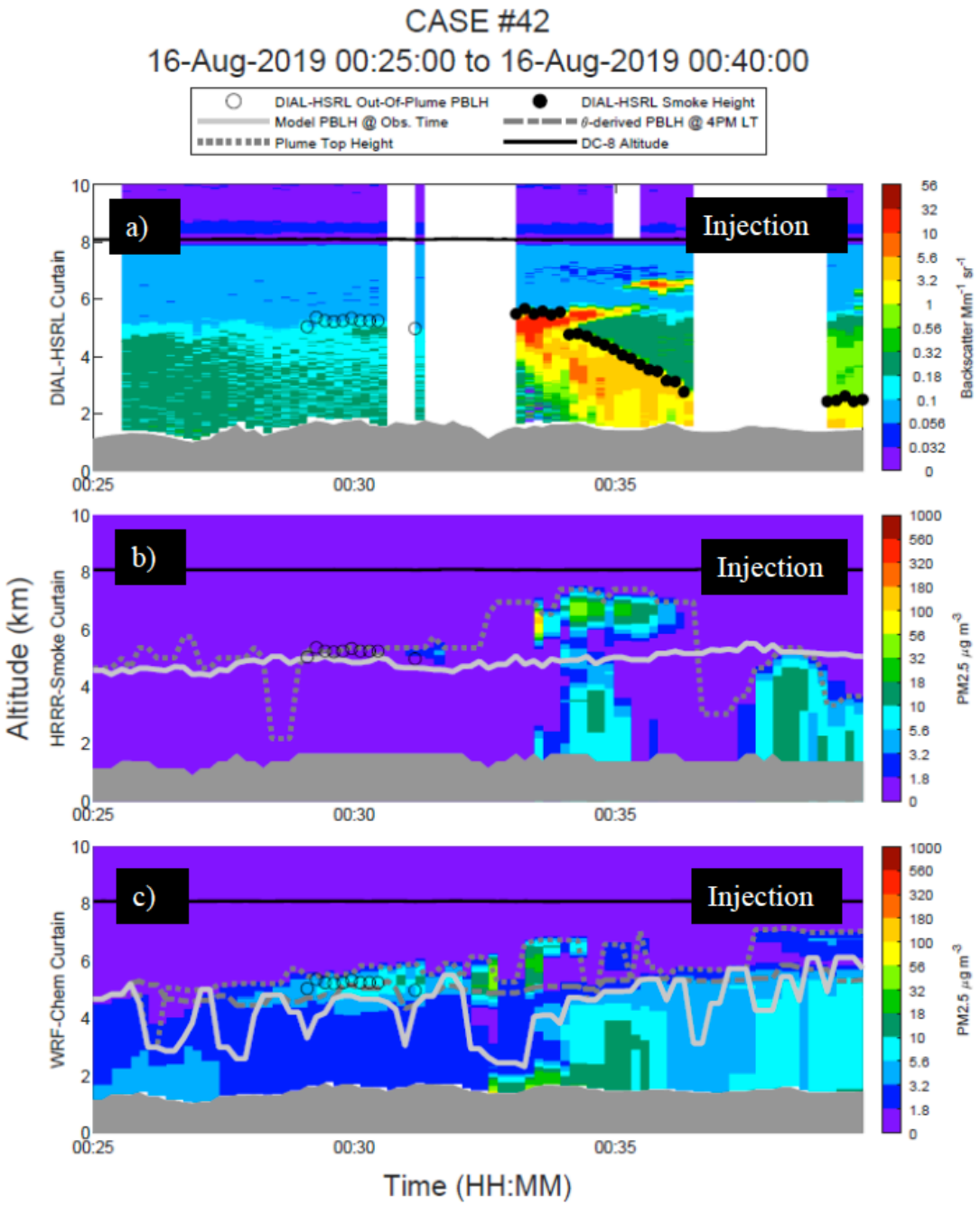

Figure A42. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheridan Fire on 2019-08-16 00:25-00:40 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

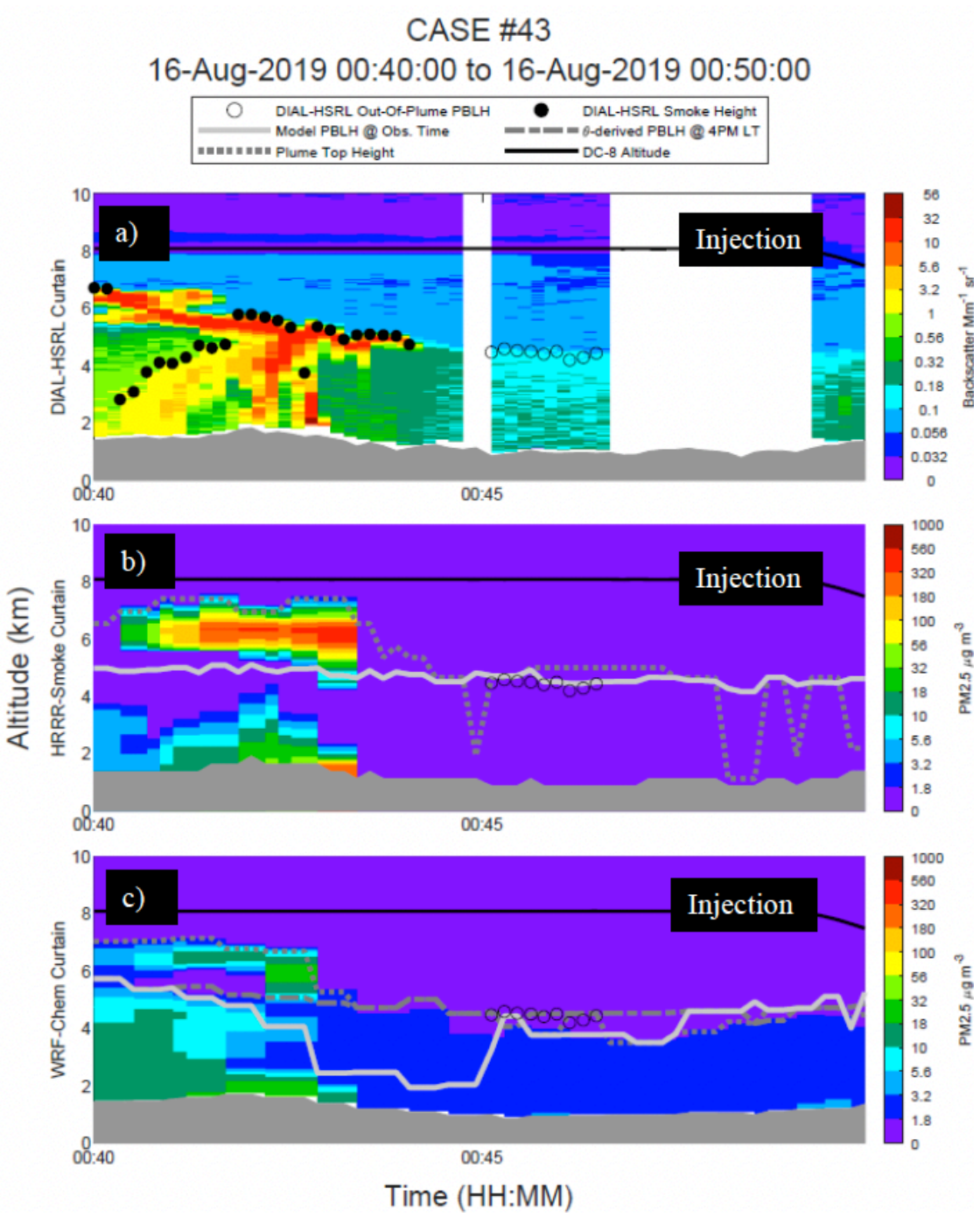

Figure A43. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheridan Fire on 2019-08-16 00:40-00:50 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

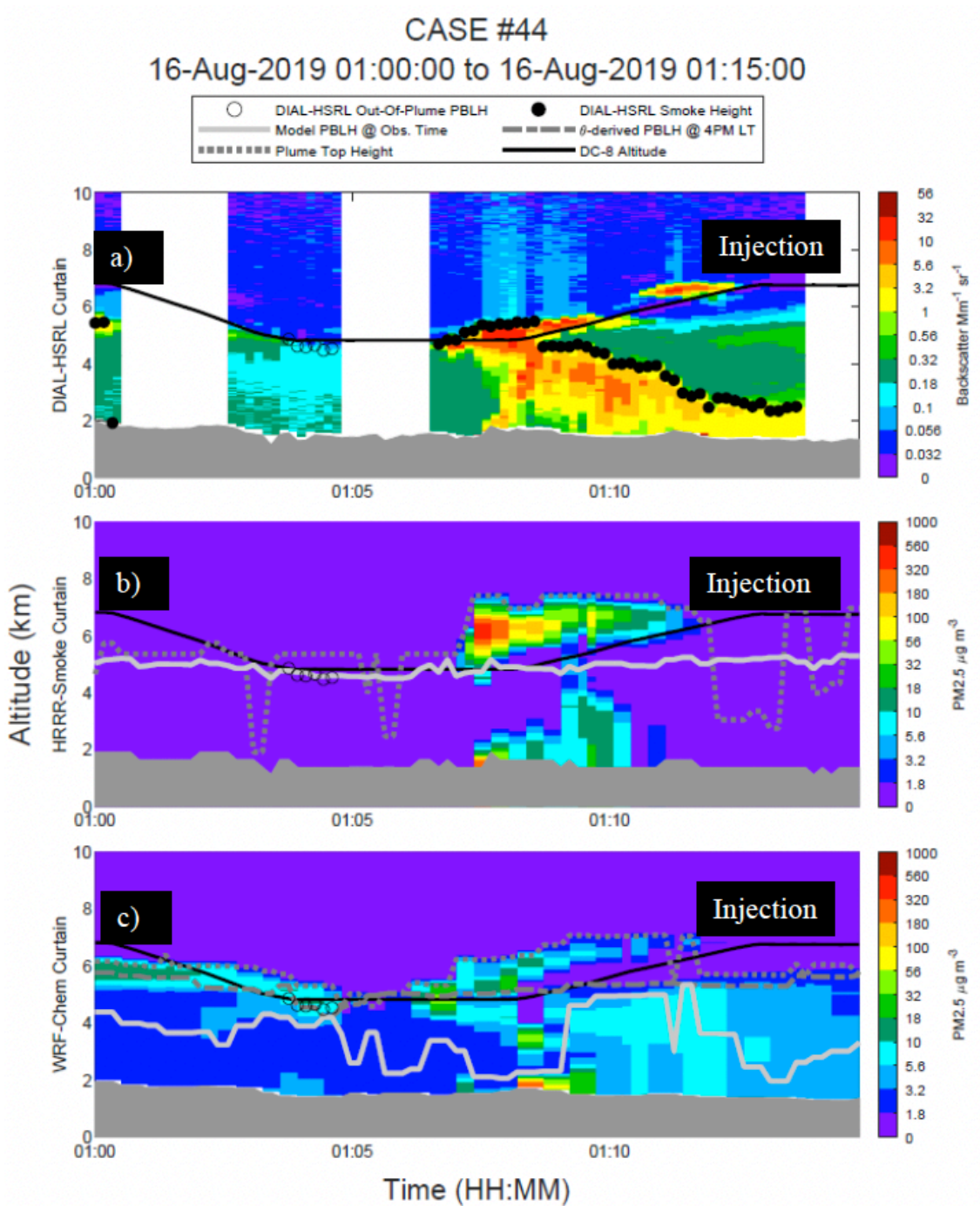

Figure A44. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheridan Fire on 2019-08-16 01:00-01:15 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

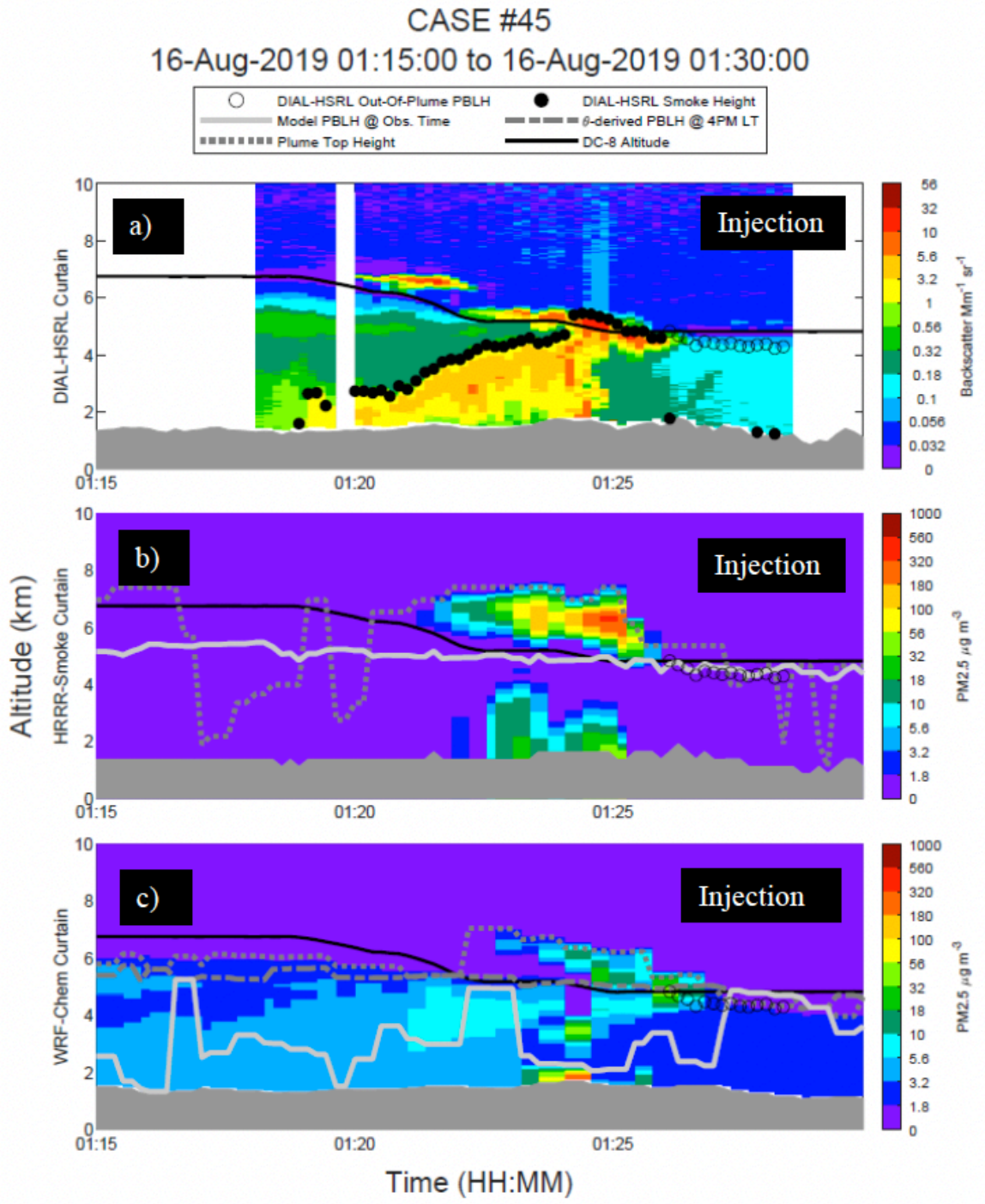

Figure A45. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b) and WRF-Chem (c) for the Sheridan Fire on 2019-08-16 01:15-01:30 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

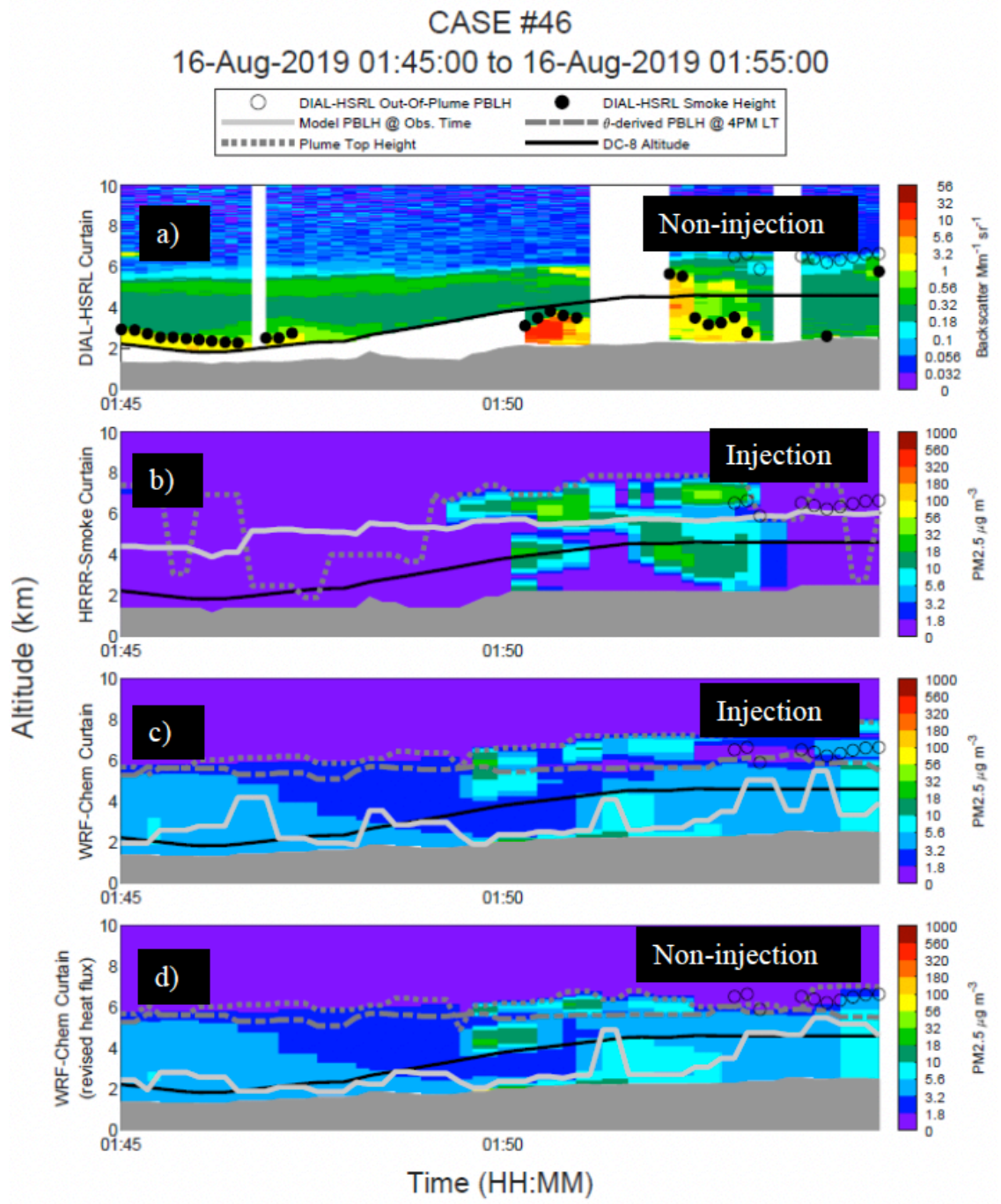

Figure A46. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Saber Fire 2019-08-16 01:45-01:55 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

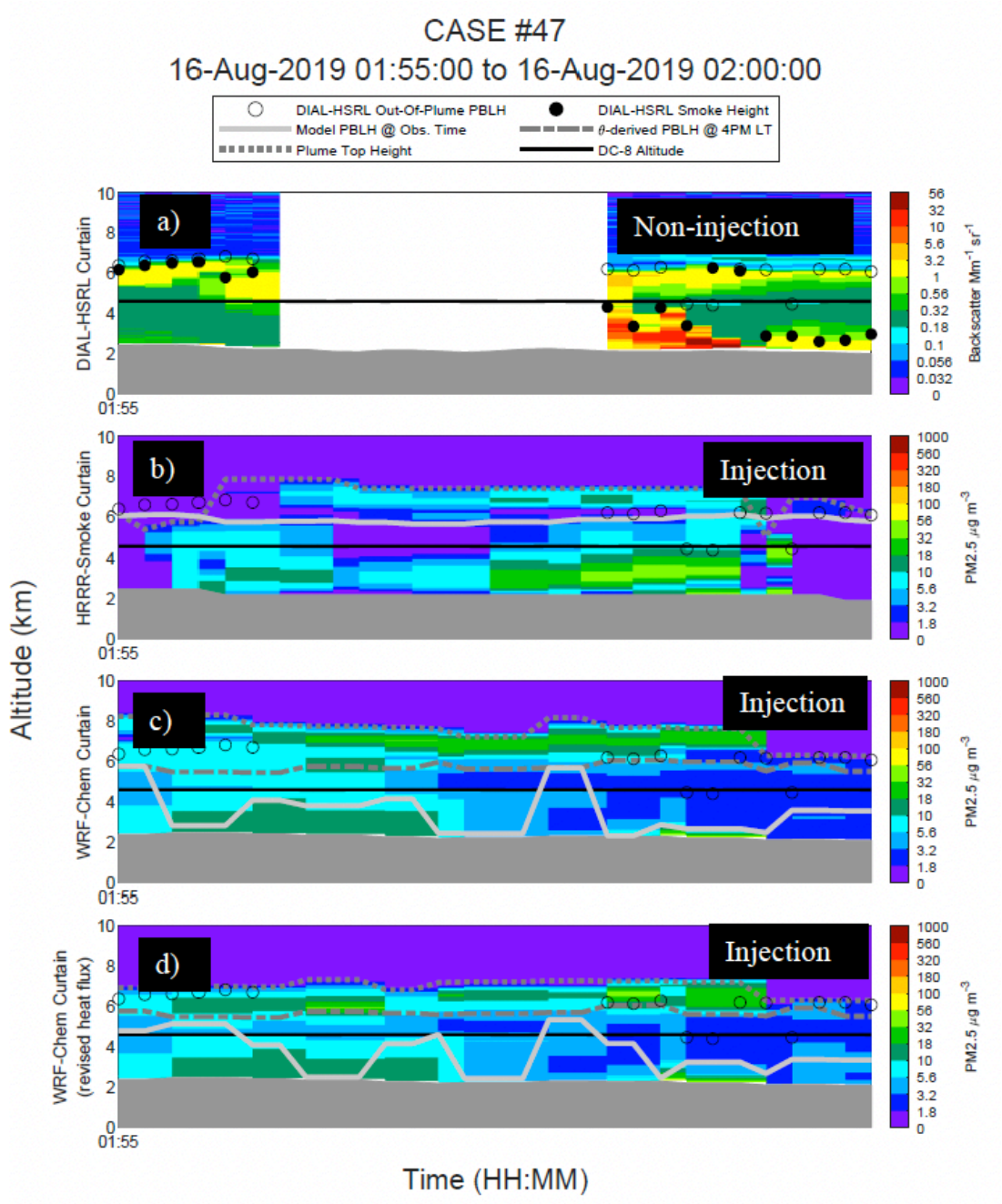

Figure A47. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Boulin Fire 2019-08-16 01:55-02:00 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

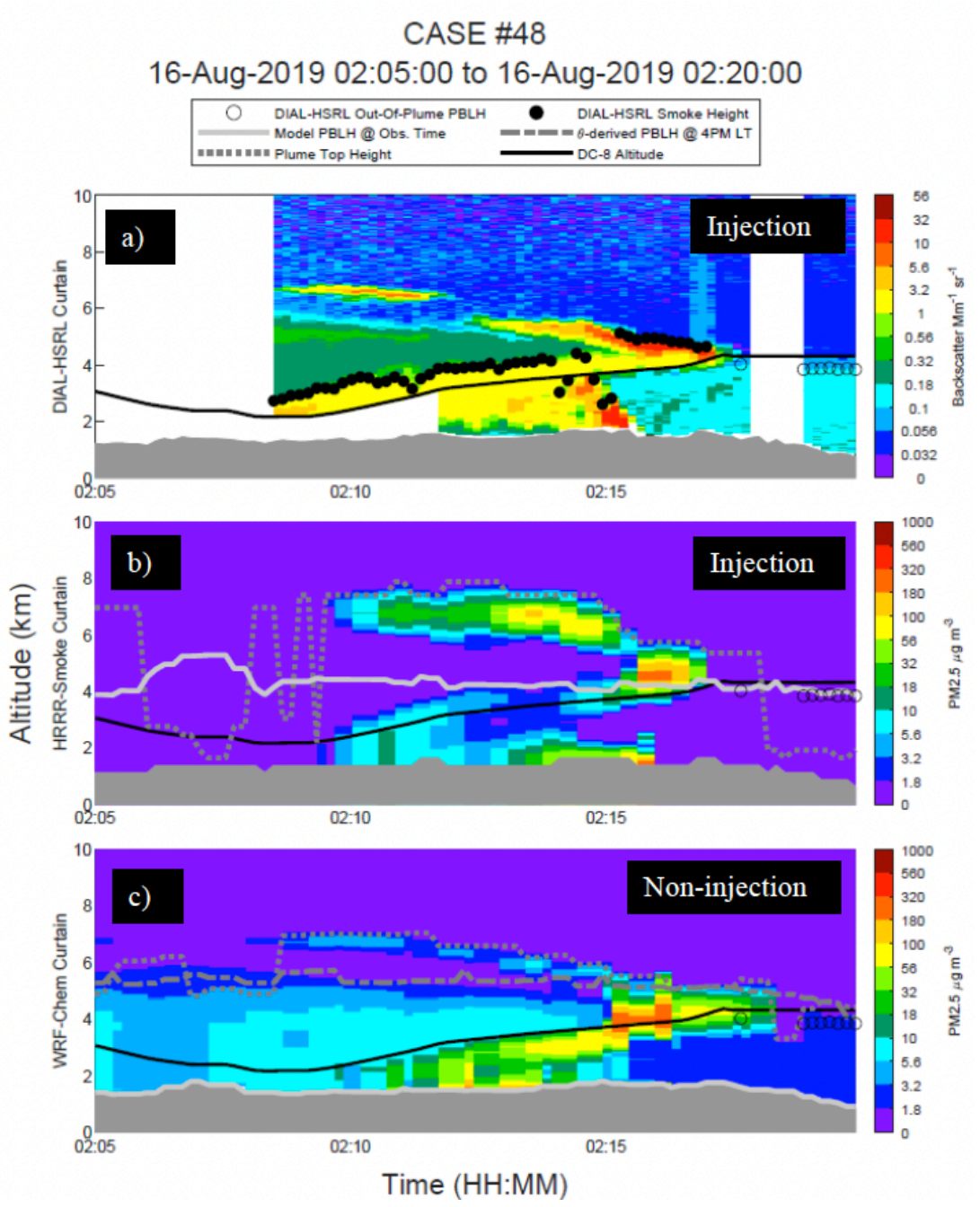

Figure A48. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Sheridan Fire 2019-08-16 02:05-02:20 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

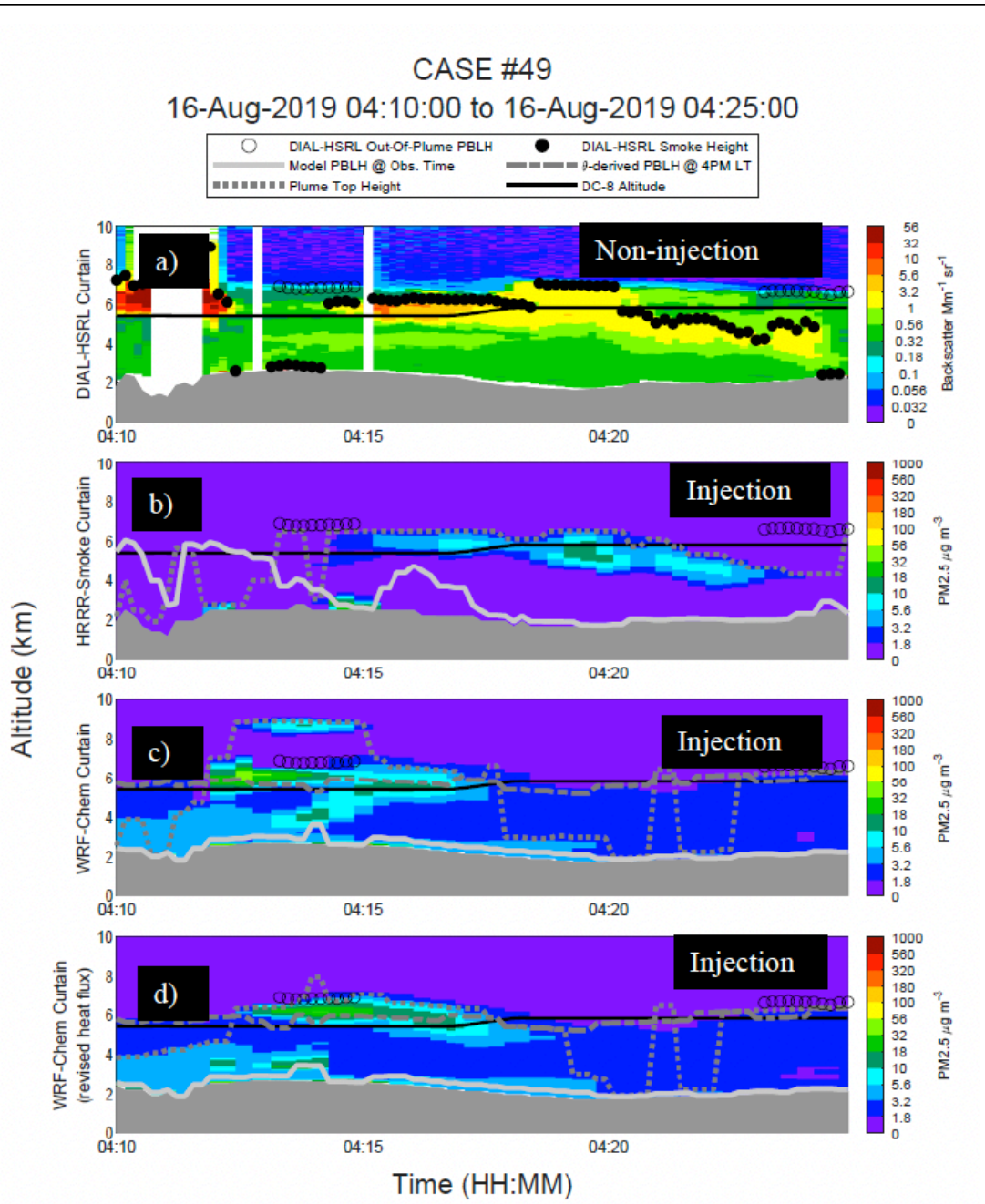

Figure A49. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Ikes Fire 2019-08-16 04:10-04:25 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

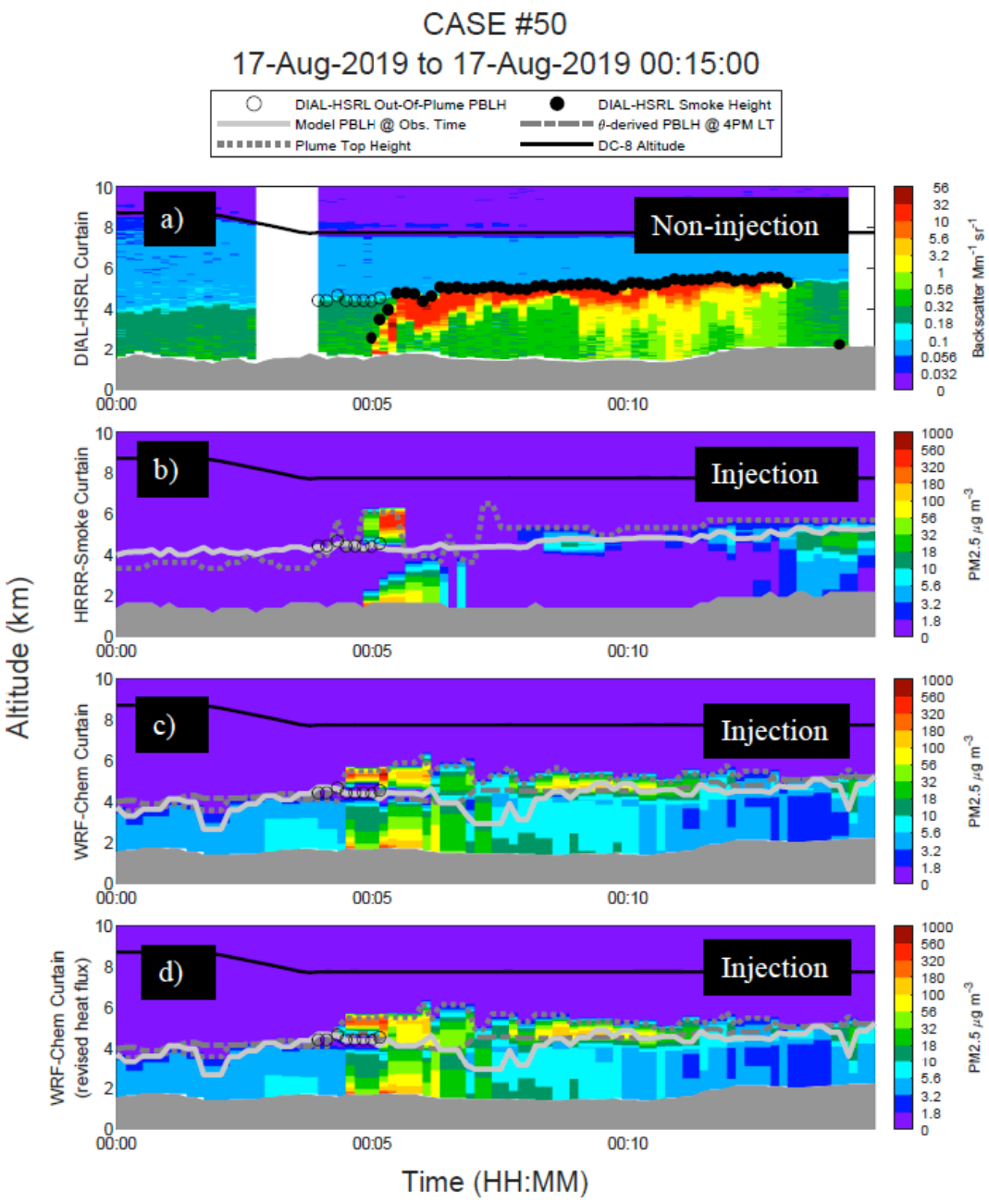

Figure A50. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Sheridan Fire 2019-08-17 00:00-00:15 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

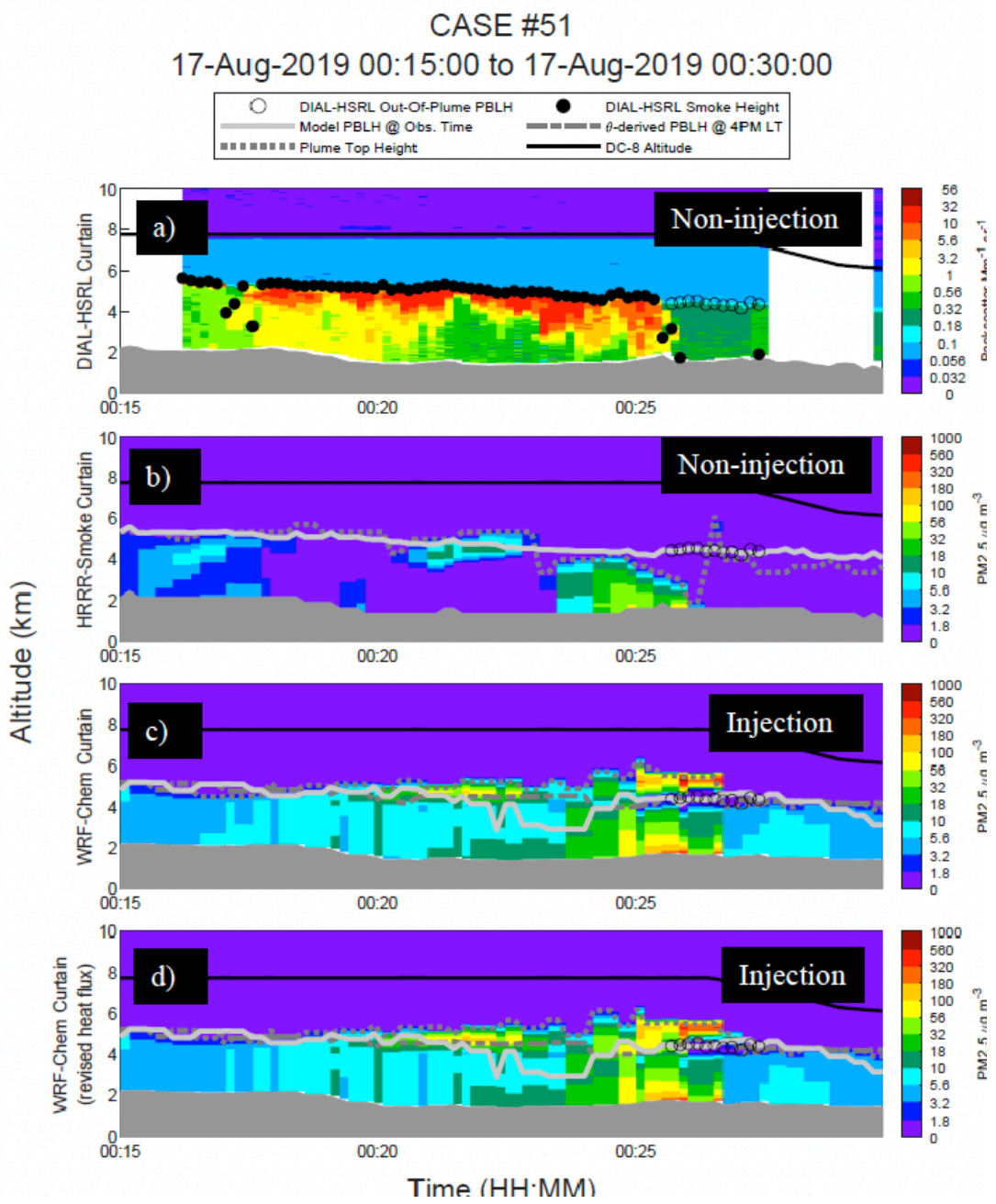

Figure A51. Vertical profiles of backscatter from the DIAL-HSRL (a) and PM2.5 from HRRR-Smoke (b), WRF-Chem with default heat fluxes (c), and WRF-Chem with adjusted heat fluxes (d) for the Sheridan Fire 2019-08-17 00:15-00:30 UTC. In panel a, open circles show the top of the PBL and filled circles are the smoke top heights. In panels b and c, the modeled PBL height is shown as the light gray line, and the smoke top height is shown as the dotted dark gray line. In panel c, the dark gray dashed line shows the PBL top at 4pm local time, assumed to be the time when the PBL height is at its maximum.

Figure A52. Comparison of observed (green) and modeled (red) planetary boundary layer height (in m above sea level) distributions for the WRF-Chem (a) and HRRR-Smoke (b) models. Cases for which these plots were generated are the WRF-Chem false positives

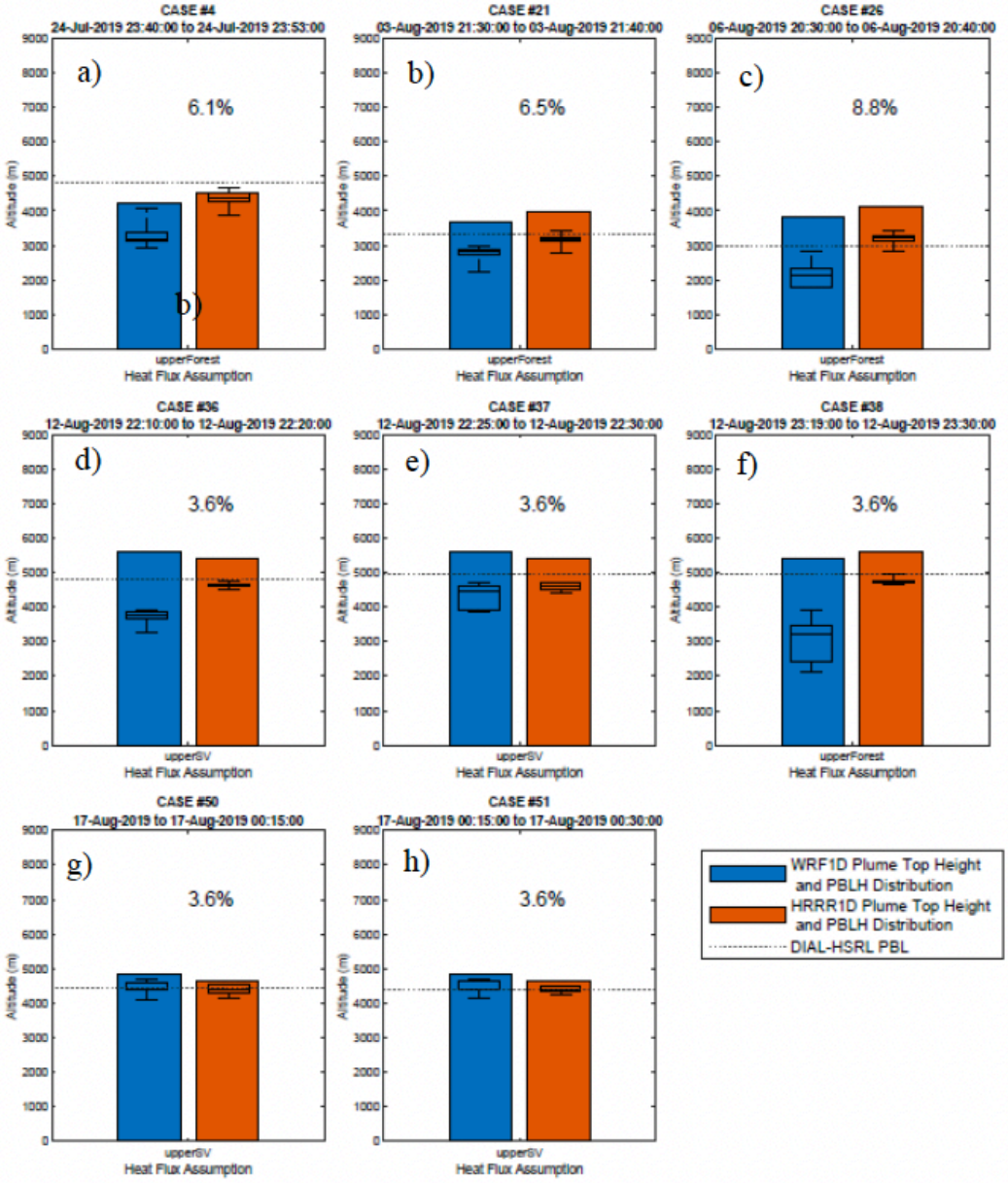

Figure A53. Plume top heights from the 1D Freitas forced by 1D WRF-Chem meteorology (blue bars), and 1D Freitas forced by 1D HRRR-Smoke meteorology (red bars) for cases where WRF-Chem underpredicted the PBLH and HRRR-Smoke accurately captured the PBLH. PBLHs associated with each meteorology configuration are overlaid as box and whisker plots over their corresponding bar, and the PBLH derived from the DIAL-HSRL is given as a dotted line. Selected fires include a) Shady, b) Mica Creek, c) Williams Flats, d-e) Spring, f) Castle, g-h) Sheridan

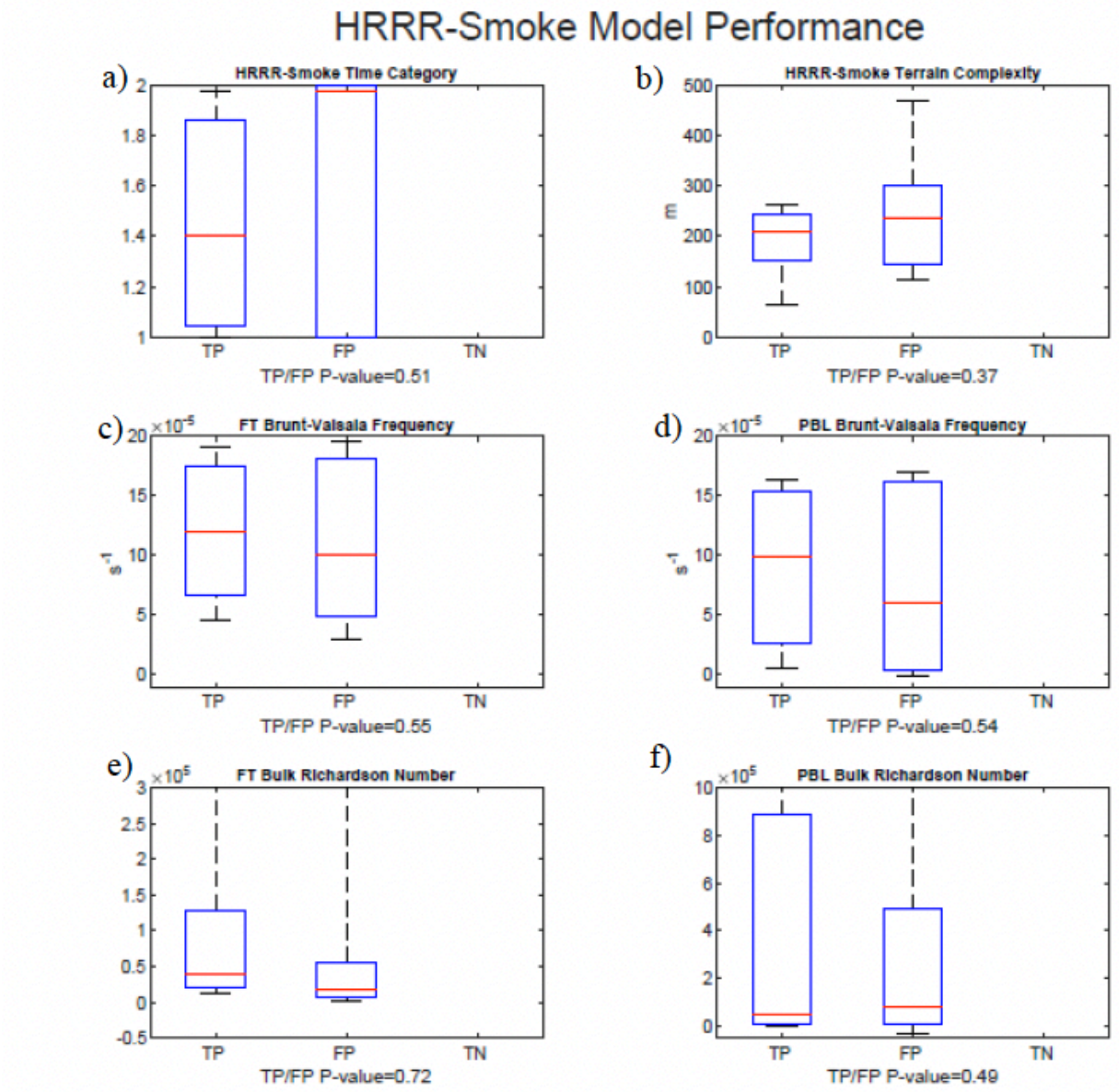

Figure A54. Box and whisker plots of (a) HRRR-Smoke time of day, (b) HRRR-Smoke terrain variability (standard deviation in 3x3 grid box around fire), (c) FT Brunt-Vaisala Frequency, (d) PBL Brunt-Vaisala Frequency (e) FT bulk Richardson number, (f) PBL bulk Richardson number. Red line denotes medians, blue box denotes the interquartile range, and whiskers denote the 1.5x interquartile range. TP, FP, and FN are defined as in Table 2.2 caption.

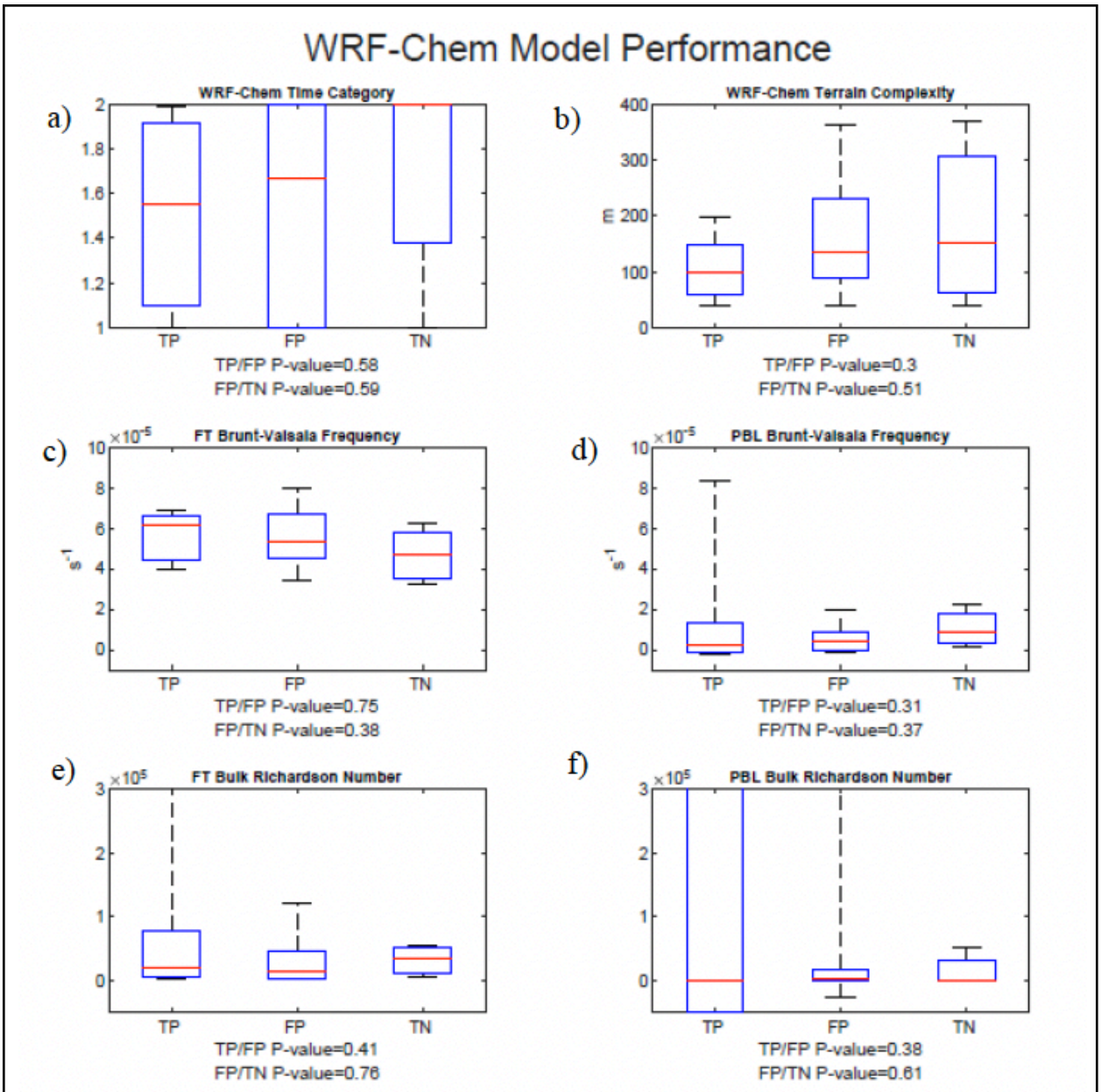

Figure A55. Box and whisker plots of (a) WRF-Chem time of day, (b) WRF-Chem terrain variability (standard deviation in 3x3 grid box around fire), (c) FT Brunt-Vaisala Frequency, (d) PBL Brunt-Vaisala Frequency (e) FT bulk Richardson number, (f) PBL bulk Richardson number. Red line denotes medians, blue box denotes the interquartile range, and whiskers denote the 1.5x interquartile range. TP, FP, and FN are defined as in Table 2.2 caption

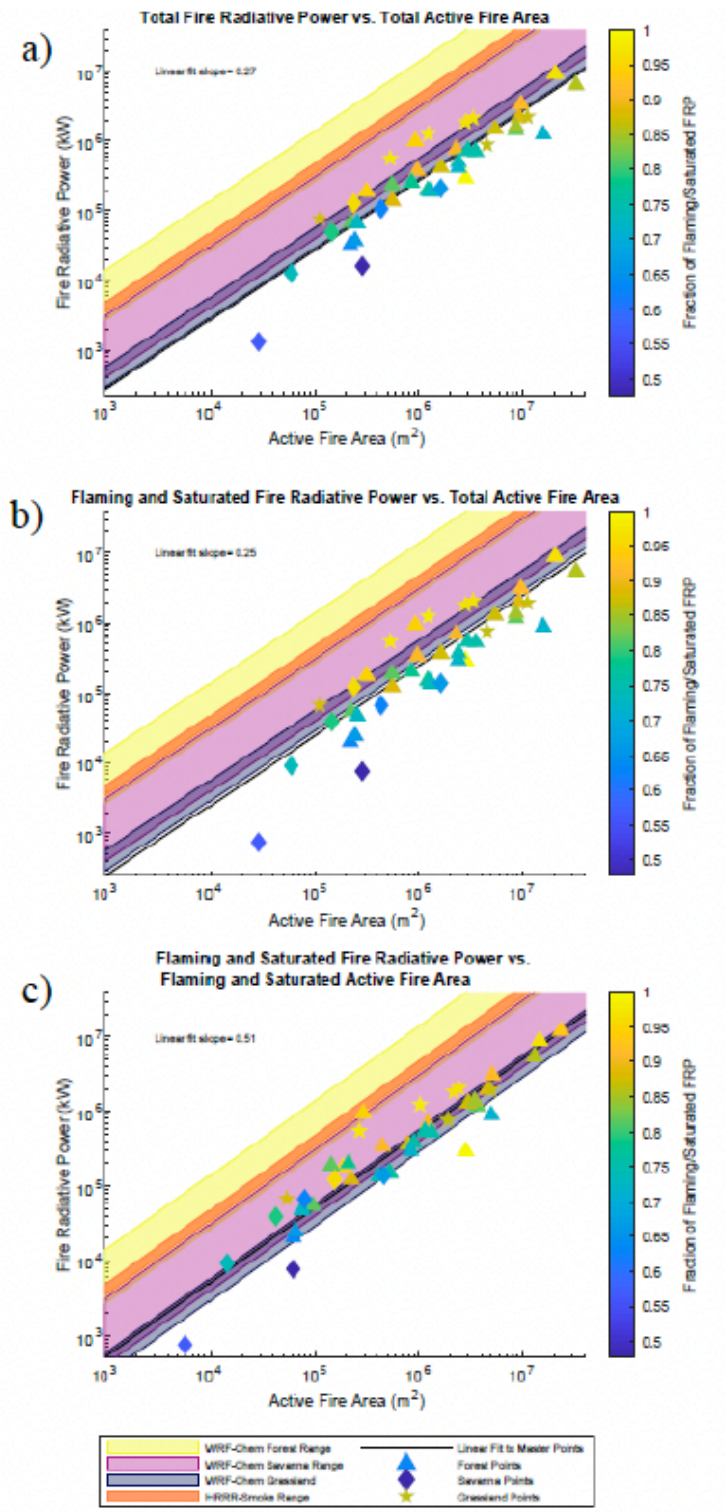

Figure A56. Analogous to Fig. 2.3, but for different combustion phases: (a) total FRP and total area, (b) flaming/saturated FRP and total area, and (c) flaming/saturated FRP and flaming/saturated area

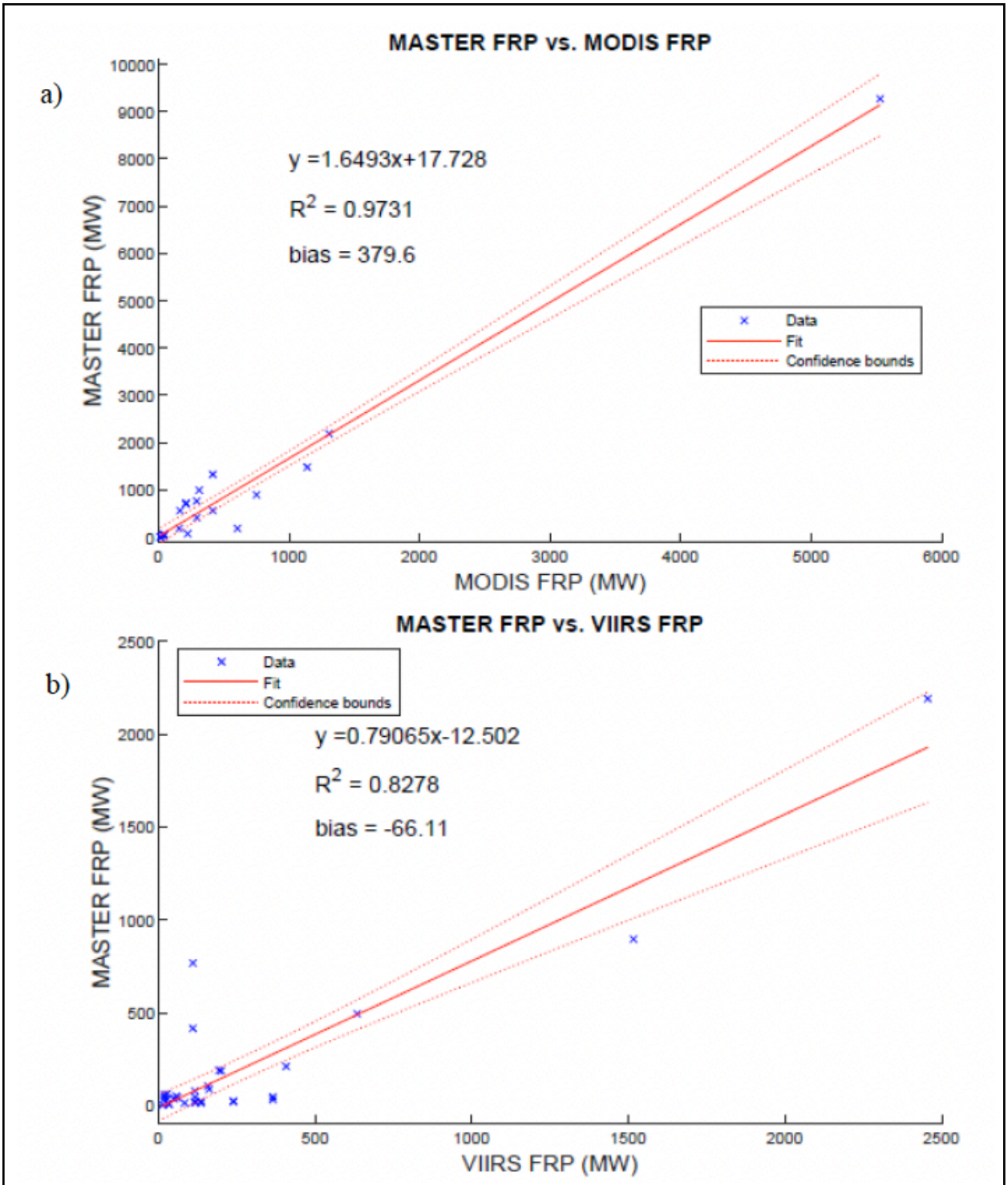

Figure A57. MASTER FRP vs (a) VIIRS FRP and (b) MODIS FRP determined using VIIRS and MODIS overpasses within +/- 2hrs from MASTER overpasses. MASTER values tend to be higher than MODIS values and lower than VIIRS values. Red dashed lines are 95% confidence bounds on the best fit line.

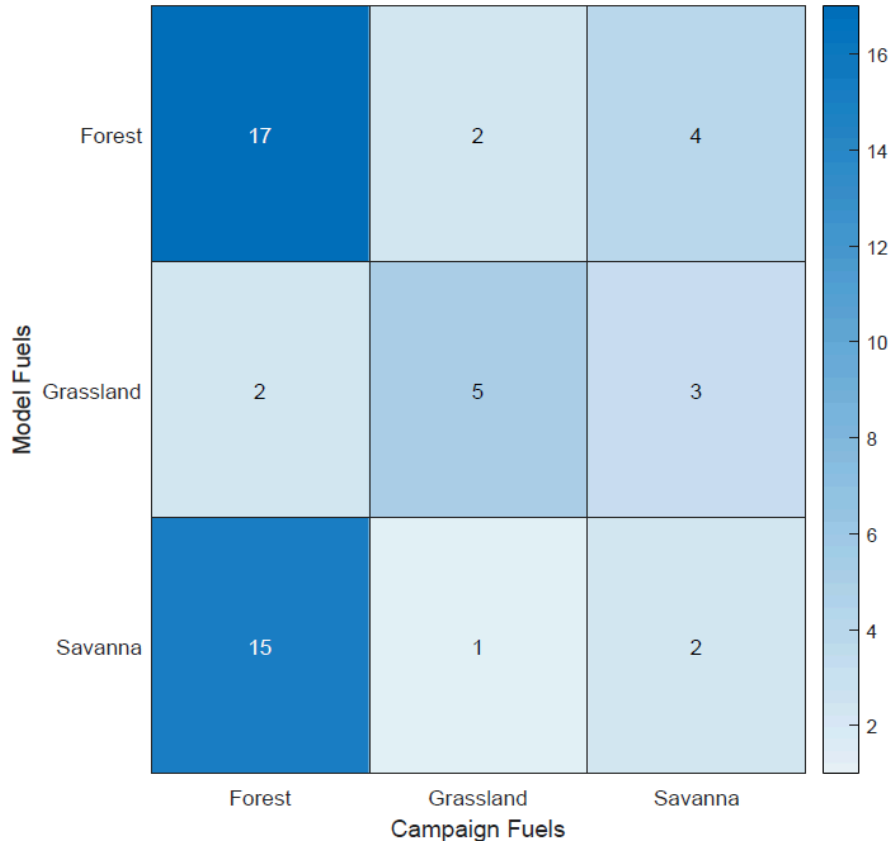

Figure A58. Comparison of campaign fuels and model fuels covering three broad fuel categories for the 51 test cases in this study: grassland, savanna, and forest. Entries along the main diagonal represent cases where the model and the campaign fuel type agreed, and off diagonals represent cases where the model and campaign fuel type disagreed.

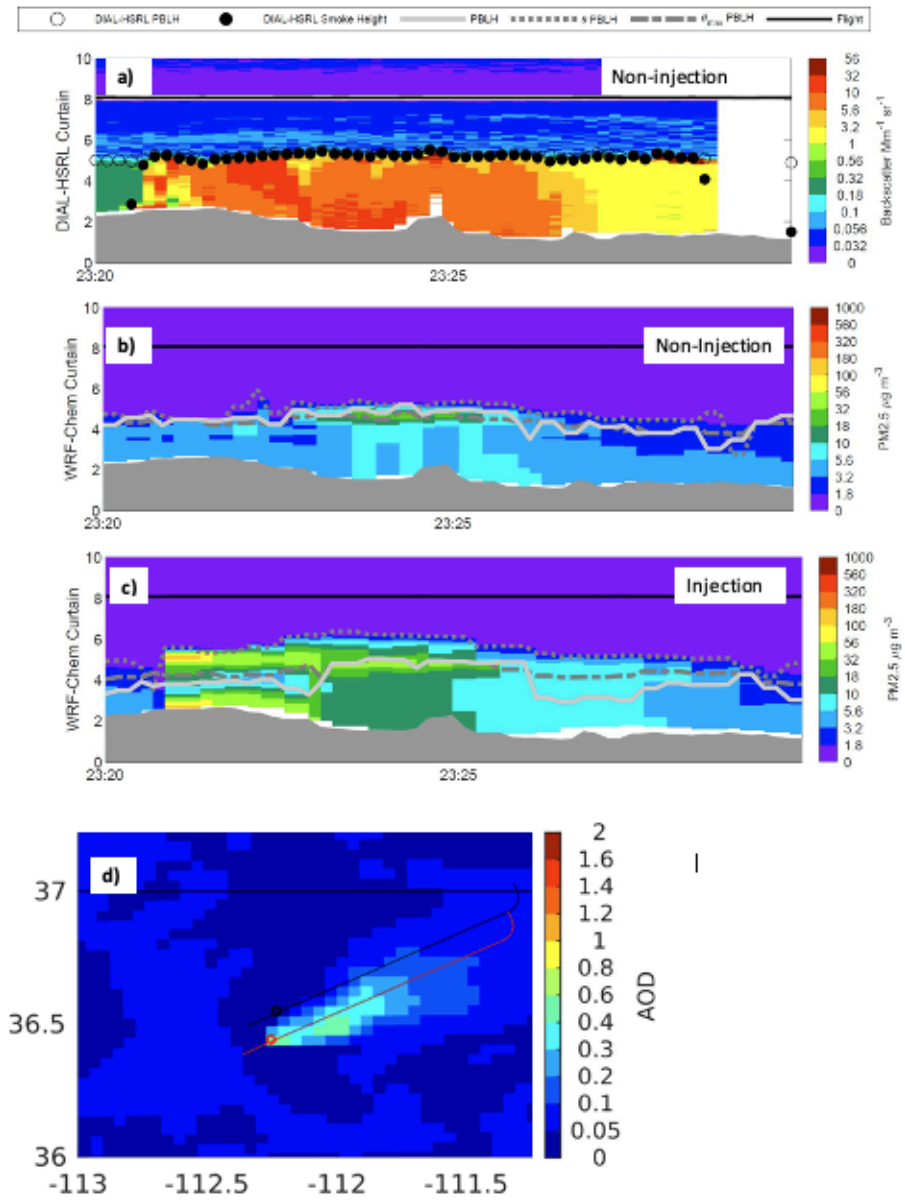

Figure A59. Illustration of the spatial shifting algorithm for the Castle fire, observed on 8/12/2019 23:20 20-23:30 UTC. (a) Observed backscatter profiles from the DIAL -HSRL. (b) Modeled PM2.5 curtain plot without shifting applied. (c) Modeled PM2.5 curtain plot with shifting applied. (d) Spatial aerosol optical depth (AOD) plot overlaid with unshifted (black) and shifted (red) flight flight track.

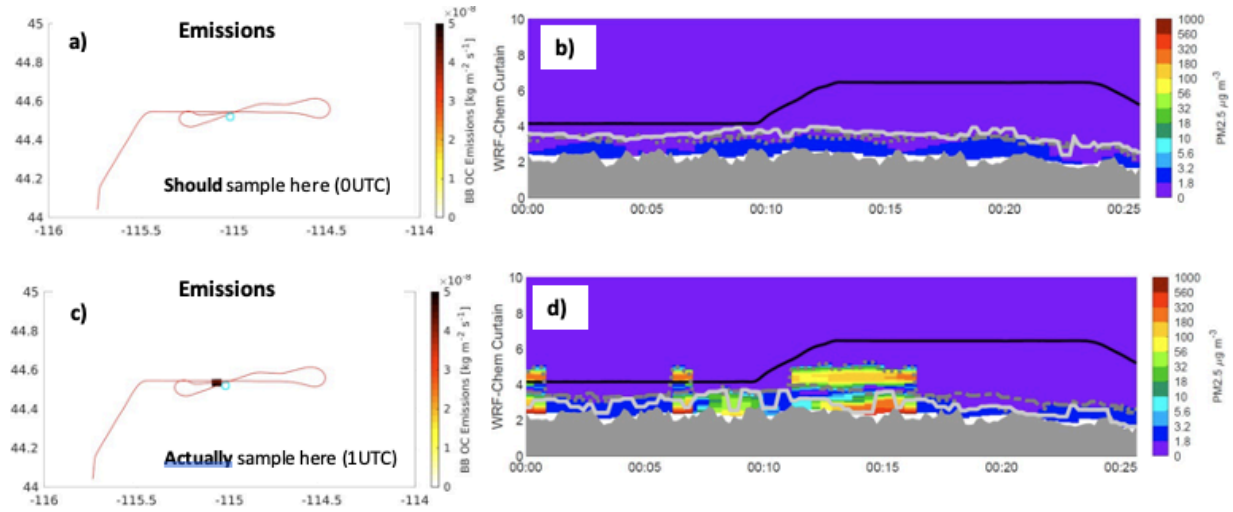

Figure A60. Illustration of time shift for the Shady Fire, observed 7/29/2019 00:00-00:38 UTC. (a) spatial emissions plot and (b) vertical curtain of PM2.5 sampled at the model time matching the observed time. Note how emissions are not present in the spatial or the vertical plot. (c) spatial emissions plot and (d) vertical curtain plot sampled at the model time shifted to one hour after the observed time. Emissions are now present in the spatial plot and vertical plume structure is developed.

		Overpass Time (UTC)		DIAL-HSRL		WRF-Chem		WRF-Chem Observed PBL		WRF-Chem Sensitivity		WRF-Chem Sensitivity Observed PBL	
Fig.	Fire	Start Time	End Time	Injection	End Node	Injection	End Node	Injection	End Node	Injection	End Node	Injection	End Node
S1	Sheep	7-24 21:35	7-24 21:45	0	B2	0	B2	1	D2	nan	nan	nan	nan
S2	Sheep	7-24 21:50	7-24 22:05	0	B2	0	B2	0	B2	nan	nan	nan	nan
S3	Sheep	7-24 22:40	7-24 23:00	0	B2	0	C2	0	C2	nan	nan	nan	nan
S4	Shady	7-24 23:40	7-24 23:53	0	B2	1	D2	0	B2	1	D2	0	B2
S5	Shady	7-24 23:55	7-25 00:07	0	B2	1	F2	0	B2	1	F2	0	B2
S6	Shady	7-25 00:03	7-25 00:10	0	B2	1	F2	0	B2	1	D2	0	B2
S7	Shady	7-25 00:10	7-25 00:20	0	B2	1	F2	0	B2	1	F2	0	B2
S8	Shady	7-25 22:30	7-25 22:45	1	D2	1	D2	1	D2	nan	nan	nan	nan
S9	Shady	7-26 01:25	7-26 01:35	1	D2	1	D2	1	D2	nan	nan	nan	nan
S10	North Hills	7-29 22:40	7-29 22:55	0	B2	0	B2	0	B2	nan	nan	nan	nan
S11	North Hills	7-29 23:00	7-29 23:15	0	B2	0	B2	0	B2	nan	nan	nan	nan
S12	Tucker	7-30 02:10	7-30 02:35	0	E2	0	E2	0	E2	nan	nan	nan	nan
S13	Tucker	7-30 04:00	7-30 04:20	0	B2	0	F1	0	B2	nan	nan	nan	nan
S14	Tucker	7-30 04:15	7-30 04:40	0	B2	0	F1	0	B2	nan	nan	nan	nan
S15	Left Hand	7-31 00:35	7-31 00:55	1	E1	1	E1	1	E1	nan	nan	nan	nan
S16	Left Hand	7-31 00:55	7-31 01:15	1	E1	1	E1	1	E1	nan	nan	nan	nan
S17	Ridge Top	8-02 22:10	8-02 22:25	0	B2	1	D2	0	B2	1	D2	0	B2
S18	Ridge Top	8-02 23:40	8-02 23:50	1	D2	1	F2	0	B2	nan	nan	nan	nan
S19	Mica/Lick Creek	8-03 00:35	8-03 00:55	0	B2	1	D2	0	B2	1	D2	0	B2
S20	Mica/Lick Creek	8-03 02:20	8-03 02:40	0	B2	1	F2	1	D2	1	F2	0	B2
S21	Mica Creek	8-03 21:30	8-03 21:40	0	B2	1	D2	1	D2	1	D2	0	B2
S22	Williams Flats	8-03 21:40	8-03 22:00	1	D2	1	D2	1	D2	nan	nan	nan	nan
S23	Williams Flats	8-04 00:05	8-04 00:25	1	D2	1	F2	1	D2	nan	nan	nan	nan
S23	Williams Flats	8-04 02:30	8-04 02:55	1	D2	1	F2	1	D2	nan	nan	nan	nan
S25	Williams Flats	8-06 18:40	8-06 19:00	0	B2	1	D2	0	B2	0	E2	0	B2
S26	Williams Flats	8-06 20:30	8-06 20:40	0	B2	1	D2	1	D2	0	E2	0	E2
S27	Williams Flats	8-06 21:45	8-06 22:00	0	B2	1	D2	0	B2	0	E2	0	B2
S28	Snow Creek	8-06 22:25	8-06 22:35	0	B2	1	D2	1	D2	1	D2	0	B2
S29	Horsefly	8-06 22:45	8-06 23:00	0	B2	1	D2	0	B2	0	E2	0	B2
S30	Horsefly	8-07 00:25	8-07 00:55	0	B2	1	D2	0	B2	0	E2	0	B2
S31	Williams Flats	8-07 23:00	8-07 23:20	1	D2	1	E1	1	E1	nan	nan	nan	nan
S32	Williams Flats	8-08 00:45	8-08 01:05	1	D2	1	D2	1	D2	nan	nan	nan	nan
S33	Williams Flats	8-08 02:20	8-08 02:40	1	D2	1	D2	1	D2	nan	nan	nan	nan
S34	Williams Flats	8-09 02:00	8-09 02:15	1	D2	1	F2	1	D2	nan	nan	nan	nan
S35	Williams Flats	8-09 02:15	8-09 02:30	1	D2	1	F2	0	B2	nan	nan	nan	nan
S36	Springs	8-12 22:10	8-12 22:20	0	B2	1	D2	1	D2	1	D2	1	D2
S37	Springs	8-12 22:25	8-12 22:30	0	B2	1	D2	0	B2	1	D2	1	D2
S38	Castle	8-12 23:19	8-12 23:30	0	B2	1	D2	1	D2	1	D2	1	D2
S39	Castle	8-12 23:30	8-12 23:50	0	B2	1	D2	1	D2	1	D2	1	D2
S40	Castle	8-13 22:30	8-13 22:45	0	C2	1	D2	1	D2	1	D2	1	D2
S41	Castle	8-13 22:40	8-13 22:55	0	C2	0	B2	1	D2	nan	nan	nan	nan
S42	Sheridan	8-16 00:25	8-16 00:40	1	D2	1	D2	1	D2	nan	nan	nan	nan
S43	Sheridan	8-16 00:40	8-16 00:50	1	D2	1	D2	1	D2	nan	nan	nan	nan
S44	Sheridan	8-16 01:00	8-16 01:15	1	D2	1	F2	1	D2	nan	nan	nan	nan
S45	Sheridan	8-16 01:15	8-16 01:30	1	D2	1	F2	1	D2	nan	nan	nan	nan
S46	Saber	8-16 01:45	8-16 01:55	0	B2	1	F2	0	B2	0	E2	0	B2
S47	Boulin	8-16 01:55	8-16 02:00	0	B2	1	F2	1	D2	1	F2	1	D2
S48	Sheridan	8-16 02:05	8-16 02:20	1	D2	0	F1	1	D2	nan	nan	nan	nan
S49	Ikes	8-16 04:10	8-16 04:25	0	B2	1	F2	0	B2	1	F2	0	B2
S50	Sheridan	8-17 00:00	8-17 00:15	0	B2	1	D2	1	D2	1	D2	1	D2
S51	Sheridan	8-17 00:15	8-17 00:30	0	B2	1	D2	1	D2	1	D2	1	D2

Table A1. Details for each case, including the figure number, fire name, date, and time of the measurements, WRF-Chem modeled and observed injection behavior (0=non-injection, 1=injection), and terminal node of the decision tree that results in each classification. Not-a-Number (Nan) means that the model missed the fire and therefore the injection behavior could not be evaluated.

		Overpass Time (UTC)		DIAL-HSRL		HRRR-Smoke		HRRR-Smoke Observed PBL	
Fig.	Fire	Start Time	End Time	Injection	End Node	Injection	End Node	Injection	End Node
S1	Sheep	7-24 21:35	7-24 21:45	0	B2	1	D2	1	D2
S2	Sheep	7-24 21:50	7-24 22:05	0	B2	1	D2	1	D2
S3	Sheep	7-24 22:40	7-24 23:00	0	B2	1	D2	1	D2
S4	Shady	7-24 23:40	7-24 23:53	0	B2	nan	nan	nan	nan
S5	Shady	7-24 23:55	7-25 00:07	0	B2	nan	nan	nan	nan
S6	Shady	7-25 00:03	7-25 00:10	0	B2	nan	nan	nan	nan
S7	Shady	7-25 00:10	7-25 00:20	0	B2	nan	nan	nan	nan
S8	Shady	7-25 22:30	7-25 22:45	1	D2	nan	nan	nan	nan
S9	Shady	7-26 01:25	7-26 01:35	1	D2	1	D2	1	D2
S10	North Hills	7-29 22:40	7-29 22:55	0	B2	1	D2	0	B2
S11	North Hills	7-29 23:00	7-29 23:15	0	B2	1	D2	0	B2
S12	Tucker	7-30 02:10	7-30 02:35	0	E2	nan	nan	nan	nan
S13	Tucker	7-30 04:00	7-30 04:20	0	B2	nan	nan	nan	nan
S14	Tucker	7-30 04:15	7-30 04:40	0	B2	nan	nan	nan	nan
S15	Left Hand	7-31 00:35	7-31 00:55	1	E1	1	E1	0	B2
S16	Left Hand	7-31 00:55	7-31 01:15	1	E1	1	E1	0	B2
S17	Ridge Top	8-02 22:10	8-02 22:25	0	B2	0	B2	0	B2
S18	Ridge Top	8-02 23:40	8-02 23:50	1	D2	1	D2	1	D2
S19	Mica/Lick Creek	8-03 00:35	8-03 00:55	0	B2	1	D2	1	D2
S20	Mica/Lick Creek	8-03 02:20	8-03 02:40	0	B2	1	D2	1	D2
S21	Mica Creek	8-03 21:30	8-03 21:40	0	B2	1	D2	1	D2
S22	Williams Flats	8-03 21:40	8-03 22:00	1	D2	1	D2	1	D2
S23	Williams Flats	8-04 00:05	8-04 00:25	1	D2	1	D2	1	D2
S23	Williams Flats	8-04 02:30	8-04 02:55	1	D2	1	D2	1	D2
S25	Williams Flats	8-06 18:40	8-06 19:00	0	B2	1	D2	0	B2
S26	Williams Flats	8-06 20:30	8-06 20:40	0	B2	1	D2	1	D2
S27	Williams Flats	8-06 21:45	8-06 22:00	0	B2	1	D2	0	B2
S28	Snow Creek	8-06 22:25	8-06 22:35	0	B2	1	D2	1	D2
S29	Horsefly	8-06 22:45	8-06 23:00	0	B2	1	D2	0	B2
S30	Horsefly	8-07 00:25	8-07 00:55	0	B2	1	D2	0	B2
S31	Williams Flats	8-07 23:00	8-07 23:20	1	D2	1	D2	1	D2
S32	Williams Flats	8-08 00:45	8-08 01:05	1	D2	1	D2	1	D2
S33	Williams Flats	8-08 02:20	8-08 02:40	1	D2	1	D2	1	D2
S34	Williams Flats	8-09 02:00	8-09 02:15	1	D2	1	D2	1	D2
S35	Williams Flats	8-09 02:15	8-09 02:30	1	D2	1	D2	0	B2
S36	Springs	8-12 22:10	8-12 22:20	0	B2	1	D2	1	D2
S37	Springs	8-12 22:25	8-12 22:30	0	B2	1	D2	1	D2
S38	Castle	8-12 23:19	8-12 23:30	0	B2	1	D2	1	D2
S39	Castle	8-12 23:30	8-12 23:50	0	B2	1	D2	1	D2
S40	Castle	8-13 22:30	8-13 22:45	0	C2	1	D2	1	D2
S41	Castle	8-13 22:40	8-13 22:55	0	C2	1	D2	1	D2
S42	Sheridan	8-16 00:25	8-16 00:40	1	D2	1	D2	1	D2
S43	Sheridan	8-16 00:40	8-16 00:50	1	D2	1	D2	1	D2
S44	Sheridan	8-16 01:00	8-16 01:15	1	D2	1	D2	1	D2
S45	Sheridan	8-16 01:15	8-16 01:30	1	D2	1	D2	1	D2
S46	Saber	8-16 01:45	8-16 01:55	0	B2	1	D2	1	D2
S47	Boulin	8-16 01:55	8-16 02:00	0	B2	1	D2	1	D2
S48	Sheridan	8-16 02:05	8-16 02:20	1	D2	1	D2	1	D2
S49	Ikes	8-16 04:10	8-16 04:25	0	B2	1	D2	0	B2
S50	Sheridan	8-17 00:00	8-17 00:15	0	B2	1	D2	1	D2
S51	Sheridan	8-17 00:15	8-17 00:30	0	B2	0	B2	0	B2

Table A2. Details for each case, including the figure number, fire name, date, and time of the measurements, HRRR-Smoke modeled and observed injection behavior (0=non-injection, 1=injection), and terminal node of the decision tree that results in each classification. Not-a-Number (Nan) means that the model missed the fire and therefore the injection behavior could not be evaluated.

Fig.	Fire	Out of plume MLH Identified?	Assumptions (unless stated differently, 0.32 is the MLH cutoff, whole air column is searched, and upwind only is used)
S1	Sheep	Yes (11pts)	Get points closest to fire, where terrain is similar to fire
S2	Sheep	Yes (16pts)	Added points post turn
S3	Sheep	Yes (9pts)	Added points post turn
S4	Shady	Yes (21pts)	Used the downwind region with similar terrain
S5	Shady	Yes (10pts)	Added points post turn
S6	Shady	Yes (11pts)	
S7	Shady	Yes (15pts)	
S8	Shady	Yes (9pts)	Not enough points, even post turn or downwind
S9	Shady	Yes (11pts)	
S10	North Hills	Yes (11pts)	Chose some upwind no clouds, some downwind
S11	North Hills	Yes (12pts)	Chose some upwind no clouds, some downwind
S12	Tucker	Yes (24pts)	
S13	Tucker	Yes (11pts)	
S14	Tucker	Yes (6pts)	Not enough points, even post turn or downwind
S15	Left Hand	Yes (10pts)	
S16	Left Hand	Yes (12pts)	
S17	Ridge Top	Yes (11pts)	Look for light green 0.56 below 11km, dark green may be pyroCu outflow?
S18	Ridge Top	Yes (11pts)	Look for light green 0.56 below 11km, dark green may be pyroCu outflow?
S19	Mica/Lick Creek	Yes (11pts)	Look for light green 0.56 below 4km, lofted smoky layer
S20	Mica/Lick Creek	Yes (11pts)	Look for light green 0.56 below 4km, lofted smoky layer
S21	Mica Creek	Yes (7pts)	Chose some upwind no clouds, some downwind
S22	Williams Flats	Yes (13pts)	Look for medium green 0.32 below 4.5km, lofted smoky layer
S23	Williams Flats	Yes (10pts)	Look for medium green 0.32 below 4.5km, lofted smoky layer
S23	Williams Flats	Yes (11pts)	Look for medium green 0.32 below 3km, lofted smoky layer
S25	Williams Flats	Yes (19pts)	Chose downwind with similar terrain
S26	Williams Flats	Yes (7pts)	Look for light green 0.56 below 11km
S27	Williams Flats	Yes (15pts)	Chose some upwind no clouds, some downwind
S28	Snow Creek	Yes (15pts)	
S29	Horsefly	Yes (11pts)	
S30	Horsefly	Yes (17pts)	Chose a smaller time window, shaved off that edge that was catching high clouds
S31	Williams Flats	Yes (17pts)	
S32	Williams Flats	Yes (8pts)	Not enough points, even post turn or downwind
S33	Williams Flats	Yes (8pts)	Not enough points, even post turn or downwind
S34	Williams Flats	Yes (11pts)	
S35	Williams Flats	Yes (23pts)	
S36	Springs	Yes (10pts)	Used the downwind region with similar terrain
S37	Springs	Yes (8pts)	Chose some upwind no clouds, some downwind
S38	Castle	Yes (8pts)	
S39	Castle	Yes (9pts)	Filtered out clouds just upwind of the fire
S40	Castle	Yes (4pts)	Not enough points, even post turn or downwind
S41	Castle	Yes (8pts)	
S42	Sheridan	Yes (10pts)	Expanded the time window, looked for blue 0.18
S43	Sheridan	Yes (9pts)	Expanded the time window, looked for blue 0.18
S44	Sheridan	Yes (6pts)	Expanded the time window, looked for blue 0.18
S45	Sheridan	Yes (14pts)	Looked for blue 0.18
S46	Saber	Yes (10pts)	Saber is observed right before Boulin, and LATE, use light blue
S47	Boulin	Yes (17pts)	Use whole thing, assume it's mixed layer
S48	Sheridan	Yes (18pts)	Looked for blue 0.18
S49	Ikes	Yes (21pts)	Chose some upwind no clouds, some downwind
S50	Sheridan	Yes (8pts)	
S51	Sheridan	Yes (11pts)	

Table A3. Assumptions for selecting the out-of-plume observed PBLH.

Appendix B: Forecasting daily fire radiative energy using data driven methods and machine learning techniques

Contents of this file

Text B1
Tables B1 to B5
Figures B1 to B16

Text B1

B1.1 Fire Location

While not directly used as an input to the machine learning or other statistical methods, fire location data allows for the tracking of wildfires in space and time and for the extraction of gridded data in only regions relevant to the wildfire.

- **Active Fire Detections.** Fire detections from the Visible Infrared Imaging Radiometer Suite (VIIRS) instruments aboard the Suomi National Polar-orbiting Partnership (SNPP) and NOAA-20 satellites (Schroeder et al., 2014) are used to track daily fire evolution.
- **Final Fire Perimeters.** Monitoring Trends in Burn Severity (MTBS) provides Landsat-based detailed final burn perimeters for fires exceeding 1000 acres in the continental US (Eidenshink et al., 2007). Some fires may take over a year to be fully processed, and in such cases MTBS provides a provisional initial assessment, which was used here to fill the gaps in the 2021 dataset. Use of the MTBS data limits our training dataset to fires larger than 1000 acres.

B1.2 Fire Intensity

- **RAVE Data**. Fire intensity measurements are obtained from the Regional ABI and VIIRS Emissions product (RAVE, Li et al., 2022). RAVE aggregates FRP retrievals from the Geostationary Operational Environmental Satellites-Advanced Baseline Imager (GOES-ABI) and the Joint Polar Satellite System (JPSS) Visible Infrared Imaging Radiometer Suite (VIIRS). RAVE also fills gaps in the FRP record (i.e. those due to cloud cover or heavy smoke (Schmidt et al., 2012) using grid cell specific climatological diurnal cycles produced for land cover and ecoregion (Li et al., 2022). Time-integrated RAVE FRP, also known as fire radiative energy (FRE), data is used as the target variable and as a feature in the random forest model (persistence FRE from the previous day). RAVE is only processed as far back as 2019, so use of this data limits our training dataset to fires occurring after 2019. FRE for all quality flags was used in this analysis.

B1.3 Fire Weather

GridMET is a climatological product derived from the high temporal North American Land Data Assimilation System Phase 2 (NLDAS-2; Mitchell et al., 2004), and the high spatial resolution Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 2008). GridMET covers the period between 1979 and present day (Abatzoglou, 2013). In this work we extract the following GridMET variables:

- **GridMET meteorology and heatwaves**. We extract vapor pressure deficit (VPD), wind speed (WS), and daily minimum and maximum temperatures (T_{min} and T_{max}). Temperature data are processed to define 90th percentiles of temperature following Perkins et al., (2012). We track the number of days a fire exceeds the historical (1979-2010) 90th percentile of only T_{max} and both T_{max} and T_{min}.
- **GridMET National Fire Danger Rating System.** The National Fire Danger Rating System (NFDRS) is the system used by the US Forest Service to anticipate fire risk and behavior (Bradshaw et al., 1984). The NFDRS uses weather and topography information along with predetermined fuel models to predict fire intensity (Burning Index (BI)) and burnable fuel (Energy Release Component (ERC)). The NFDRS also produces class ratings, which range from low to extreme and describe likely fire behaviors given ignition under various weather conditions. GridMET calculates the above NFDRS components, assuming a conifer forest fuel model for the BI and the ERC.

The High Resolution Rapid Refresh (HRRR) model is a regional cloud-resolving model that has been operational since 2014 (Dowell et al., 2022). HRRR forecasts are initiated every hour and cover 18-48 hours. At each restart, HRRR performs data assimilation of weather observations and 3D radar reflectivity, which make the HRRR 0-hour forecasts comparable to a meteorological reanalysis. For this work we pull operational HRRR (https://rapidrefresh.noaa.gov/hrrr/) 0-hour forecasts and calculate the following fire weather indices, which track the influence of surface conditions and vertical stability on fire growth:

- **HRRR meteorology**. We extract temperature, relative humidity, and WS from the lowest 500m HRRR model. Temperature and relative humidity are used to calculate VPD for the HRRR model. The max VPD and WS over the lowest 500m of the HRRR model are provided as inputs to the models tested.
- **Hot-Dry-Windy.** Defined as the product of the maximum VPD and WS over the lowest 500m of the atmosphere, the hot-dry-windy index (HDW) is an indicator of the effect of local meteorology on wildfire behavior that can be calculated in any location and from any meteorological dataset that tracks relative humidity and wind speed (Srock et al., 2018). We developed a version of HDW based on HRRR that puts a lag of up to 3 days on the VPD used to calculate the HDW. This version is denoted HDmW0, where m is a number denoting the lag applied to the VPD. If m=0 then VPD and wind from the same

day were used. Fuels take time to dry out, and the lag was used to investigate the impact of enhanced VPD on fuel curing.

- **Hourly Wildfire Potential.** Hourly Wildfire Potential (HWP) is a new diagnostic product developed for the HRRR model which uses surface air temperature, humidity, wind gust potential, and soil moisture model output and is tuned to be an indicator of wildfire intensity in a given region, given a source of ignition. HWP is calculated following the formulation developed by Eric James (Equation 1).

$$(1)\ HWP = 0.237 * V * G^{1.11} * D^{0.92} * M^{6.95} * S$$

Where V=binary vegetation term (0 for urban, barren, snow/ice, water land cover, 1 for all other land covers); G=10-m gust maximum term (maximum of [3m/s, HRRR surface wind gust output]); D=dewpoint depression term (maximum of [15K, HRRR 2m dewpoint depression]); M=moisture availability term (fraction of maximum surface soil moisture 0.01*(100-HRRR surface moisture)); S=binary snow cover term (1 when HRRR snow water equivalent is greater than 25mm, 0 otherwise).

- **Continuous Haines Index.** The Continuous Haines Index (CHI) quantifies the impact of vertical changes in moisture and temperature on atmospheric stability (Dowdy & Pepler, 2018; Pinto et al., 2020). It was developed to extend the Haines Index, which was known to saturate for large wildfires (Jenkins, 2002; Potter, 2018; Potter, 2005). CHI is calculated using HRRRv4 temperature at 850 and 700 mb as well as dewpoint depression at 850mb following Pinto et al., (2020).
- **Pyrocumulonimbus Firepower Threshold.** Also a measure of atmospheric stability on wildfire behavior, the pyrocumulonimbus (pyroCb) firepower threshold (PFT) is a measure of the fire energy (in GW) needed to generate a pyroCb in a given thermodynamic environment (Tory & Kepert, 2021). The Tory & Kepert (2021) formulation is based on the Briggs equations of plume rise, which allows PFT to be calculated as a function of free convection height, wind speed, and potential temperature. These quantities may all be derived from a sounding, which is assembled using HRRRv4 columns of temperature, dewpoint, specific humidity, and u- and v-winds, and grid cell values of terrain and boundary layer height.

The Global Fire WEather Database (GFWED) is a global implementation of the widely-used Canadian Forest Fire Danger Rating System (CFFDRS, Wagner, 1987).

- **CFFDRS Components.** For this work, we pull Field's (2015) formulation of the system, which uses MERRA-2 renalysis and IMERG precipitation (Field, 2020a, 2020b) to calculate indices describing fire intensity. The buildup index (BUI) describes heat release from large fuel, and the fire weather index (FWI) describes fire intensity with influence from fine fuels and weather conditions. Calculating CFFDRS components from

reanalysis data allows greater spatial and temporal coverage than the original formulation based on in situ observations, but biases in MERRA-2 temperature and relative humidity have been shown to introduce bias into final FWI calculations (Field, 2020a) .

B1.4 Moisture Products

- **GridMET fuel moisture.** The NFDRS also tracks the moisture of various time lagged fuels. Fuels with a 100hr timelag are 1-3in thick or ¾-4in deep. Fuels with a 1000hr timelag are 3-8in thick or 4in deep (Bradshaw et al., 1984). GridMET derives the moisture values for fuel model G (conifer forest)
- **GFWED Fuel Moisture.** The CFFDRS tracks the moisture of fuels at various depths as moisture codes, where higher values mean drier fuels. The fine fuel moisture code (FFMC) represents the top 1 cm of organic litter and is equivalent to fuels with a 16 hour timelag. The duff moisture code (DMC) represents the top 7cm of organic litter and is equivalent to fuels with a 360 hour timelag. Finally, drought code (DC) represents the top 18cm of organic litter and is equivalent to fuels with a 1272 hour timelag.
- **NOAA Soil Moisture Operational Products System.** The NOAA Soil Moisture Operational Products System (SMOPS) is a multi-instrument soil moisture retrieval based on microwave brightness temperature from seven satellite sensors which gives the volume percentage of water in the top 1-5 cm of soil (Liu et al., 2016). This work utilizes the Blended_SM variable from the SMOPS version 3 product, which has been shown to have very good agreement with in-situ soil moisture measurements (Yin et al., 2020).
- **NCAR Fuel Moisture.** These are machine learning-derived estimates of living and dead fuel moisture based on MODIS reflectances, WRF-Hydro model outputs, and terrain information (McCandless et al., 2020). In this dataset, dead fuel moisture (FMCG2D) corresponds to the 10 hr fuel moisture, which represents fractional moisture of fuels 0.25-1 inches thick or below the surface. Living moisture from this dataset (FMCGLH2D) represents the fractional moisture of living plants.
- **HRRR Soil Moisture.** Soil moisture at depths of the surface, 1cm, 4cm, 10cm and 30m are extracted from the HRRR model.
- **Plant Water Sensitivity.** The plant water sensitivity (PWS) indicates the degree to which living plant ecosystems are buffered against climate-driven changes in water availability, with high values of PWS indicating that living plants rapidly respond to changes in water availability and low values of PWS indicating a water-retaining ecosystem (Rao et al., 2022). Rao (2022) shows that high PWS values are associated with greater wildfire burned area during dry periods than low PWS values.
- **Evaporative Stress Index.** The evaporative stress index (ESI, Anderson et al., 2011) uses geostationary land surface temperature data from GOES to quantify anomalies in the ratio of actual to potential evapotranspiration, an indicator of drought. Positive values of ESI indicate above normal water usage and negative values of ESI indicate below normal water usage.

B1.5 Fuel Loading

Even though vegetative fuels are necessary for fire, this does not guarantee their predictive potential for large fires that continue to burn, because these fires are already fueled by dry vegetation. Here, we use the mass of fuels and the amount that could burn (under low to extreme fire weather) to test their predictive potential. Fuel mass was derived using the 30 m Fuel Characteristic Classification System (FCCS) that represents the structure and composition of wildland fuels to support a wide range of fire behavior (surface, crown, smoldering, flaming, residual) using the six horizontal fuel layers (strata) including the canopy, shrubs, nonwoody, woody, litter-lichen-moss, and duff (Ottmar et al., 2007). Then we calculate the potential for the fuelbeds in each strata to burn based on the 5 classes of NFDRS fire danger classifications (low, moderate, high, very high and extreme). Fuel maps were established at 30 meters across the contiguous United States for every fire danger classification, based on previous work (Soja et al., 2004), for the FIREX-AQ field campaign (Peterson et al., 2022; Warneke et al., 2023).

B1.6 Slope and Elevation

Land surface slope in degrees and elevation in meters was obtained at 30m resolution (LANDFIRE, 2016) and averaged to 990m resolution to allow for faster processing. Standard deviations of both slope and elevation were extracted to track terrain.

B1.7 Firefighting Reports

The SIT-209 system aggregates national situation (SIT)- and incident-level (ICS-209) wildfire management information including fire location, start and end dates, and containment and resources applied to fires through their lifetimes (Jamieson, 2005). In this work, SIT-209 time, location, and incident type data are used to define the wildfire incidents and provide a starting point to define fire polygons. Containment percentage, numbers of resource types deployed, personnel assigned to each resource type, and affected structures are provided at roughly daily resolution and are used as predictors. Resource types considered include: crews, engines, aircraft, construction, and overhead (unassigned floater personnel). Structures affected are split as: structures damaged, structures destroyed, and structures threatened in the next 72 hours. The numbers of resources and affected structures are at times best estimates by incident management and dispatch teams.

B1.8 Population

One key objective of fire management agencies is to protect life and property, so it is reasonable to assume that proximity to population centers would be a predictor of fire suppression activities and thus total wildfire heat output. For this work, Gridded Population of the World, version 4 (GPWv4) population density was pulled (CIESIN-Columbia University, 2018).

Date	irwinID	δBI (1-day)
2020-08-26 00:00:00	2B8959C3-D33F-4B9E-9433-ADB64A57E3E9	0.007206
2020-07-22 00:00:00	795DCBB1-F228-4F7B-9C92-14999DC4B397	24.448432
2020-08-21 00:00:00	7FFA8FCA-1A10-42A6-AB5C-C4CF78119F7D	0.000829
2020-09-12 00:00:00	BCE8E7E2-9777-4C55-837E-A5404927E420	0.003473
2020-09-13 00:00:00	BCE8E7E2-9777-4C55-837E-A5404927E420	36.997390
2020-09-08 00:00:00	C8AD0C86-8BBA-4259-B993-A07DB688C3E1	0.029320
2020-08-24 00:00:00	D4175FA1-8BB6-4E9E-9A8E-C737957CBF67	0.022953

Table B1. δBI Outliers. Values of δBI used to generate the 1-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Date	irwinID	δBI (2-day)
2020-11-04 00:00:00	33CBB9DC-6983-4F47-B821-9C9A6CAC381D	0.038242
2020-07-22 00:00:00	795DCBB1-F228-4F7B-9C92-14999DC4B397	22.683319
2020-11-03 00:00:00	843FC86F-D6D2-4E49-93EA-96F6B0720C3D	0.030518
2020-09-13 00:00:00	BCE8E7E2-9777-4C55-837E-A5404927E420	29.801551
2020-08-21 00:00:00	DD23418F-7040-439F-8024-1441931BA3F0	0.004292

Table B2. δBI Outliers. Values of δBI used to generate the 2-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Date	irwinID	δHWP (one day)
2020-09-10 00:00:00	ECE62D4A-E790-496E-8C53-991E9E7028D5	0.031978
2020-09-11 00:00:00	ECE62D4A-E790-496E-8C53-991E9E7028D5	28.748383

Table B3. δHWP Outliers. Values of δHWP used to generate the 1-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

Date	irwinID	δHWP (two day)
2020-09-11 00:00:00	ECE62D4A-E790-496E-8C53-991E9E7028D5	75.092339

Table B4. δHWP Outliers. Values of δHWP used to generate the 2-day forecast from day 0 FRE which are related to unphysical increases or decreases in fire behavior.

a)

	FRE_R2	FRE_R2_adj
persistence	0.016	0.015
features_all	0.479	0.447
features_no_persistence	0.451	0.418
features_no_surface_weather	0.302	0.277
features_only_raw_gridmet	0.389	0.366
features_only_raw_hrrrmet	0.355	0.33
features_only_temperatures_gridmet	0.31	0.284
features_only_heatwave	0.297	0.271
features_only_hd0w0	0.418	0.397
features_only_hd1w0	0.404	0.382
features_only_hd2w0	0.375	0.352
features_only_hd3w0	0.393	0.37
features_only_hwp	0.474	0.454
features_only_cffdrs	0.444	0.423
features_only_nfdrs	0.429	0.407
features_no_stability	0.485	0.455
features_only_pft	0.483	0.452
features_only_chi	0.479	0.448
features_no_living_moisture	0.478	0.446
features_no_dead_moisture	0.476	0.449
features_only_noaa_sm	0.485	0.458
features_only_ncar_sm	0.478	0.451
features_only_cffdrs_sm	0.478	0.45
features_only_nfdrs_sm	0.478	0.451
features_only_hrrr_sm	0.477	0.448
features_no_pws	0.473	0.441
features_no_esi	0.485	0.454
features_no_loading	0.479	0.449
features_only_LowN	0.484	0.454
features_only_ModerateN	0.483	0.453
features_only_HighN	0.481	0.451
features_only_VeryHighN	0.48	0.45
features_only_ExtremeN	0.479	0.449
features_no_terrain	0.479	0.449
features_only_slope	0.483	0.452
features_only_elevation	0.478	0.447
features_no_doy	0.479	0.448
features_no_population	0.48	0.448
features_no_containment	0.475	0.443
features_no_structures	0.479	0.448
features_only_destroyed_structures	0.479	0.447
features_only_damaged_structures	0.479	0.447
features_only_threatened_structures	0.479	0.448
features_no_resource_quantity	0.478	0.448
features_only_crew_quantity	0.477	0.446
features_only_engine_quantity	0.479	0.448
features_only_aircraft_quantity	0.478	0.448
features_only_construction_quantity	0.478	0.447
features_no_resource_personnel	0.481	0.451
features_only_crew_personnel	0.481	0.451
features_only_engine_personnel	0.481	0.45
features_only_aircraft_personnel	0.48	0.45
features_only_construction_personnel	0.481	0.451
features_only_overhead_personnel	0.477	0.447

b)

	FRE_R2	FRE_R2_adj
persistence	0.016	0.015
features_all	0.48	0.448
features_hrrrmet	0.388	0.381
features_gridmet	0.422	0.415
features_hwp	0.466	0.461
features_hd0w0	0.461	0.456
features_hrrrmet_plus_hwp	0.485	0.478
features_hrrrmet_plus_hd0w0	0.414	0.407
features_cffdrs	0.484	0.478
features_nfdrs	0.446	0.439
ens	0.483	0.451

c)

	FRE_R2	FRE_R2_adj
persistence	-0.618	-0.619
features_all	0.251	0.199
features_hrrrmet	0.131	0.12
features_gridmet	0.169	0.158
features_hwp	0.306	0.299
features_hd0w0	0.309	0.302
features_hrrrmet_plus_hwp	0.343	0.334
features_hrrrmet_plus_hd0w0	0.268	0.257
features_cffdrs	0.092	0.079
features_nfdrs	0.195	0.185
ens	0.272	0.221

d)

	FRE_R2	FRE_R2_adj
SF_pers1	0.016	0.015
dvpd_gridmet_pred1	0.128	0.127
dwind_speed_gridmet_pred1	0.191	0.19
dvpd_hrrrmet_pred1	0.156	0.155
dwind_speed_hrrrmet_pred1	0.345	0.344
dhwp_pred1	0.437	0.436
dhd0w0_pred1	0.418	0.418
dhd1w0_pred1	0.345	0.344
dhd2w0_pred1	0.315	0.314
dhd3w0_pred1	0.332	0.33
dIMERG_FINAL_v6_FWI_pred1	0.264	0.263
dburning_index_g_pred1	0.183	0.182
SF_pred1	0.48	0.448

e)

	FRE_R2	FRE_R2_adj
SF_pers2	-0.618	-0.619
dvpd_gridmet_pred2	-0.356	-0.358
dwind_speed_gridmet_pred2	-0.398	-0.4
dvpd_hrrrmet_pred2	-0.434	-0.436
dwind_speed_hrrrmet_pred2	0.006	0.004
dhwp_pred2	0.262	0.261
dhd0w0_pred2	0.213	0.212
dhd1w0_pred2	0.242	0.241
dhd2w0_pred2	0.111	0.109
dhd3w0_pred2	-0.215	-0.217
dIMERG_FINAL_v6_FWI_pred2	-0.304	-0.306
dburning_index_g_pred2	-0.219	-0.221
SF_pred2	0.251	0.199

Table B5. Comparison of R^2 and R^2_{adj} for a) sensitivity analysis, b) 1-day optimal subsets, c) 2-day optimal subsets, d) 1-day RF (all features) and e) 2-day RF (all features)

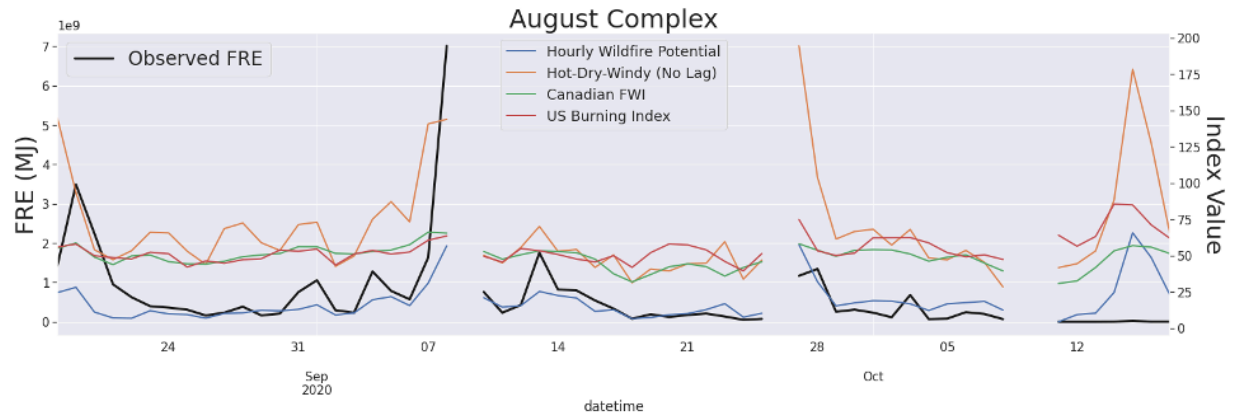

Figure B1. Time series of FRE (black), hourly wildfire potential (blue), hot-dry-windy index (orange), Canadian Fire Weather Index (green), US Burning Index (red) for the August Complex fire.

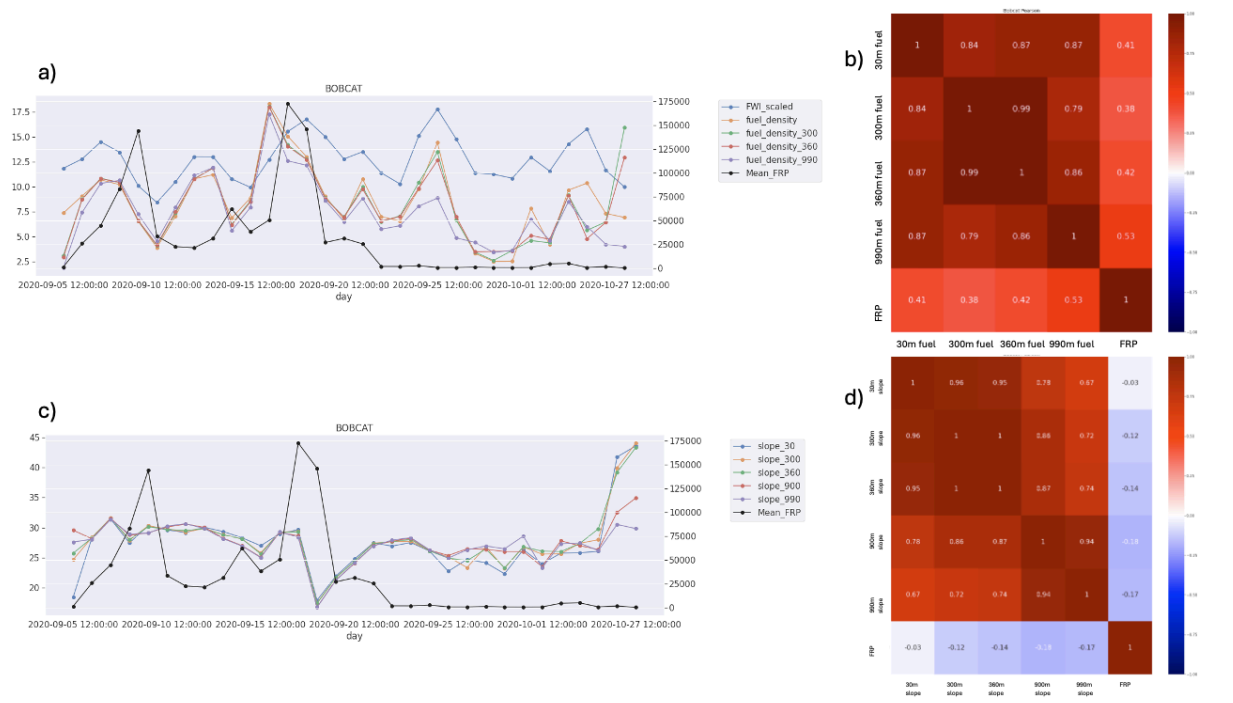

Figure B2. Example of the impact of grid resolution on polygon-averaged fuel density (a-b) and slope (c-d). a) and c) show a time series of FRP (black) and fuel density/slope at different grid resolutions (colored lines) for the Bobcat Fire. b) and d) show pearson correlation for fuel density/slope at different grid resolutions.

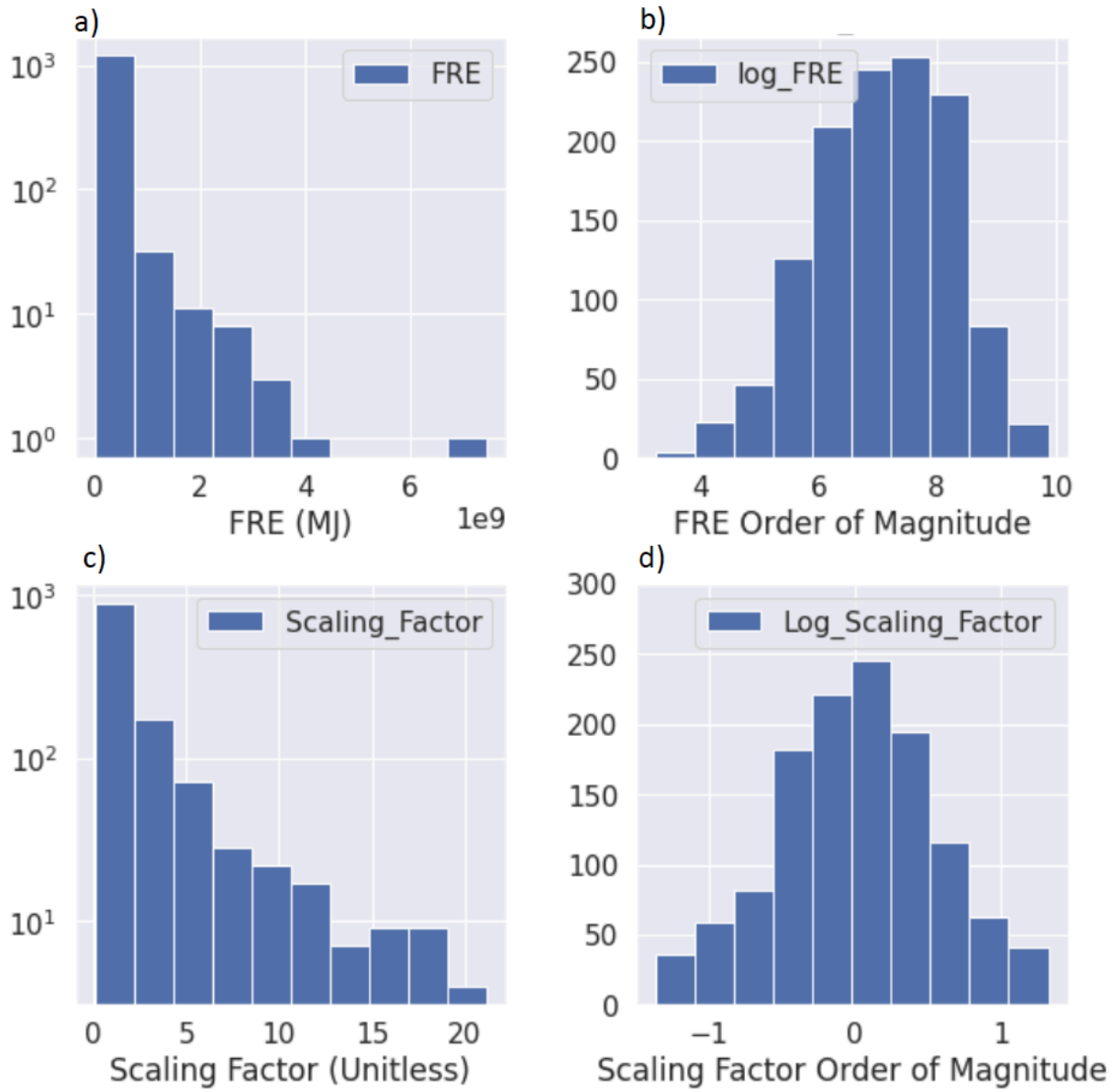

Figure B3. Distributions of a) total daily FRE, b) log(FRE), and c) δFRE, and d) log(δFRE) for the training years, 2019 and 2021. a) and c) use a log scale for the y axis, c) and d) use a linear scale. Outliers which lie more than 1.5 interquartile ranges from the median of the log of δFRE are removed.

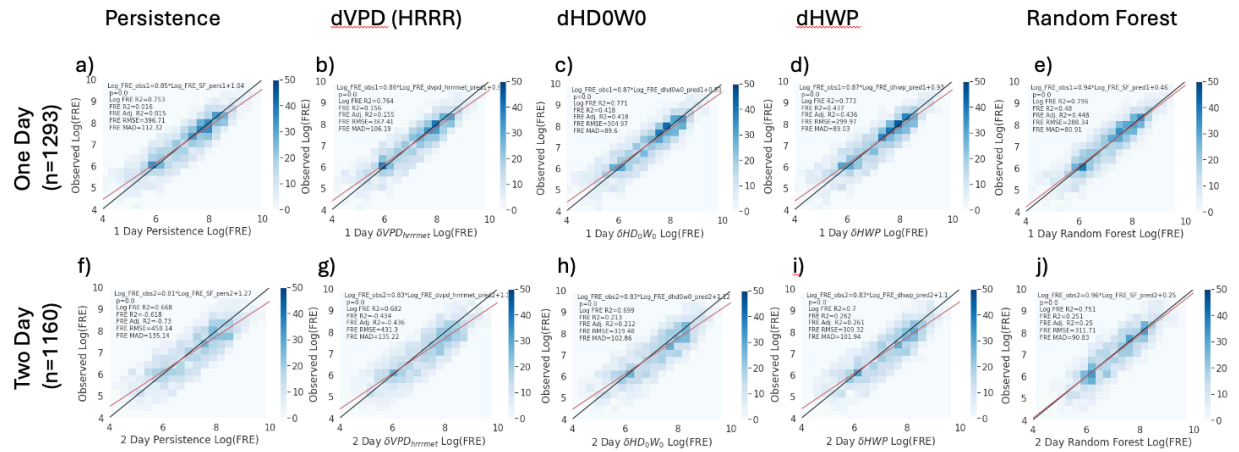

Figure B4. Density plots of observed and modeled FRE as predicted by a,f) persistence, b,g) scaled VPD from HRRR, c,h) scaled HD0W0, d,i) scaled HWP, e,j) the random forest trained on all variables. The red line is the best fit to the log of the data and the black line is the 1:1 line. The best fit equation, R^2, RMSE, and MAD are also provided for each model. The top row is for the 1-day forecast and the bottom row is for the 2-day forecast.

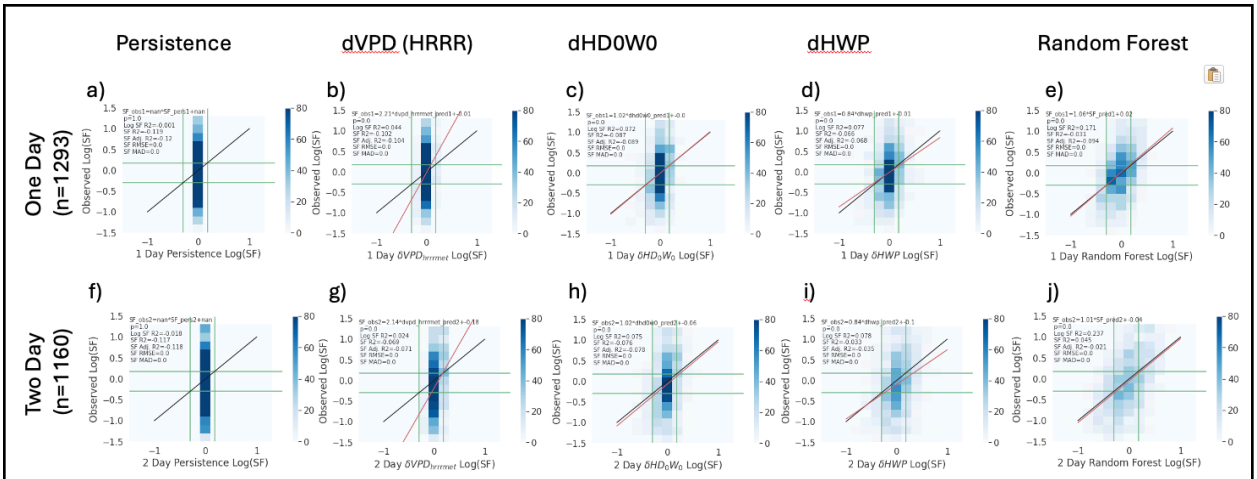

Figure B5. Density plots of observed and modeled scaling factors as predicted by a,f) persistence, b,g) scaled VPD from HRRR, c,h) scaled HD0W0, d,i) scaled HWP, e,j) the random forest trained on all variables. The red line is the best fit to the log of the data and the black line is the 1:1 line. The best fit equation, R^2, RMSE, and MAD are also provided for each model. The top row is for the 1-day forecast and the bottom row is for the 2-day forecast.

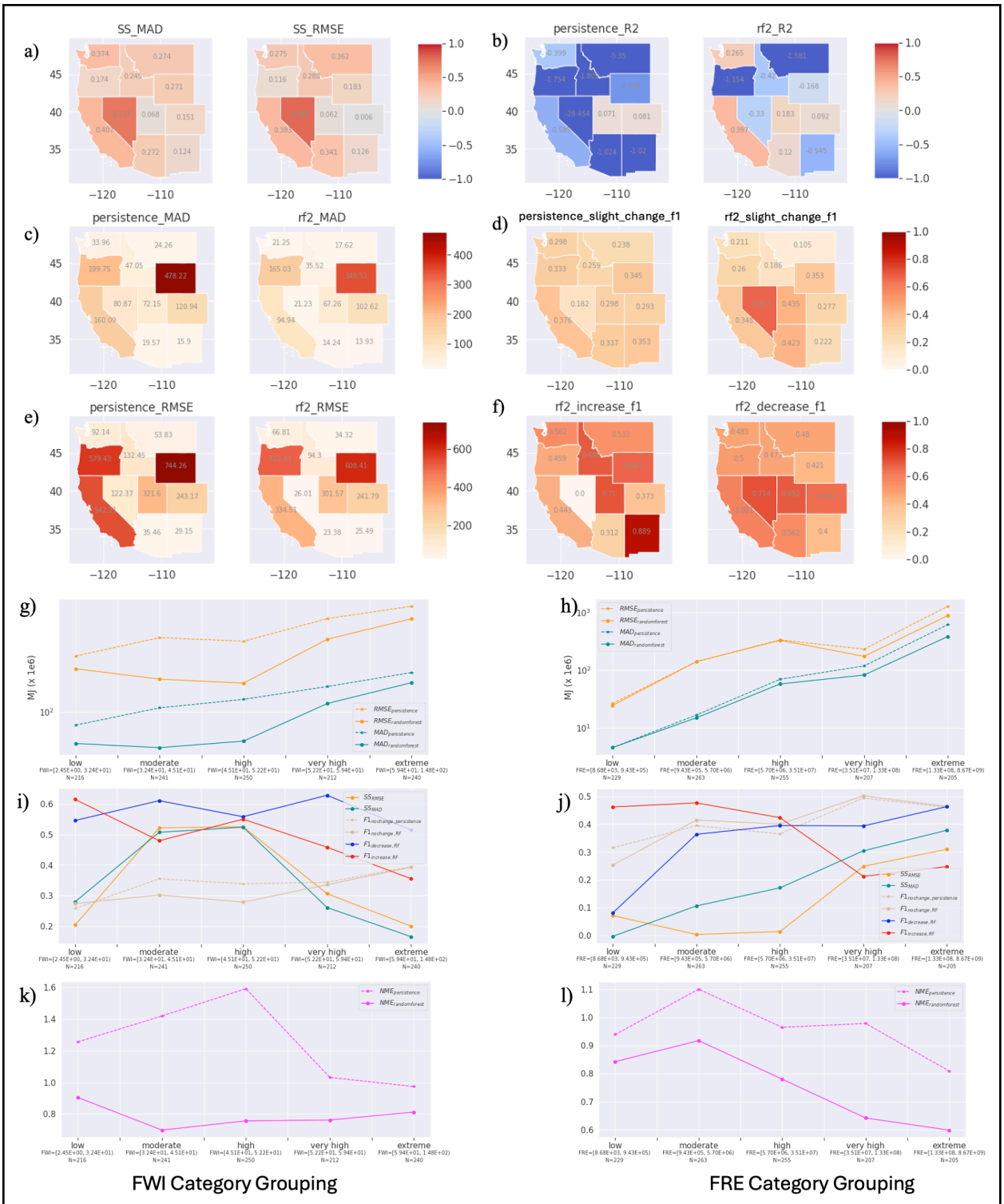

Figure B6. Random forest evaluation metrics by state (a-f) and by fire severity (g-j) for the 2-day RF and persistence predictions. a) SS_{MAD} and SS_{RMSE} for the RF model; b) R^2 for the persistence and RF models; c) MAD for the persistence and RF models; d) $F1_{slight\_change}$ for the persistence and RF models; e) RMSE for the persistence and RF models; f) $F1_{increase}$ and $F1_{decrease}$ for the RF model; g) RMSE (yellow) and MAD (teal) for the persistence (dotted) and

RF (solid) models split by 20th-percentiles of FWI; h) as in f but split by 20th-percentiles of FRE. i) SS_{MAD} (teal), SS_{RMSE} (yellow), $F1_{increase}$ (red), $F1_{decrease}$ (blue), and $F1_{slight\_change}$ (tan) for the RF models with $F1_{slight\_change}$ (tan dotted) for the persistence model split by 20th percentiles of FWI. j) as in i but split by 20th percentiles of FRE.

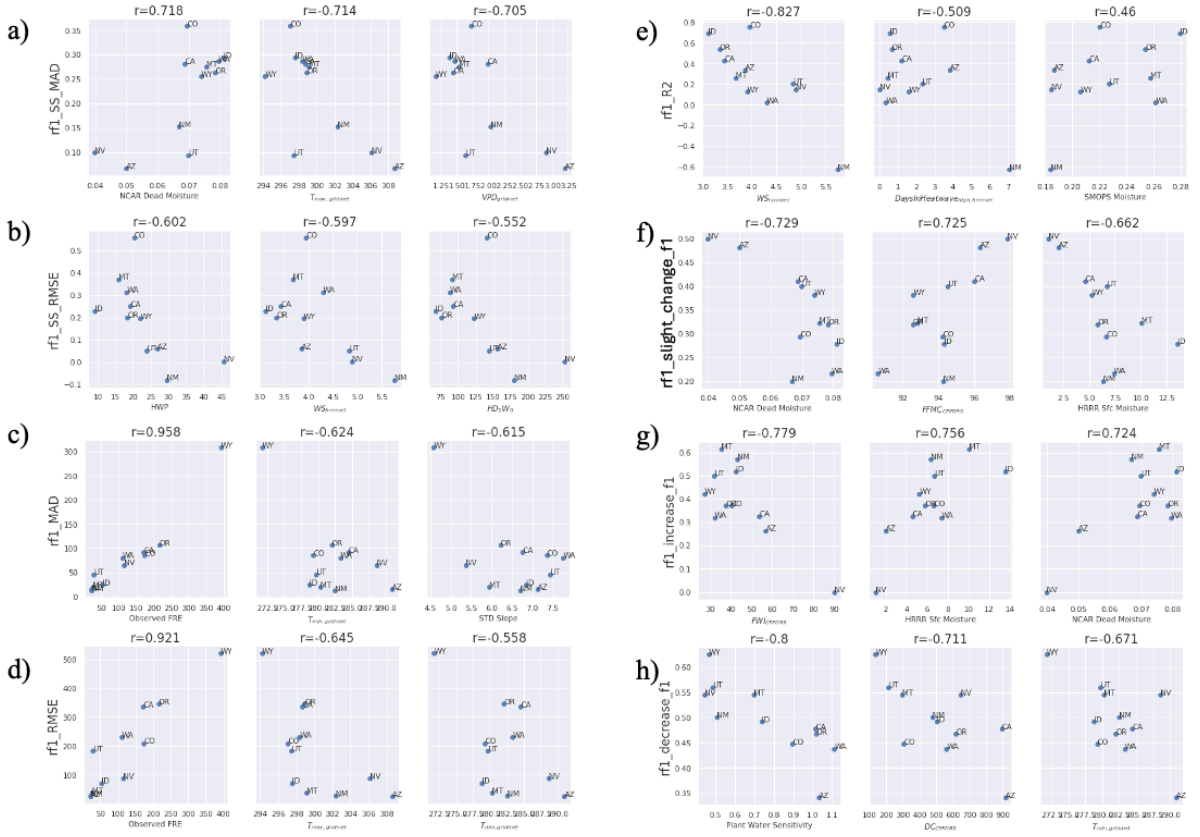

Figure B7. Scatter plots of average error vs average values of predictor by state for the 1-day random forest predictions. Shown are the top 3 predictors whose state-averaged values had the highest person correlation (r^2) with each error type. The metrics presented (y-axes) include a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{no\ change}$, g) $F1_{increase}$, and h) $F1_{decrease}$. The x-axes are labeled with the predictor whose average values are plotted.

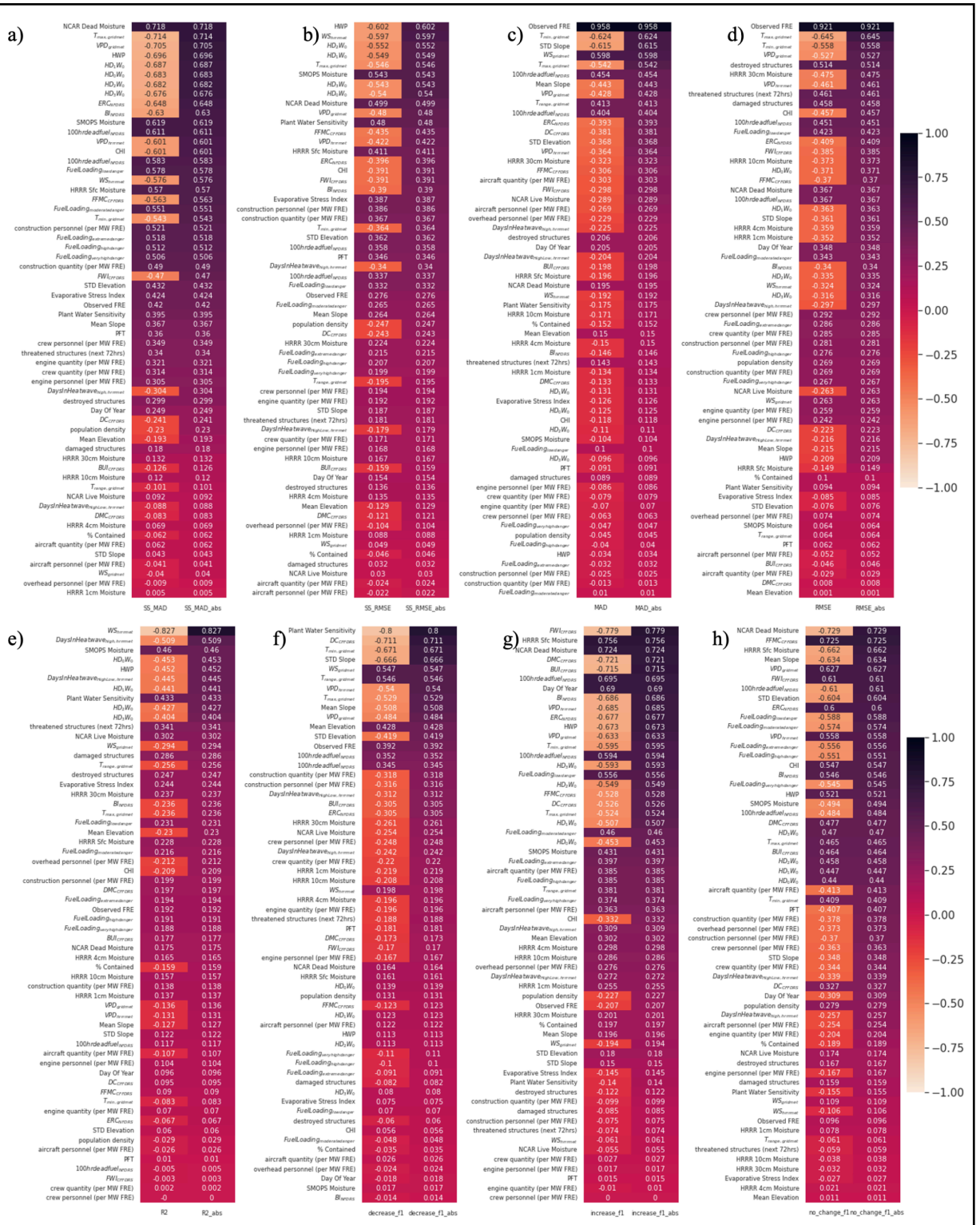

Figure B8. Pearson correlation (r^2) and correlation strength (absolute value of r^2, shading, annotated numbers) between error metrics and state-average random forest input variables (row labels) for the 1-day forecasts. Pearson correlations are shown for a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{decrease}$, g) $F1_{increase}$, h) $F1_{no\ change}$

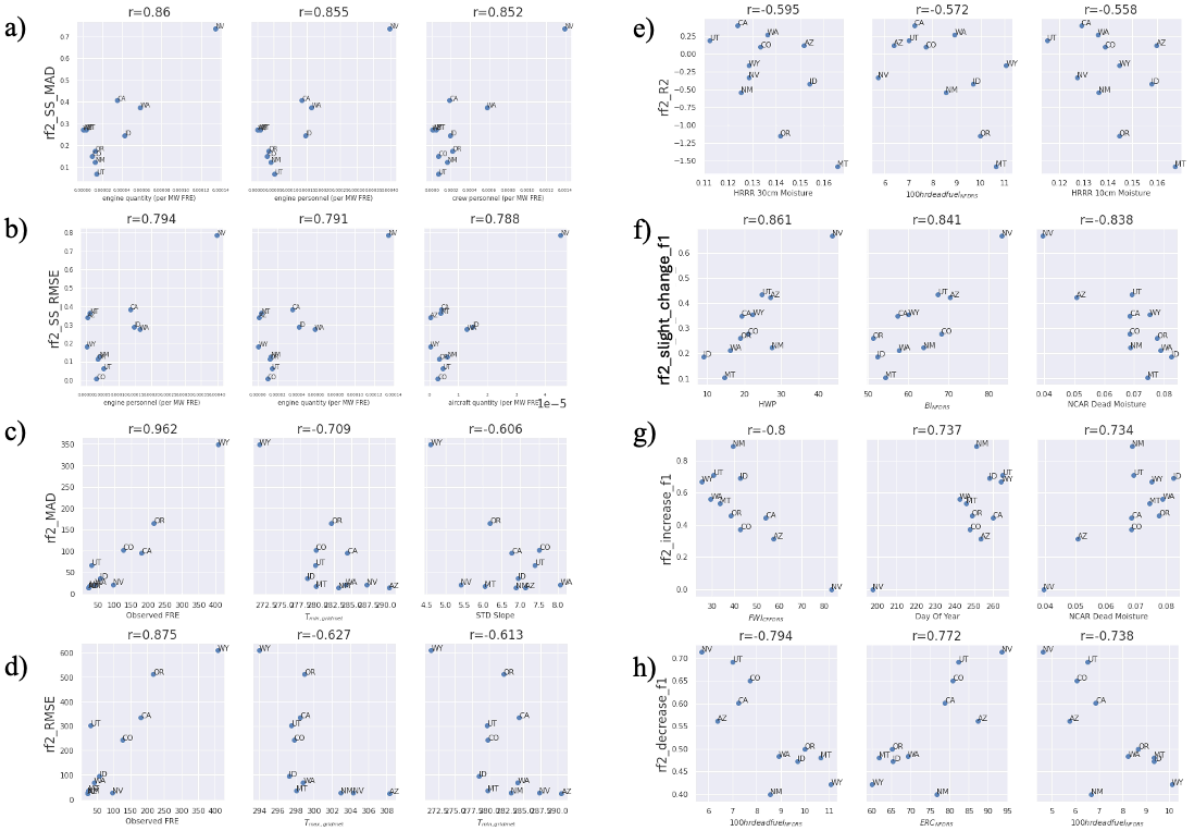

Figure B9. Scatter plots of average error vs average values of predictor by state for the 2-day random forest predictions. Shown are the top 3 predictors whose state-averaged values had the highest person correlation (r^2) with each error type.The metrics presented (y-axes) include a) SS_{MAD} , b) SS_{RMSE} , c) MAD, d) RMSE, e) R^2, f) $F1_{no\ change}$, g) $F1_{increase}$, and h) $F1_{decrease.}$ The x-axes are labeled with the predictor whose average values are plotted.

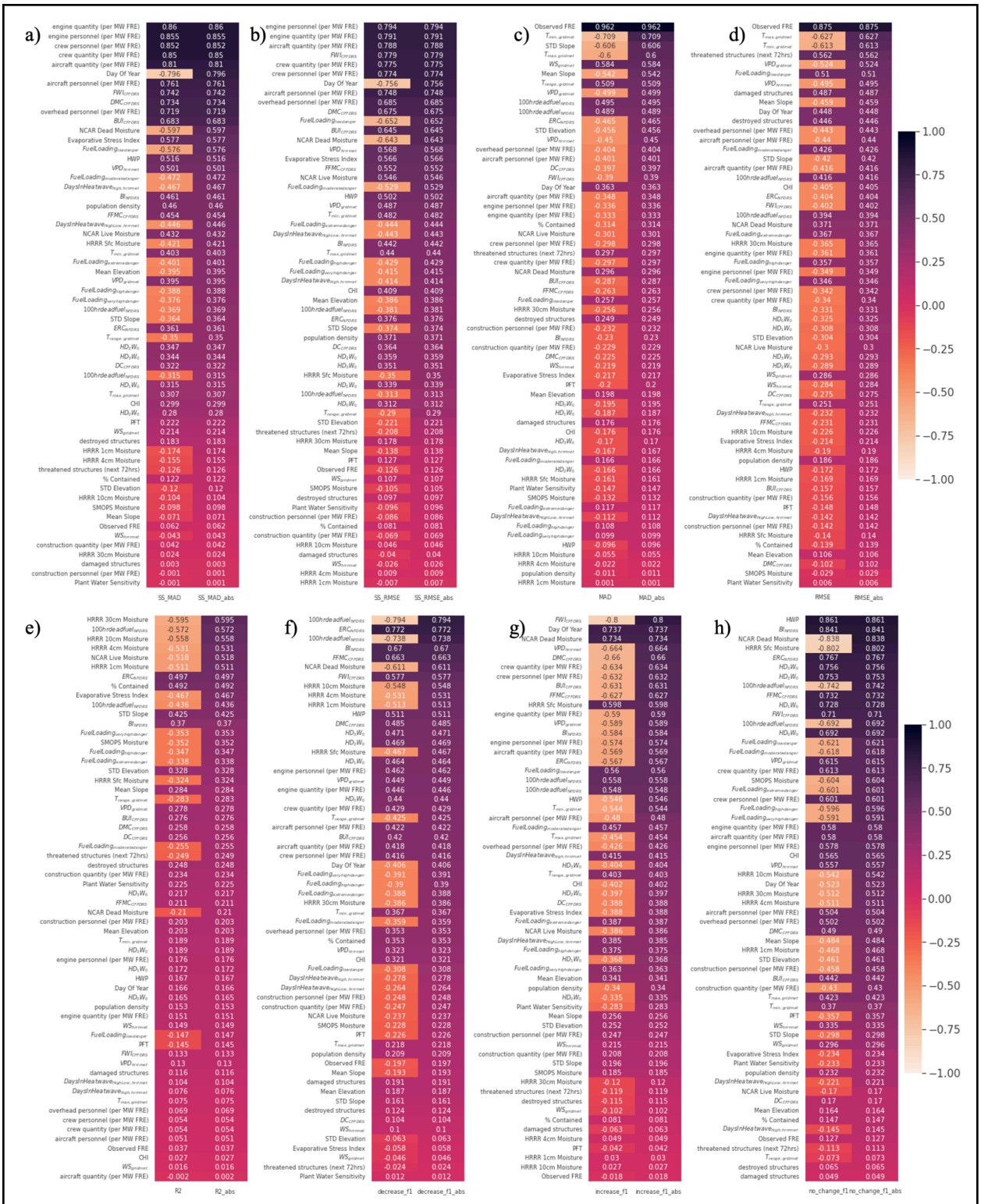

Figure B10. Pearson correlation (r^2) and correlation strength (absolute value of r^2, shading, annotated numbers) between error metrics and state-average random forest input variables (row labels) for the 2-day forecasts. Pearson correlations are shown for a) SS_{MAD}, b) SS_{RMSE}, c) MAD, d) RMSE, e) R^2, f) $F1_{decrease}$, g) $F1_{increase}$, h) $F1_{no\ change}$

Figure B11. Time series of total FRE (black line, left axis) and model 1-day forecast error (teal and gold lines) for the 2020 fire season. a) shows daily total FRE over the western US (black) along with RMSE (gold, right axis) and MAD (green, right axis) for the random forest (solid) and persistence (dashed) forecast methods. b) shows daily total FRE over the western US along with the SS_{RMSE}(gold) and SS_{MAD} (teal) compared with persistence. The dashed black line is at 0 on the skill score axis; points above this line are days where the random forest improves on persistence on average and points below this line are days where the random forest is worse than persistence on average. In a) and b) the size of the stars is proportional to the number of fires burning in the Western US domain on that day.

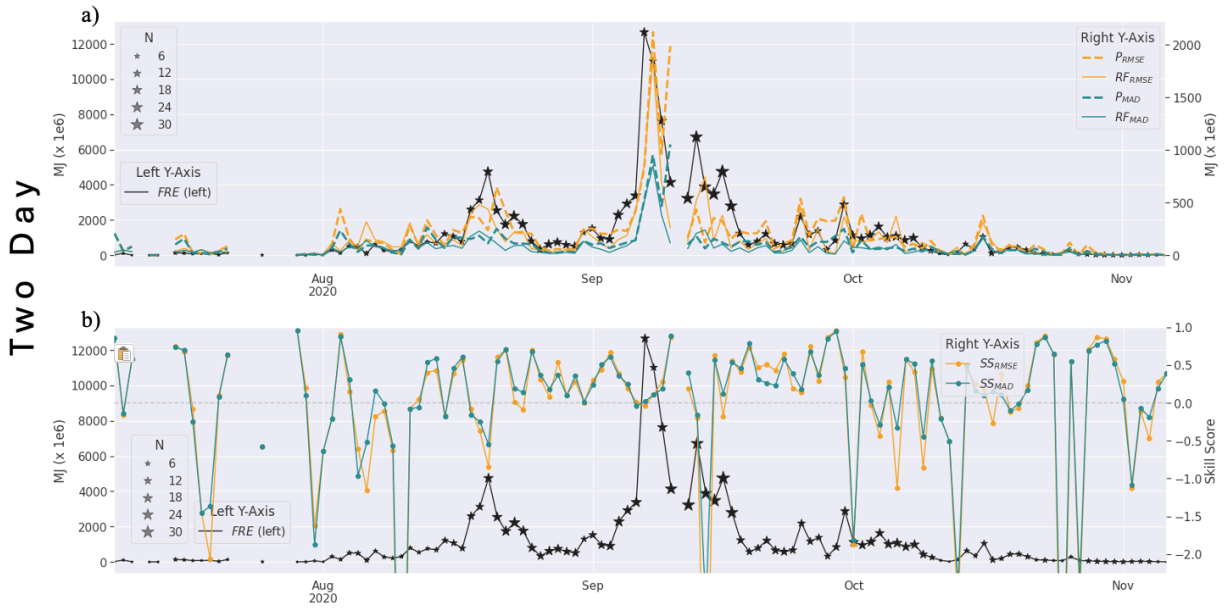

Figure B12. Time series of total FRE (black line, left axis) and model 2-day forecast error (teal and gold lines) for the 2020 fire season. a) shows daily total FRE over the western US (black) along with RMSE (gold, right axis) and MAD (green, right axis) for the random forest (solid) and persistence (dashed) forecast methods. b) shows daily total FRE over the western US along with the SS_{RMSE}(gold) and SS_{MAD} (teal) compared with persistence. The dashed black line is at 0 on the skill score axis; points above this line are days where the random forest improves on persistence on average and points below this line are days where the random forest is worse than persistence on average. In a) and b) the size of the stars is proportional to the number of fires burning in the Western US domain on that day.

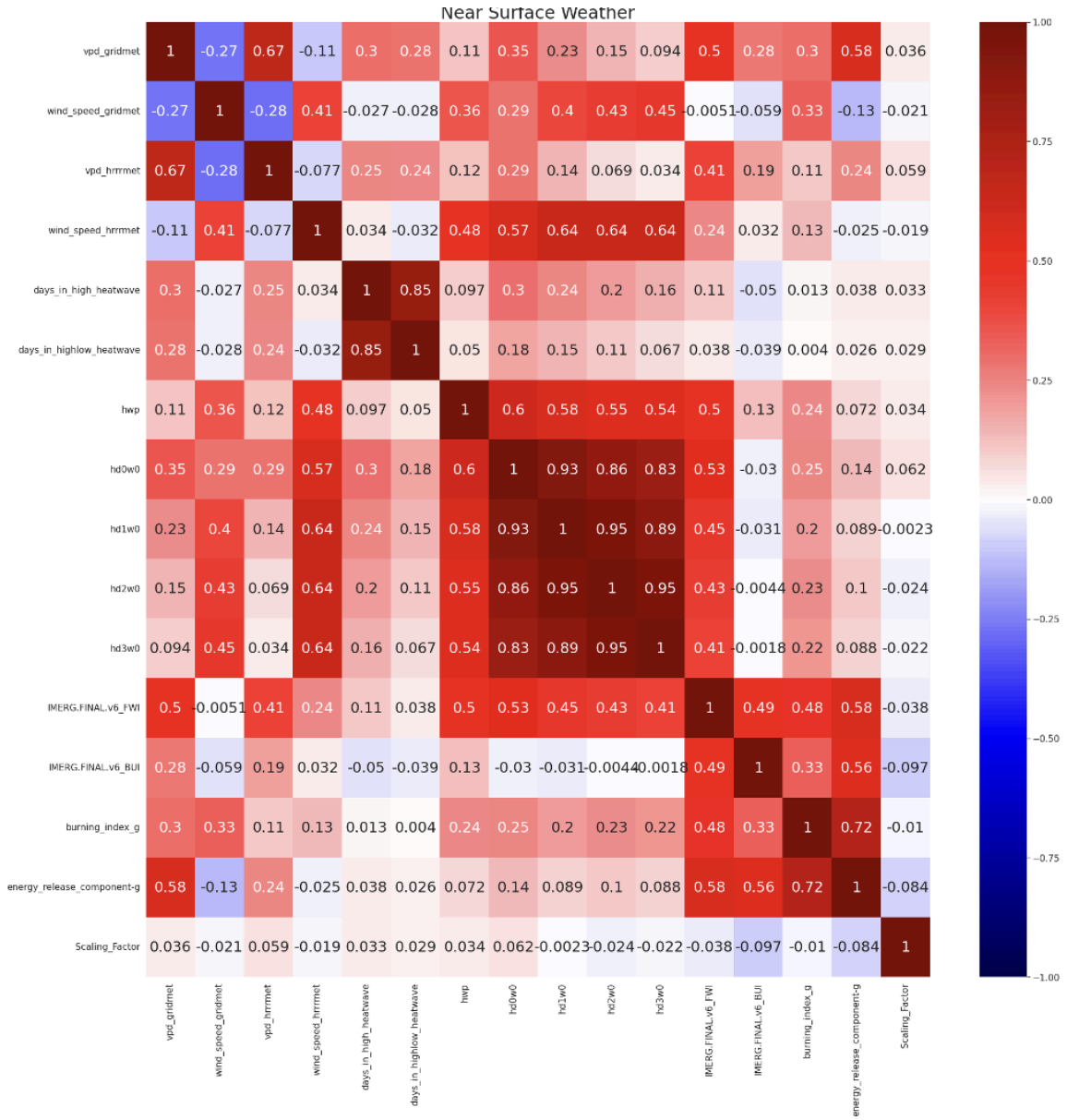

Figure B13. Correlation heatmap for near-surface weather variables. Scaled weather variables are not shown.

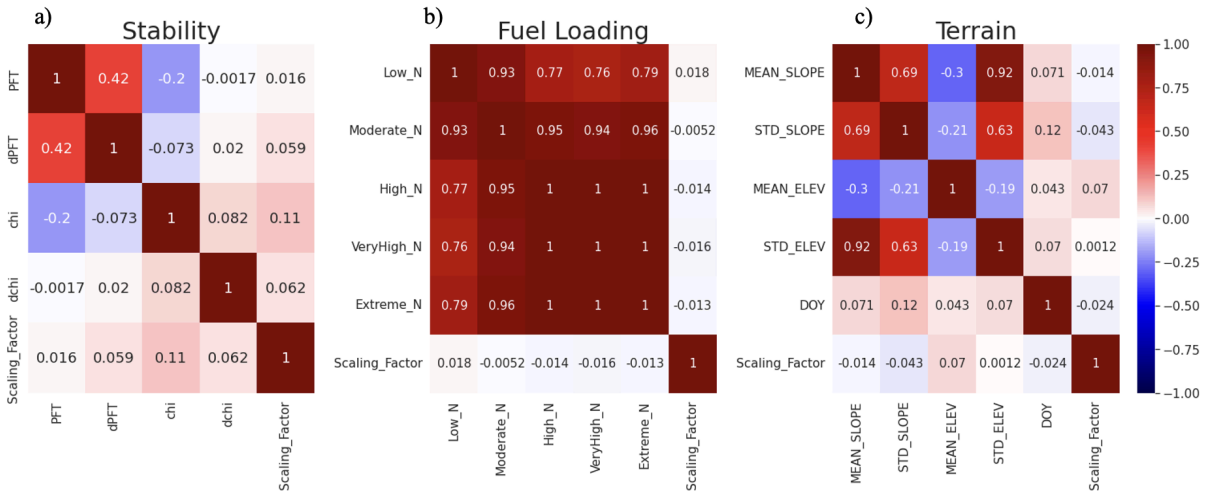

Figure B14 Correlation heatmap for a) stability, b) fuel loading, and c) terrain variables.

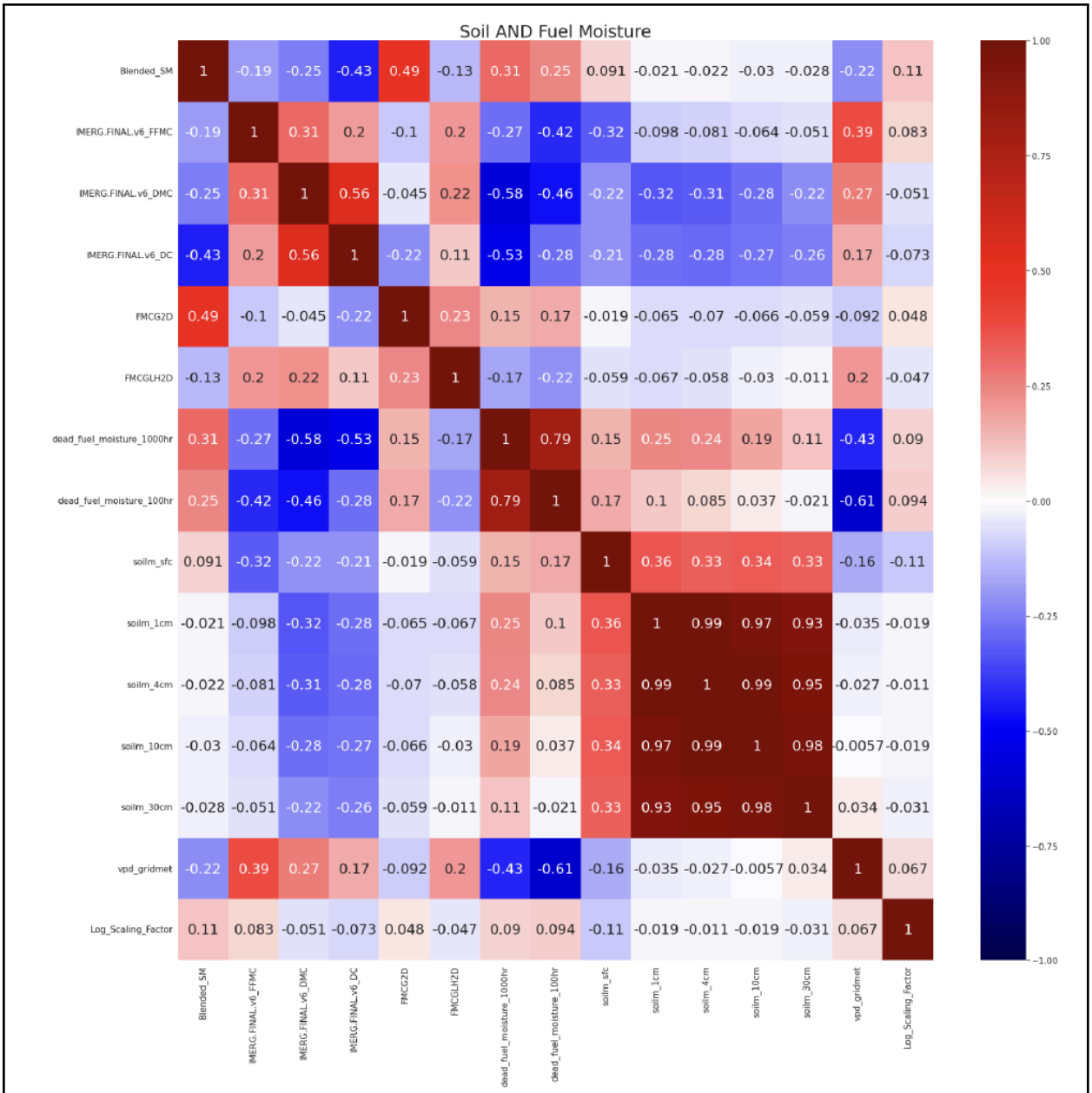

Figure B15. Correlation heatmap for moisture variables.

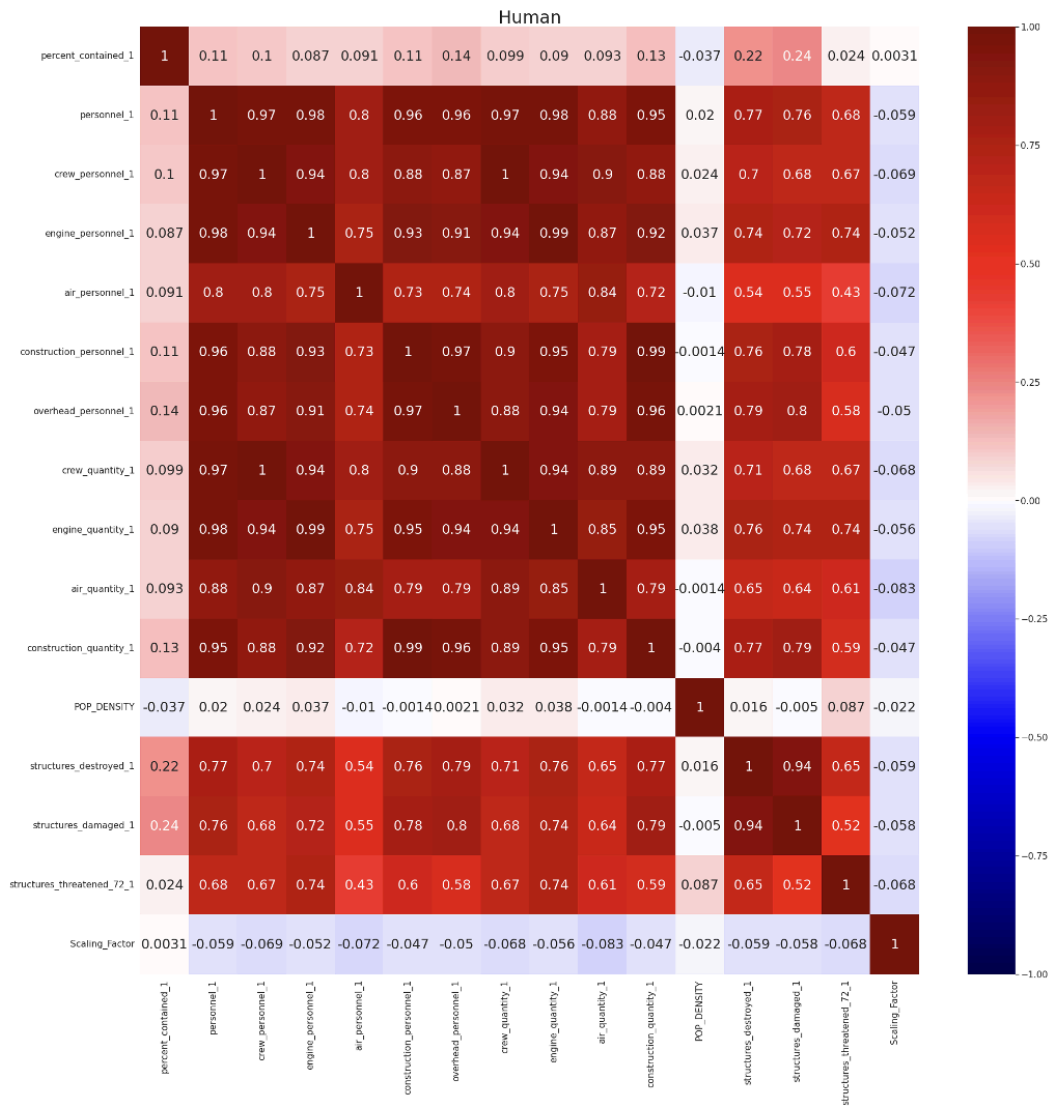

Figure B16. Correlation heatmap for human variables.

Appendix C: Air quality model sensitivity to machine learning-derived fire emissions

Contents of this file

Figures C1 to C27

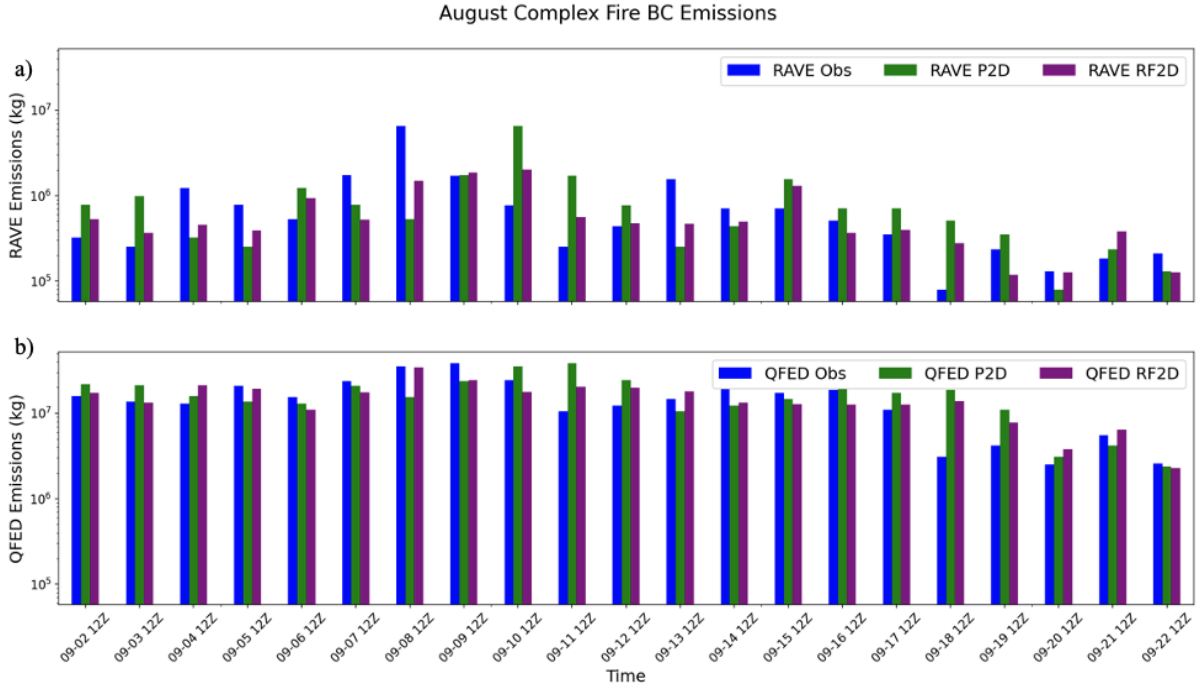

Figure C1. Black carbon (BC) emissions from August Complex Fire from the a) RAVE and b) QFED emissions inventories for September 2nd-September 22nd, 2020. Reference emissions are plotted in blue, and *2-day* persistence and random forest emissions are plotted in green and purple respectively. Note that the persistence assumption dictates that the amount of emissions one day (green bar) is the same as the emissions amount the day before (previous day blue bar).

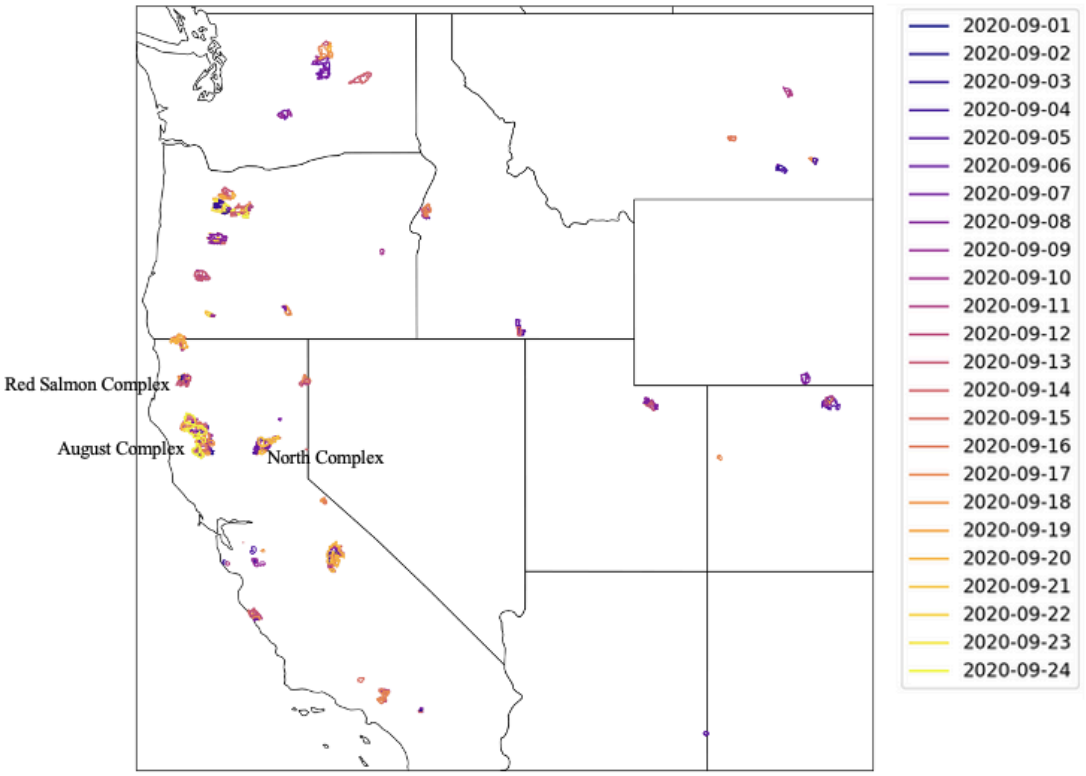

Figure C2. Map of fires burning in September 2020, with colored shading representing the day of burning, defined as 12Z on the labeled calendar day to 12Z on the following calendar day. Certain fires are noted by name if mentioned in the text or figures.

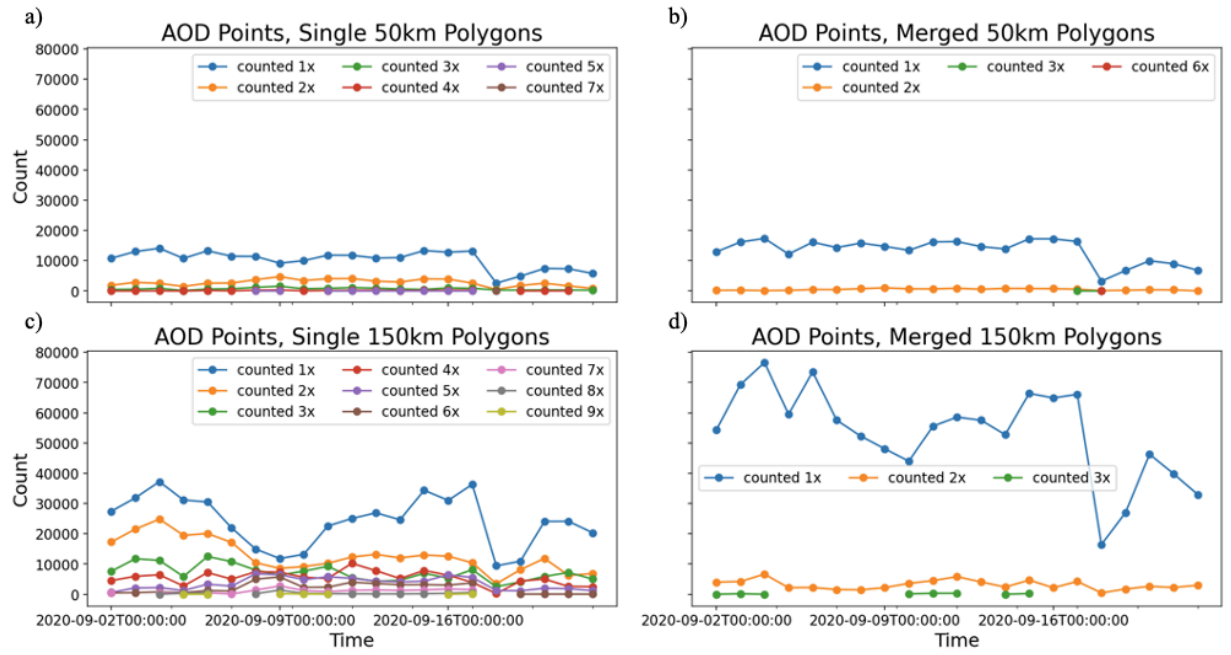

Figure C3. Time series of the number of AOD grid cells which are counted once (blue line) and more than once (other colored lines) for 50 km (a-b) and 150 km (c-d) buffers. a,c) show the number of multi-counts for polygons based on individual fires and b,d) show the number of multi counts for merged polygons.

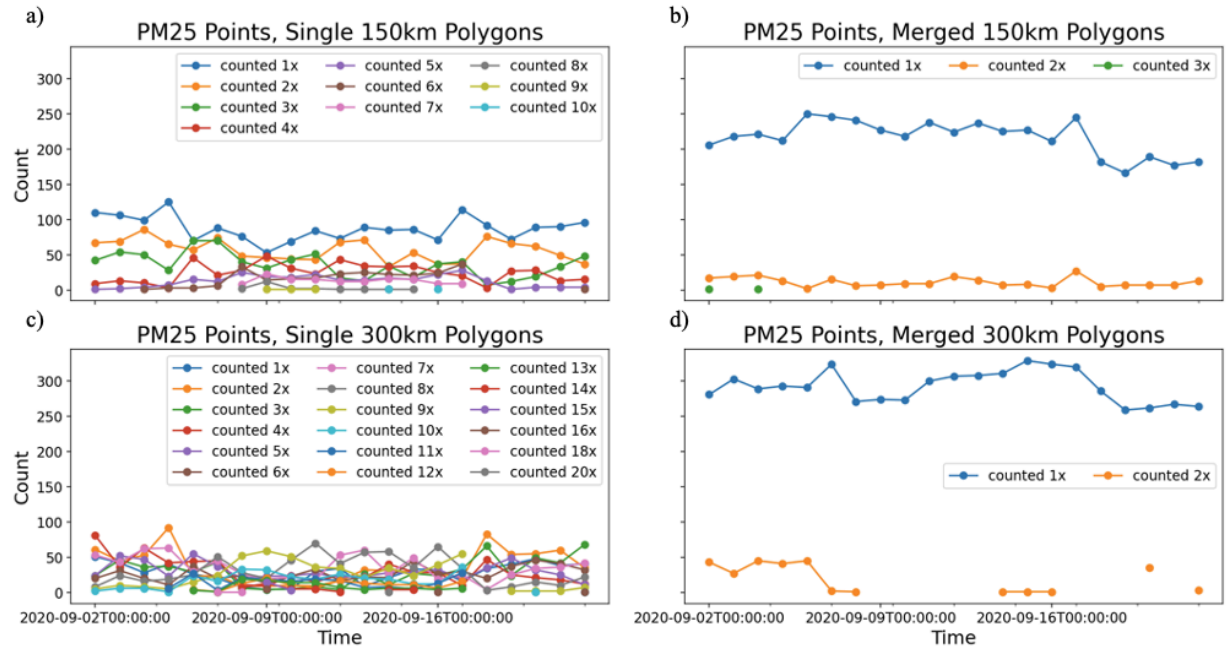

Figure C4. Time series of the number of PM2.5 stations which are counted once (blue line) and more than once (other colored lines) for 150 km (a-b) and 300 km (c-d) buffers. a,c) show the number of multi-counts for polygons based on individual fires and b,d) show the number of

multi counts for merged polygons.

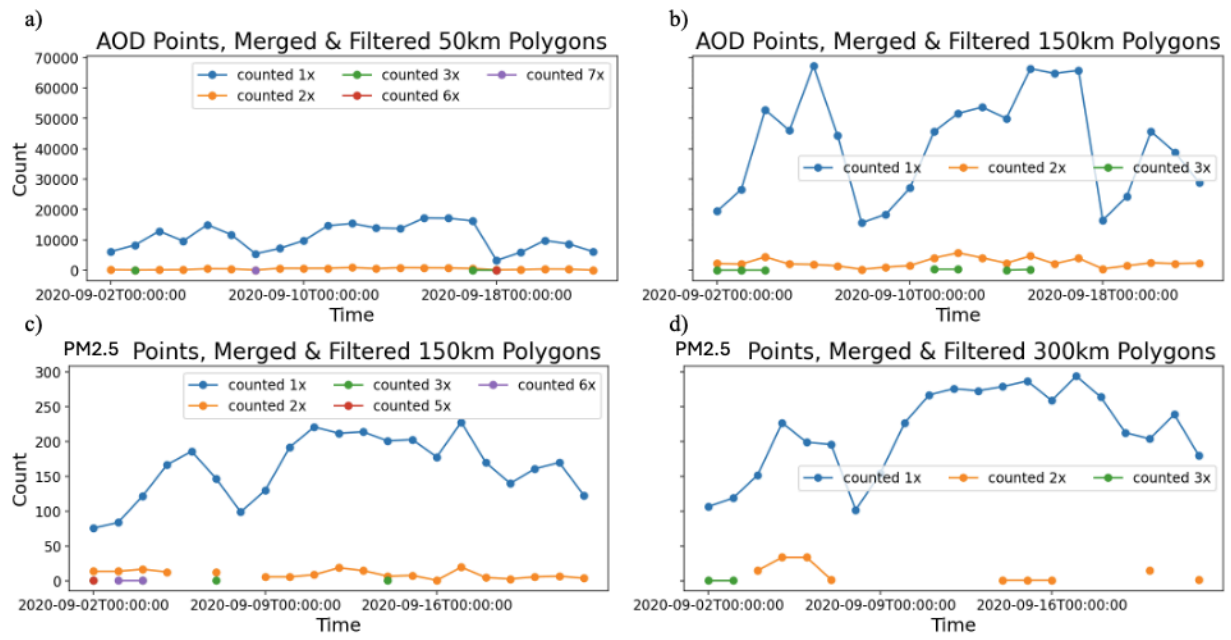

Figure C5. Time series of the number of AOD grid cells (a,b) and PM2.5 stations (c,d) which are counted once (blue line) and more than once (other colored lines) for merged polygons. a,c) show the number of multi-counts for unfiltered data and b,d) show the number of multi counts for filtered data.

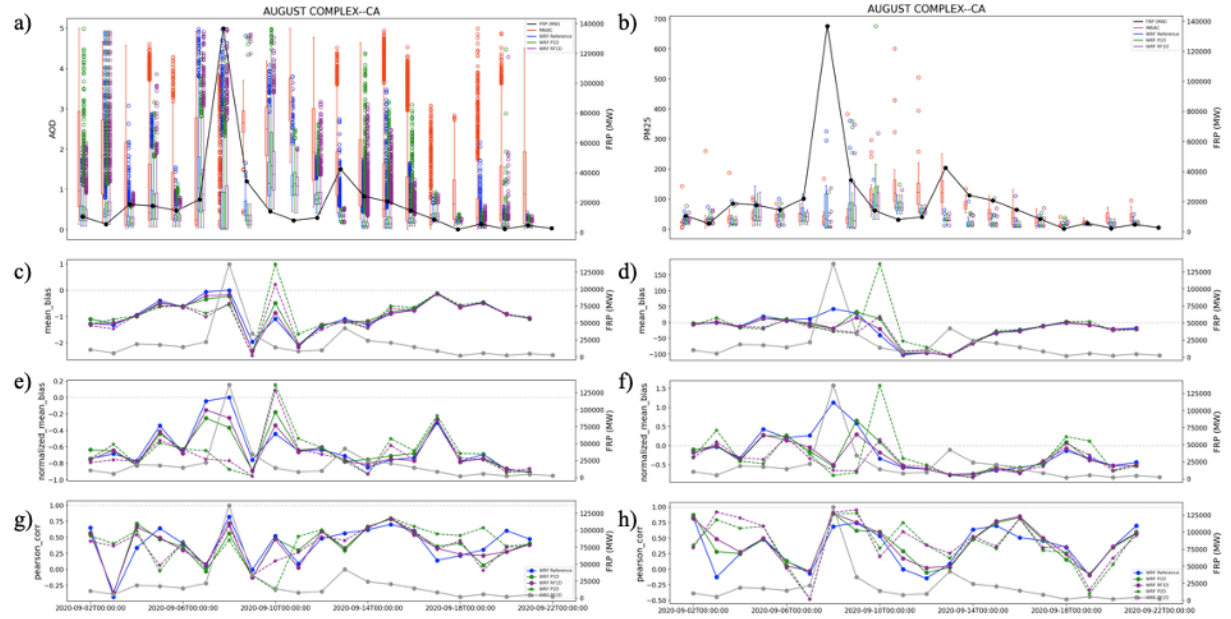

Figure C6. August Complex Fire time series of AOD (a,c,e,g) and PM2.5 (b,d,f,h) predictions, observations and error metrics. a,b) are a time series of observed AOD (a)/PM2.5 (b) (red), reference model (blue), 1-day persistence (green), and 1-day random forest (purple) distributions. The other time series show mean bias (c,d), normalized mean bias (e,f) and pearson correlation (g,h) for one day (line with filled circle markers) and two day (line with star markers) predictions. (a,c,e,g) is a repeat of Figure 2 g-j)

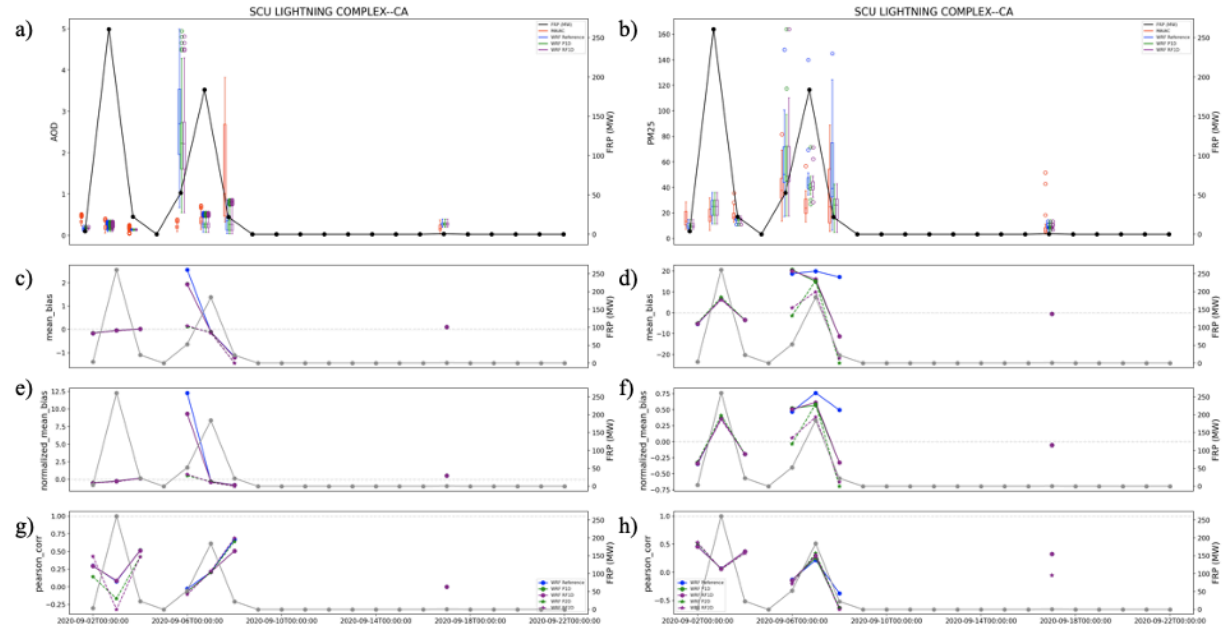

Figure C7. As in S6 for the SCU Lightning Complex fire in CA

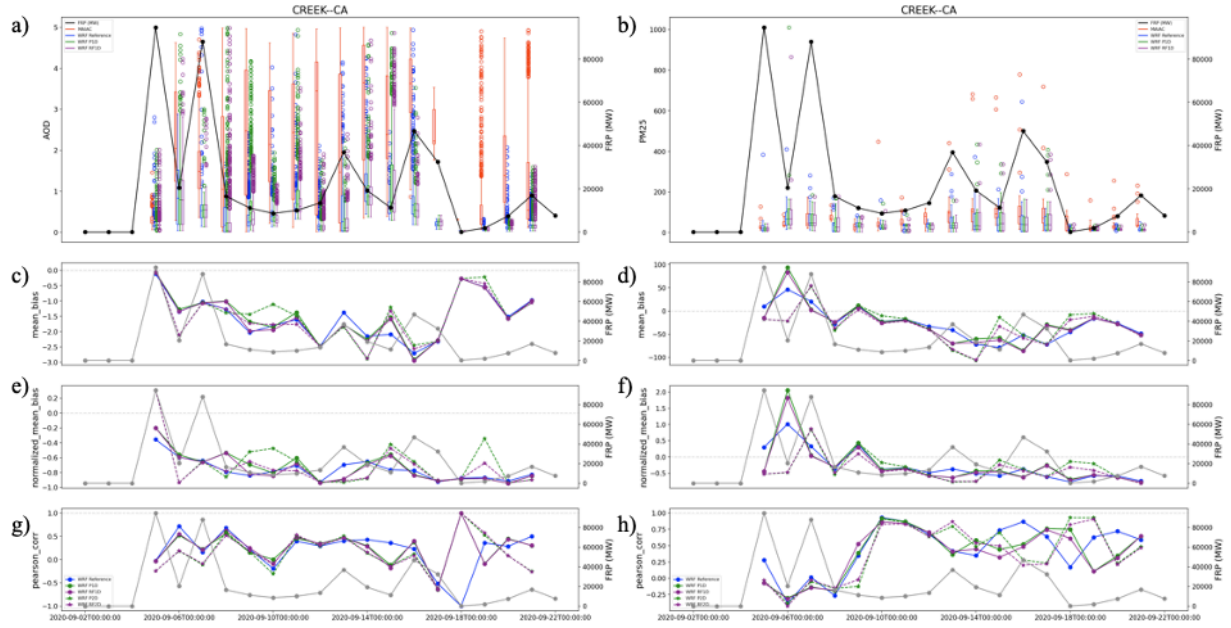

Figure C8. As in S6 for the Creek fire in CA

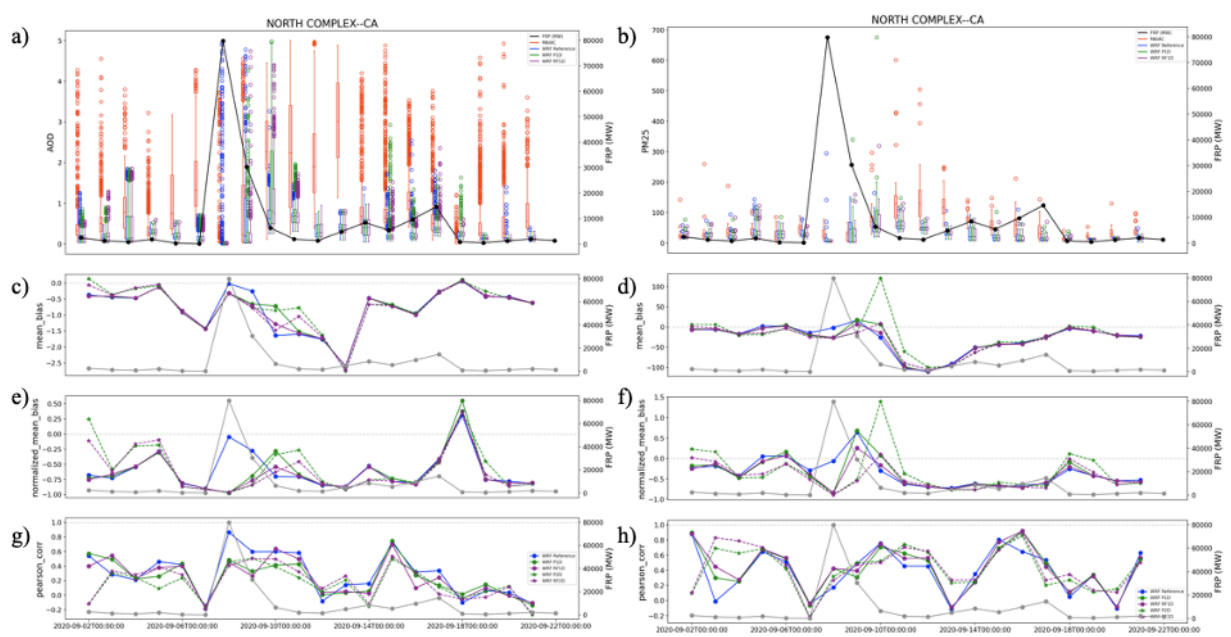

Figure C9. As in S6 for the North Complex fire in CA

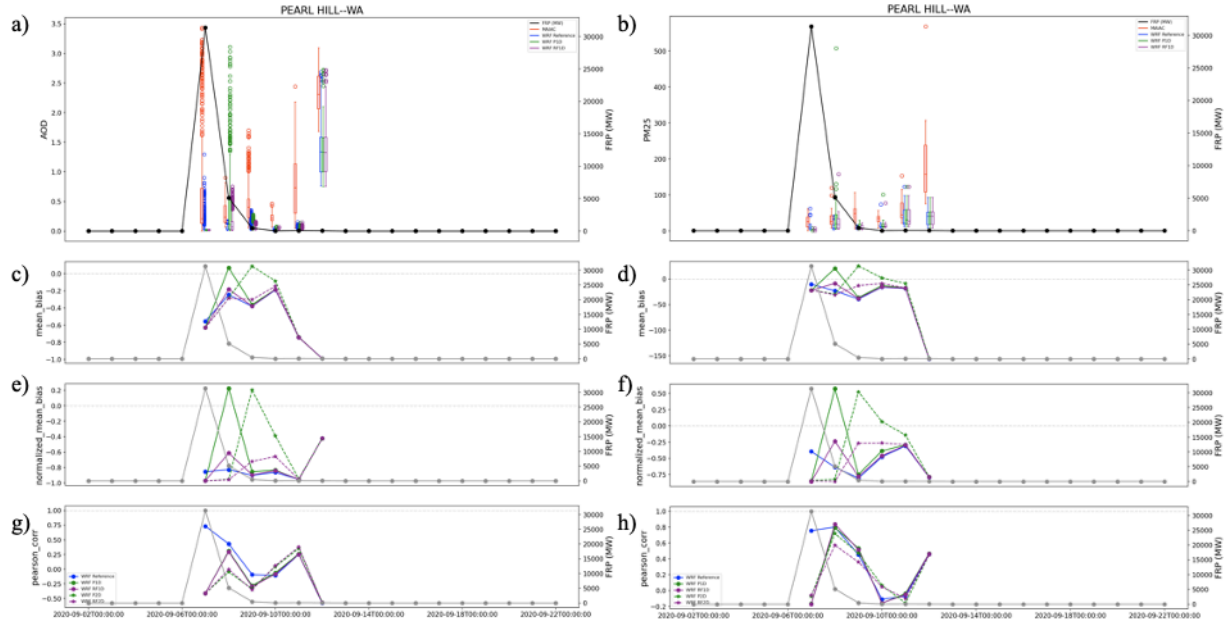

Figure C10. As in S6 for the Pearl Hill fire in WA

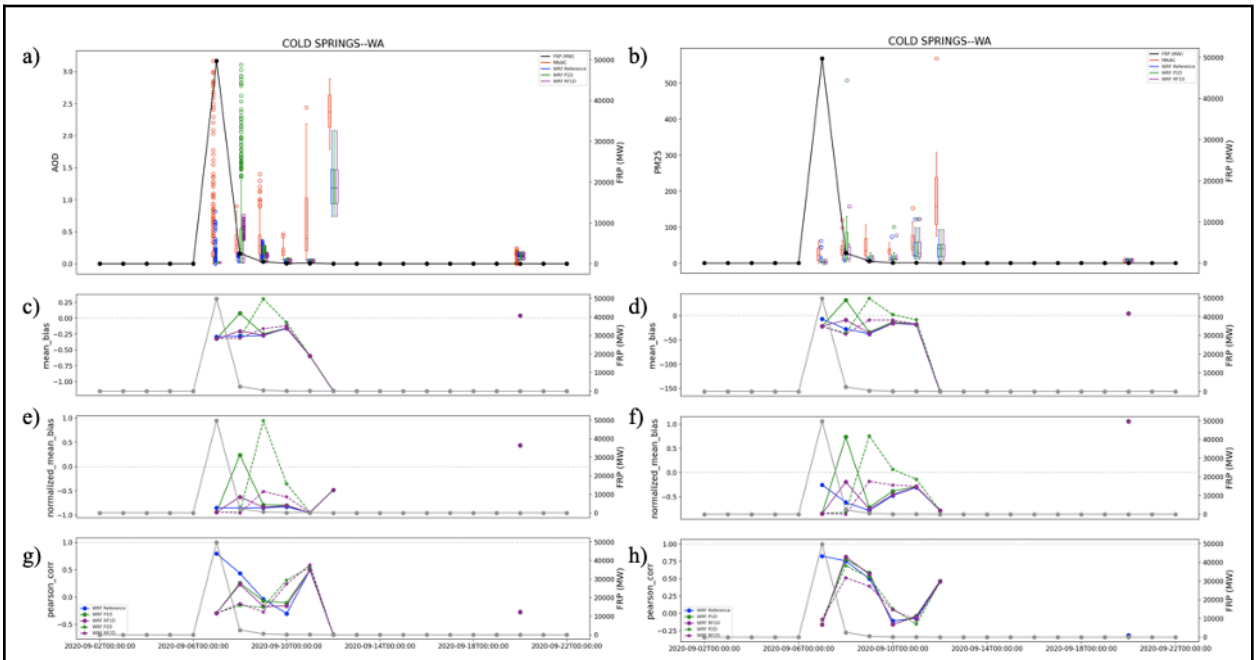

Figure C11. As in S6 for the Cold Springs fire in WA

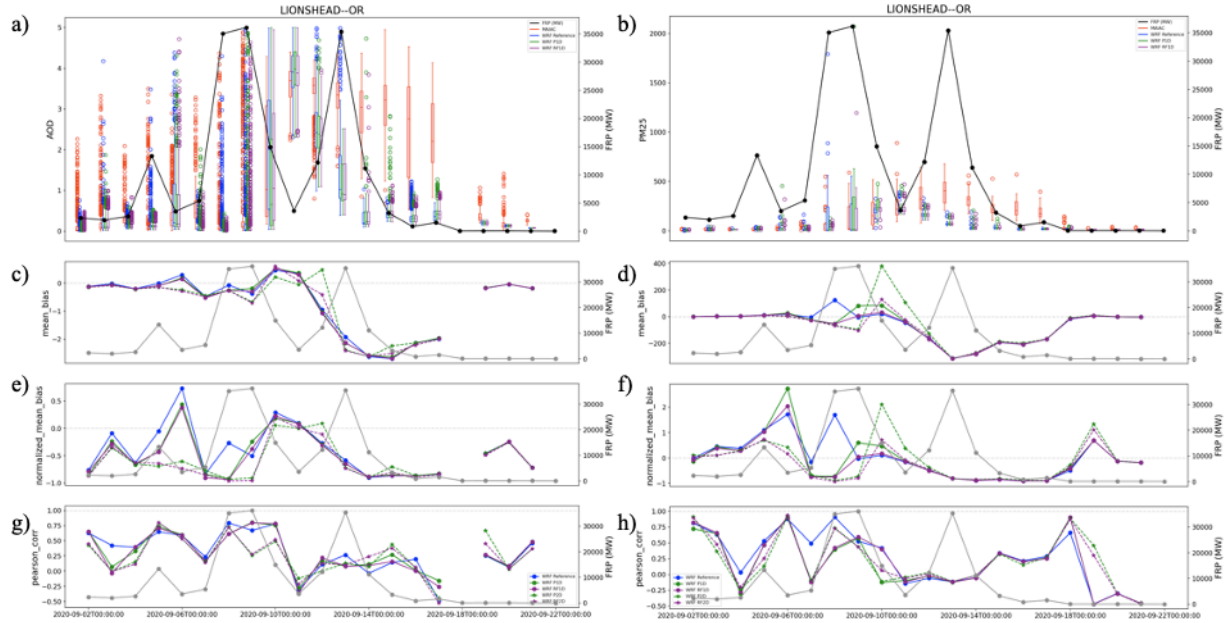

Figure C12. As in S6 for the Lionshead fire in OR

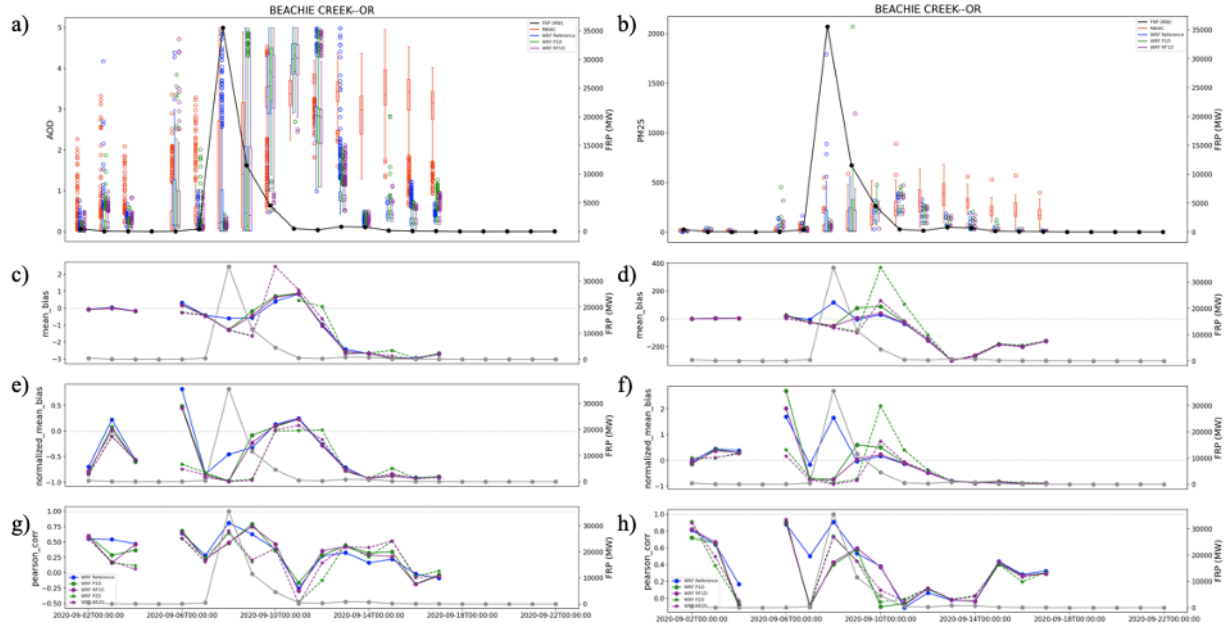

Figure C13. As in S6 for the Beachie Creek fire in OR

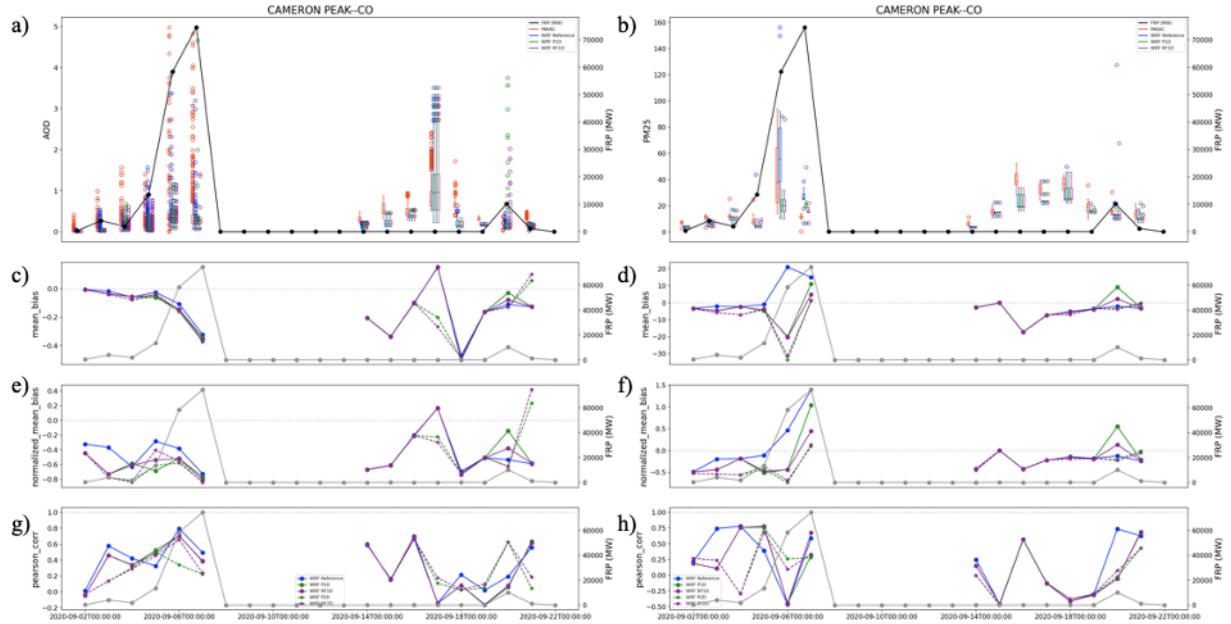

Figure C14. As in S6 for the Cameron Peak fire in CO

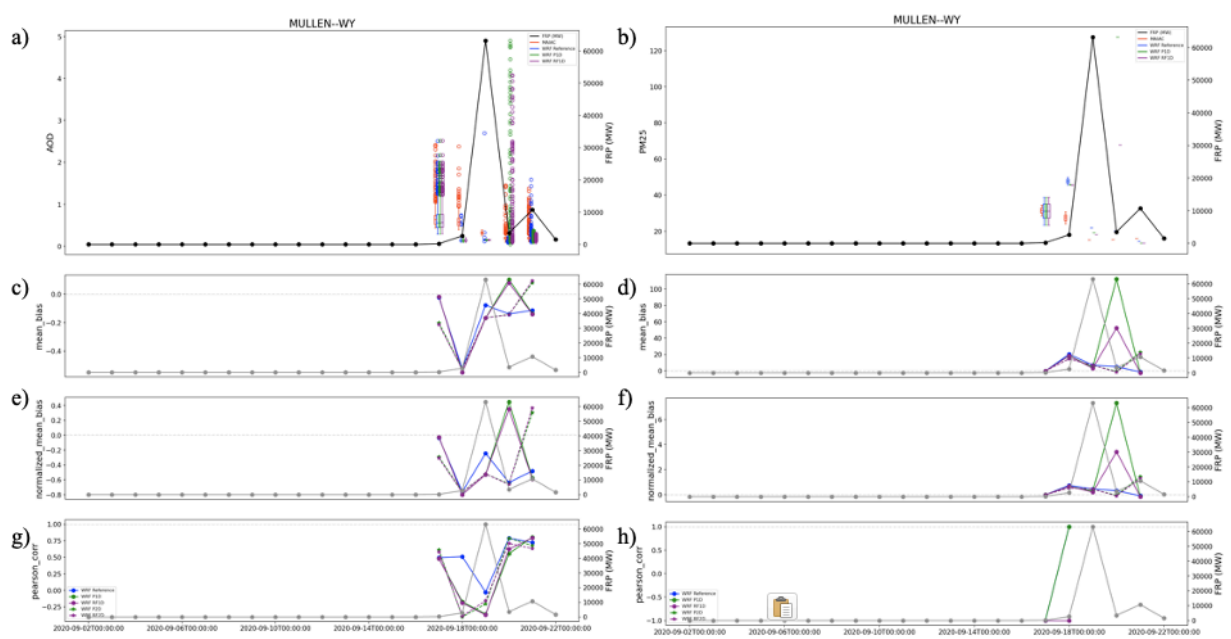

Figure C15. As in S6 for the Mullen fire in WY

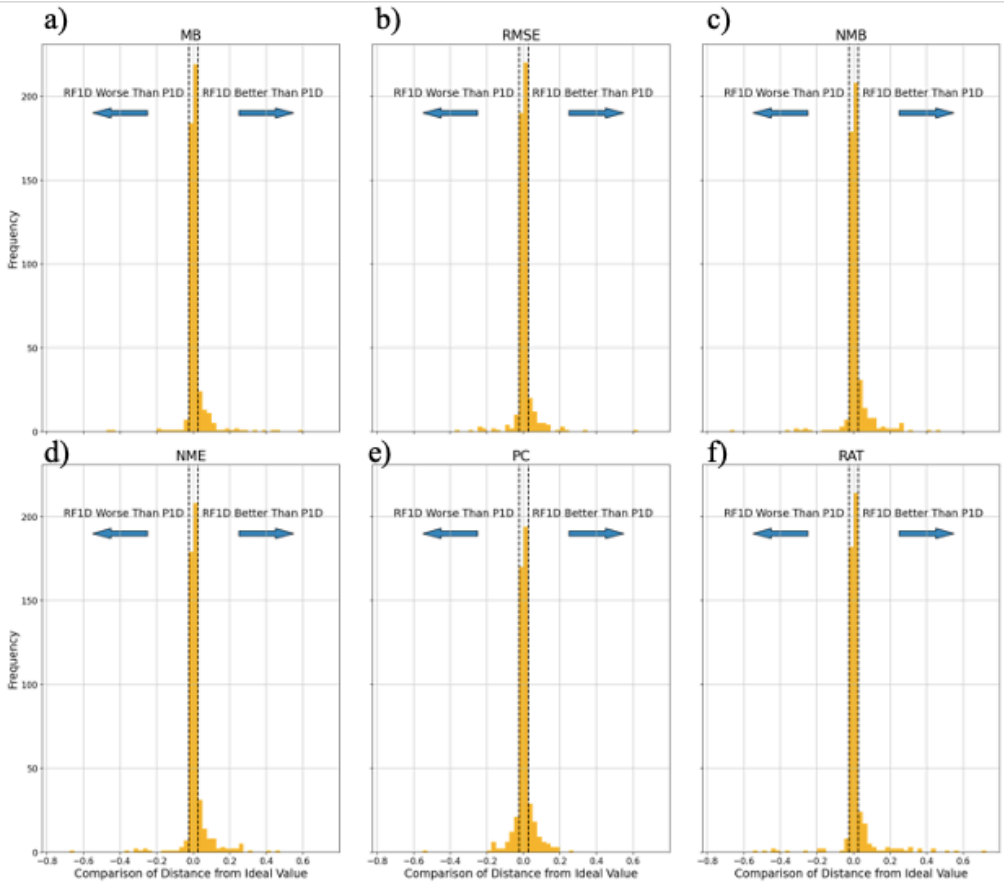

Figure C16. Distributions of the value which compares the distances of the persistence and random forest of AOD (e.g.) from its ideal value. If the random forest gets the statistic closer to its ideal value, this number is positive, and if the random forest gets the statistic further from its ideal value this number is negative. This value was computed for a) mean bias, b) root-mean-squared error, c) normalized mean bias, d) normalized mean error, e) pearson correlation, and f) ratio. The black dotted lines are a +/- 0.025, representing the cutoffs for the 3 categories of model performance: RF better than P, RF about the same as P, and RF worse than P. The values plotted here are an example computed from the filtered AOD over merged 50 km buffer polygons.

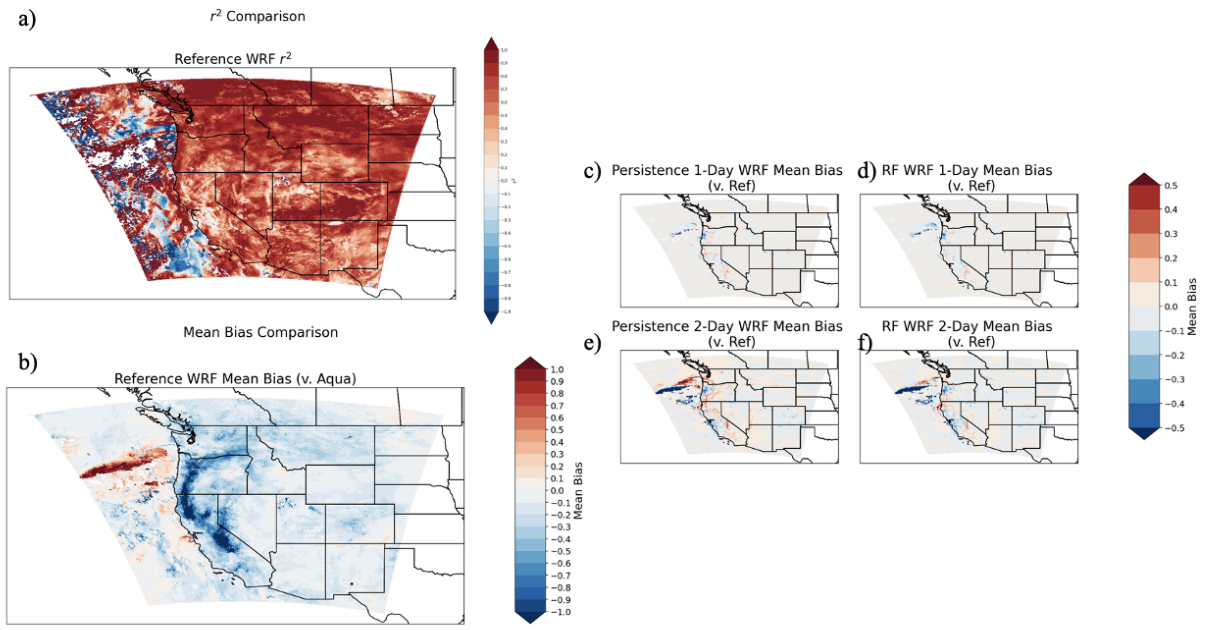

Figure C17. Sample maps of r^2 and mean bias in AOD over the WRF Chem model domain. a) and b) show the two metrics for the whole reference model period, with the comparison done at daily resolution. In a) red means a positive correlation in time and blue means a negative correlation in time. In b) red corresponds to overestimated AOD blue corresponds to underestimated AOD. c-f) show the difference in mean bias between the reference simulation and the c) 1-day persistence, d) 1-day random forest, e) 2-day persistence, and f) 2-day random forest predictions. In c-f) red means that higher AOD than the reference simulation is predicted, and blue means that lower AOD than the reference simulation is predicted.

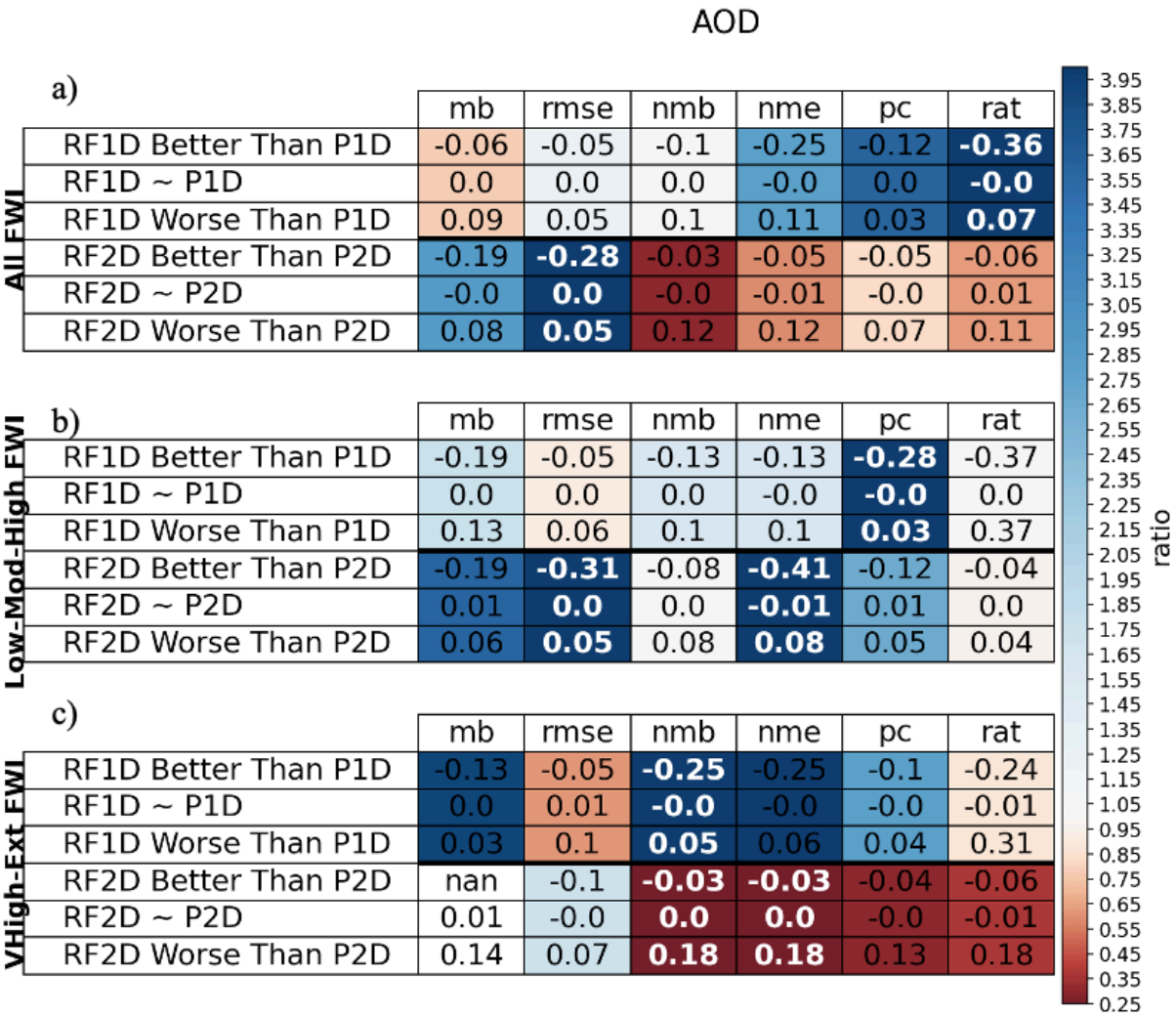

Figure C18. Comparison of median error metrics for AOD in smoky regions downstream of fires for a) all fire days, b) the subset of fire days showing low, moderate, or high fire weather index (FWI) values, and c) the subset of fire days showing very high or high extreme FWI values. Numbers in the grid cells are the distance between the median of the error distribution and the ideal value for the metric (1 for rat and pc, 0 for all others). The boxes are shaded blue if the random forest emissions get the median AOD much closer to the ideal value than persistence, and red if vice versa.

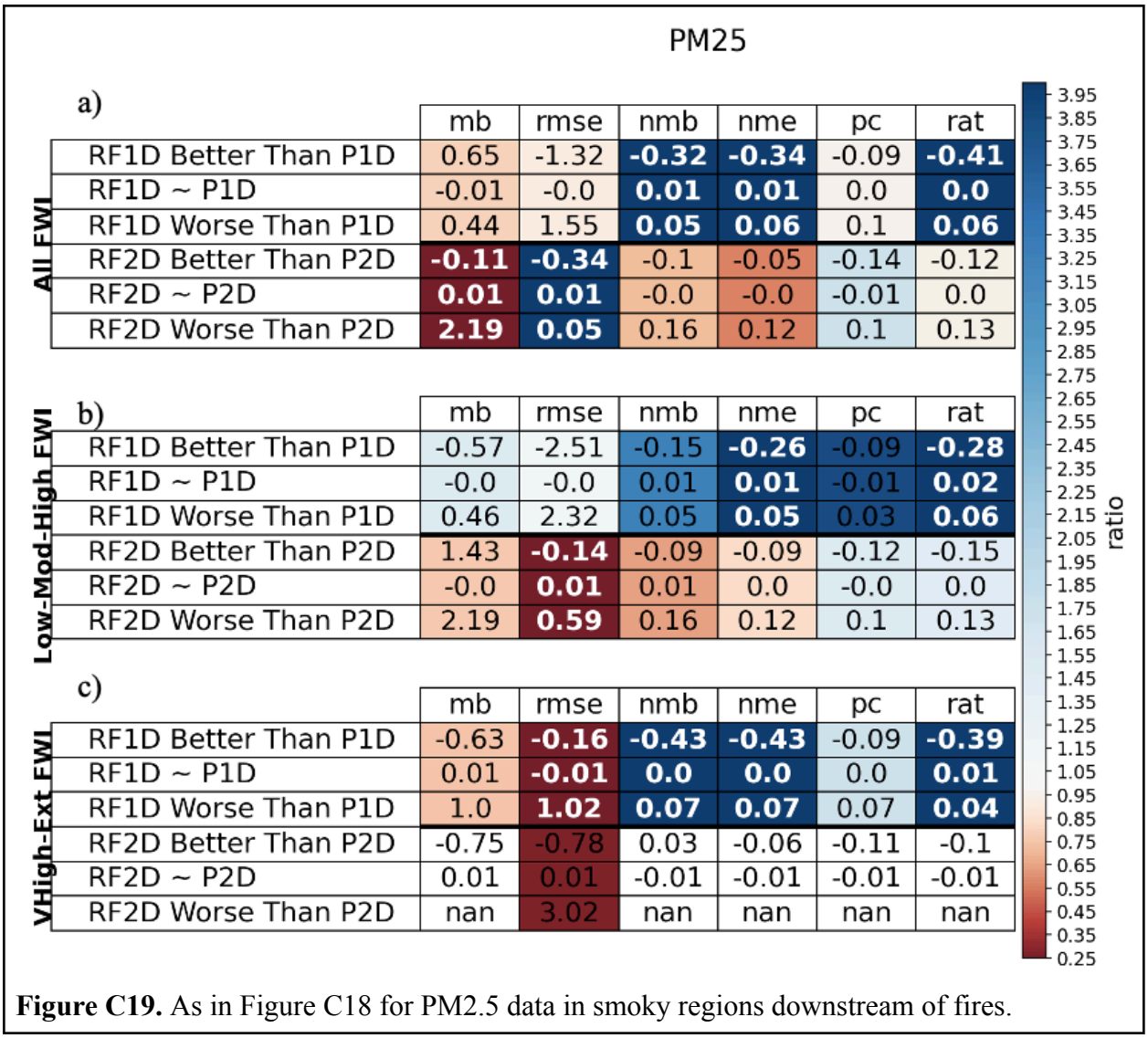

Figure C19. As in Figure C18 for PM2.5 data in smoky regions downstream of fires.

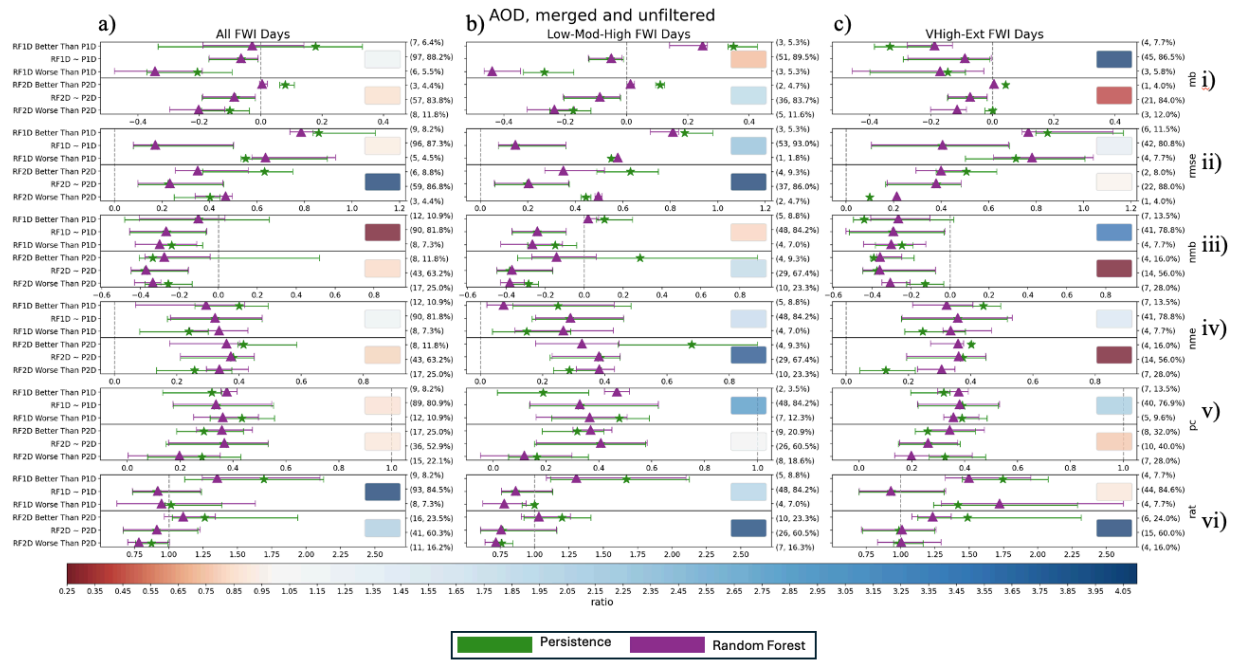

Figure C20. Comparison of the distribution of the median (stars* and triangles^) and spread (bars) of AOD error metrics for a) all fire-days, b) days with low-mod-high FWI, and c) days with very high-extreme FWI. AOD statistics were computed over merged polygons and included smoke and non-smoke values. Error metrics plotted include i) mean bias, ii) root-mean-squared error, iii) normalized mean bias, iv) normalized mean error, v) pearson correlation, and vi) ratio. Shaded boxes show the ratio between the change in the persistence (*) and random forest (^) medians when RF is better than P and the change in the persistence (*) and random forest (^) medians when RF is worse than P. Blue means RF increases overall forecast skill in terms of the metric, and red means RF decreases skill in terms of the metric.

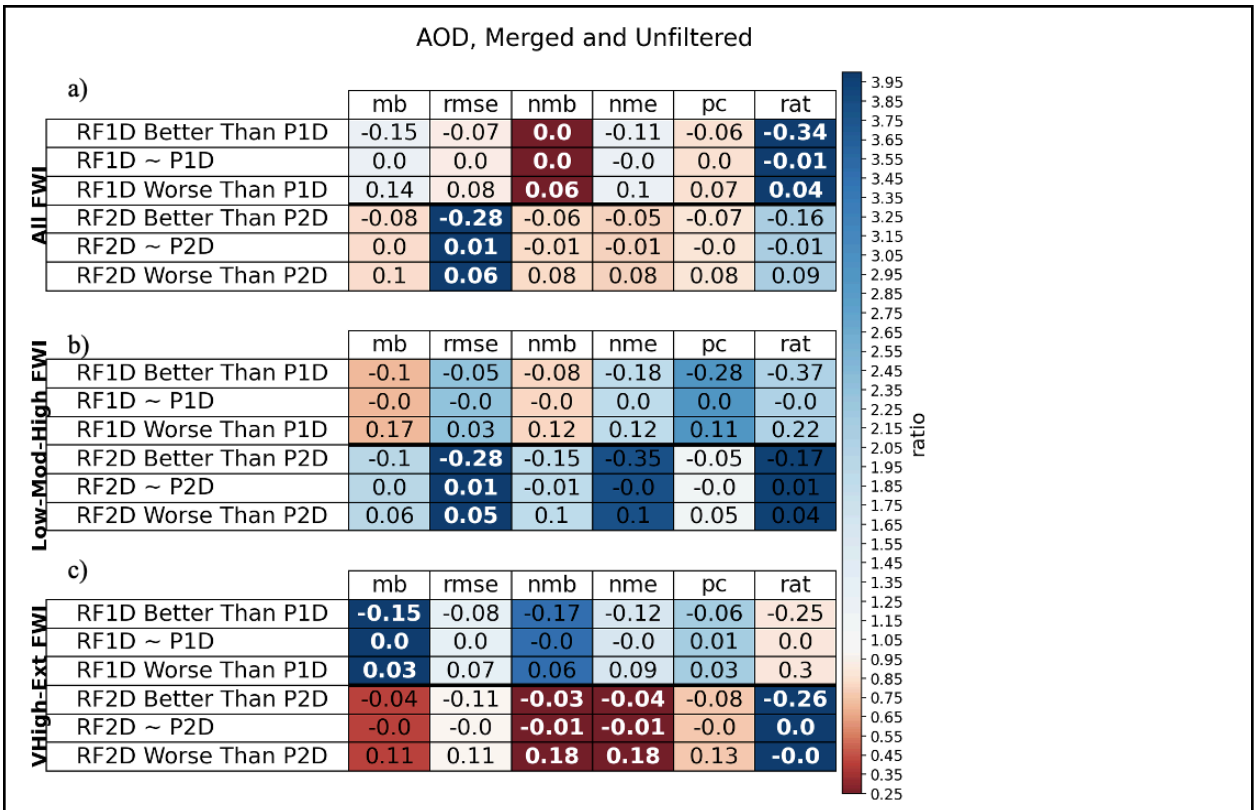

Figure C21. As in Figure C18 for AOD statistics computed over merged polygons, smoke and non-smoke values

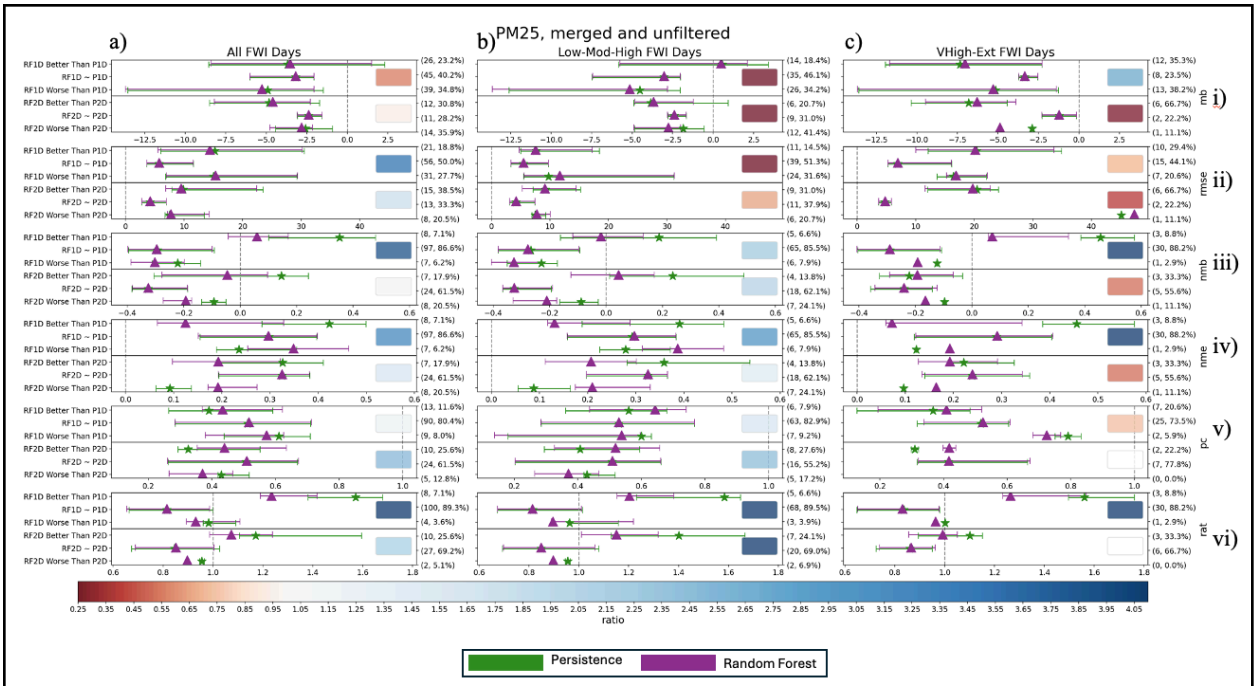

Figure C22. As in Figure C20 for PM2.5 computed over merged polygons, smoke and non-smoke AOD values included

PM25, merged and unfiltered

a)

All FWI	mb	rmse	nmb	nme	pc	rat
RF1D Better Than P1D	-0.16	-0.88	-0.29	-0.3	-0.04	**-0.34**
RF1D ~ P1D	0.0	-0.02	0.0	-0.0	-0.0	**0.0**
RF1D Worse Than P1D	0.36	0.29	0.08	0.11	0.04	**0.05**
RF2D Better Than P2D	-0.22	-0.34	-0.1	-0.13	-0.11	-0.1
RF2D ~ P2D	0.0	-0.0	-0.0	-0.0	-0.0	-0.01
RF2D Worse Than P2D	0.24	0.23	0.1	0.1	0.06	0.06

b)

Low-Mod-High FWI	mb	rmse	nmb	nme	pc	rat
RF1D Better Than P1D	**-0.08**	**-0.07**	-0.17	-0.26	-0.08	**-0.38**
RF1D ~ P1D	**-0.0**	**0.0**	0.01	0.0	0.0	**0.0**
RF1D Worse Than P1D	**0.64**	**1.88**	0.1	0.11	0.06	**0.07**
RF2D Better Than P2D	**-0.17**	-0.13	-0.19	-0.15	-0.11	**-0.25**
RF2D ~ P2D	**0.0**	-0.0	-0.0	-0.0	-0.0	**0.0**
RF2D Worse Than P2D	**0.92**	0.23	0.12	0.12	0.06	**0.06**

c)

VHigh-Ext FWI	mb	rmse	nmb	nme	pc	rat
RF1D Better Than P1D	-0.3	-0.25	**-0.38**	**-0.38**	-0.04	**-0.3**
RF1D ~ P1D	0.0	0.0	**0.0**	**0.0**	-0.01	**0.01**
RF1D Worse Than P1D	0.14	0.43	**0.07**	**0.07**	0.07	**0.03**
RF2D Better Than P2D	-0.52	-0.75	-0.03	-0.03	-0.11	-0.09
RF2D ~ P2D	0.01	0.0	0.0	0.0	-0.0	-0.0
RF2D Worse Than P2D	2.0	2.21	0.07	0.07	nan	nan

ratio: 3.95, 3.85, 3.75, 3.65, 3.55, 3.45, 3.35, 3.25, 3.15, 3.05, 2.95, 2.85, 2.75, 2.65, 2.55, 2.45, 2.35, 2.25, 2.15, 2.05, 1.95, 1.85, 1.75, 1.65, 1.55, 1.45, 1.35, 1.25, 1.15, 1.05, 0.95, 0.85, 0.75, 0.65, 0.55, 0.45, 0.35, 0.25

Figure C23. As in Figure C18 for PM2.5 statistics computed over merged polygons, smoke and non-smoke values

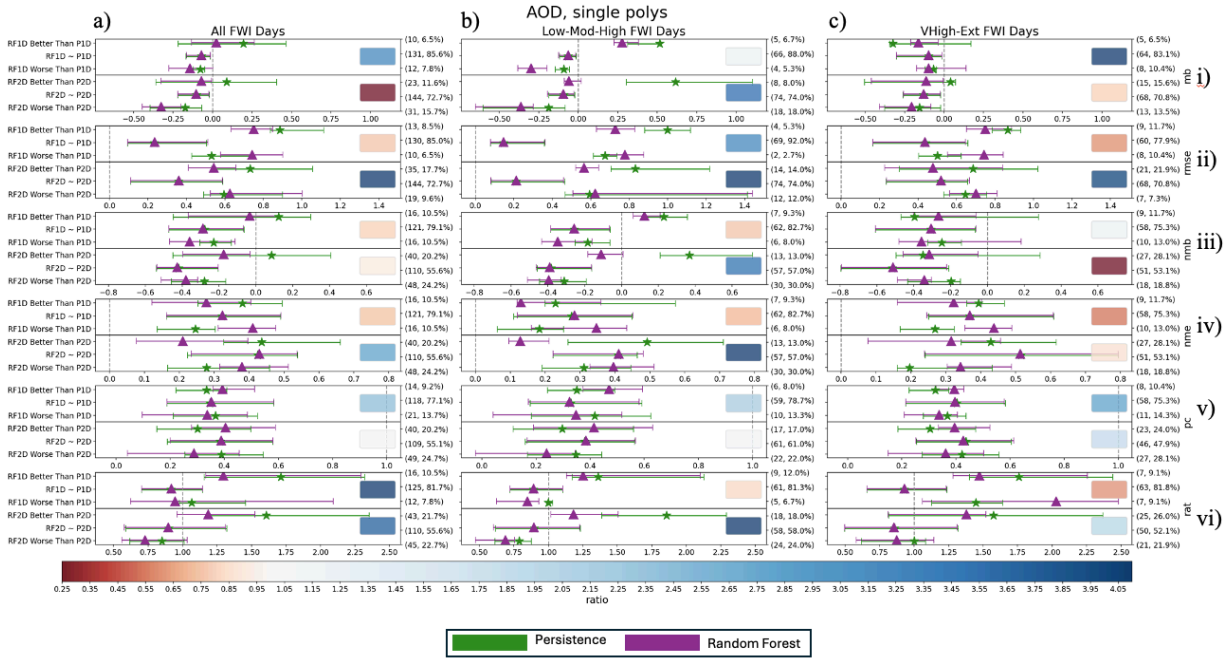

Figure C24. As in C20 for AOD computed over single polygons, smoke and non-smoke included

AOD, single polys

a) All FWI

	mb	rmse	nmb	nme	pc	rat
RF1D Better Than P1D	-0.17	-0.14	-0.09	-0.1	-0.06	**-0.42**
RF1D ~ P1D	0.0	0.0	0.01	0.0	-0.01	**-0.0**
RF1D Worse Than P1D	0.07	0.21	0.13	0.16	0.03	**-0.01**
RF2D Better Than P2D	**-0.02**	**-0.19**	0.09	-0.23	-0.1	-0.42
RF2D ~ P2D	**0.0**	**0.0**	0.01	0.0	-0.0	0.01
RF2D Worse Than P2D	**0.15**	**0.03**	0.1	0.1	0.1	0.12

b) Low-Mod-High FWI

	mb	rmse	nmb	nme	pc	rat
RF1D Better Than P1D	-0.24	-0.27	-0.11	-0.1	-0.12	-0.11
RF1D ~ P1D	0.0	-0.0	0.0	0.01	0.01	0.0
RF1D Worse Than P1D	0.21	0.1	0.16	0.16	0.07	0.15
RF2D Better Than P2D	-0.56	**-0.27**	-0.26	**-0.36**	-0.12	**-0.68**
RF2D ~ P2D	0.0	**0.0**	0.01	**0.0**	-0.0	**0.0**
RF2D Worse Than P2D	0.18	**0.03**	0.08	**0.08**	0.11	**0.1**

c) VHigh-Ext FWI

	mb	rmse	nmb	nme	pc	rat
RF1D Better Than P1D	**-0.16**	-0.12	-0.13	-0.07	-0.07	-0.29
RF1D ~ P1D	**0.0**	-0.0	-0.0	0.0	0.01	0.0
RF1D Worse Than P1D	**0.03**	0.24	0.11	0.17	0.03	0.58
RF2D Better Than P2D	0.07	-0.21	**-0.03**	-0.11	-0.09	-0.2
RF2D ~ P2D	0.0	0.0	**-0.0**	-0.0	0.01	0.01
RF2D Worse Than P2D	0.05	0.05	**0.15**	0.15	0.06	0.13

ratio: 3.95 3.85 3.75 3.65 3.55 3.45 3.35 3.25 3.15 3.05 2.95 2.85 2.75 2.65 2.55 2.45 2.35 2.25 2.15 2.05 1.95 1.85 1.75 1.65 1.55 1.45 1.35 1.25 1.15 1.05 0.95 0.85 0.75 0.65 0.55 0.45 0.35 0.25

Figure C25. As in Figure C18 for AOD statistics computed over single polygons, smoke and

non-smoke values

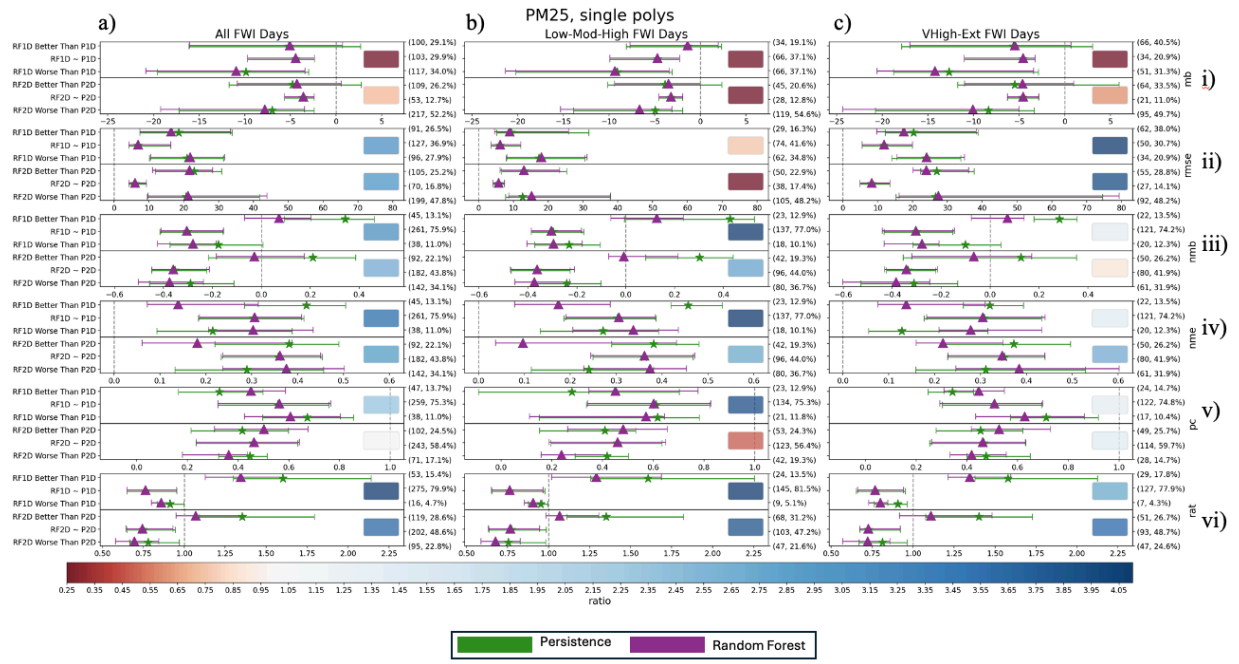

Figure C26. As in C20 for PM2.5 computed over single polygons, smoke and non-smoke included

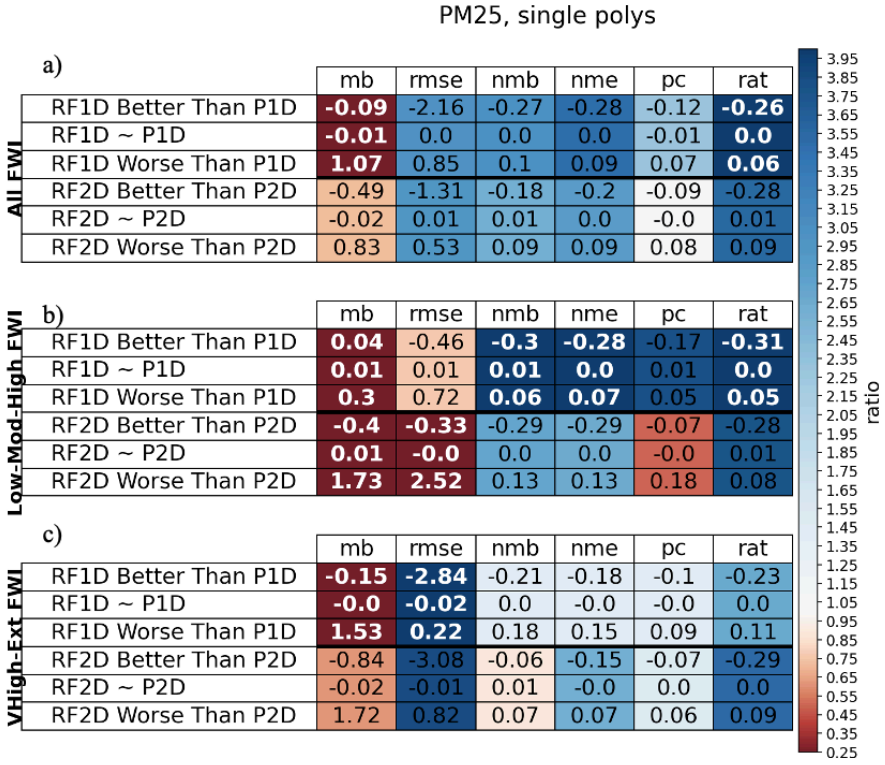

Figure C27. As in Figure C18 for PM2.5 statistics computed over single polygons, smoke and non-smoke values

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, *33*(1), 121–131. https://doi.org/10.1002/joc.3413

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., & Shankar, U. (1998). Modal aerosol dynamics model for Europe: Development and first applications. *Atmospheric Environment*, *32*(17), 2981–2999. https://doi.org/10.1016/S1352-2310(98)00006-5

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., & Wennberg, P. O. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics*, *11*(9), 4039–4072. https://doi.org/10.5194/acp-11-4039-2011

Ahmadov, R., Grell, G., James, E., Csiszar, I., Tsidulko, M., Pierce, B., McKeen, S., Benjamin, S., Alexander, C., Pereira, G., Freitas, S., & Goldberg, M. (2017). Using VIIRS fire radiative power data to simulate biomass burning emissions, plume rise and smoke transport in a real-time air quality modeling system. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2017-July*, 2806–2808. https://doi.org/10.1109/IGARSS.2017.8127581

Ahmadov, R., McKeen, S. A., Robinson, A. L., Bahreini, R., Middlebrook, A. M., de Gouw, J. A., Meagher, J., Hsie, E.-Y., Edgerton, E., Shaw, S., & Trainer, M. (2012). A volatility basis set model for summertime secondary organic aerosols over the eastern United States in 2006. *Journal of Geophysical Research: Atmospheres*, *117*(D6). https://doi.org/10.1029/2011JD016831

Alapaty, K., Herwehe, J. A., Otte, T. L., Nolte, C. G., Bullock, O. R., Mallard, M. S., Kain, J. S., & Dudhia, J. (2012). Introducing subgrid-scale cloud feedbacks to radiation for regional meteorological and climate modeling. *Geophysical Research Letters*, *39*(24). https://doi.org/10.1029/2012GL054031

Allen, R. J., Amiri-Farahani, A., Lamarque, J.-F., Smith, C., Shindell, D., Hassan, T., & Chung, C. E. (2019). Observationally constrained aerosol–cloud semi-direct effects. *Npj Climate and Atmospheric Science*, *2*(1), 1–12. https://doi.org/10.1038/s41612-019-0073-9

Anderson, M. C., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J. R., & Kustas, W. P. (2011). Evaluation of Drought Indices Based on Thermal Remote Sensing of Evapotranspiration over the Continental United States. *Journal of Climate*, *24*(8), 2025–2044. https://doi.org/10.1175/2010JCLI3812.1

Andreae, M. O. (2019). Emission of trace gases and aerosols from biomass burning – an updated assessment. *Atmospheric Chemistry and Physics*, *19*(13), 8523–8546. https://doi.org/10.5194/acp-19-8523-2019

Arienti, M. C., Cumming, S. G., & Boutin, S. (2006). Empirical models of forest fire initial attack success probabilities: The effects of fuels, anthropogenic linear features, fire weather, and management. *Canadian Journal of Forest Research*, *36*(12), 3155–3166. https://doi.org/10.1139/x06-188

Baker, K. R., Woody, M. C., Tonnesen, G. S., Hutzell, W., Pye, H. O. T., Beaver, M. R., Pouliot, G., & Pierce, T. (2016). Contribution of regional-scale fire events to ozone and PM2.5 air quality estimated by photochemical modeling approaches. *Atmospheric Environment*, *140*, 539–554. https://doi.org/10.1016/j.atmosenv.2016.06.032

Bakhshaii, A., & Johnson, E. A. (2019). A review of a new generation of wildfire–atmosphere

modeling. Canadian Journal of Forest Research, 49(6), 565–574. https://doi.org/10.1139/cjfr-2018-0138

Berman, M. T., Ye, X., Thapa, L. H., Peterson, D. A., Hyer, E. J., Soja, A. J., Gargulinski, E. M., Csiszar, I., Schmidt, C. C., & Saide, P. E. (2023). Quantifying burned area of wildfires in the western United States from polar-orbiting and geostationary satellite active-fire detections. *International Journal of Wildland Fire*, *32*(5), 665–678. https://doi.org/10.1071/WF22022

Bertoletti, A. Z., Phan, T., & Campos do Prado, J. (2022). Wildfire Smoke, Air Quality, and Renewable Energy—Examining the Impacts of the 2020 Wildfire Season in Washington State. *Sustainability*, *14*(15), Article 15. https://doi.org/10.3390/su14159037

Bradshaw, L. S., Deeming, J. E., Burgan, R. E., & compilers.Cohen, J. D. (1984). The 1978 National Fire-Danger Rating System: Technical documentation. *General Technical Report INT-169. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. 44 p.*, *169*. https://doi.org/10.2737/INT-GTR-169

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burke, M., Driscoll, A., Heft-Neal, S., Xue, J., Burney, J., & Wara, M. (2021). The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences*, *118*(2), Article 2. https://doi.org/10.1073/pnas.2011048118

Burke, M., Childs, M. L., de la Cuesta, B., Qiu, M., Li, J., Gould, C. F., Heft-Neal, S., & Wara, M. (2023). The contribution of wildfire to PM2.5 trends in the USA. *Nature*, *622*(7984), Article 7984. https://doi.org/10.1038/s41586-023-06522-6

Carlsen, A. (2020, September 23). 1 In 7 Americans Have Experienced Dangerous Air Quality Due To Wildfires This Year. *NPR*. https://www.npr.org/2020/09/23/915723316/1-in-7-americans-have-experienced-dangerous-air-quality-due-to-wildfires-this-ye

Carter, T. S., Heald, C. L., Jimenez, J. L., Campuzano-Jost, P., Kondo, Y., Moteki, N., Schwarz, J. P., Wiedinmyer, C., Darmenov, A. S., Da Silva, A. M., & Kaiser, J. W. (2020). How emissions uncertainty influences the distribution and radiative impacts of smoke from fires in North America. Atmos. Chem. Phys, 20. https://doi.org/10.5194/acp-20-2073-2020

Center For International Earth Science Information Network-CIESIN-Columbia University. (2018). *Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets*. https://doi.org/10.7927/H45Q4T5F

Chen, F., & Dudhia, J. (2001). Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*(4), Article 4. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Chen, Y., Hantson, S., Andela, N., Coffield, S. R., Graff, C. A., Morton, D. C., Ott, L. E., Foufoula-Georgiou, E., Smyth, P., Goulden, M. L., & Randerson, J. T. (2022). California wildfire spread derived using VIIRS satellite observations and an object-based tracking system. *Scientific Data*, *9*(1), 249. https://doi.org/10.1038/s41597-022-01343-0

Coen, J. L., Cameron, M., Michalakes, J., Patton, E. G., Riggan, P. J., & Yedinak, K. M. (2013). WRF-Fire: Coupled Weather–Wildland Fire Modeling with the Weather Research and Forecasting Model. *Journal of Applied Meteorology and Climatology*, *52*(1), 16–38.

https://doi.org/10.1175/JAMC-D-12-023.1

Collow, A. B., Colarco, P. R., da Silva, A. M., Buchard, V., Bian, H., Chin, M., Das, S., Govindaraju, R., Kim, D., & Aquila, V. (2024). Benchmarking GOCART-2G in the Goddard Earth Observing System (GEOS). *Geoscientific Model Development*, *17*(3), 1443–1468. https://doi.org/10.5194/gmd-17-1443-2024

Countryman, C. M. (1972). *The Fire Environment Concept*. 15.

Cromar, K., Gladson, L., Gohlke, J., Li, Y., Tong, D., & Ewart, G. (2024). Adverse Health Impacts of Outdoor Air Pollution, Including from Wildland Fires, in the United States: "Health of the Air," 2018–2020. *Annals of the American Thoracic Society*, *21*(1), 76–87. https://doi.org/10.1513/AnnalsATS.202305-455OC

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*(15), 2031–2064. https://doi.org/10.1002/joc.1688

Darmenov, A. S., and da Silva, A. M. (2015). Technical Report Series on Global Modeling and Data Assimilation, Volume 38 The Quick Fire Emissions Dataset (QFED):Documentation of Versions 2.1, 2.2 and 2.4. http://www.sti.nasa.gov

Das, S., Colarco, P. R., Oman, L. D., Taha, G., & Torres, O. (2021). The long-term transport and radiative impacts of the 2017 British Columbia pyrocumulonimbus smoke aerosols in the stratosphere. *Atmospheric Chemistry and Physics*, *21*(15), 12069–12090. https://doi.org/10.5194/acp-21-12069-2021

De Vasconcelos, M. P., Silva, S., Tome, M., Alvim, M., & Pereira, J. C. (2001). Spatial prediction of fire ignition probabilities: comparing logistic regression and neural

networks. Photogrammetric engineering and remote sensing, 67(1), 73-81.

Di Giuseppe, F., Rémy, S., Pappenberger, F., & Wetterhall, F. (2017). Improving Forecasts of Biomass Burning Emissions with the Fire Weather Index. Journal of Applied Meteorology and Climatology, 56(10), 2789–2799. https://doi.org/10.1175/JAMC-D-16-0405.1

Di Giuseppe, F., Rémy, S., Pappenberger, F., & Wetterhall, F. (2018). Using the Fire Weather Index (FWI) to improve the estimation of fire emissions from fire radiative power (FRP) observations. Atmospheric Chemistry and Physics, 18(8), 5359–5370. https://doi.org/10.5194/acp-18-5359-2018

Dowdy, A. J., & Pepler, A. (2018). Pyroconvection Risk in Australia: Climatological Changes in Atmospheric Stability and Surface Fire Weather Conditions. *Geophysical Research Letters*, *45*(4), 2005–2013. https://doi.org/10.1002/2017GL076654

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., Blake, B. T., Brown, J. M., Olson, J. B., Hu, M., Smirnova, T. G., Ladwig, T., Kenyon, J. S., Ahmadov, R., Turner, D. D., Duda, J. D., & Alcott, T. I. (2022). The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description. *Weather and Forecasting*, *37*(8), 1371–1395. https://doi.org/10.1175/WAF-D-21-0151.1

Earl, N., & Simmonds, I. (2018). Spatial and Temporal Variability and Trends in 2001–2016 Global Fire Activity. *Journal of Geophysical Research: Atmospheres*, *123*(5), 2524–2536. https://doi.org/10.1002/2017JD027749

Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z.-L., Quayle, B., & Howard, S. (2007). A Project for Monitoring Trends in Burn Severity. Fire Ecology, 3(1), Article 1.

https://doi.org/10.4996/fireecology.0301003

Field, R. D. (2020a). Evaluation of Global Fire Weather Database reanalysis and short-term forecast products. *Natural Hazards and Earth System Sciences*, *20*(4), 1123–1147. https://doi.org/10.5194/NHESS-20-1123-2020

Field, R. D. (2020b). Using Satellite Estimates of Precipitation for Fire Danger Rating. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement: Volume 2* (pp. 1131–1154). Springer International Publishing. https://doi.org/10.1007/978-3-030-35798-6_33

Field, R. D., Spessa, A. C., Aziz, N. A., Camia, A., Cantin, A., Carr, R., de Groot, W. J., Dowdy, A. J., Flannigan, M. D., Manomaiphiboon, K., Pappenberger, F., Tanpipat, V., & Wang, X. (2015). Development of a Global Fire Weather Database. *Natural Hazards and Earth System Sciences*, *15*(6), 1407–1423. https://doi.org/10.5194/nhess-15-1407-2015

Finney, M. A. (1998). FARSITE: Fire Area Simulator-model development and evaluation. *Res. Pap. RMRS-RP-4, Revised 2004. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 47 p.*, *4*. https://doi.org/10.2737/RMRS-RP-4

Freitas, S. R., Longo, K. M., Chatfield, R., Latham, D., Dias, M. A. F. S., Andreae, M. O., Prins, E., Santos, J. C., Gielow, R., & Carvalho, J. A. (2007). Atmospheric Chemistry and Physics Including the sub-grid scale plume rise of vegetation fires in low resolution atmospheric transport models. In *Atmos. Chem. Phys* (Vol. 7, pp. 3385–3398). www.atmos-chem-phys.net/7/3385/2007/

Freitas, S. R., Longo, K. M., Trentmann, J., & Latham, D. (2010). Technical Note: Sensitivity of 1-D smoke plume rise models to the inclusion of environmental wind drag. *Atmospheric Chemistry and Physics*, *10*(2), Article 2. https://doi.org/10.5194/acp-10-585-2010

Freeborn, P. H., Wooster, M. J., Hao, W. M., Ryan, C. A., Nordgren, B. L., Baker, S. P., & Ichoku, C. (2008). Relationships between energy release, fuel mass loss, and trace gas and aerosol emissions during laboratory biomass fires. *Journal of Geophysical Research*, *113*(D1), Article D1. https://doi.org/10.1029/2007JD008679

García-Llamas, P., Suárez-Seoane, S., Taboada, A., Fernández-Manso, A., Quintano, C., Fernández-García, V., Fernández-Guisuraga, J. M., Marcos, E., & Calvo, L. (2019). Environmental drivers of fire severity in extreme fire events that affect Mediterranean pine forest ecosystems. *Forest Ecology and Management*, *433*, 24–32. https://doi.org/10.1016/j.foreco.2018.10.051

Garcia-Menendez, F., Hu, Y., & Odman, M. T. (2014). Simulating smoke transport from wildland fires with a regional-scale air quality model: Sensitivity to spatiotemporal allocation of fire emissions. *Science of The Total Environment*, *493*, 544–553. https://doi.org/10.1016/j.scitotenv.2014.05.108

Giglio, L., Descloitres, J., Justice, C. O., & Kaufman, Y. J. (2003). An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment*, *87*(2), 273–282. https://doi.org/10.1016/S0034-4257(03)00184-6

Gillies, S. & et al. (2007). *Shapely: Manipulation and analysis of geometric objects*. toblerity.org. https://github.com/Toblerity/Shapely

Graff, C. A., Coffield, S. R., Chen, Y., Foufoula-Georgiou, E., Randerson, J. T., & Smyth, P. (2020). Forecasting Daily Wildfire Activity Using Poisson Regression. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(7), 4837–4851. https://doi.org/10.1109/TGRS.2020.2968029

Gray, M. E., Zachmann, L. J., & Dickson, B. G. (2018). A weekly, continually updated dataset of

the probability of large wildfires across western US forests and woodlands. *Earth System Science Data*, *10*(3), 1715–1727. https://doi.org/10.5194/essd-10-1715-2018

Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, *14*(10), Article 10. https://doi.org/10.5194/acp-14-5233-2014

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & Geron, C. (2006). Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmospheric Chemistry and Physics*, *6*(11), 3181–3210. https://doi.org/10.5194/acp-6-3181-2006

Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., & Hovis, F. E. (2008). Airborne High Spectral Resolution Lidar for profiling Aerosol optical properties. *Applied Optics*, *47*(36), Article 36. https://doi.org/10.1364/AO.47.006734

Hernandez, C., Keribin, C., Drobinski, P., & Turquety, S. (2015). Statistical modelling of wildfire size and intensity: A step toward meteorological forecasting of summer extreme fire risk. Annales Geophysicae, 33(12), 1495–1506. https://doi.org/10.5194/angeo-33-1495-2015

Herron-Thorpe, F. L., Mount, G. H., Emmons, L. K., Lamb, B. K., Jaffe, D. A., Wigder, N. L., Chung, S. H., Zhang, R., Woelfle, M. D., & Vaughan, J. K. (2014). Air quality simulations of wildfires in the Pacific Northwest evaluated with surface and satellite observations during the summers of 2007 and 2008. *Atmospheric Chemistry and Physics*, *14*(22), 12533–12551. https://doi.org/10.5194/acp-14-12533-2014

Hook, S. J., Myers, J. J., Thome, K. J., Fitzgerald, M., & Kahle, A. B. (2001). The MODIS/ASTER airborne simulator (MASTER)—A new instrument for earth science

studies. *Remote Sensing of Environment*, *76*(1), Article 1. https://doi.org/10.1016/S0034-4257(00)00195-4

Howes, C. (2024). *Improving Biomass-Burning Smoke Properties, Evolution, Emissions, and Air Quality Forecasts in Models* [Ph.D., University of California, Los Angeles]. https://www.proquest.com/docview/3067681395/abstract/2027006552684DB1PQ/1

Hu, X., & Ntaimo, L. (2009). Integrated simulation and optimization for wildfire containment. *ACM Transactions on Modeling and Computer Simulation*, *19*(4), 1–29. https://doi.org/10.1145/1596519.1596524

Hu, X.-M., Nielsen-Gammon, J. W., & Zhang, F. (2010). Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model. *Journal of Applied Meteorology and Climatology*, *49*(9), 1831–1844. https://doi.org/10.1175/2010JAMC2432.1

Hung, W.-T., Lu, C.-H. (Sarah), Shrestha, B., Lin, H.-C., Lin, C.-A., Grogan, D., Hong, J., Ahmadov, R., James, E., & Joseph, E. (2020). The impacts of transported wildfire smoke aerosols on surface air quality in New York State: A case study in summer 2018. *Atmospheric Environment*, *227*, 117415. https://doi.org/10.1016/j.atmosenv.2020.117415

Huot, F., Hu, R. L., Goyal, N., Sankar, T., Ihme, M., & Chen, Y.-F. (2022). Next Day Wildfire Spread: A Machine Learning Dataset to Predict Wildfire Spreading From Remote-Sensing Data. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–13. https://doi.org/10.1109/TGRS.2022.3192974

Iacono, M. J. et al. Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *J. Geophys. Res.* **113**, D13103 (2008).

Ichoku, C., Martins, J. V., Kaufman, Y. J., Wooster, M. J., Freeborn, P. H., Hao, W. M., Baker, S., Ryan, C. A., & Nordgren, B. L. (2008). Laboratory investigation of fire radiative energy

and smoke aerosol emissions. *Journal of Geophysical Research: Atmospheres*, *113*(D14). https://doi.org/10.1029/2007JD009659

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., & Suttie, M. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Jaffe, D. A., O'Neill, S. M., Larkin, N. K., Holder, A. L., Peterson, D. L., Halofsky, J. E., & Rappold, A. G. (2020). Wildfire and prescribed burning impacts on air quality in the United States. *Journal of the Air & Waste Management Association*, *70*(6), 583–615. https://doi.org/10.1080/10962247.2020.1749731

Jamieson, G. NIMS and the Incident Command System. In International Oil Spill Conference 291–294 (2005a).

Jamieson, G. (2005b). The SIT-209 Database [Dataset]. https://www.wildfire.gov/application/sit209. Wildland Fire Application Information Portal. Accesed 2022.

Jang, E., Kang, Y., Im, J., Lee, D.-W., Yoon, J., & Kim, S.-K. (2019). Detection and Monitoring of Forest Fires Using Himawari-8 Geostationary Satellite Data in South Korea. *Remote Sensing*, *11*(3), Article 3. https://doi.org/10.3390/rs11030271

Jenkins, M. A. (2002). An examination of the sensitivity of numerically simulated wildfires to low-level atmospheric stability and moisture, and the consequences for the Haines Index. *International Journal of Wildland Fire*, *11*(4), 213–232. https://doi.org/10.1071/wf02006

Jiang, Y., Lu, Z., Liu, X., Qian, Y., Zhang, K., Wang, Y., & Yang, X.-Q. (2016). Impacts of

global open-fire aerosols on direct radiative, cloud and surface-albedo effects simulated with CAM5. *Atmospheric Chemistry and Physics*, *16*(23), 14805–14824. https://doi.org/10.5194/acp-16-14805-2016

Jolly, W. M., Freeborn, P. H., Page, W. G., & Butler, B. W. (2019). Severe Fire Danger Index: A Forecastable Metric to Inform Firefighter and Community Wildfire Risk Management. Fire, 2(3), Article 3. https://doi.org/10.3390/fire2030047

Juliano, T. W., Jiménez, P. A., Kosović, B., Eidhammer, T., Thompson, G., Berg, L. K., Fast, J., Motley, A., & Polidori, A. (2022). Smoke from 2020 United States wildfires responsible for substantial solar energy forecast errors. *Environmental Research Letters*, *17*(3), 034010. https://doi.org/10.1088/1748-9326/ac5143

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., & van der Werf, G. R. (2012). Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. Biogeosciences, 9(1), 527–554. https://doi.org/10.5194/bg-9-527-2012

Kahn, R. A., Chen, Y., Nelson, D. L., Leung, F.-Y., Li, Q., Diner, D. J., & Logan, J. A. (2008). Wildfire smoke injection heights: Two perspectives from space. *Geophysical Research Letters*, *35*(4), Article 4. https://doi.org/10.1029/2007GL032165

Kahn, R. A., Li, W.-H., Moroney, C., Diner, D. J., Martonchik, J. V., & Fishbein, E. (2007). Aerosol source plume physical characteristics from space-based multiangle imaging. *Journal of Geophysical Research*, *112*(D11), Article D11. https://doi.org/10.1029/2006JD007647

Kelp, M. M., Jacob, D. J., Lin, H., & Sulprizio, M. P. (2022). An Online-Learned Neural Network Chemical Solver for Stable Long-Term Global Simulations of Atmospheric

Chemistry. Journal of Advances in Modeling Earth Systems, 14(6), e2021MS002926. https://doi.org/10.1029/2021MS002926

Kochanski, A. K., Herron-Thorpe, F., Mallia, D. V., Mandel, J., & Vaughan, J. K. (2021). Integration of a Coupled Fire-Atmosphere Model Into a Regional Air Quality Forecasting System for Wildfire Events. *Frontiers in Forests and Global Change*, *4*. https://doi.org/10.3389/ffgc.2021.728726

Kochanski, A. K., Jenkins, M. A., Mandel, J., Beezley, J. D., Clements, C. B., & Krueger, S. (2013). Evaluation of WRF-SFIRE performance with field observations from the FireFlux experiment. *Geoscientific Model Development*, *6*(4), 1109–1126. https://doi.org/10.5194/gmd-6-1109-2013

Kochanski, A. K., Jenkins, M. A., Yedinak, K., Mandel, J., Beezley, J., & Lamb, B. (2015). Toward an integrated system for fire, smoke and air quality simulations. *International Journal of Wildland Fire*, *25*(5), 534–546. https://doi.org/10.1071/WF14074

Kochanski, A. K., Mallia, D. V., Fearon, M. G., Mandel, J., Souri, A. H., & Brown, T. (2019). Modeling Wildfire Smoke Feedback Mechanisms Using a Coupled Fire-Atmosphere Model With a Radiatively Active Aerosol Scheme. *Journal of Geophysical Research: Atmospheres*, *124*(16), 9099–9116. https://doi.org/10.1029/2019JD030558

Krishna, M., Saide, P. E., Ye, X., Turney, F. A., Hair, J. W., Fenn, M., & Shingler, T. (2024). Evaluation of Wildfire Plume Injection Heights Estimated from Operational Weather Radar Observations Using Airborne Lidar Retrievals. *Journal of Geophysical Research: Atmospheres*, *129*(9), e2023JD039926. https://doi.org/10.1029/2023JD039926

Krueger, E. S., Levi, M. R., Achieng, K. O., Bolten, J. D., Carlson, J. D., Coops, N. C., Holden, Z. A., Magi, B. I., Rigden, A. J., Ochsner, T. E., Krueger, E. S., Levi, M. R., Achieng, K.

O., Bolten, J. D., Carlson, J. D., Coops, N. C., Holden, Z. A., Magi, B. I., Rigden, A. J., & Ochsner, T. E. (2022). Using soil moisture information to better understand and predict wildfire danger: A review of recent developments and outstanding questions. International Journal of Wildland Fire, 32(2), 111–132. https://doi.org/10.1071/WF22056

LANDFIRE. (2016). *Slope Degrees*. https://www.landfire.gov/version_download.php#

Larsen, A. E., Reich, B. J., Ruminski, M., & Rappold, A. G. (2018). Impacts of fire smoke plumes on regional air quality, 2006–2013. *Journal of Exposure Science & Environmental Epidemiology*, *28*(4), 319–327. https://doi.org/10.1038/s41370-017-0013-x

Lassman, W., Mirocha, J. D., Arthur, R. S., Kochanski, A. K., Farguell Caus, A., Bagley, A. M., Carreras Sospedra, M., Dabdub, D., & Barbato, M. (2023). Using Satellite-Derived Fire Arrival Times for Coupled Wildfire-Air Quality Simulations at Regional Scales of the 2020 California Wildfire Season. *Journal of Geophysical Research: Atmospheres*, *128*(6), e2022JD037062. https://doi.org/10.1029/2022JD037062

Li, F., Zhang, X., Kondragunta, S., Lu, X., Csiszar, I., & Schmidt, C. C. (2022). Hourly biomass burning emissions product from blended geostationary and polar-orbiting satellites for air quality forecasting applications. *Remote Sensing of Environment*, *281*, 113237–113237. https://doi.org/10.1016/J.RSE.2022.113237

Li, Y., Tong, D., Makkaroon, P., DelSole, T., Tang, Y., Campbell, P., Baker, B., Cohen, M., Darmenov, A., Ahmadov, R., James, E., Hyer, E., & Xian, P. (2024). Multi-Agency Ensemble Forecast of Wildfire Air Quality in the United States: Toward Community Consensus of Early Warning. *Bulletin of the American Meteorological Society*, *1*(aop). https://doi.org/10.1175/BAMS-D-23-0208.1

Liu, J., Zhan, X., Hain, C., Yin, J., Fang, L., Li, Z., & Zhao, L. (2016). NOaa Soil Moisture Operational Product System (SMOPS) and its validations. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 3477–3480. https://doi.org/10.1109/IGARSS.2016.7729899

Liu, T., Randerson, J. T., Chen, Y., Morton, D. C., Wiggins, E. B., Smyth, P., Foufoula-Georgiou, E., Nadler, R., & Nevo, O. (2023). *Systematically tracking the hourly progression of large wildfires using GOES satellite observations* [Preprint]. ESSD – Land/Energy and Emissions. https://doi.org/10.5194/essd-2023-389

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, *11*(10), 5741–5765. https://doi.org/10.5194/amt-11-5741-2018

Makkaroon, P., Tong, D. Q., Li, Y., Hyer, E. J., Xian, P., Kondragunta, S., Campbell, P. C., Tang, Y., Baker, B. D., Cohen, M. D., Darmenov, A., Lyapustin, A., Saylor, R. D., Wang, Y., & Stajner, I. (2023). Development and Evaluation of a North America Ensemble Wildfire Air Quality Forecast: Initial Application to the 2020 Western United States "Gigafire." *Journal of Geophysical Research: Atmospheres*, *128*(22), e2022JD037298. https://doi.org/10.1029/2022JD037298

Mallia, D. V., Kochanski, A. K., Urbanski, S. P., & Lin, J. C. (2018). Optimizing smoke and plume rise modeling approaches at local scales. *Atmosphere*, *9*(5), Article 5. https://doi.org/10.3390/atmos9050166

Mallia, D. V., Kochanski, A. K., Kelly, K. E., Whitaker, R., Xing, W., Mitchell, L. E., Jacques, A., Farguell, A., Mandel, J., Gaillardon, P.-E., Becnel, T., & Krueger, S. K. (2020). Evaluating Wildfire Smoke Transport Within a Coupled Fire-Atmosphere Model Using a

High-Density Observation Network for an Episodic Smoke Event Along Utah's Wasatch Front. *Journal of Geophysical Research: Atmospheres*, *125*(20), e2020JD032712. https://doi.org/10.1029/2020JD032712

Mandel, J., Vejmelka, M., Kochanski, A., Farguell, A., Haley, J., Mallia, D., & Hilburn, K. (2019). An Interactive Data-Driven HPC System for Forecasting Weather, Wildland Fire, and Smoke. *2019 IEEE/ACM HPC for Urgent Decision Making (UrgentHPC)*, 35–44. https://doi.org/10.1109/UrgentHPC49580.2019.00010

Mason, R. (2017). *2017 National Emissions Inventory: January 2021 Updated Release, Technical Support Document*.

McCandless, T. C., Kosovic, B., & Petzke, W. (2020). Enhancing wildfire spread modelling by building a gridded fuel moisture content product with machine learning. *Machine Learning: Science and Technology*, *1*(3), 035010. https://doi.org/10.1088/2632-2153/aba480

McCarter, R. J. & Broido, A. Radiative and convective energy from wood crib fires. *Pyrodynamics* **2**, 65–85 (1965).

McClure, C. D., Pavlovic, N. R., Huang, S., Chaveste, M., & Wang, N. (2023). Consistent, high-accuracy mapping of daily and sub-daily wildfire growth with satellite observations. *International Journal of Wildland Fire*, *32*(5), 694–708. https://doi.org/10.1071/WF22048

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., & Shi, W. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360.

https://doi.org/10.1175/BAMS-87-3-343

Mirzaei, M., Bertazzon, S., & Couloigner, I. (2018). Modeling Wildfire Smoke Pollution by Integrating Land Use Regression and Remote Sensing Data: Regional Multi-Temporal Estimates for Public Health and Exposure Models. *Atmosphere*, *9*(9), Article 9. https://doi.org/10.3390/atmos9090335

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003JD003823

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Monthly Weather Review*, *137*(3), 991–1007. https://doi.org/10.1175/2008MWR2556.1

Mukunga, T., Forkel, M., Forrest, M., Zotta, R.-M., Pande, N., Schlaffer, S., & Dorigo, W. (2023). Effect of Socioeconomic Variables in Predicting Global Fire Ignition Occurrence. *Fire*, *6*(5), Article 5. https://doi.org/10.3390/fire6050197

Nakanishi, M., & Niino, H. (2004). An Improved Mellor–Yamada Level-3 Model with Condensation Physics: Its Design and Verification. *Boundary-Layer Meteorology*, *112*(1), 1–31. https://doi.org/10.1023/B:BOUN.0000020164.04146.98

Nicodemus, K. K., & Malley, J. D. (2009). Predictor correlation impacts machine learning

algorithms: Implications for genomic studies. Bioinformatics, 25(15), 1884–1890. https://doi.org/10.1093/bioinformatics/btp331

Ottmar, R., Sandberg, D., Riccardi, C., & Prichard, S. (2007). An overview of the Fuel Characteristic Classification System—Quantifying, classifying, and creating fuelbeds for resource planningThis article is one of a selection of papers published in the Special Forum on the Fuel Characteristic Classification System. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere - CAN J FOREST RES*, *37*. https://doi.org/10.1139/X07-077

Pan, X., Ichoku, C., Chin, M., Bian, H., Darmenov, A., Colarco, P., Ellison, L., Kucsera, T., da Silva, A., Wang, J., Oda, T., & Cui, G. (2020). Six global biomass burning emission datasets: Intercomparison and application in one global aerosol model. *Atmospheric Chemistry and Physics*, *20*(2), 969–994. https://doi.org/10.5194/acp-20-969-2020

Partanen, T. M., & Sofiev, M. (2022). Forecasting the regional fire radiative power for regularly ignited vegetation fires. Natural Hazards and Earth System Sciences, 22(4), 1335–1346. https://doi.org/10.5194/nhess-22-1335-2022

Paugam, R., Wooster, M., Atherton, J., Freitas, S. R., Schultz, M. G., & Kaiser, J. W. (2015). Development and optimization of a wildfire plume rise model based on remote sensing data inputs – Part 2. *Atmospheric Chemistry and Physics Discussions*, *15*(6), 9815–9895. https://doi.org/10.5194/acpd-15-9815-2015

Paugam, R., Wooster, M., Freitas, S., & Val Martin, M. (2016). A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models. *Atmospheric Chemistry and Physics*, *16*(2), Article 2. https://doi.org/10.5194/acp-16-907-2016

Pavlovic, R., Chen, J., Anderson, K., Moran, M. D., Beaulieu, P.-A., Davignon, D., & Cousineau, S. (2016). The FireWork air quality forecast system with near-real-time biomass burning emissions: Recent developments and evaluation of performance for the 2015 North American wildfire season. *Journal of the Air & Waste Management Association*, *66*(9), 819–841. https://doi.org/10.1080/10962247.2016.1158214

Perkins, S. E., Alexander, L. V., & Nairn, J. R. (2012). Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophysical Research Letters*, *39*(20). https://doi.org/10.1029/2012GL053361

Peterson, D., Hyer, E., & Wang, J. (2013). A short-term predictor of satellite-observed fire activity in the North American boreal forest: Toward improving the prediction of smoke emissions. *Atmospheric Environment*, *71*, 304–310. https://doi.org/10.1016/j.atmosenv.2013.01.052

Peterson, D. A., Thapa, L. H., Saide, P. E., Soja, A. J., Gargulinski, E. M., Hyer, E. J., Weinzierl, B., Dollner, M., Schöberl, M., Papin, P. P., Kondragunta, S., Camacho, C. P., Ichoku, C., Moore, R. H., Hair, J. W., Crawford, J. H., Dennison, P. E., Kalashnikova, O. V., Bennese, C. E., … Xu, C. (2022). Measurements from inside a Thunderstorm Driven by Wildfire: The 2019 FIREX-AQ Field Experiment. Bulletin of the American Meteorological Society, 103(9), E2140–E2167. https://doi.org/10.1175/BAMS-D-21-0049.1

Petters, M. D., Carrico, C. M., Kreidenweis, S. M., Prenni, A. J., DeMott, P. J., Collett Jr., J. L., & Moosmüller, H. (2009). Cloud condensation nucleation activity of biomass burning aerosol. *Journal of Geophysical Research: Atmospheres*, *114*(D22). https://doi.org/10.1029/2009JD012353

Pinto, M. M., DaCamara, C. C., Hurduc, A., Trigo, R. M., & Trigo, I. F. (2020). Enhancing the fire weather index with atmospheric instability information. *Environmental Research Letters*, *15*(9), 0940b7. https://doi.org/10.1088/1748-9326/ab9e22

Potter, B. (2018). The Haines Index – it's time to revise it or replace it. *International Journal of Wildland Fire*, *27*(7), 437. https://doi.org/10.1071/WF18015

Potter, B. E. (2005). The role of released moisture in the atmospheric dynamics associated with wildland fires. *International Journal of Wildland Fire*, *14*(1), 77–84. https://doi.org/10.1071/WF04045

Powers, D. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. Mach. Learn. Technol., 2.

Preisler, H. K., & Westerling, A. L. (2007). Statistical Model for Forecasting Monthly Large Wildfire Events in Western United States. *Journal of Applied Meteorology and Climatology*, *46*(7), 1020–1030. https://doi.org/10.1175/JAM2513.1

Prichard, S. J., Sandberg, D. V., Ottmar, R. D., Eberhardt, E., Andreu, A., Eagle, P., & Swedin, K. (2013). *United States Department of Agriculture Forest Service DEPA R TMEN T OF AGRICU LT URE Fuel Characteristic Classification System Version 3.0: Technical Documentation*.

Rao, K., Williams, A. P., Diffenbaugh, N. S., Yebra, M., & Konings, A. G. (2022). Plant-water sensitivity regulates wildfire vulnerability. *Nature Ecology & Evolution*, *6*(3), Article 3. https://doi.org/10.1038/s41559-021-01654-2

Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., Raffuse, S. M., & Balmes, J. R. (2015). Spatiotemporal Prediction of Fine Particulate Matter During the 2008 Northern California Wildfires Using Machine Learning. *Environmental Science*

& Technology, *49*(6), 3887–3896. https://doi.org/10.1021/es505846r

Rodrigues, M., Alcasena, F., Gelabert, P., & Vega-García, C. (2020). Geospatial Modeling of Containment Probability for Escaped Wildfires in a Mediterranean Region. *Risk Analysis*, *40*(9), 1762–1779. https://doi.org/10.1111/risa.13524

Rooney, B., Wang, Y., Jiang, J. H., Zhao, B., Zeng, Z.-C., & Seinfeld, J. H. (2020). Air quality impact of the Northern California Camp Fire of November 2018. *Atmospheric Chemistry and Physics*, *20*(23), 14597–14616. https://doi.org/10.5194/acp-20-14597-2020

Rothermel, R. C. (1972). A mathematical model for predicting fire spread in wildland fuels. *INT*, *115*, 40.

Rothermel, R. C. (1991). *Predicting Behavior and Size of Crown Fires in the Northern Rocky Mountains*. U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Roy, B., Pouliot, G. A., Gilliland, A., Pierce, T., Howard, S., Bhave, P. V., & Benjey, W. (2007). Refining fire emissions for air quality modeling with remotely sensed fire counts: A wildfire case study. *Atmospheric Environment*, *41*(3), 655–665. https://doi.org/10.1016/j.atmosenv.2006.08.037

Roy, A., Choi, Y., Souri, A. H., Jeon, W., Diao, L., Pan, S., & Westenbarger, D. (2018). Effects of Biomass Burning Emissions on Air Quality Over the Continental USA: A Three-Year Comprehensive Evaluation Accounting for Sensitivities Due to Boundary Conditions and Plume Rise Height. In *Energy, Environment, and Sustainability* (pp. 245–278). Springer Nature. https://doi.org/10.1007/978-981-10-7332-8_12

Saide, P. E., Peterson, D. A., da Silva, A., Anderson, B., Ziemba, L. D., Diskin, G., Sachse, G., Hair, J., Butler, C., Fenn, M., Jimenez, J. L., Campuzano-Jost, P., Perring, A. E., Schwarz,

J. P., Markovic, M. Z., Russell, P., Redemann, J., Shinozuka, Y., Streets, D. G., … Carmichael, G. R. (2015). Revealing important nocturnal and day-to-day variations in fire smoke emissions through a multiplatform inversion. *Geophysical Research Letters*, *42*(9), 3609–3618. https://doi.org/10.1002/2015GL063737

Saide, P. E., Thapa, L. H., Ye, X., Pagonis, D., Campuzano-Jost, P., Guo, H., Schuneman, M. L., Jimenez, J.-L., Moore, R., Wiggins, E., Winstead, E., Robinson, C., Thornhill, L., Sanchez, K., Wagner, N. L., Ahern, A., Katich, J. M., Perring, A. E., Schwarz, J. P., … Shingler, T. J. (2022). Understanding the Evolution of Smoke Mass Extinction Efficiency Using Field Campaign Measurements. *Geophysical Research Letters*, *49*(18), e2022GL099175. https://doi.org/10.1029/2022GL099175

Saide, P. E., Krishna, M., Ye, X., Thapa, L. H., Turney, F., Howes, C., & Schmidt, C. C. (2023). Estimating Fire Radiative Power Using Weather Radar Products for Wildfires. *Geophysical Research Letters*, *50*(21), e2023GL104824. https://doi.org/10.1029/2023GL104824

Sazib, N., Bolten, J. D., & Mladenova, I. E. (2022). Leveraging NASA Soil Moisture Active Passive for Assessing Fire Susceptibility and Potential Impacts Over Australia and California. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 779–787. https://doi.org/10.1109/JSTARS.2021.3136756

Schmidt, C. C., Ssec, U.-M., Hoffman, J., Ssec, U.-M., Prins, E., Ssec, U.-M., Lindstrom, S., & Ssec, U.-M. (2012). *GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Fire / Hot Spot Characterization*.

Schroeder, W., Oliva, P., Giglio, L., & Csiszar, I. A. (2014). The New VIIRS 375m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of*

Environment, *143*, 85–96. https://doi.org/10.1016/j.rse.2013.12.008

Schum, S. K., Zhang, B., Dzepina, K., Fialho, P., Mazzoleni, C., & Mazzoleni, L. R. (2018). Molecular and physical characteristics of aerosol at a remote free troposphere site: Implications for atmospheric aging. *Atmospheric Chemistry and Physics*, *18*(19), Article 19. https://doi.org/10.5194/acp-18-14017-2018

Seiler, W., & Crutzen, P. J. (1980). Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. Climatic Change, 2(3), 207–247. https://doi.org/10.1007/BF00137988

Sessions, W. R., Fuelberg, H. E., Kahn, R. A., & Winker, D. M. (2011). An investigation of methods for injecting emissions from boreal wildfires using WRF-Chem during ARCTAS. *Atmospheric Chemistry and Physics*, *11*(12), Article 12. https://doi.org/10.5194/acp-11-5719-2011

Shi, H., Jiang, Z., Zhao, B., Li, Z., Chen, Y., Gu, Y., Jiang, J. H., Lee, M., Liou, K.-N., Neu, J. L., Payne, V. H., Su, H., Wang, Y., Witek, M., & Worden, J. (2019). Modeling Study of the Air Quality Impact of Record-Breaking Southern California Wildfires in December 2017. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6554–6570. https://doi.org/10.1029/2019JD030472

Shin, H. H., Hong, S.-Y., & Dudhia, J. (2012). Impacts of the Lowest Model Level Height on the Performance of Planetary Boundary Layer Parameterizations. *Monthly Weather Review*, *140*(2), 664–682. https://doi.org/10.1175/MWR-D-11-00027.1

Sofiev, M., Ermakova, T., & Vankevich, R. (2012). Evaluation of the smoke-injection height from wild-land fires using remote-sensing data. *Atmospheric Chemistry and Physics*, *12*(4), Article 4. https://doi.org/10.5194/acp-12-1995-2012

Soja, A., Choi, H., Gargulinski, E., Tackett, J., Vaughan, M., & Winker, D. (2021). *Defining the vertical structure of smoke aerosols: CALIPSO-based smoke plume detrainment heights*. *2021*, A13B-05.

Soja, A. J., Cofer, W. R., Shugart, H. H., Sukhinin, A. I., Stackhouse Jr., P. W., McRae, D. J., & Conard, S. G. (2004). Estimating fire emissions and disparities in boreal Siberia (1998–2002). *Journal of Geophysical Research: Atmospheres*, *109*(D14). https://doi.org/10.1029/2004JD004570

Srock, A. F., Charney, J. J., Potter, B. E., & Goodrick, S. L. (2018). The Hot-Dry-Windy Index: A New Fire Weather Index. *Atmosphere*, *9*(7), Article 7. https://doi.org/10.3390/atmos9070279

Stocks, B. J. & Kauffman, J. B. in *Sediment Records of Biomass Burning and Global Change* 169–188 (Springer Berlin Heidelberg, 1997).

Stockwell, W. R., Kirchner, F., Kuhn, M., & Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research: Atmospheres*, *102*(D22), 25847–25879. https://doi.org/10.1029/97JD00849

Strahler, A., Prepared, D., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E., & Moody, A. (1999). *MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD) Version 5.0 MODIS Land Cover and Land-Cover Change Principal Investigator*.

Suarez, M. J. (1999). *Technical Report Series on Global Modeling and Data Assimilation*. *15*.

Tao, Z., He, H., Sun, C., Tong, D., & Liang, X.-Z. (2020). Impact of Fire Emissions on U.S. Air Quality from 1997 to 2016–A Modeling Study in the Satellite Era. *Remote Sensing*, *12*(6), Article 6. https://doi.org/10.3390/rs12060913

Thapa, L. H., Ye, X., Hair, J. W., Fenn, M. A., Shingler, T., Kondragunta, S., Ichoku, C.,

Dominguez, R., Ellison, L., Soja, A. J., Gargulinski, E., Ahmadov, R., James, E., Grell, G. A., Freitas, S. R., Pereira, G., & Saide, P. E. (2022). Heat flux assumptions contribute to overestimation of wildfire smoke injection into the free troposphere. *Communications Earth & Environment*, *3*(1), Article 1. https://doi.org/10.1038/s43247-022-00563-x

Thapa, L. H., Saide, P.E., Bortnik, J., Berman, M.T., da Silva, A., Peterson, D.A., Li, F., Kondragunta, S., Ahmadov, R., James, E., Ye, X., Soja, A.J., Wiggins, E.B., Gargulinski, E. (Accepted). Forecasting daily fire radiative energy using data driven methods and machine learning techniques. *Journal of Geophysical Research-Atmospheres*

Tory, K. J., & Kepert, J. D. (2021). Pyrocumulonimbus Firepower Threshold: Assessing the Atmospheric Potential for pyroCb. *Weather and Forecasting*, *36*(2), 439–456. https://doi.org/10.1175/WAF-D-20-0027.1

Turney, F. A., Saide, P. E., Jimenez Munoz, P. A., Muñoz-Esparza, D., Hyer, E. J., Peterson, D. A., Frediani, M. E., Juliano, T. W., DeCastro, A. L., Kosović, B., Ye, X., & Thapa, L. H. (2023). Sensitivity of Burned Area and Fire Radiative Power Predictions to Containment Efforts, Fuel Density, and Fuel Moisture Using WRF-Fire. *Journal of Geophysical Research: Atmospheres*, *128*(18), e2023JD038873. https://doi.org/10.1029/2023JD038873

Val Martin, M., Logan, J. A., Kahn, R. A., Leung, F. Y., Nelson, D. L., & Diner, D. J. (2010). Smoke injection heights from fires in North America: Analysis of 5 years of satellite observations. *Atmospheric Chemistry and Physics*, *10*(4), Article 4. https://doi.org/10.5194/acp-10-1491-2010

Val Martin, M., Kahn, R. A., Logan, J. A., Paugam, R., Wooster, M., & Ichoku, C. (2012). Space-based observational constraints for 1-D fire smoke plume-rise models. *Journal of*

Geophysical Research: Atmospheres, *117*(D22), Article D22. https://doi.org/10.1029/2012JD018370

Val Martin, M., Kahn, R., & Tosca, M. (2018). A Global Analysis of Wildfire Smoke Injection Heights Derived from Space-Based Multi-Angle Imaging. *Remote Sensing*, *10*(10), Article 10. https://doi.org/10.3390/rs10101609

Viegas, D. X. (2004). Slope and wind effects on fire propagation. *International Journal of Wildland Fire*, *13*(2), 143. https://doi.org/10.1071/WF03046

Wagner, C. E. V. (1987). *Development and structure of the Canadian Forest Fire Weather Index System*. Minister of Supply and Services Canada.

Walter, C., Freitas, S. R., Kottmeier, C., Kraut, I., Rieger, D., Vogel, H., & Vogel, B. (2016). The importance of plume rise on the concentrations and atmospheric impacts of biomass burning aerosol. *Atmospheric Chemistry and Physics*, *16*(14), Article 14. https://doi.org/10.5194/acp-16-9201-2016

Wang, S. S.-C., Qian, Y., Leung, L. R., & Zhang, Y. (2022). Interpreting machine learning prediction of fire emissions and comparison with FireMIP process-based models. Atmospheric Chemistry and Physics, 22(5), 3445–3468. https://doi.org/10.5194/acp-22-3445-2022

Warneke, C., Schwarz, J. P., Dibb, J., Kalashnikova, O., Frost, G., Al-Saad, J., Brown, S. S., Brewer, Wm. A., Soja, A., Seidel, F. C., Washenfelder, R. A., Wiggins, E. B., Moore, R. H., Anderson, B. E., Jordan, C., Yacovitch, T. I., Herndon, S. C., Liu, S., Kuwayama, T., … Team, T. F.-A. S. (2023). Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ). *Journal of Geophysical Research: Atmospheres*, *128*(2), e2022JD037758. https://doi.org/10.1029/2022JD037758

Watson, G. L., Telesca, D., Reid, C. E., Pfister, G. G., & Jerrett, M. (2019). Machine learning models accurately predict ozone exposure during wildfire events. *Environmental Pollution*, *254*, 112792. https://doi.org/10.1016/j.envpol.2019.06.088

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase Western U.S. forest wildfire activity. Science, 313(5789), 940–943. https://doi.org/10.1126/science.1128834

Western Regional Air Partnership. *2002 Fire Emission Inventory for the WRAP Region—Phase II* Project No. 178–6 (Air Sciences, INC, 2005)

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geosci. Model Dev*, *4*, 625–641. https://doi.org/10.5194/gmd-4-625-2011

Wilcox, E. M. (2012). Direct and semi-direct radiative forcing of smoke aerosols over clouds. *Atmospheric Chemistry and Physics*, *12*(1), 139–149. https://doi.org/10.5194/acp-12-139-2012

Wilmot, T. Y., Mallia, D. V., Hallar, A. G., & Lin, J. C. (2022). Wildfire plumes in the Western US are reaching greater heights and injecting more aerosols aloft as wildfire activity intensifies. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-16607-3

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, 79–82. https://doi.org/10.3354/cr030079

Wooster, M. J., Roberts, G., Perry, G. L. W., & Kaufman, Y. J. (2005). Retrieval of biomass

combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *Journal of Geophysical Research: Atmospheres*, *110*(D24). https://doi.org/10.1029/2005JD006318

Wooster, M. J., Zhukov, B., & Oertel, D. (2003). Fire radiative energy for quantitative study of biomass burning: Derivation from the BIRD experimental satellite and comparison to MODIS fire products. *Remote Sensing of Environment*, *86*(1), 83–107. https://doi.org/10.1016/S0034-4257(03)00070-1

Yao, J., Brauer, M., & Henderson, S. B. (2013). Evaluation of a Wildfire Smoke Forecasting System as a Tool for Public Health Protection. *Environmental Health Perspectives*. https://doi.org/10.1289/ehp.1306768

Yao, J., Raffuse, S. M., Brauer, M., Williamson, G. J., Bowman, D. M. J. S., Johnston, F. H., & Henderson, S. B. (2018). Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite. *Remote Sensing of Environment*, *206*, 98–106. https://doi.org/10.1016/j.rse.2017.12.027

Ye, X., Arab, P., Ahmadov, R., James, E., Grell, G. A., Pierce, B., Kumar, A., Makar, P., Chen, J., Davignon, D., Carmichael, G. R., Ferrada, G., McQueen, J., Huang, J., Kumar, R., Emmons, L., Herron-Thorpe, F. L., Parrington, M., Engelen, R., … Saide, P. E. (2021). Evaluation and intercomparison of wildfire smoke forecasts from multiple modeling systems for the 2019 Williams Flats fire. *Atmospheric Chemistry and Physics*, *21*(18), 14427–14469. https://doi.org/10.5194/acp-21-14427-2021

Ye, X., Deshler, M., Lyapustin, A., Wang, Y., Kondragunta, S., & Saide, P. (2022a). Assessment of Satellite AOD during the 2020 Wildfire Season in the Western U.S. *Remote Sensing*,

14(23), Article 23. https://doi.org/10.3390/rs14236113

Ye, X., Saide, P. E., Hair, J., Fenn, M., Shingler, T., Soja, A., Gargulinski, E., & Wiggins, E. (2022b). Assessing Vertical Allocation of Wildfire Smoke Emissions Using Observational Constraints From Airborne Lidar in the Western U.S. *Journal of Geophysical Research: Atmospheres*, *127*(21), e2022JD036808. https://doi.org/10.1029/2022JD036808

Yin, J., Zhan, X., & Liu, J. (2020). NOAA Satellite Soil Moisture Operational Product System (SMOPS) Version 3.0 Generates Higher Accuracy Blended Satellite Soil Moisture. *Remote Sensing*, *12*(17), Article 17. https://doi.org/10.3390/rs12172861

Young, J. D., Thode, A. E., Huang, C.-H., Ager, A. A., & Fulé, P. Z. (2019). Strategic application of wildland fire suppression in the southwestern United States. *Journal of Environmental Management*, *245*, 504–518. https://doi.org/10.1016/j.jenvman.2019.01.003

Yu, P., Toon, O. B., Bardeen, C. G., Bucholtz, A., Rosenlof, K. H., Saide, P. E., Da Silva, A., Ziemba, L. D., Thornhill, K. L., Jimenez, J., Campuzano‐Jost, P., Schwarz, J. P., Perring, A. E., Froyd, K. D., Wagner, N. L., Mills, M. J., & Reid, J. S. (2016). Surface dimming by the 2013 Rim Fire simulated by a sectional aerosol model. *Journal of Geophysical Research: Atmospheres*, *121*(12), Article 12. https://doi.org/10.1002/2015JD024702

Zhuang, Y., Fu, R., Santer, B. D., Dickinson, R. E., & Hall, A. (2021). Quantifying contributions of natural variability and anthropogenic forcings on increased fire weather risk over the western United States. *Proceedings of the National Academy of Sciences*, *118*(45), Article 45. https://doi.org/10.1073/pnas.2111875118

Zou, Y., O'Neill, S. M., Larkin, N. K., Alvarado, E. C., Solomon, R., Mass, C., Liu, Y., Odman, M. T., & Shen, H. (2019). Machine learning-based integration of high-resolution wildfire

smoke simulations and observations for regional health impact assessment. *International Journal of Environmental Research and Public Health*, *16*(12), Article 12. https://doi.org/10.3390/ijerph16122137