Research Report 238

Ambient Air Pollution and COVID-19 in California

Michael Kleeman et al.

Appendix E. Supplemental Information for Chapter 7: Association Between Air Pollution and Post-Acute Sequelae of SARS-CoV-2 Infection

---

Appendix E was reviewed by the HEI Review Committee and has been lightly edited for spelling, grammar, punctuation, and cross-references to the main report.

Correspondence may be addressed to Dr. Michael Kleeman, University of California, Davis, Department of Civil and Environmental Engineering, 1 Shields Avenue, Davis, CA 95616; email: *mjkleeman@ucdavis.edu*.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83998101 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not necessarily reflect the views of the Agency; thus, no official endorsement by it should be inferred. It also has not been reviewed by private-party institutions, including those that support the Health Effects Institute, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

CONTENTS

Random Forest Regression Datasets and Methods 2
Results 6
Statistical Analysis 6
Time Series Analysis 14
Annual Average 21
Summary 32
PASC Disease Groupings, ICD Names, ICD-10 Codes, and Example Diagnoses 33
Multipollutant Associations with PASC Disease Categories at 3 Months and 12 Months 39
References 43

Random Forest Regression Datasets and Methods

The statistical metrics chosen to evaluate the performance of random forest regression (RFR) training in the current study are listed in Table E1.

Table E1. Statistical Measures and Benchmarks for Model Performance Evaluation Discussed in This Work

Statistics/Abbreviation	Definition[a]	Benchmarks[b]
Mean Fractional Bias (MFB)	$MFB = \frac{2}{N} \times \sum \frac{(P_i - O_i)}{(P_i + O_i)}$	• 24-hr total and speciated PM Goal ≤ ±0.3; Criteria ≤ ±0.6
Mean Fractional Error (MFE)	$MFE = \frac{2}{N} \times \sum \frac{\lvert P_i - O_i \rvert}{(P_i + O_i)}$	• 24-hr total and speciated PM Goal ≤ ±0.5; Criteria ≤ ±0.75
Normalized Mean Bias (NMB)	$NMB = \frac{1}{N} \times \sum \frac{(P_i - O_i)}{(O_i)}$	• 24-hr $PM_{2.5}$, SO_4, NH_4 Goal < ±0.1; Criteria ≤ ±0.3 • 24-hr NO_3 Goal < ±0.15; Criteria ≤ ±0.65 • 24-hr OC Goal < ±0.15; Criteria ≤ ±0.5 • 24-hr EC Goal < ±0.2; Criteria ≤ ±0.4
Normalized Mean Error (NME)	$NME = \frac{1}{N} \times \sum \frac{\lvert P_i - O_i \rvert}{(O_i)}$	• 24-hr $PM_{2.5}$, SO_4, NH_4 Goal < 0.35; Criteria ≤ 0.5 • 24-hr NO_3 Goal < 0.65; Criteria ≤ 1.15 • 24-hr OC Goal < 0.45; Criteria ≤ 0.65 • 24-hr EC Goal < 0.5; Criteria ≤ 0.75

Statistics/Abbreviation	Definition[a]	Benchmarks[b]
Correlation coefficient (r)	$r = \frac{\sum[(P_i - \overline{P}) \times (O_i - \overline{O})]}{\sqrt{\sum(P_i - \overline{P})^2 \times \sum(O_i - \overline{O})^2}}$	• 24-hr $PM_{2.5}$, SO_4, NH_4 Goal > 0.7; Criteria > 0.4

NH_4 = ammonium; NO_3 = nitrate; SO_4 = sulfur trioxide.
[a] Observations (O) and model predictions (P).
[b] Benchmarks for photochemical performance, suggested by Emery et al.[1] "Goals" are met by one-third of top-performing models, while "Criteria" are met by two-thirds of models.

Four major support elements were used in the RFR approach: (1) surface monitoring data from the US EPA and PurpleAir, (2) Moderate Resolution Imaging Spectroradiometer aerosol optical depth (AOD) retrievals, (3) meteorology data from Weather Research and Forecast, and (4) CTM results from the UCD/CIT model. CTM predictions for the years 2016 and 2020 were corrected using the RFR approach, as these represent the chronic and acute exposure fields in the current study.

Figure E1 illustrates the basic steps of how the RFR technique was employed in this study, using $PM_{2.5}$ mass as an example. The first step was calculated $FB_{PM2.5\ mass}$ based on UCD/CIT $PM_{2.5}$ mass predictions and EPA daily average $PM_{2.5}$ mass measurements. The data for each month were randomly split into a training set (75%) and a test set (25%), with training features listed in Figure E1. The second step trained the RFR model. The RFR algorithm constructs a large number of decision trees and then combines the predictions from all the trees to arrive at a final prediction for the test data. To evaluate model accuracy, the predictions are compared to the measured values in the test dataset. In the third step, the trained model is used to predict the $FB_{PM2.5\ mass}$ values for every grid cell in the modeling region, using the support variables listed in Table 40. The RFR predictions are independent of the original FB equations listed in Table 39; therefore, any extreme FB values must be limited to the range between +2 and -2.

Table E2. Data and Variables Used in the RFR Training for Southern California, 2016 and 2020

Data Source	Variables Used [a]	2016	2020
EPA Air Quality System	Daily Average $PM_{2.5}$ mass, $PM_{2.5}$ OC, $PM_{2.5}$ EC, $PM_{2.5}$ N(_III), $PM_{2.5}$ N(V), $PM_{2.5}$ S(VI)	Y	Y
PurpleAir Sensor	Daily Average $PM_{2.5}$ (PurpleAir_$PM_{2.5}$ mass)		Y
MODIS	AOD	Y	Y
WRF Simulations	Surface air temperature at 2 m Relative humidity at 2 m Precipitation rates at the surface Planetary boundary layer heights (PBL) Surface wind speed and directions at 10 m (U, V, W)	Y	Y

Data Source	Variables Used [a]	2016	2020
UCD/CIT Simulations	*$PM_{2.5}$Tracer1~9*, *$PM_{2.5}$ mass*, *$PM_{2.5}$ OC*, *$PM_{2.5}$ EC*, *$PM_{2.5}$ N(_III)*, *$PM_{2.5}$ N(V)*, *$PM_{2.5}$ S(VI)*	Y	Y

[a] $PM_{2.5}$ species: $PM_{2.5}$ OC (organic compounds), $PM_{2.5}$ EC (elemental carbon), $PM_{2.5}$ N(_III) (ammonium ion), $PM_{2.5}$ N(V) (nitrate) and $PM_{2.5}$ S(VI) (sulfate).

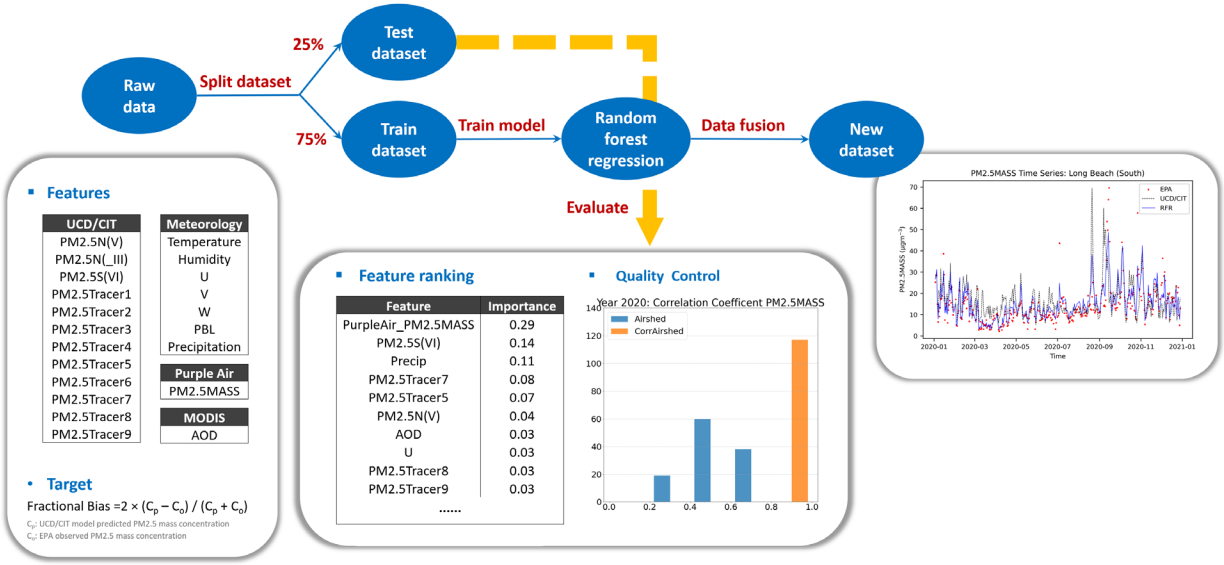

Figure E1. Flow chart of random forest algorithm.

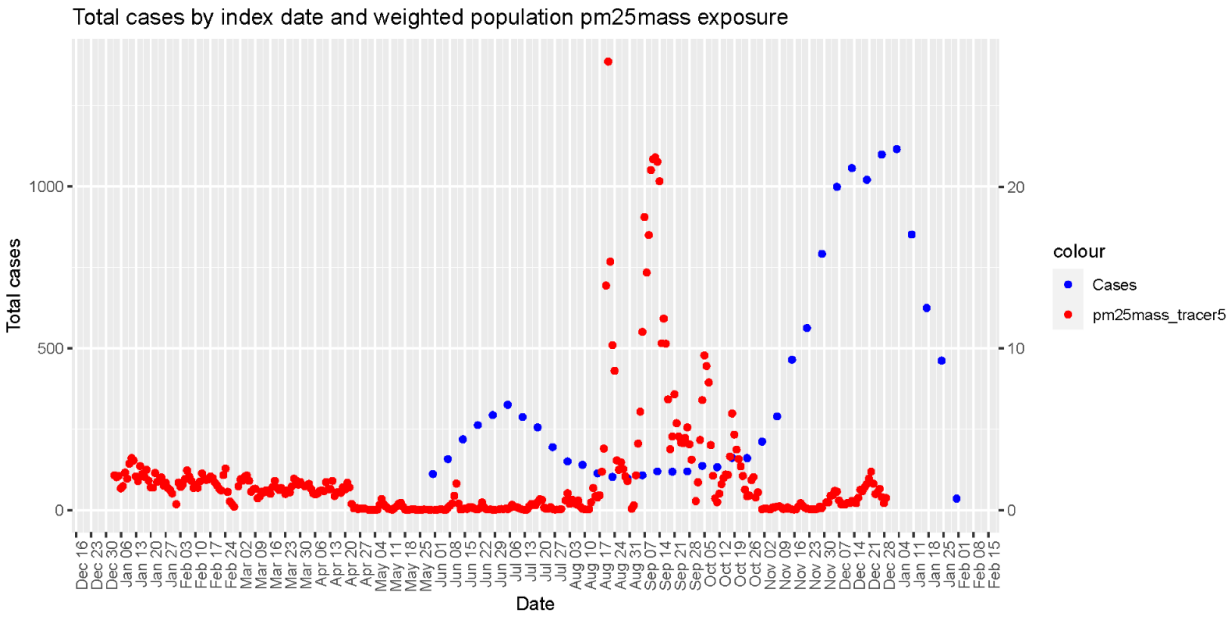

Figure E2. Time history of COVID-19 total cases and wildfire population weighted-average exposure concentration (labelled as tracer5). Case count (left axis) has units of cases/day, and exposure concentration (right axis) has units of µg m-3

Results

The model performance of the original UCD/CIT model simulations and the RFR corrected predictions are evaluated using three separate methods: 1) statistical analysis based on correlation coefficient (*r*), NME, and NMB calculated at EPA measurement sites; 2) time series analysis of predicted and measured concentrations at selected EPA sites; and 3) comparison of annul average concentration fields before and after RFR processing. Results are presented for the years 2020 and 2016, as both datasets were used for different portions of the COVID-19 health effects analysis.

Statistical Analysis

Year 2020

Figure E3, Figure E4, and Figure E5 illustrate the number of comparison sites that fall into different performance bins for correlation coefficient (*r*), NME, and NMB, respectively, during simulation for the year 2020. All comparisons are based on daily average concentrations throughout the year. The target level of performance is indicated by the green background shading in each figure. Darker green corresponds to performance goals (the best a CTM can achieve), and the lighter green corresponds to performance criteria (typical CTM performance) as defined by Emery et al.[1] Application of the RFR technique improves the CTM performance for $PM_{2.5}$ mass, OC, EC, and N(V) (nitrate). In many cases, the original CTM predictions met the performance criteria, and the application of the RFR improved performance at the comparison sites, such that the new results meet performance goals.

The RFR scheme implemented in the current study weights the pollutants with higher concentrations more heavily than pollutants with lower concentrations; thus, the improvements for $PM_{2.5}$ mass are greater than the improvements for components of $PM_{2.5}$ mass, such as OC, EC, and N(V). Improvements for $PM_{2.5}$ S(VI) (sulfate) and N(_III) (ammonium ion) are even more muted, as seen in Figure E3, Figure E4, and Figure E5. $PM_{2.5}$ S(VI) (sulfate) concentrations are generally low in the study domain, and the CTM struggles to accurately predict the seasonal trends for this component. Sulfate is hygroscopic and nonvolatile. The amount of sulfate that condenses on particles often influences the predicted amount of particle-phase ammonium nitrate. The uncertainties in these predicted concentrations are coupled and difficult to correct using the sparse network of measured concentrations in the study region. This lack of measurement support poses similar problems for LUR models that seek to predict exposure fields for air pollution studies.

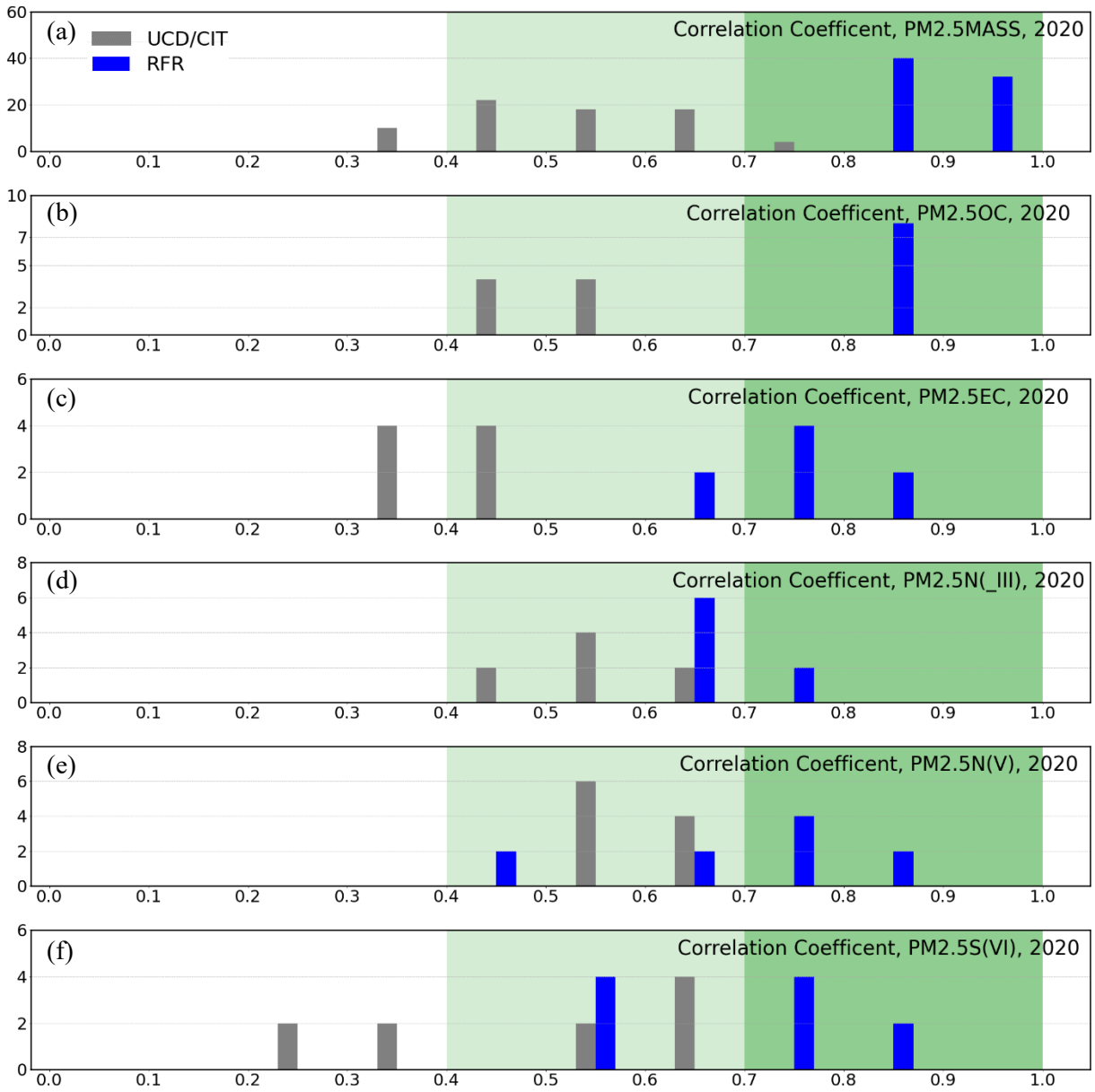

Figure E3. Number of sites in Correlation Coefficient bins for 2020. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ OC, (c) $PM_{2.5}$ EC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI). Target performance goals are shaded in darker green.

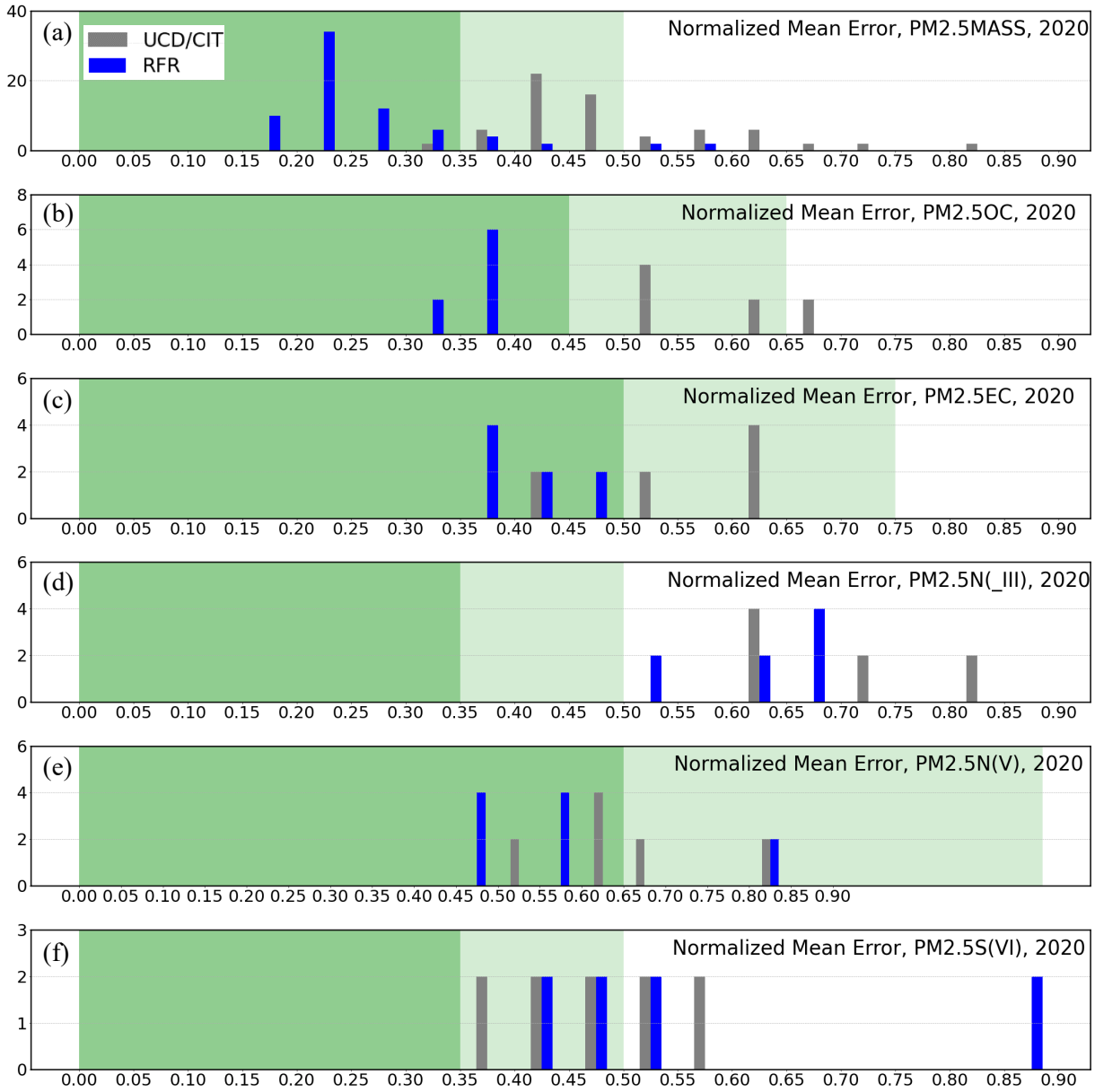

Figure E4. Number of sites in NME bins for 2020. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ OC, (c) $PM_{2.5}$ EC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI). Target performance goals are shaded in darker green.

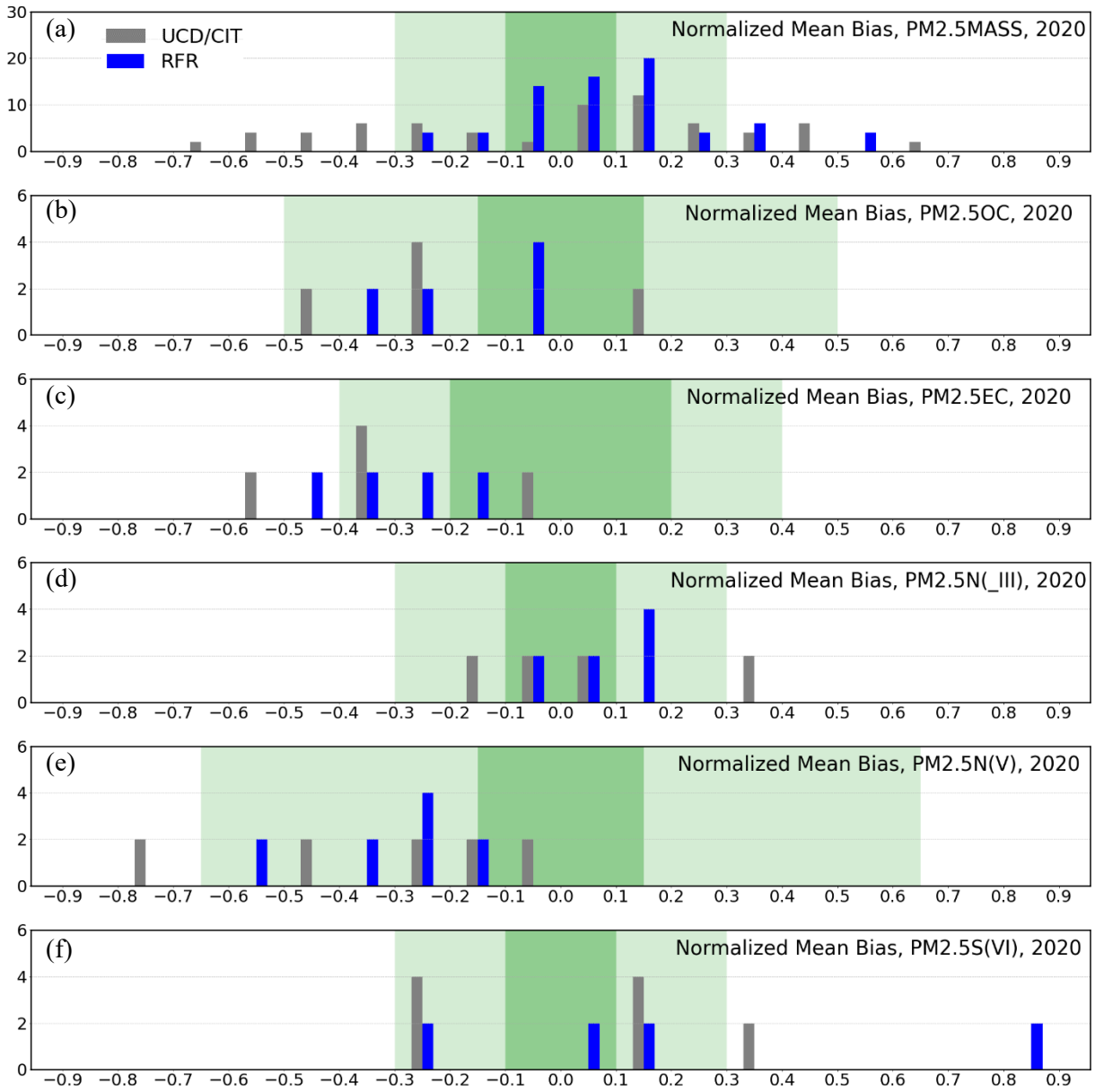

Figure E5. Number of sites in NMB bins for 2020. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ OC, (c) $PM_{2.5}$ EC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI). Target performance goals are shaded in darker green.

Year 2016

Figure E6, Figure E7, and Figure E8 illustrate the number of comparison sites that fall into different performance bins for correlation coefficient *r*, NME, and NMB, respectively, during simulations for the year 2016. Results from an earlier bias correction (BC) approach based on a constrained multilinear regression analysis are shown (Bias_Corr). All comparisons shown in Figure E6, Figure E7, and Figure E8 are based on monthly average concentrations throughout the year. Concentration fields for 2016 were used to characterize chronic exposures during the current study; thus, monthly average concentrations were used to characterize the seasonal cycle of pollutant exposures. It should be noted that the support variables included in the 2016 analysis do not include the measurements from the PurpleAir network that started reporting data in the year 2017.

The results from the RFR approach improve all the performance metrics for predicted $PM_{2.5}$ mass in the year 2016, compared to the original CTM predictions and the original BC predictions. NMB and NME also improve for $PM_{2.5}$ EC and $PM_{2.5}$ N(V) with the use of the RFR method, but improvements relative to the original BC predictions for other PM species are less obvious. The original CTM predictions for 2016 were biased high (Figure E8a) due to an underprediction of wind speeds. Both the RFR method and the original constrained multilinear regression (MLR) method are effective at removing this bias for $PM_{2.5}$ mass, but the magnitude of the correction in 2016 increases the difficulty of accurately adjusting concentrations for $PM_{2.5}$ species components that are present at lower concentrations. The overall performance across species is similar between the RFR method and the original BC procedure in the year 2016.

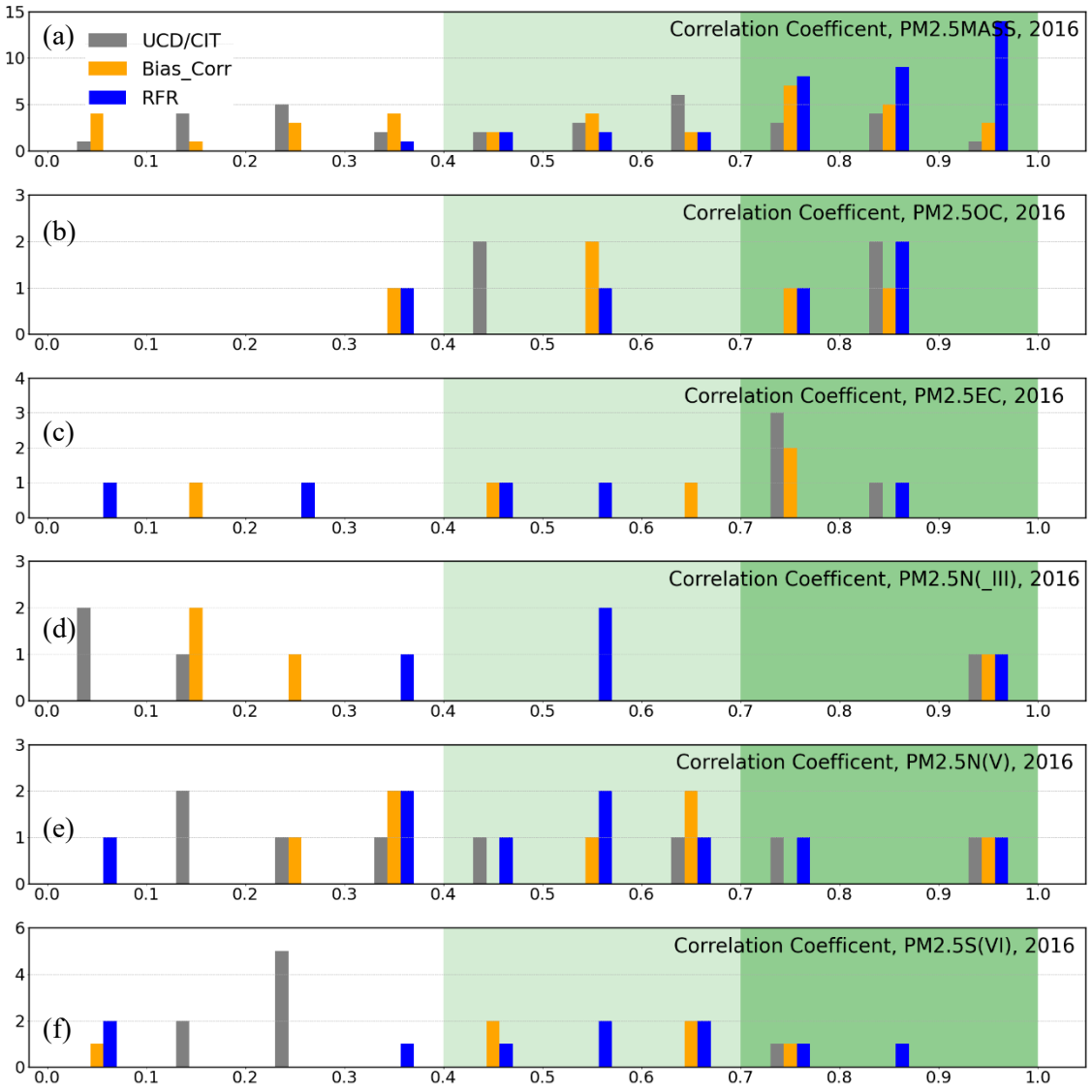

Figure E6. Number of sites in Correlation Coefficient bins for 2016. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ OC, (c) $PM_{2.5}$ EC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

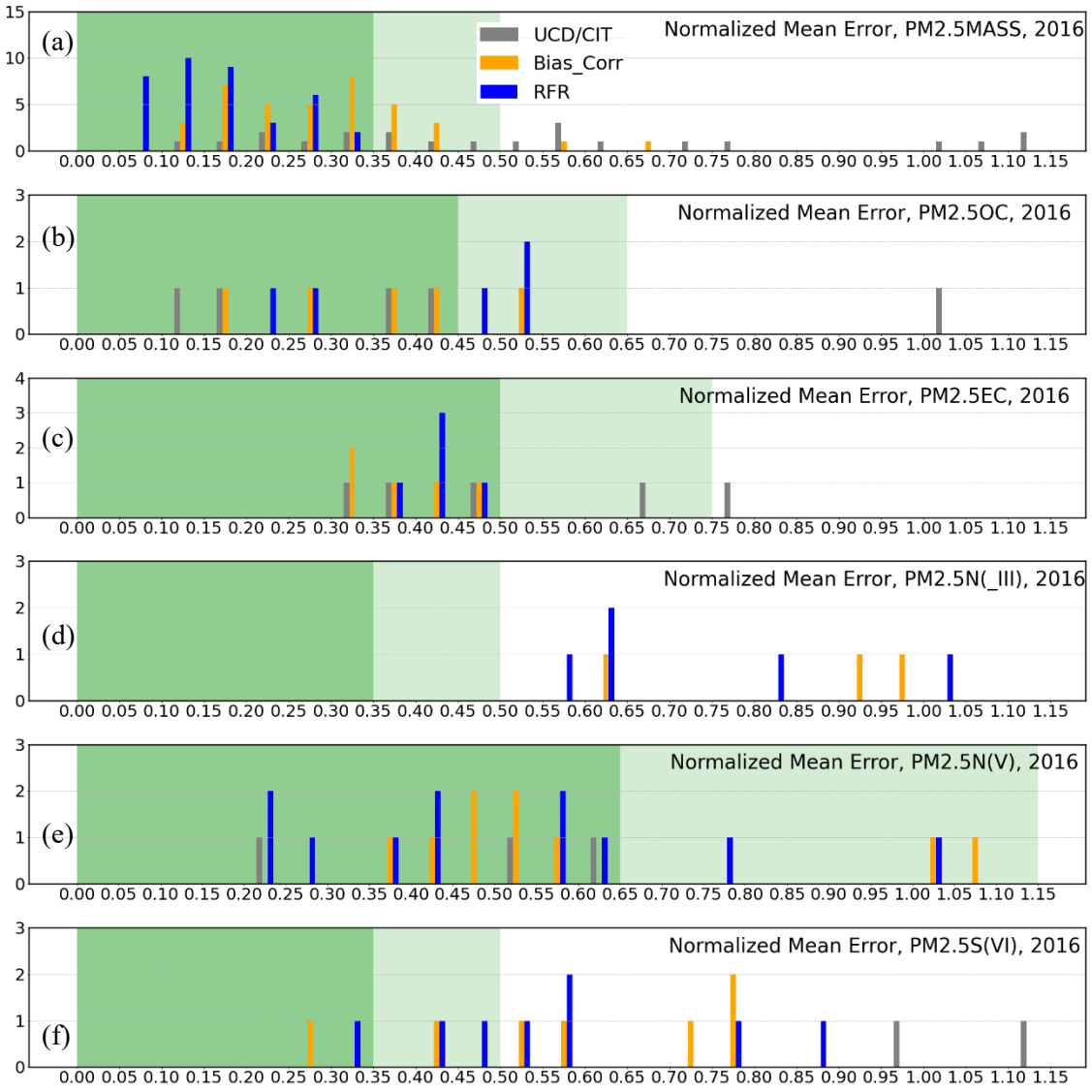

Figure E7. Number of sites in NME bins for 2016. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ OC, (c) $PM_{2.5}$ EC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

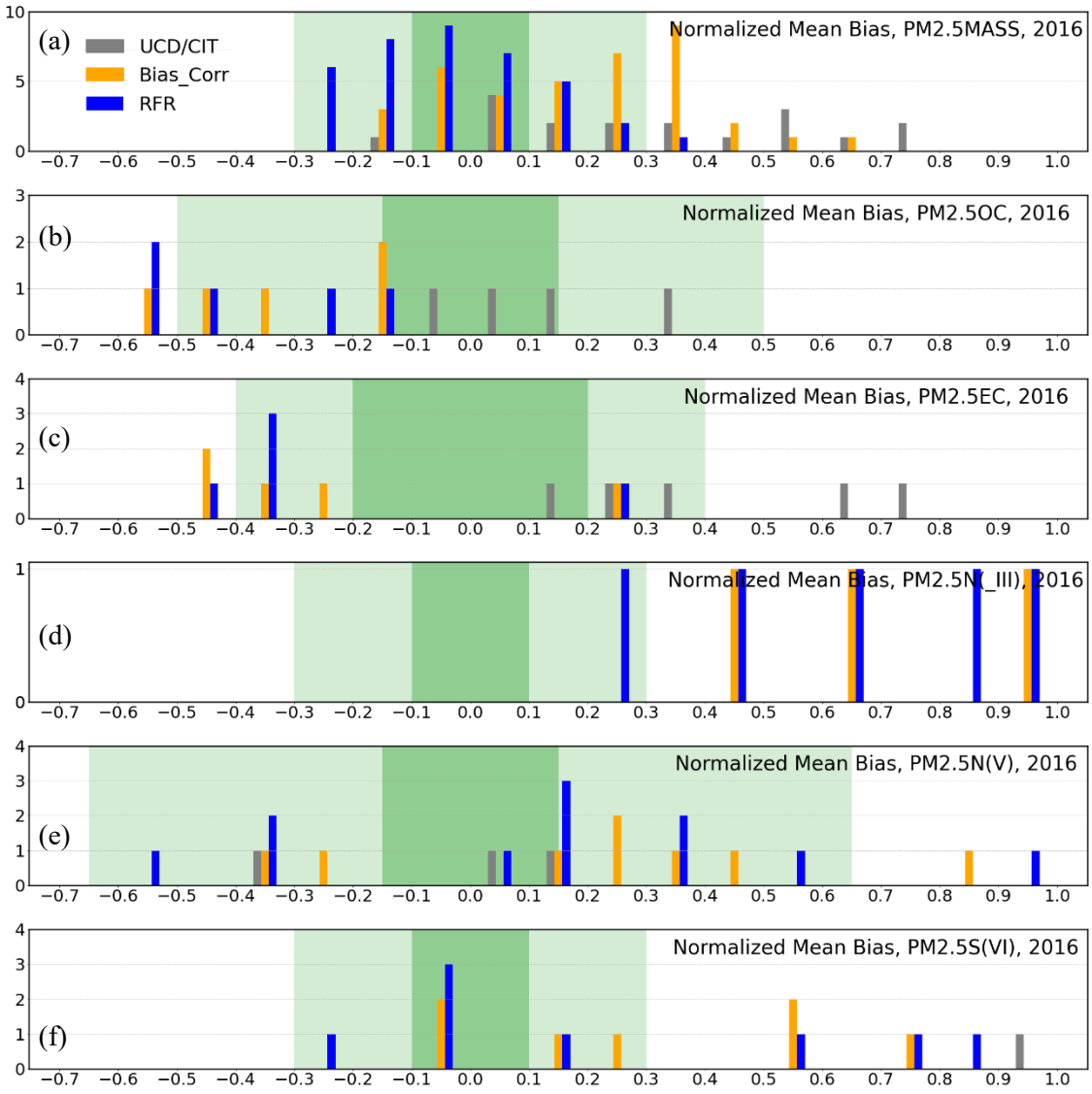

Figure E8. Number of sites in NMB bins for 2016. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ OC, (c) $PM_{2.5}$ EC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

Time Series Analysis

Los Angeles

Figure E9 and Figure E10 illustrate the time series of predicted and measured concentrations in Los Angeles, California, in 2020 and 2016, respectively. Measured concentrations are illustrated as red dots; CTM predictions are illustrated as dashed lines, and RFR predictions are shown as blue lines. $PM_{2.5}$ mass predictions in the year 2020 (Figure E9a) are improved in two important ways after RFR corrections are applied. The reduced $PM_{2.5}$ mass concentrations that occurred in March and April 2020 as a result of "shelter-in-place" orders are more accurately simulated using the RFR procedure compared to the overpredictions from the original CTM. These improvements in predicting low concentrations are also obvious for $PM_{2.5}$ EC (Figure E9b) and $PM_{2.5}$ OC (Figure E9c), $PM_{2.5}$ N(_III) (Figure E9d), and $PM_{2.5}$ N(V) (Figure E9e) The peaks in the $PM_{2.5}$ mass concentrations measured during the fall and early winter of 2020 are also predicted more accurately by the RFR method compared to the original CTM predictions. These improvements are not obvious in the statistical metrics highlighted in Figure E3, Figure E4, and Figure E5 because the affected time period only accounts for 2 months out of the year. The reduced exposures during this time period are an important perturbation that can be examined in the epidemiological analysis.

Improvements in predicted $PM_{2.5}$ concentrations for the year 2016 are obvious in Figure E10, given the overprediction bias in CTM calculations associated with underpredicted wind speeds during those simulations. The RFR method improves on the original constrained MLR approach, yielding improved performance statistics for $PM_{2.5}$ mass, relative to the original constrained MLR approach (Figure E6, Figure E7, and Figure E8).

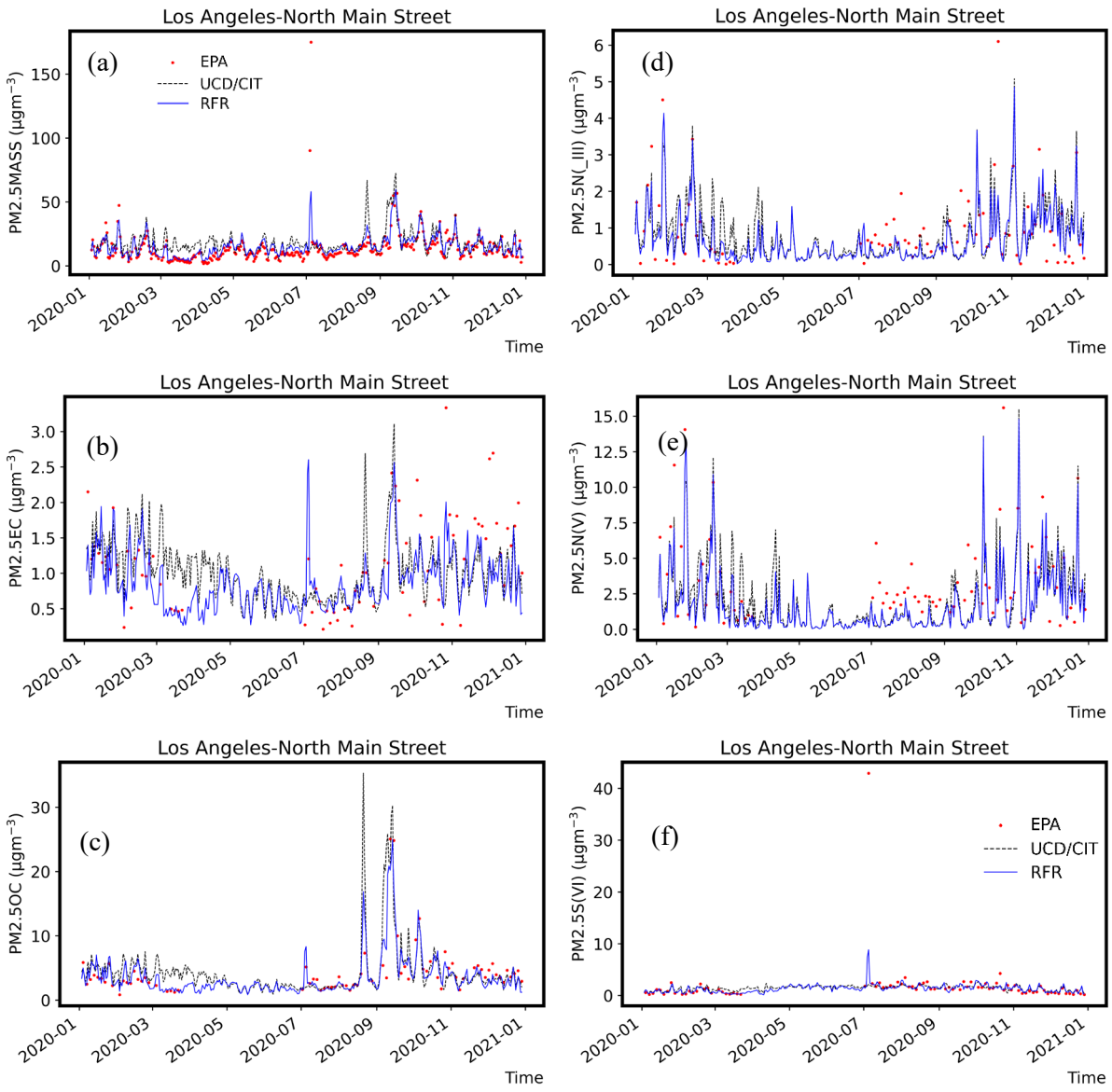

Figure E9. Time series of EPA observations, UCD/CIT model results, and RFR model results for Los Angeles, 2020. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ EC, (c) $PM_{2.5}$ OC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

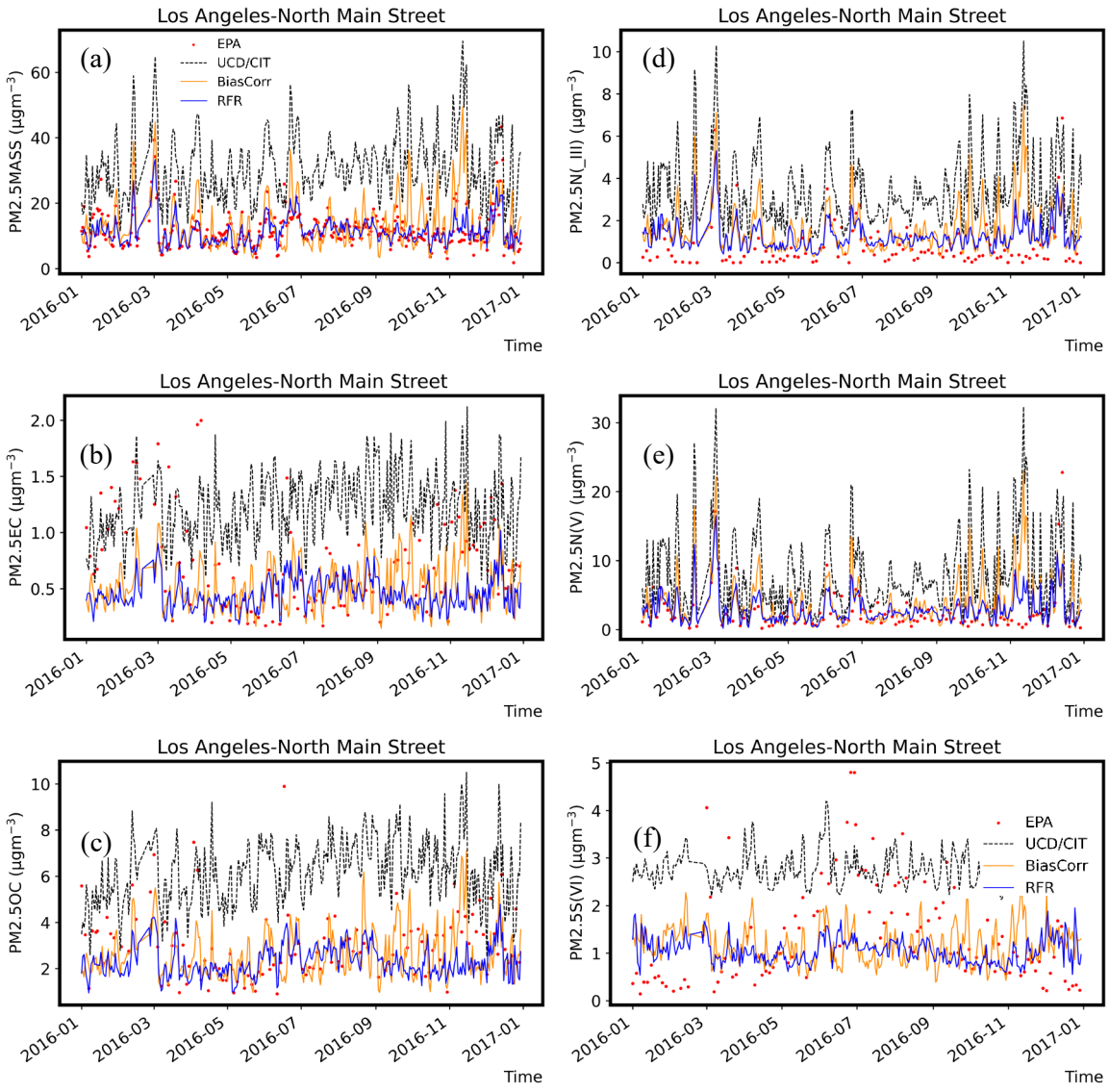

Figure E10. Time series of EPA observations, UC/CIT model results, MLR bias corrected UCD/CIT model results, and RFR corrected UCD/CIT model results for Los Angeles, 2016. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ EC, (c) $PM_{2.5}$ OC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

Bakersfield

Figure E11 and Figure E12 illustrate the time series of predicted and measured concentrations at Bakersfield, California, in 2020 and 2016, respectively. RFR corrections to the original CTM predictions are subtle in 2020 but more pronounced in 2016, due to the low bias in wind speeds during the earlier simulations. Corrected concentrations generally follow the seasonal trends of measured concentrations in Bakersfield during both simulated years.

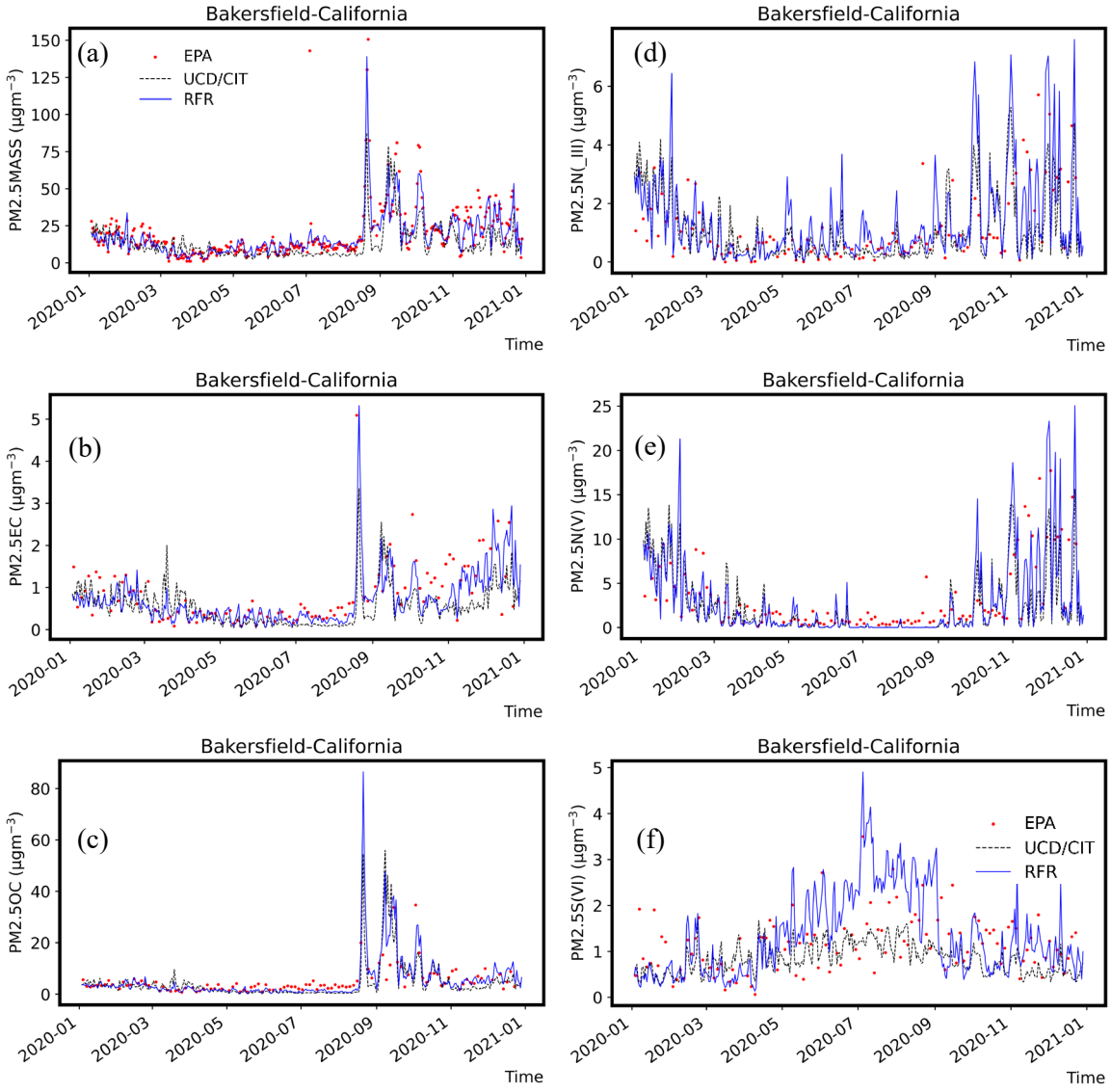

Figure E11. Time series of EPA observations and results of the UCD/CIT and RFR models for Bakersfield, 2020. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ EC, (c) $PM_{2.5}$ OC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

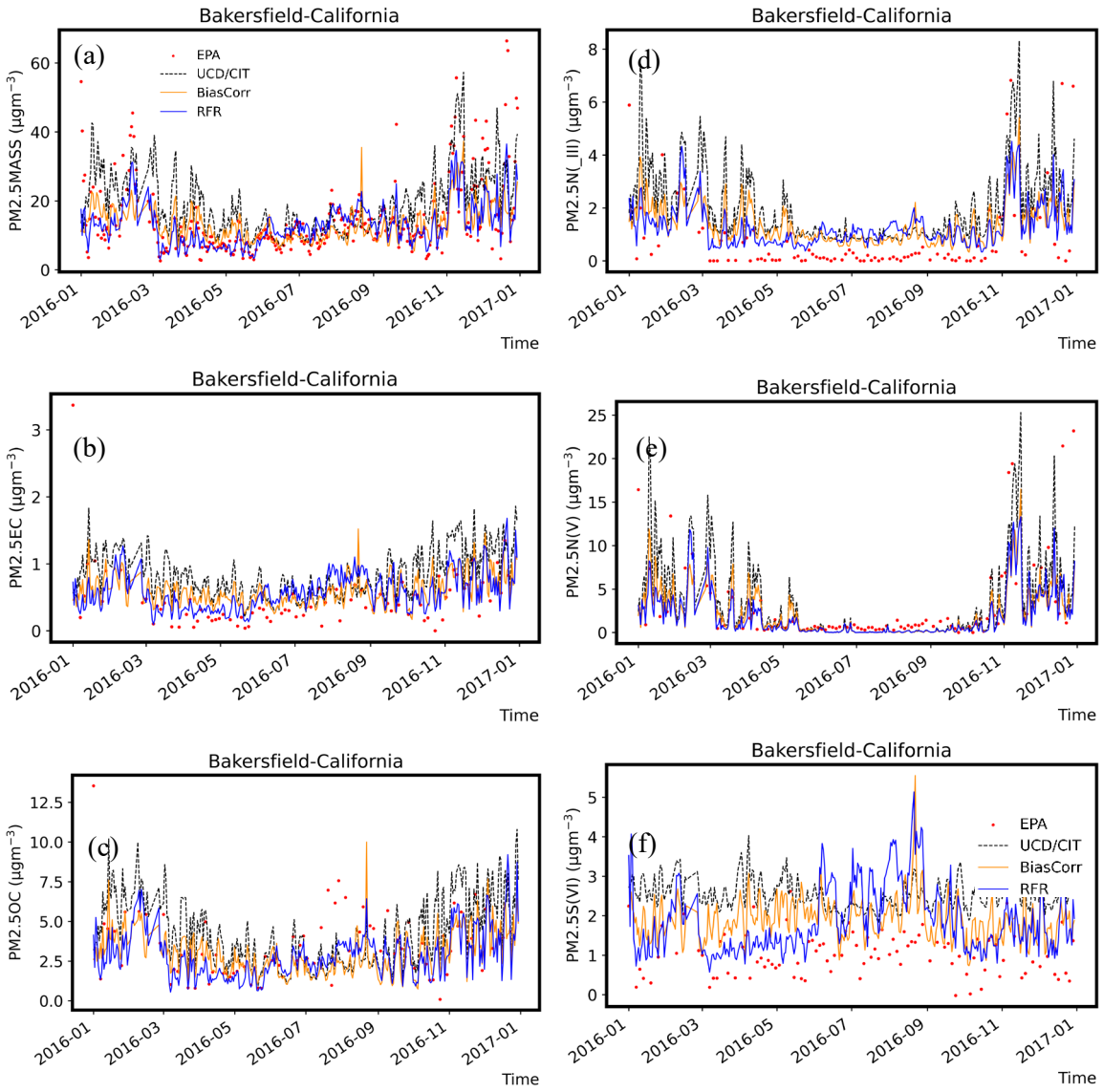

Figure E12. Time series of EPA observations, UC/CIT model results, MLR bias corrected UCD/CIT model results, and RFR corrected UCD/CIT model results for Bakersfield, 2016. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ EC, (c) $PM_{2.5}$ OC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

Rubidoux

Figure E13 and Figure E14 illustrate the time series of predicted and measured concentrations at Rubidoux, California, in 2020 and 2016, respectively. RFR corrections to CTM predictions are relatively modest in 2020 but more definite in 2016, given the need for greater BC in that year.

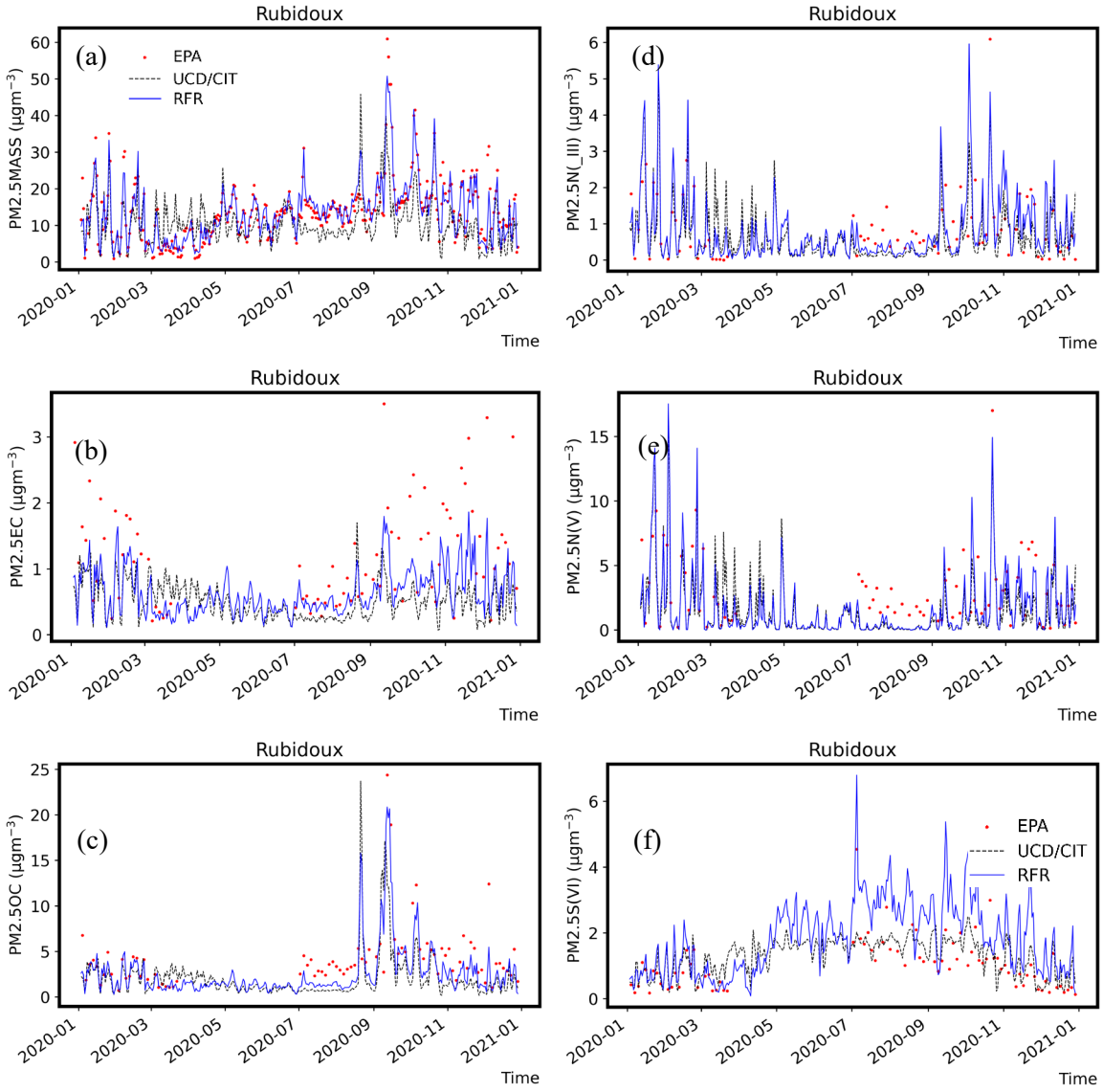

Figure E13. Time series of EPA observations, UCD/CIT model results, and RFR model results for Rubidoux, 2020. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ EC, (c) $PM_{2.5}$ OC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

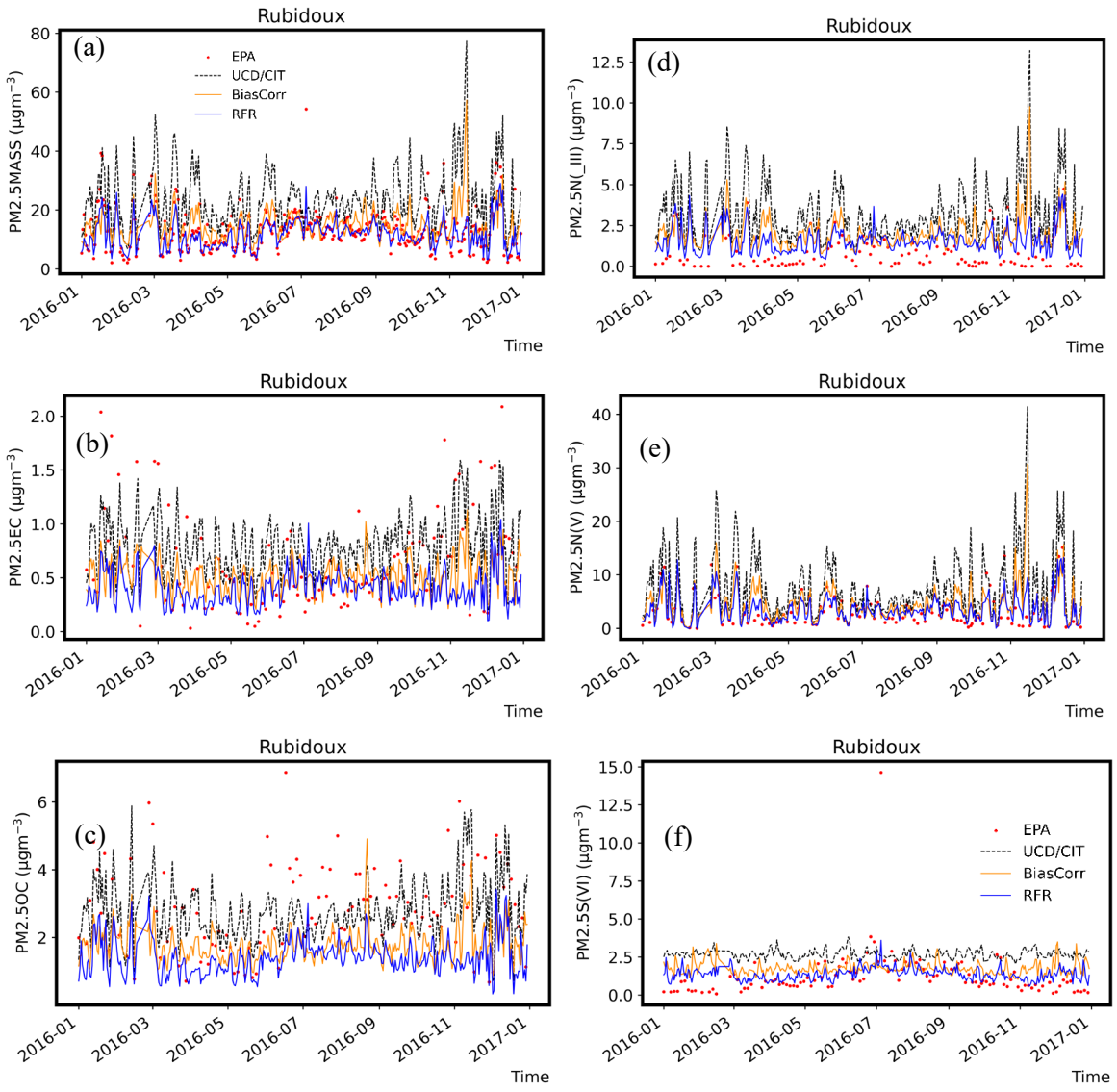

Figure E14. Time series of EPA observations, UC/CIT model results, MLR bias corrected UCD/CIT model results, and RFR corrected UCD/CIT model results for Rubidoux, 2016. (a) $PM_{2.5}$ mass, (b) $PM_{2.5}$ EC, (c) $PM_{2.5}$ OC, (d) $PM_{2.5}$ N(_III), (e) $PM_{2.5}$ N(V), (f) $PM_{2.5}$ S(VI).

Annual Average

Year 2020

Figure E15 through Figure E20 illustrate predicted annual average concentration fields for $PM_{2.5}$ mass and $PM_{2.5}$ species components in the year 2020. Each figure is organized in three panels that display the original CTM prediction (a), the adjusted CTM prediction using the RFR approach (b), and the difference between the original and adjusted concentrations (c). Spatial patterns are generally similar across all species, as the RFR method applied the same weighted correction factor to $PM_{2.5}$ mass and all $PM_{2.5}$ species, such that the particle composition remained thermodynamically balanced. The RFR method generally predicted decreased concentrations in regions along the coastline that had the highest concentrations in the original CTM predictions. The RFR method generally predicted increased concentrations at inland regions that had lower concentrations in the original CTM predictions. Some of the highest increases in predicted concentrations occur in the region surrounding Bakersfield in the San Joaquin Valley of California. The time series plot shown in Figure E11 indicates that the majority of this concentration increase occurs during the fall and winter months.

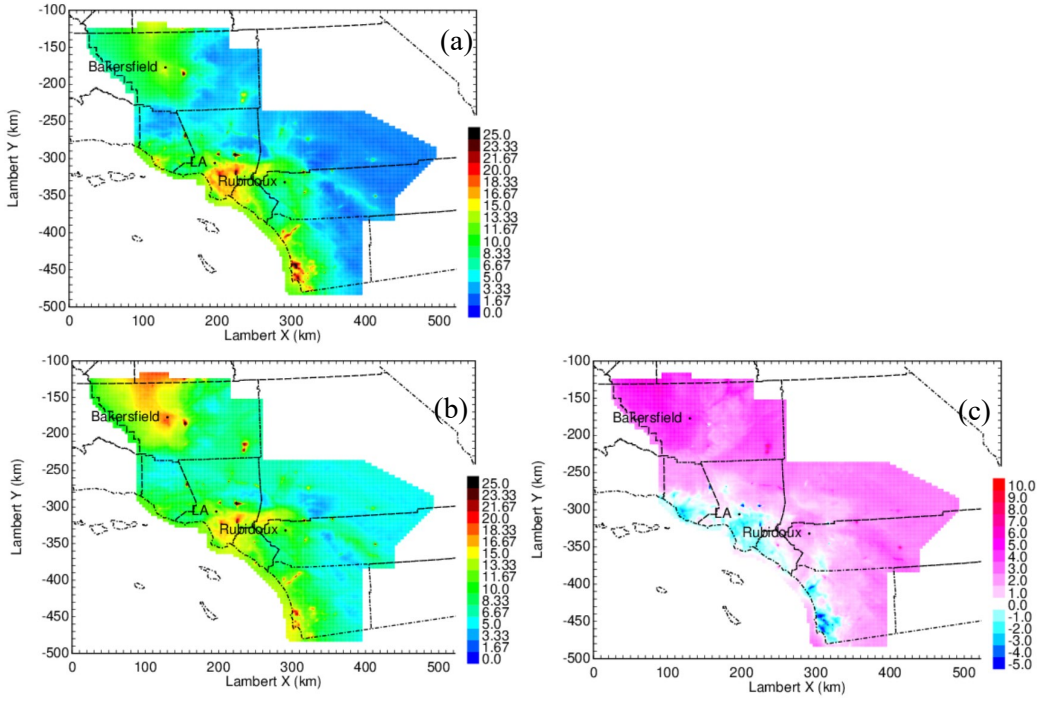

Figure E15. 2020 Annual average $PM_{2.5}$ mass concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT.

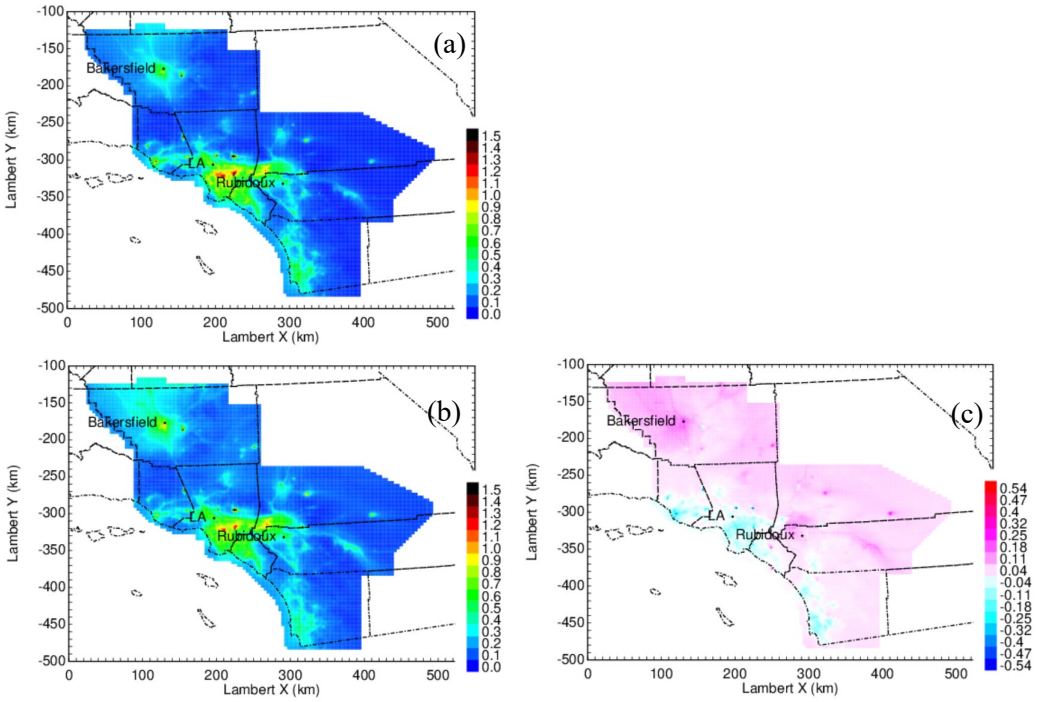

Figure E16. 2020 Annual average $PM_{2.5}$ EC concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT.

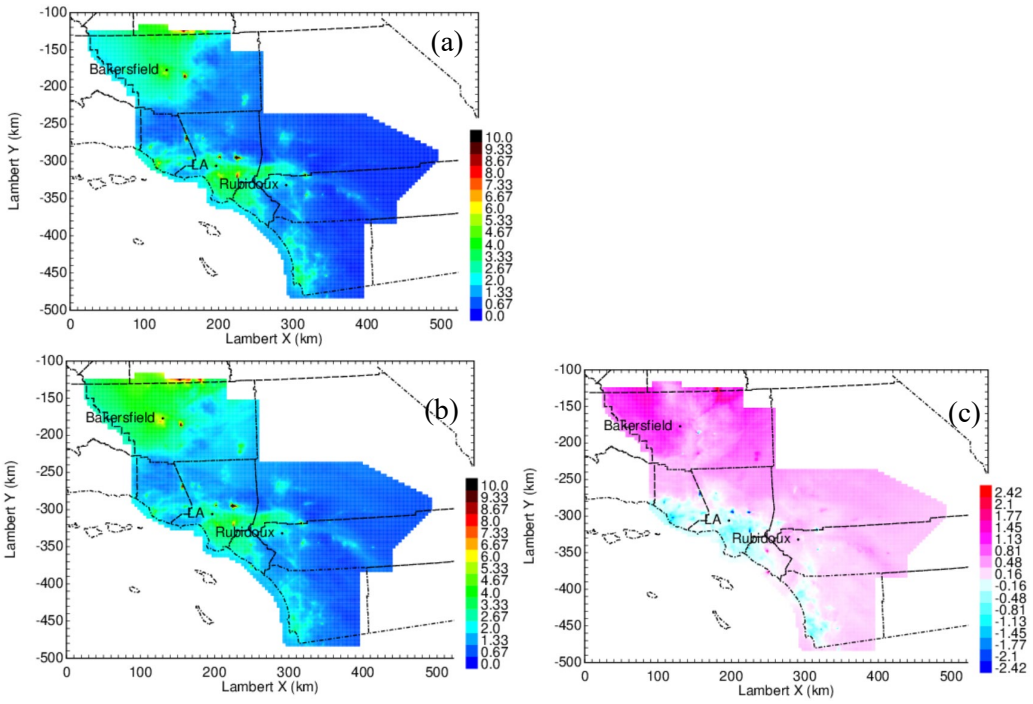

Figure E17. 2020 Annual average $PM_{2.5}$ OC concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT.

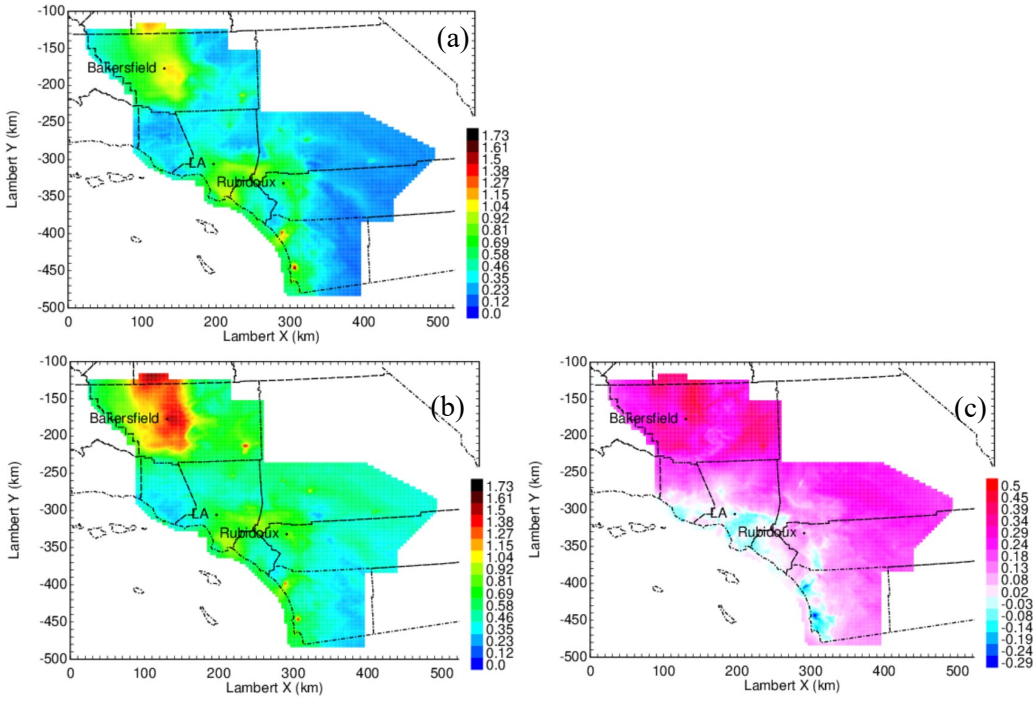

Figure E18. 2020 Annual average $PM_{2.5}$ N(_III) concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT.

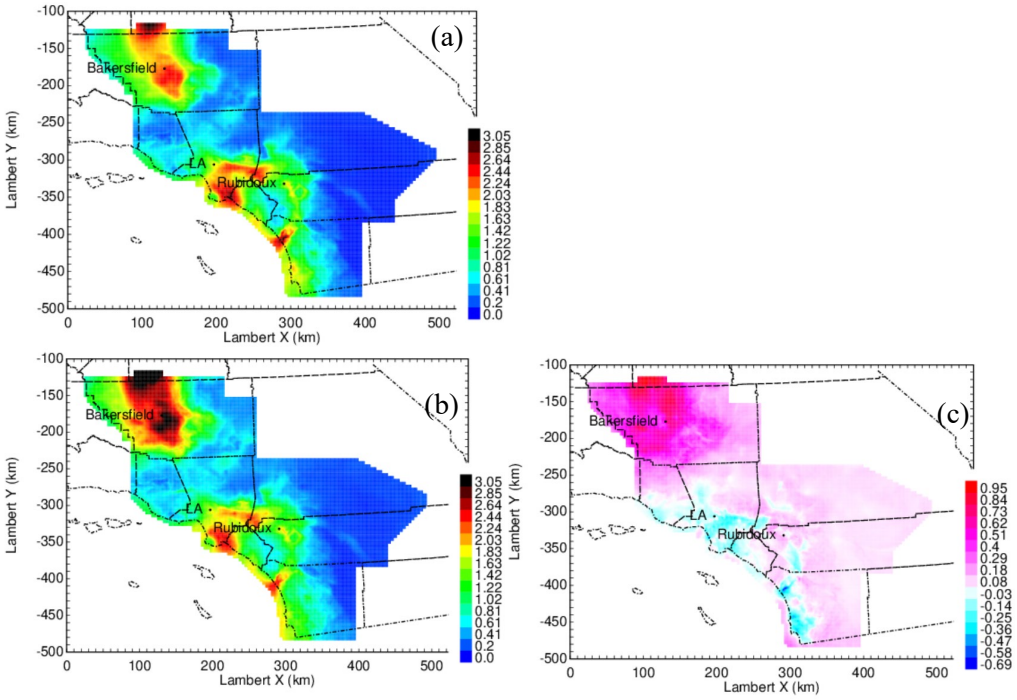

Figure E19. 2020 Annual average $PM_{2.5}$ N(V) concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT.

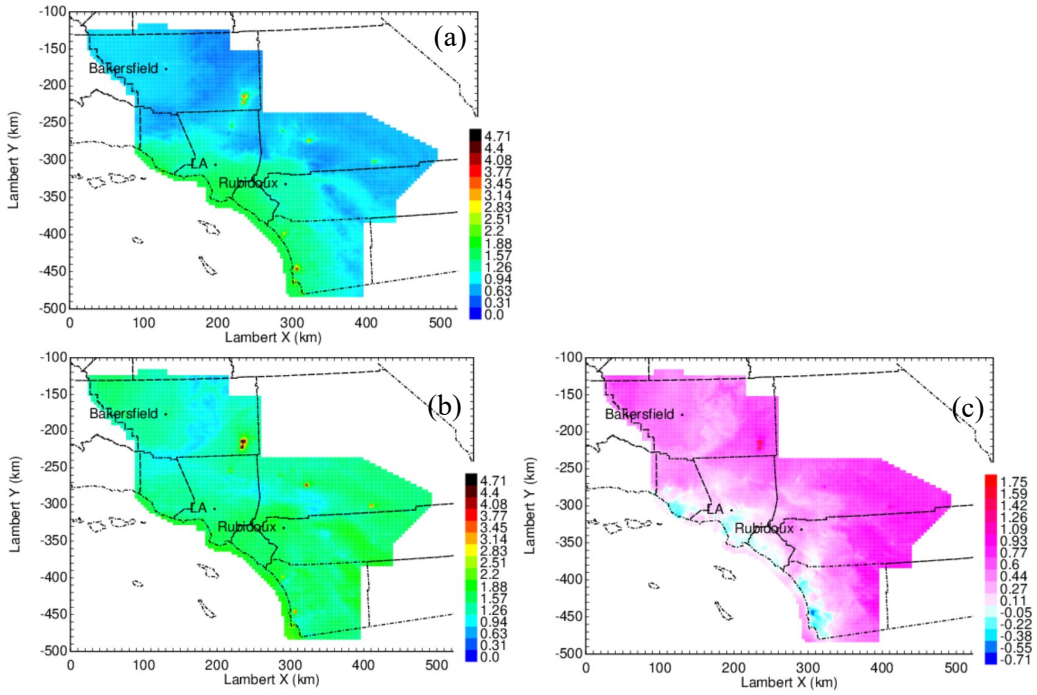

Figure E20. 2020 Annual average $PM_{2.5}$ S(VI) concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT.

Year 2016

Figure E21 through Figure E26 illustrate predicted annual average concentration fields for $PM_{2.5}$ mass and $PM_{2.5}$ species components in the year 2016. Each figure is organized in five panels that display the original CTM prediction (a), the adjusted CTM prediction using the RFR approach (b), the difference between the original and RFR adjusted concentrations (c), the adjusted CTM prediction using the original constrained MLR approach (d), and the difference between the original and MLR adjusted concentrations (e). As noted in the previous discussion, a single weighted-average correction factor was applied for $PM_{2.5}$ mass and $PM_{2.5}$ species components at each location, and so the spatial patterns for all $PM_{2.5}$ plots are similar. The corrections to the original CTM calculations are larger in 2016 than in 2020, given a low bias in predicted wind speeds in 2016. Consistent with the trends shown in the time series plots (Figure E10, Figure E12, and Figure E14Concentrations in polluted regions along the California coast are adjusted downward by almost 50% in both the RFR method and the original constrained MLR method. Adjustments to concentrations in inland regions are more modest in both methods. The RFR method predicts very little concentration increase at inland regions, whereas the constrained MLR method predicts $PM_{2.5}$ mass concentration increases of 1-2 μg m^{-3} at these locations, with proportional changes for $PM_{2.5}$ subcomponents of mass.

Spatial patterns in the exposure fields were a key driver of the impacts of chronic air pollution on COVID-19 outcomes in our analysis to date. The consistency in the RFR and MLR corrected concentration fields is illustrated in Figure E21 through Figure E26 builds confidence in the robustness of the epidemiological results.

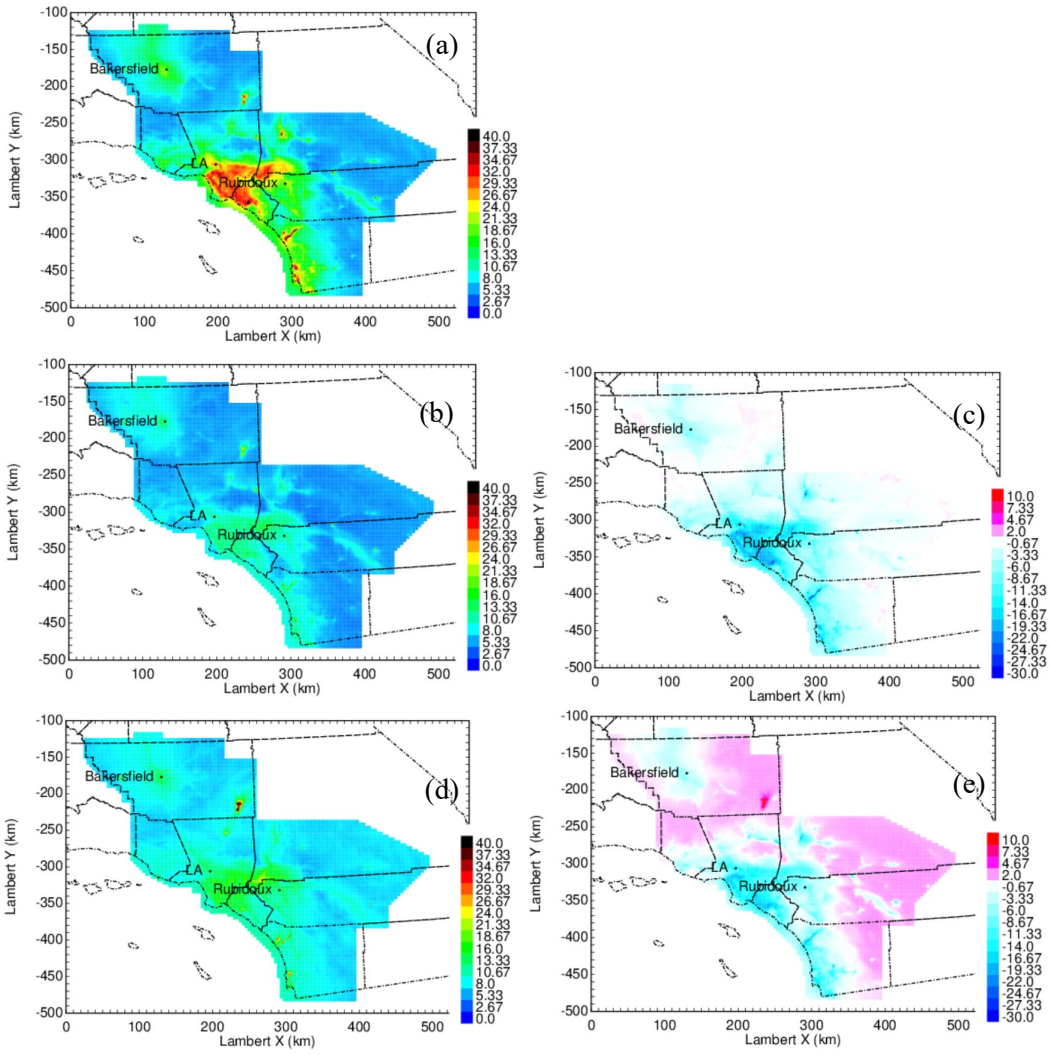

Figure E21. 2016 Annual average $PM_{2.5}$ mass concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT, d) BC, e) BC − UCD/CIT.

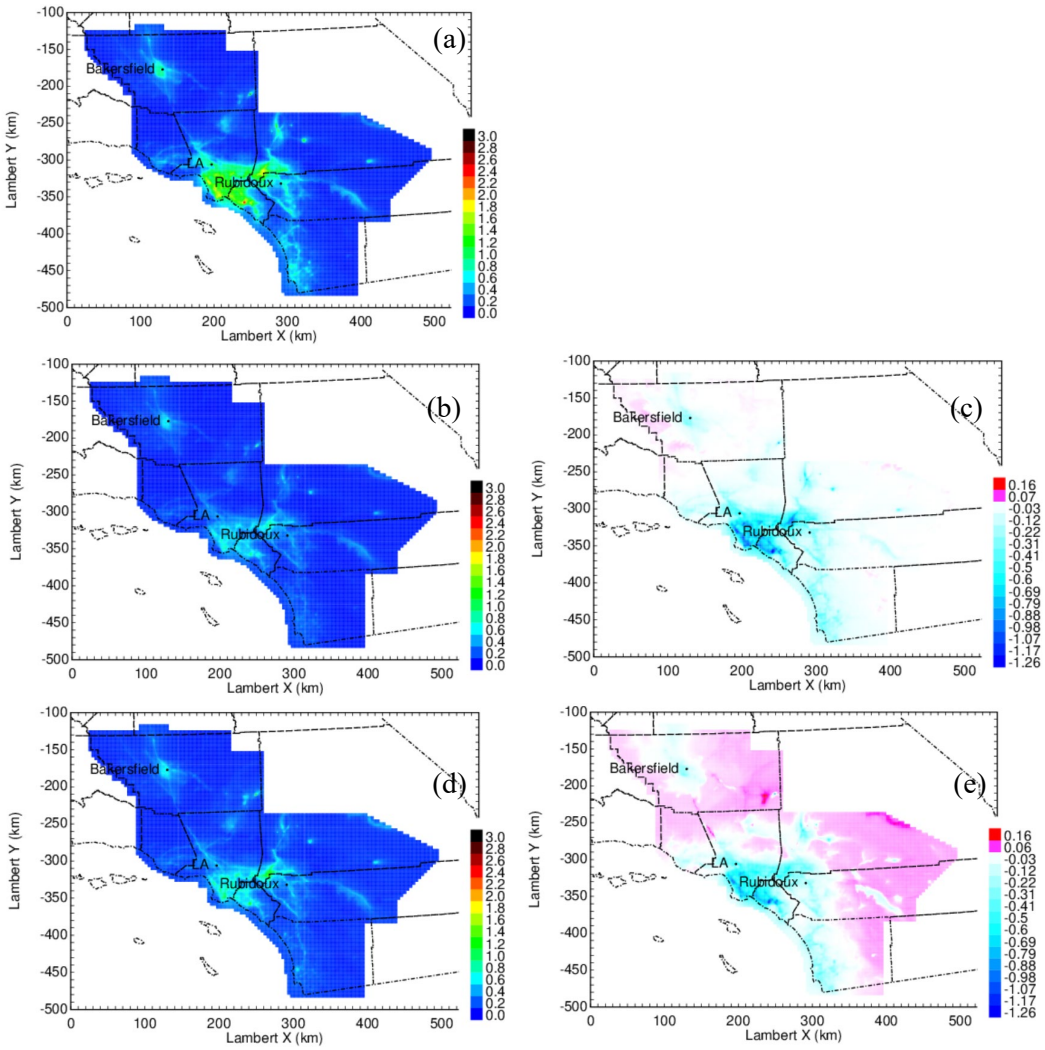

Figure E22. 2016 Annual average $PM_{2.5}$ EC concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT, d) BC, e) BC − UCD/CIT.

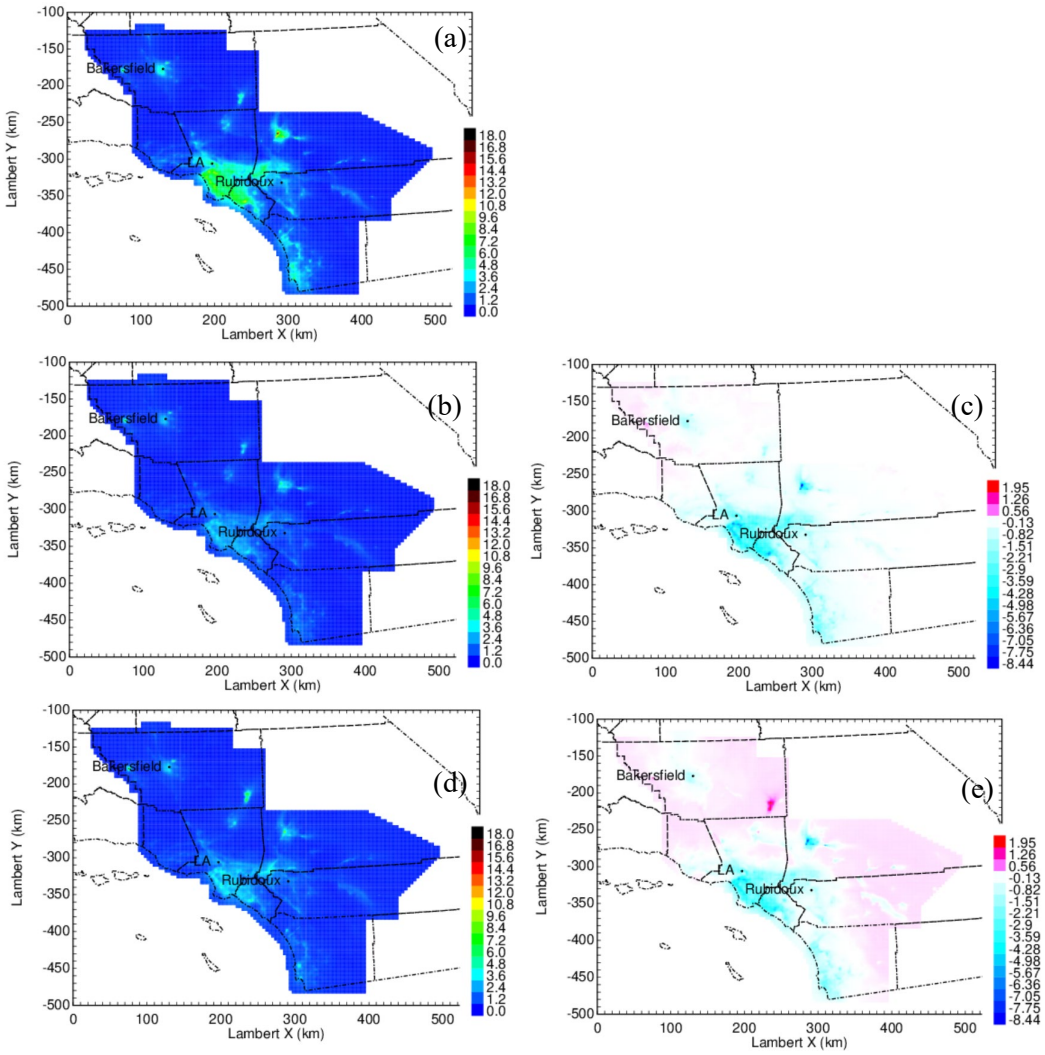

Figure E23. 2016 Annual average $PM_{2.5}$ OC concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT, d) BC, e) BC − UCD/CIT.

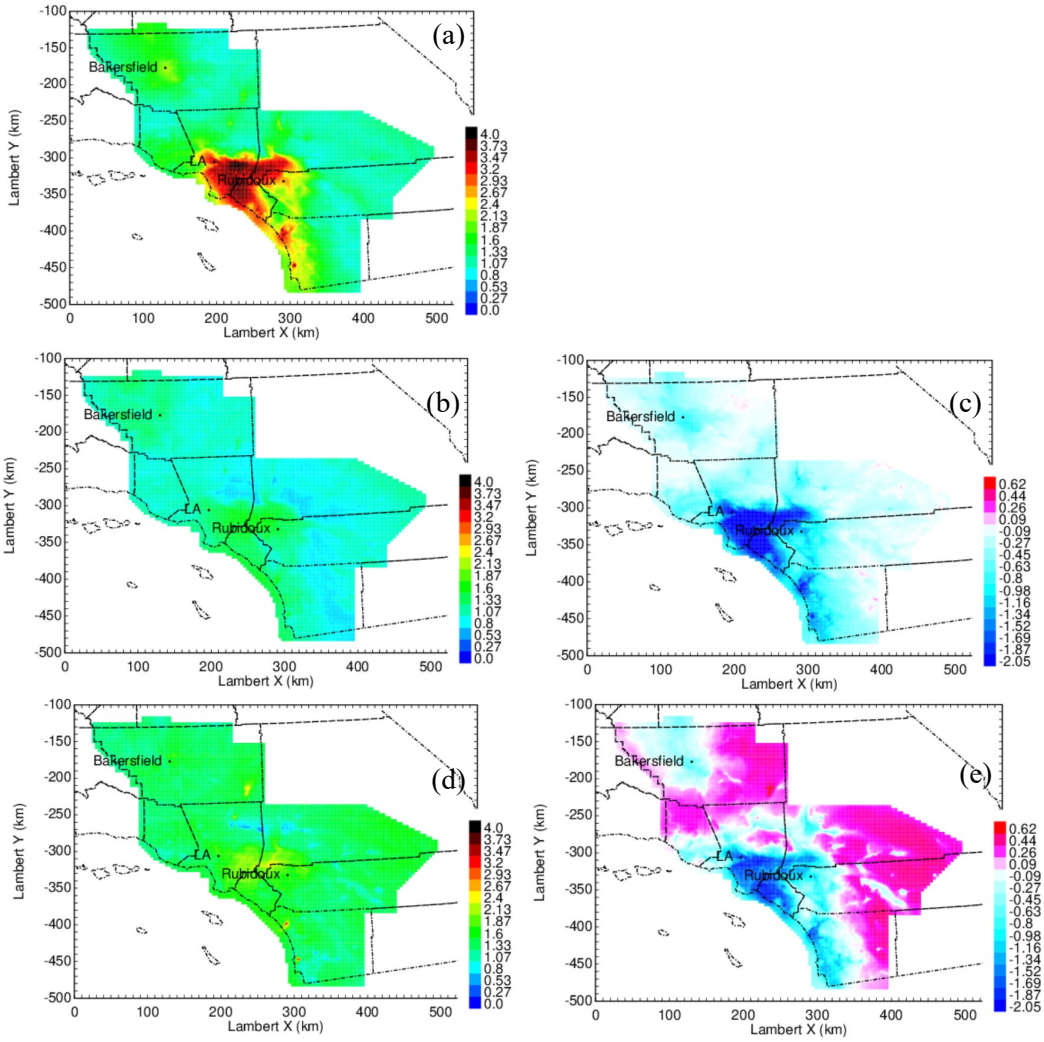

Figure E24. 2016 Annual average $PM_{2.5}$ N(_III) concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT, d) BC, e) BC − UCD/CIT.

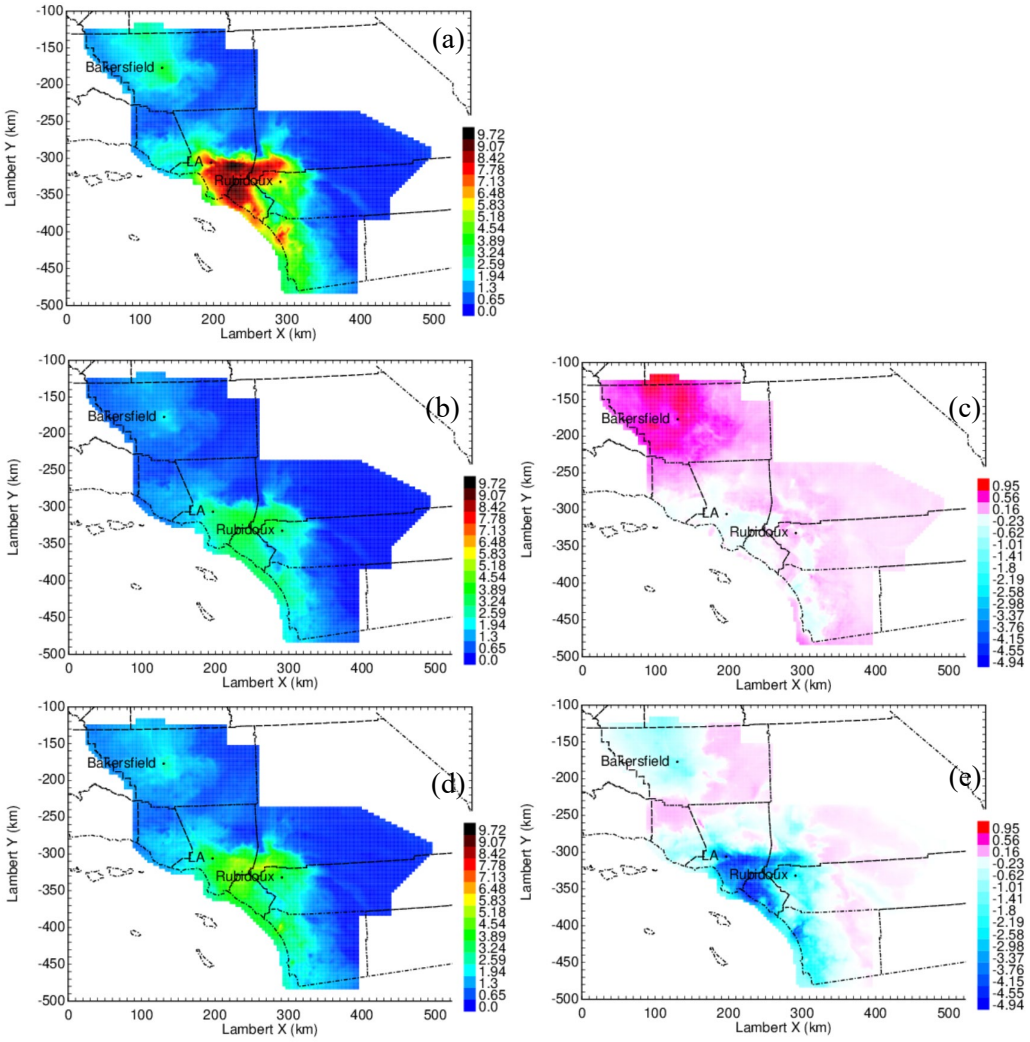

Figure E25. 2016 Annual average $PM_{2.5}$ N(V) concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT, d) BC, e) BC − UCD/CIT.

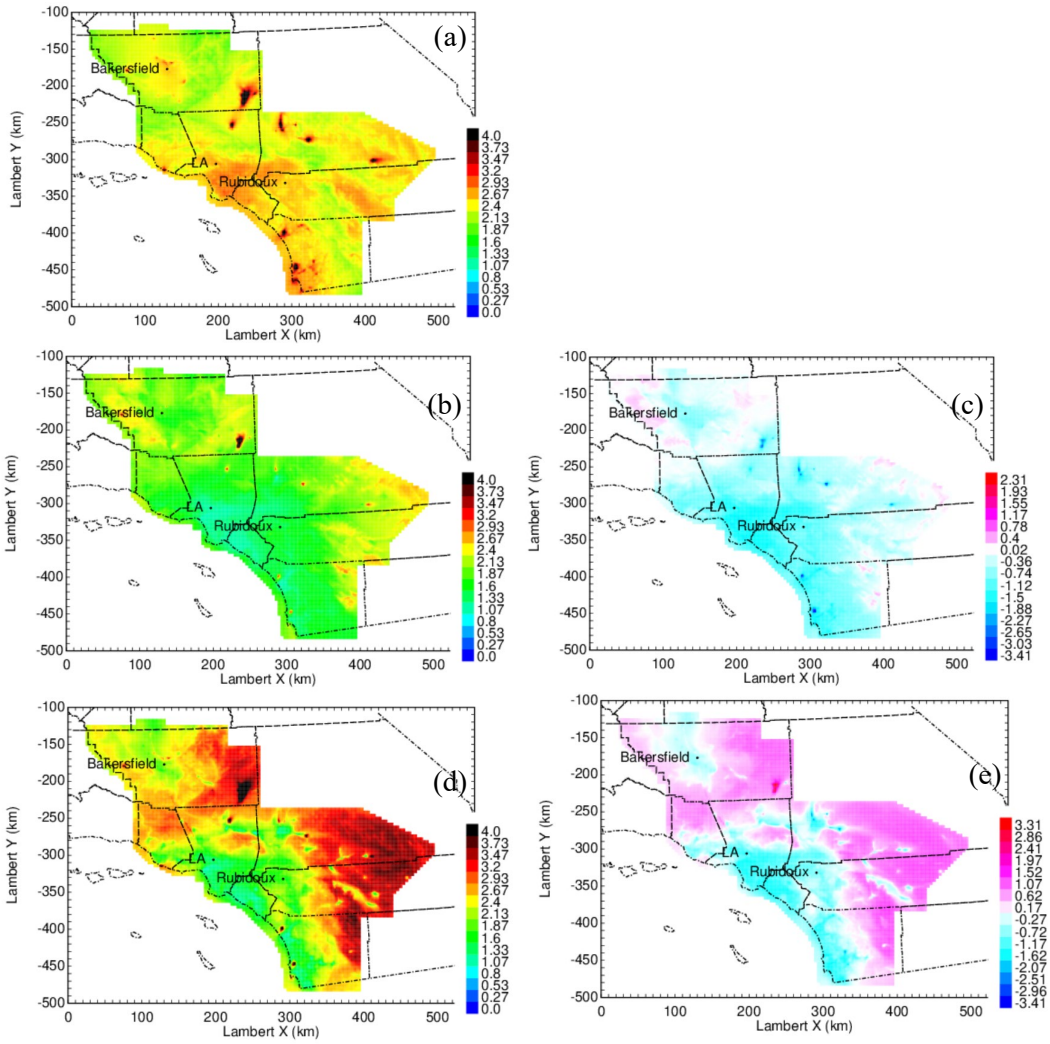

Figure E26. 2016 Annual average $PM_{2.5}$ S(VI) concentrations from model predictions. a) UCD/CIT, b) RFR, c) RFR − UCD/CIT, d) BC, e) BC − UCD/CIT.

Summary

The RFR method developed in the current study improves the accuracy of the air pollution exposure fields predicted by CTM calculations for Southern California for the years 2020 and 2016. The RFR method uses support variables, including ground-based measurements, satellite measurements, meteorological predictions, and source-oriented tracer concentrations, to reduce the bias in CTM predictions. The combination of the RFR method and the CTM predictions retains the rich data describing species and sources in the CTM fields while compensating for some of the random and systematic errors in the CTM input data that produce errors in the raw CTM output fields.

The RFR method improved CTM performance in cases in which the bias in the original CTM fields was small (2020) or the bias in the original CTM fields was large (2016). The RFR method improved predictions during periods when "shelter-in-place" orders reduced ambient concentrations and during periods when wildfires generated high-concentration events in the year 2020. The use of ground-level measurements made by low-cost sensors (i.e., PurpleAir Network) improved the ability of the RFR method to accurately adjust CTM concentrations. These measurements are generally available for years beginning in 2017.

Predicted CTM concentration fields for the year 2016 that were adjusted using the RFR method are similar to concentration fields adjusted using an earlier constrained MLR approach. These findings suggest that epidemiological results generated using previous concentration fields will be consistent with the updated RFR fields.

All indications suggest that adjustment of predicted CTM concentration fields using RFR methods will improve the accuracy of the exposure fields used in epidemiological studies.

Table E3. PASC Disease Groupings and ICD Names, ICD-10 Codes, and Example Diagnoses

PASC Diagnosis Group	Disease Group	ICD Name	ICD-10 Code	Examples of Diagnoses Listed Under the Parent ICD-10 Code
Cardiac	Arrhythmias	Abnormalities of heartbeat	R00	Tachycardia, bradycardia
	Arrhythmias	Paroxysmal tachycardia	I47	Supraventricular tachycardia, ventricular tachycardia, paroxysmal tachycardia unspecified
	Arrhythmias	Atrial fibrillation and flutter	I48	
	Arrhythmias	Other cardiac arrhythmias	I49	Ventricular fibrillation, other and unspecified premature depolarization, cardiac arrhythmia unspecified
	Myocarditis/pericarditis	Acute pericarditis	I30	
	Myocarditis/pericarditis	Acute myocarditis	I40	
	Myocarditis/pericarditis	Myocarditis, unspecified	I51.4	
	Myocarditis/pericarditis	Viral carditis, unspecified	B33	Viral carditis, viral endocarditis, viral myocarditis, viral pericarditis, viral cardiomyopathy
	Stress cardiomyopathy	Takotsubo syndrome	I51.81	
Constitutional and Lymphatic	Constitutional	Fever of other and unknown origin	R50	Fever unspecified
	Constitutional	Generalized hyperhidrosis	R61	Night sweats
	Constitutional	Malaise and fatigue	R53	Weakness, malaise, fatigue
	Constitutional	Post-viral fatigue syndrome	G93.3	Specific ICD code used here because G93 also includes cerebral edema, brain death, among others
	Lymphadenopathy	Enlarged lymph nodes	R59	
Cardiometabolic Diseases	Arrhythmias	Abnormalities of heartbeat	R00	Tachycardia, bradycardia
	Arrhythmias	Paroxysmal tachycardia	I47	Supraventricular tachycardia, ventricular tachycardia, paroxysmal tachycardia unspecified
	Arrhythmias	Atrial fibrillation and flutter	I48	
	Arrhythmias	Other cardiac arrhythmias	I49	Ventricular fibrillation, other and unspecified premature depolarization, cardiac arrhythmia unspecified
	Diabetes	Diabetes mellitus, type 1	E10	
	Diabetes	Diabetes mellitus, type 2	E11	
	Renal disease	Chronic kidney disease	N18	
	Renal disease	Unspecified kidney failure	N19	
Dermatological	Skin	Unspecified viral infection with skin and mucous membrane lesions	B09	Viral exanthema NOS
	Skin	Disturbances of skin sensation	R20	Hypoesthesia of skin, paresthesia of skin
	Skin	Rash and other nonspecific skin eruptions	R21	Rash NOS
	Skin	Other skin changes	R23	Cyanosis, flushing, pallor

	Diabetes	Diabetes mellitus, type 1	E10	
	Diabetes	Diabetes mellitus, type 2	E11	
	Thyroid	Other hypothyroidism	E03	Post-infectious hypothyroidism, hypothyroidism unspecified
Endocrine	Thyroid	Thyroiditis	E06	Autoimmune thyroiditis, thyroiditis unspecified
	ENT	Conductive and sensorineural hearing loss	H90	
	ENT	Other and unspecified hearing loss	H91	
	ENT	Otalgia and effusion of the ear	H92	
	ENT	Other disorders of the ear, not elsewhere classified	H93	Hyperacusis, tinnitus
	ENT	Chronic rhinitis, nasopharyngitis, and pharyngitis	J31	
	ENT	Disturbances of smell and taste	R43	Anosmia, paralgesia, and other disturbances of smell and taste
Ear, Nose, and Throat	ENT	Aphagia and dysphagia	R13	
	Abdominal pain	Abdominal and pelvic pain	R10	
	Change in bowel habits	Irritable bowel syndrome	K58	
	Change in bowel habits	Other functional intestinal disorders	K59	Constipation
	Change in bowel habits	Viral and other specified intestinal infections	A08	Other viral enteritis, viral intestinal infection unspecified
	Change in bowel habits	Infectious gastroenteritis and colitis, unspecified	A09	
	Change in bowel habits	Change in bowel habit	R19.4	
	Change in bowel habits	Diarrhea, unspecified	R19.7	
Gastrointestinal	Nausea/vomiting	Nausea and vomiting	R11	
	Cytopenias	Coagulation defects, purpura, and other hemorrhagic conditions	D69	Secondary thrombocytopenia, thrombocytopenia unspecified, ITP
Hematological	Cytopenias	Decreased white blood cell count	D72.81	Lymphocytopenia, other decreased white blood cell count
	Myalgia/arthralgia	Post-infective and reactive arthropathies	M02	
	Myalgia/arthralgia	Other joint disorder, not elsewhere classified	M25	Pain in [insert joint], stiffness of [insert joint]
Myalgia	Myalgia/arthralgia	Other and unspecified soft tissue disorders, not elsewhere classified	M79	Myalgias, pain in [insert specific limb]
	Ataxia/trouble walking	Abnormalities of gait and mobility	R26	Ataxic gait, unsteadiness on feet, difficulty in walking, not elsewhere classified
	Ataxia/trouble walking	Other lack of coordination	R27	Ataxia unspecified, repeated falls
	Ataxia/trouble walking	Extrapyramidal and movement disorders in diseases classified elsewhere	G26	
Neurological	Autonomic dysfunction	Orthostatic hypotension	I95.1	

Autonomic dysfunction	Other cardiac arrhythmias	I49	Other specified cardiac arrhythmias (POTS often diagnosed under this code)
Autonomic dysfunction	Disorders of the autonomic nervous system	G90	Autonomic dysreflexia, disorders of the autonomic nervous system, unspecified
Autonomic dysfunction	Syncope and collapse	R55	
Delirium or encephalopathy	Delirium due to a known physiological condition	F05	Delirium not superimposed on dementia, delirium superimposed on dementia, delirium unspecified
Delirium or encephalopathy	Somnolence	R40.0	
Delirium or encephalopathy	Other symptoms and signs involving cognitive functions and awareness	R41	Altered mental status unspecified, disorientation unspecified, other symptoms and signs involving cognitive functions and awareness
Delirium or encephalopathy	Other symptoms and signs involving general sensations and perceptions	R44	Auditory hallucinations, visual hallucinations
Dementia	Vascular dementia	F01	Vascular dementia of acute onset, multi-infarct dementia, vascular dementia unspecified
Dementia	Dementia in other diseases classified elsewhere	F02	
Dementia	Unspecified dementia	F03	Unspecified dementia with or without behavioral disturbance
Dementia	Other degenerative diseases of the nervous system, not elsewhere classified	G31	Mild cognitive impairment, so stated, dementia with Lewy bodies, and frontotemporal dementia
Encephalitis	Other viral encephalitis, not elsewhere classified	A85	Other specified viral encephalitis
Encephalitis	Unspecified viral encephalitis	A86	Viral encephalomyelitis NOS
Encephalitis	Encephalitis, myelitis, and encephalomyelitis	G04	ADEM, other encephalitis
Encephalitis	Encephalitis, myelitis, and encephalomyelitis in other diseases	G05	
Encephalitis	Other symptoms and signs involving the nervous and musculoskeletal systems	R29	Meningismus, abnormal reflexes
Headache	Migraine	G43	
Headache	Other headache syndromes	G44	Cluster headache and other trigeminal autonomic cephalalgias, drug-induced headache, and complicated headache syndromes
Headache	Headache	R51	Headache with orthostatic component, headache unspecified
Myoneural disorders	Other and unspecified myopathies	G72	Critical illness myopathy, myopathy unspecified
Myoneural disorders	Myositis	M60	
Ophthalmologic conditions following stroke	Visual disturbances	H53	
Ophthalmologic conditions following stroke	Blindness and low vision	H54	

Parkinsonism and other Extrapyramidal syndromes	Secondary parkinsonism	G21	
Parkinsonism and other Extrapyramidal syndromes	Dystonia	G24	
Parkinsonism and other Extrapyramidal syndromes	Other extrapyramidal and movement disorders	G25	Myoclonus, other forms of tremor, and other chorea
Peripheral nerve disorders	Disorders of the trigeminal nerve	G50	
Peripheral nerve disorders	Facial nerve disorders	G51	
Peripheral nerve disorders	Disorders of other cranial nerves	G52	
Peripheral nerve disorders	Cranial nerve disorders in diseases classified elsewhere	G53	
Peripheral nerve disorders	Nerve root and plexus disorders	G54	
Peripheral nerve disorders	Nerve root and plexus compressions in diseases classified elsewhere	G55	
Peripheral nerve disorders	Mononeuropathies of the upper limb	G56	
Peripheral nerve disorders	Mononeuropathies of the lower limb	G57	
Peripheral nerve disorders	Other mononeuropathies	G58	Mononeuritis multiplex, mononeuropathy unspecified
Peripheral nerve disorders	Mononeuropathy in diseases classified elsewhere	G59	
Peripheral nerve disorders	Inflammatory polyneuropathy	G61	Guillain-Barré syndrome, chronic inflammatory demyelinating polyradiculoneuropathy
Peripheral nerve disorders	Other and unspecified polyneuropathies	G62	Critical illness polyneuropathy, Polyneuropathy unspecified
Peripheral nerve disorders	Other disorders of the peripheral nervous system	G64	Disorder of the peripheral nervous system NOS
Peripheral nerve disorders	Sequelae of inflammatory and toxic polyneuropathies	G65	Sequelae of Guillain-Barré syndrome, sequelae of other inflammatory polyneuropathy
Seizures	Epilepsy and recurrent seizures	G40	
Seizures	Status epilepticus	G41	
Stroke	Stroke, not specified as hemorrhage or infarction	I64	
Stroke	Sequelae of cerebrovascular disease	I69	Sequelae of nontraumatic intracerebral hemorrhage, sequelae of cerebral infarction
Stroke	Transient cerebral ischemic attacks and related syndromes	G45	
Stroke	Vascular syndromes of the brain in cerebrovascular diseases	G46	Cerebellar stroke syndrome, brain stem stroke syndrome
Stroke (intracranial hemorrhage)	Subarachnoid hemorrhage	I60	
Stroke (intracranial hemorrhage)	Intracerebral hemorrhage	I61	
Stroke (intracranial hemorrhage)	Other and unspecified nontraumatic intracranial hemorrhage	I62	

	Stroke (ischemic)	Cerebral infarction	I63	
	Vertigo	Other viral infections of the central nervous system, not elsewhere classified	A88	Epidemic vertigo
	Vertigo	Disorders of vestibular function	H81	Other peripheral vertigo
	Vertigo	Dizziness and giddiness (including lightheadedness and vertigo)	R42	
Other	Hair loss	Alopecia areata	L63	
	Hair loss	Other nonscarring hair loss	L65	
	Infectious	Sequelae of other and unspecified infectious and parasitic diseases	B94	Viral hepatitis, viral encephalitis
	Weight loss	Symptoms and signs concerning food and fluid intake	R63	Abnormal weight loss
	Weight loss	Cachexia	R64	
Psychological	Anxiety	Phobic anxiety disorder	F40	
	Anxiety	Other anxiety disorders	F41	Generalized anxiety disorder, anxiety disorder unspecified
	Anxiety	Obsessive-compulsive disorder	F42	
	Anxiety	Reaction to severe stress and adjustment disorders	F43	PTSD, adjustment disorder
	Anxiety	Dissociative and conversion disorders	F44	
	Anxiety	Somatoform disorders	F45	
	Anxiety	Other neurotic disorders	F48	Neurasthenia
	Anxiety	Symptoms and signs involving emotional state	R45	Restlessness and agitation, anhedonia
	Mood disorders	Manic episode	F30	
	Mood disorders	Bipolar affective disorder	F31	
	Mood disorders	Major depressive disorder, single episode	F32	
	Mood disorders	Major depressive disorder, recurrent	F33	
	Mood disorders	Persistent mood [affective] disorders	F34	Dysthymic disorder
	Mood disorders	Other mood [affective] disorders	F38	Recurrent brief depressive episodes
	Mood disorders	Unspecified mood [affective] disorder	F39	
	Psychosis	Schizophrenia	F20	
	Psychosis	Schizotypal disorder	F21	
	Psychosis	Persistent delusional disorders	F22	
	Psychosis	Acute and transient psychotic disorders	F23	
	Psychosis	Shared psychotic disorder	F24	
	Psychosis	Schizoaffective disorders	F25	

	Psychosis	Other psychotic disorder not due to a substance or known physiological condition	F28	Other specified schizophrenia spectrum and other psychotic disorder
	Psychosis	Unspecified psychosis not due to a substance or known physiological condition	F29	Psychosis NOS
	Sleep disorders	Sleep disorders such as insomnia and hypersomnia	G47	Insomnia, hypersomnia, sleep apnea
	Sleep disorders	Sleep disorder not from a physiologic condition	F51	Primary insomnia, adjustment insomnia
Pulmonary	Bronchitis	Acute bronchitis	J20	
	Bronchitis	Bronchitis, not specified as acute or chronic	J40	
	Bronchitis	Simple and mucopurulent chronic bronchitis	J41	
	Bronchitis	Unspecified chronic bronchitis	J42	
	Chest/throat	Pain in the throat and chest	R07	Chest pain, pleurodynia, intercostal pain, pain in throat
	Cough	Cough	R05	
	Dyspnea	Abnormalities of breathing	R06	Dyspnea, shortness of breath, tachypnea
	Hypoxemia	Other symptoms and signs involving the circulatory and respiratory system	R09	Hypoxemia
	ILD	Other interstitial pulmonary diseases	J84	Pulmonary fibrosis unspecified, cryptogenic organizing pneumonia, interstitial pulmonary disease unspecified
	PE/DVT	Pulmonary embolism	I26	
	PE/DVT	Other venous embolism and thrombosis	I82	Acute embolism and thrombosis of deep veins of the lower extremity, of the femoral vein, etc.
	Pulmonary edema	Pulmonary edema	J81	
Renal	Renal disease	Chronic kidney disease	N18	
	Renal disease	Unspecified kidney failure	N19	

ADEM = acute disseminated encephalomyelitis; ENT = ear, nose, and throat; ITP = immune thrombocytopenic purpura; NOS = not otherwise specified; POTS = postural orthostatic tachycardia syndrome; PTSD = posttraumatic stress disorder.

Multipollutant Associations with PASC Disease Categories at 3 Months and 12 Months

Table E4. Significant Multipollutant Associations with PASC Disease Categories at 3 Months and 12 Months after Hospital Discharge

PASC Group	Main Pollutant	Co-Pollutants	Outcome Time (months)	Estimate (95% CI)
Cardiac	$PM_{0.1}$	O_3	3	1.123* (1.012, 1.246)
Cardiac	$PM_{0.1}$	$O_3 + NO_2$	3	1.122* (1.004, 1.255)
Cardiac	$PM_{0.1}$	$O_3 + NO_2$ (LUR)	3	1.118* (1.006, 1.243)
Cardiac	$PM_{0.1}$	NO_2 (LUR)	3	1.113* (1.004, 1.234)
Cardiometabolic/diabetes	$PM_{0.1}$	NO_2	3	1.120* (1.026, 1.223)
Cardiometabolic/diabetes	$PM_{0.1}$	NO_2 (LUR)	3	1.128** (1.035, 1.228)
Cardiometabolic/diabetes	$PM_{0.1}$	O_3	3	1.135** (1.041, 1.238)
Cardiometabolic/diabetes	$PM_{0.1}$	$O_3 + NO_2$	3	1.123* (1.022, 1.233)
Cardiometabolic/diabetes	$PM_{0.1}$	$O_3 + NO_2$ (LUR)	3	1.125** (1.029, 1.229)
Pulmonary	NO_2	$PM_{2.5}$	3	1.115* (1.009, 1.232)
Pulmonary	NO_2	$O_3 + PM_{2.5}$	3	1.146** (1.034, 1.269)
Pulmonary	NO_2	$O_3 + PM_{2.5}$	12	1.097* (1.001, 1.203)
Pulmonary	NO_2	$O_3 + PM_{2.5}$ (without Tracer 5)	3	1.122* (1.013, 1.244)
Pulmonary	O_3	$NO_2 + PM_{2.5}$	3	1.104* (1.018, 1.198)
Pulmonary	O_3	$PM_{0.1}$	3	1.080* (1.002, 1.165)
Pulmonary	O_3	$PM_{0.1}$	12	1.081* (1.009, 1.157)
Pulmonary	O_3	$NO_2 + PM_{0.1}$	12	1.094* (1.013, 1.183)
Pulmonary	O_3	$NO_2 + PM_{0.1}$	3	1.102* (1.012, 1.200)
Pulmonary	O_3	$PM_{2.5}$ (without Tracer 5)	3	1.088* (1.003, 1.180)
Pulmonary	O_3	$NO_2 + PM_{2.5}$ (without Tracer 5)	3	1.106* (1.019, 1.201)
Pulmonary	O_3	$PM_{2.5}$ (LUR)	3	1.099* (1.020, 1.183)
Pulmonary	O_3	$PM_{2.5}$ (LUR)	12	1.085* (1.014, 1.160)
Pulmonary	O_3	NO_2 (LUR) + $PM_{2.5}$ (LUR)	3	1.137** (1.046, 1.238)
Pulmonary	O_3	NO_2 (LUR) + $PM_{2.5}$ (LUR)	12	1.092* (1.011, 1.179)
Pulmonary	O_3	NO_2 (LUR) + $PM_{0.1}$	3	1.108* (1.017, 1.207)
Pulmonary	O_3	NO_2 (LUR) + $PM_{0.1}$	12	1.093* (1.012, 1.182)
Pulmonary	$PM_{0.1}$	NO_2	3	1.063* (1.008, 1.121)
Pulmonary	$PM_{0.1}$	NO_2 (LUR)	3	1.062* (1.009, 1.118)
Pulmonary	$PM_{2.5}$	NO_2	3	0.885** (0.814, 0.962)
Pulmonary	$PM_{2.5}$	NO_2	12	0.891** (0.827, 0.960)
Pulmonary	$PM_{2.5}$	$O_3 + NO_2$	3	0.903* (0.830, 0.982)
Pulmonary	$PM_{2.5}$	$O_3 + NO_2$	12	0.903** (0.838, 0.974)
Pulmonary	$PM_{2.5}$ (without Tracer 5)	NO_2	12	0.890** (0.823, 0.963)
Pulmonary	$PM_{2.5}$ (without Tracer 5)	NO_2	3	0.904* (0.828, 0.987)
Pulmonary	$PM_{2.5}$ (without Tracer 5)	$O_3 + NO_2$	12	0.906* (0.835, 0.983)
Renal	$PM_{0.1}$	$O_3 + NO_2$ (LUR)	12	1.164* (1.000, 1.354)

*$P \leq 0.05$; **$P \leq 0.01$.

Table E5. Average Exposure 30 Days and 365 Days Before Hospitalization, Normalized by IQR

	30 Days			365 Days		
Characteristic	Median (IQR)	Mean (SD)	Range	Median (IQR)	Mean (SD)	Range
NO_2	1.18 (0.80, 1.80)	1.30 (0.63)	0.08, 3.57	1.81 (1.27, 2.27)	1.78 (0.67)	0.13, 4.20
O_3	3.79 (3.41, 4.41)	4.08 (0.97)	2.57, 8.51	4.82 (4.34, 5.34)	4.84 (0.59)	3.07, 7.19
$PM_{0.1}$ mass	2.44 (1.95, 2.95)	2.56 (0.98)	0.42, 12.59	4.36 (3.76, 4.76)	4.24 (0.76)	0.79, 17.67
$PM_{2.5}$ EC	1.39 (0.94, 1.94)	1.45 (0.66)	0.06, 9.12	2.30 (1.68, 2.68)	2.21 (0.71)	0.25, 5.30
$PM_{2.5}$ mass	2.37 (1.87, 2.87)	2.40 (0.85)	0.37, 17.25	4.72 (4.15, 5.15)	4.61 (0.78)	1.49, 10.26
$PM_{2.5}$ nitrate	0.86 (0.40, 1.40)	0.94 (0.59)	0.00, 4.26	2.62 (2.03, 3.03)	2.52 (0.71)	0.22, 8.27
$PM_{2.5}$ OC	1.42 (0.93, 1.93)	1.58 (1.07)	0.05, 24.85	2.75 (2.16, 3.16)	2.69 (0.76)	0.25, 9.63
$PM_{2.5}$ Tracer5	0.55 (0.17, 1.17)	1.52 (3.66)	0.02, 115.21	2.21 (1.47, 2.47)	2.07 (1.09)	0.07, 20.54
LUR NO_2	1.87 (1.38, 2.38)	1.88 (0.66)	0.00, 4.64	2.34 (1.75, 2.75)	2.29 (0.66)	0.00, 6.01
LUR $PM_{2.5}$	3.63 (3.14, 4.14)	3.72 (0.97)	0.00, 8.98	4.80 (4.29, 5.29)	4.77 (0.86)	0.00, 8.50
$PM_{2.5}$ without Tracer5	2.44 (1.90, 2.90)	2.39 (0.72)	0.38, 5.12	4.97 (4.38, 5.38)	4.79 (0.78)	1.39, 8.43

Table E6. Akaike Information Criterion Estimates for Models with Different Exposure Windows: 30-Day, 365-Day, and Deviation from the Mean Models (30-day – 365-day + 365-day)

PASC Group	*Pollutant*	*Outcome Time (months)*	*AIC (NLL) 30-Day*	*AIC (NLL) 365-Day*	*AIC (NLL) 30-Day − 365-Day & 365-Day*
Cardiac	PM0.1	3	3723.406* (1851.703)	3726.108 (1853.054)	3725.266 (1851.633)
Cardiac	PM2.5 NV	12	5032.716* (2506.358)	5034.559 (2507.280)	5034.564 (2506.282)
Cardiometabolic/ diabetes	PM2.5 NV	3	5126.294 (2553.147)	5124.923* (2552.461)	5126.561 (2552.28)
Cardiometabolic/ diabetes	PM0.1	3	5123.140* (2551.570)	5130.069 (2555.035)	5125.026 (2551.513)
Cardiometabolic/ diabetes	PM2.5 NV	12	7109.248* (3544.624)	7110.040 (3545.020)	7110.704 (3544.352)
Pulmonary	PM0.1	3	12263.590* (6121.796)	12267.240 (6123.618)	12265.430 (6121.716)
Pulmonary	O3	3	12262.880* (6121.440)	12264.190 (6122.095)	12264.350 (6121.176)
Pulmonary	O3	12	14250.180 (7115.091)	14246.460* (7113.228)	14248.290 (7113.146)

NLL = negative log likelihood.

*Lowest AIC value among three models examined.

Table E7. Sandwich Estimator Sensitivity Analysis

PASC Group	Pollutant	Outcome Time (months)	Estimate (95% CI) 30-Day	Estimate (95% CI) 365-Day	Estimate (95% CI) 30-Day Deviation from 365-Day	Estimate (95% CI) 365-Day Deviation Model
Cardiac	$PM_{0.1}$	3	1.111*	1.077	1.106	1.093
			(1.010,	(0.961,	(0.986,	(0.979,
			1.222)	1.207)	1.241)	1.219)
Cardiac	$PM_{2.5}$	12	1.194*	1.109	1.148	1.073
	NV		(1.031,	(0.995,	(1.005,	(0.952,
			1.384)	1.237)	1.313)	1.210)
Cardiometabolic/diabetes	$PM_{2.5}$	3	1.176*	1.142**	1.046	1.126*
	NV		(1.013,	(1.025,	(0.904,	(1.000,
			1.365)	1.274)	1.211)	1.268)
Cardiometabolic/diabetes	$PM_{0.1}$	3	1.124**	1.044	1.146**	1.069
			(1.037,	(0.950,	(1.038,	(0.966,
			1.220)	1.149)	1.266)	1.183)
Cardiometabolic/diabetes	$PM_{2.5}$	12	1.150*	1.095*	1.071	1.072
	NV		(1.020,	(1.006,	(0.952,	(0.977,
			1.297)	1.192)	1.205)	1.176)
Pulmonary	$PM_{0.1}$	3	1.052*	1.034	1.051*	1.043
			(1.009,	(0.986,	(1.000,	(0.994,
			1.098)	1.085)	1.105)	1.094)
Pulmonary	O_3	3	1.083*	1.074*	1.087	1.092**
			(1.016,	(1.006,	(0.964,	(1.018,
			1.156)	1.146)	1.226)	1.170)
Pulmonary	O_3	12	1.061*	1.082*	1.018	1.086**
			(1.006,	(1.025,	(0.921,	(1.025,
			1.120)	1.142)	1.126)	1.151)

*$P \leq 0.05$; **$P \leq 0.01$.

References

1. Emery C, Liu Z, Russell AG, Odman MT, Yarwood G, Kumar N. Recommendations on statistics and benchmarks to assess photochemical model performance. *J Air Waste Manag Assoc*. 2017;67(5):582-598. doi:10.1080/10962247.2016.1265027