AGU PUBLICATIONS

Earth and Space Science

Supporting Information for

Sub-city Scale Hourly Air Quality Forecasting by Combining Models, Satellite Observations, and Ground Measurements

C. Malings[1,2,3], K. E. Knowland[2,3], C. A. Keller[2,3], and S. E. Cohn[2]

[1]NASA Postdoctoral Program Fellow, Goddard Space Flight Center, Greenbelt, MD, USA.
[2]Global Modeling and Assimilation Office, Goddard Space Flight Center, Greenbelt, MD, USA.
[3]Universities Space Research Association, Columbia, MD, USA.
Corresponding Author: Carl Malings (carl.a.malings@nasa.gov)

Contents of this file

Figures S1 to S62
Tables S1 to S2

Additional Supporting Information (Files uploaded separately)

All codes used for data analysis and prediction, as well as related outputs, are available at https://doi.org/10.5281/zenodo.4581090.

Introduction

This document contains supplementary figures and tables related to the main publications. These include supplementary background material (Section S1, Figures S1 and S2) and detailed examinations of certain results (Section S2, Tables S1 and S2 and Figures S3 to S62).

S1 Background Information

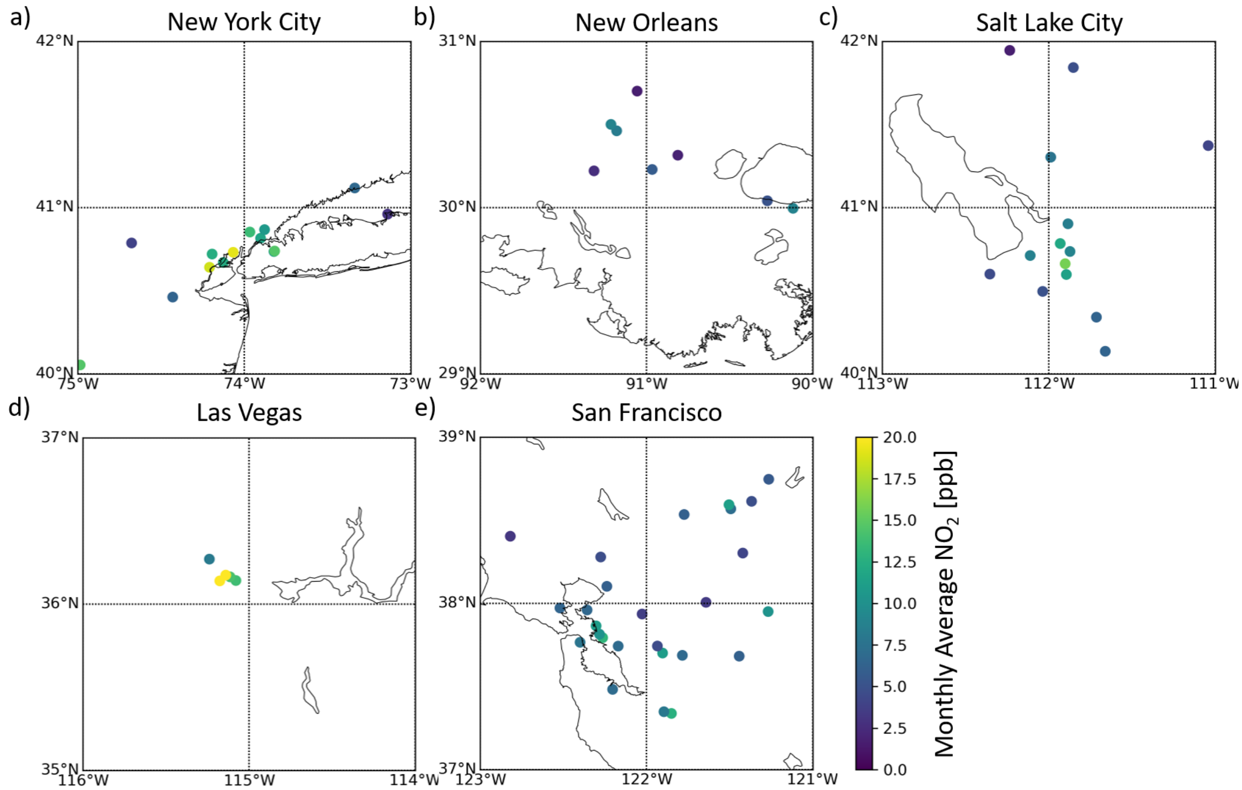

Figure S1. Overview of areas of interest. Points indicate locations of EPA ground monitoring stations for NO_2 in each given region, with the average monthly concentrations during September 2019 for each site indicated by the color.

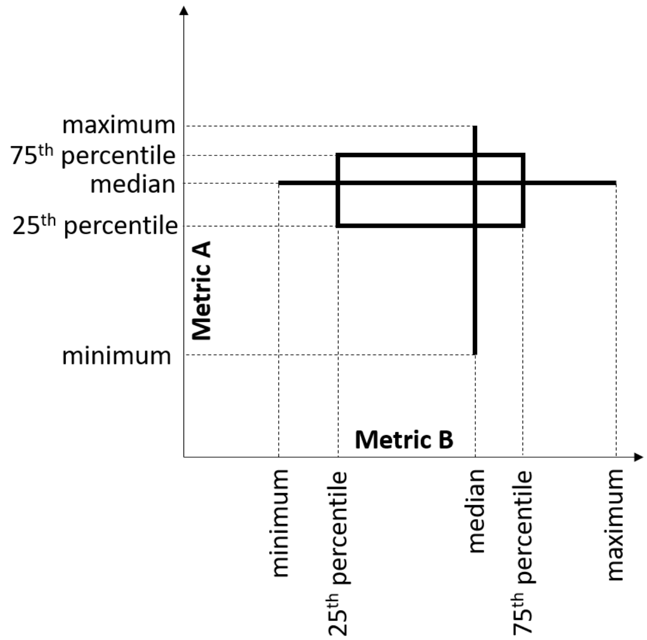

Figure S2. A guide to interpreting two-dimensional boxplots. The position of the cross indicates the median for the metrics of both axes, and the lengths of the arms of the cross indicate the spread in values for both metrics. The box surrounding the cross denotes the inter-quartile range (25^{th} percentile to 75^{th} percentile) of both metrics.

S2 Detailed Results

	Values				
Parameter	Las Vegas	New Orleans	New York City	Salt Lake City	San Francisco
λ_{space}	10 kilometers	25 kilometers	35 kilometers	20 kilometers	30 kilometers
λ_{time}	5 hours	5 hours	5 hours	4 hours	6 hours
$\sigma_{variability}$	10 ppb	6 ppb	8 ppb	5 ppb	6 ppb
$\sigma_{measure}$	0.1 ppb	0.06 ppb	0.08 ppb	0.05 ppb	0.06 ppb

Table S1. Parameters estimated for spatial and temporal covariance definition in different application areas. Temporal and spatial correlation scale parameters λ_{time} and λ_{space} are roughly estimated by inspection from plots like Figure 5 of the main paper generated for each city. Residual variability $\sigma_{variability}$ is chosen to roughly match the RMSE values of the method applied without updating (based on results presented in Section 5 of the main paper). Measurement noise $\sigma_{measure}$ is assumed to be quite small (1%) relative to the residual variability, implying precise measurements.

S2.1 Comparison of approach settings in different areas

	Model Downscaling (Section 3.1)	**Pattern Extraction (Section 3.2)**	**Pattern Integration (Section 3.3)**	**Regression Weighting (Section 3.5)**

Overall	Slight preference for linear interpolation	Slight preference for extraction at satellite overpass times (Equation 3) only	Slight preference for pattern combination via addition (Equation 6)	Slight preference for decaying periodic weighting (Equation 11)
New York City	Slight improvement with linear interpolation	Similar performance of both methods	Similar performance of both methods	Periodic (Equation 9) and Decaying Periodic weightings have highest correlation, time-of-day weighting has lowest bias
New Orleans	Slight improvement with linear interpolation	Correlations improved by using overpass times only (Equation 3)	Adding in patterns improves correlation, but regression using patterns reduces bias	Periodic and decaying periodic weightings have best performance overall
Las Vegas	Linear interpolation improved most performance metrics	Accuracy and bias improved through use of full calibration period patterns (Equation 2)	Adding in patterns reduces bias and error	Decaying periodic weighting has highest correlation, time-of-day weighting has lowest error and bias
Salt Lake City	Slight improvement with linear interpolation	Similar performance of both methods	Similar performance of both methods	Time-of-day weighting is best by all metrics
San Francisco	Slight improvement with linear interpolation	Restricted patterns (Equation 3) improve correlation but full patterns (Equation 2) reduce bias	Adding patterns (Equation 6) improves correlations	Decaying (Equation 10) and decaying periodic weightings have highest correlation, time-of-day

				weighting has least bias

Table S2. A qualitative summary of the observed effects of taking different approaches during the various phases of downscaling, typical pattern extraction and combination, and regression weighting.

S2.1.1 New York City

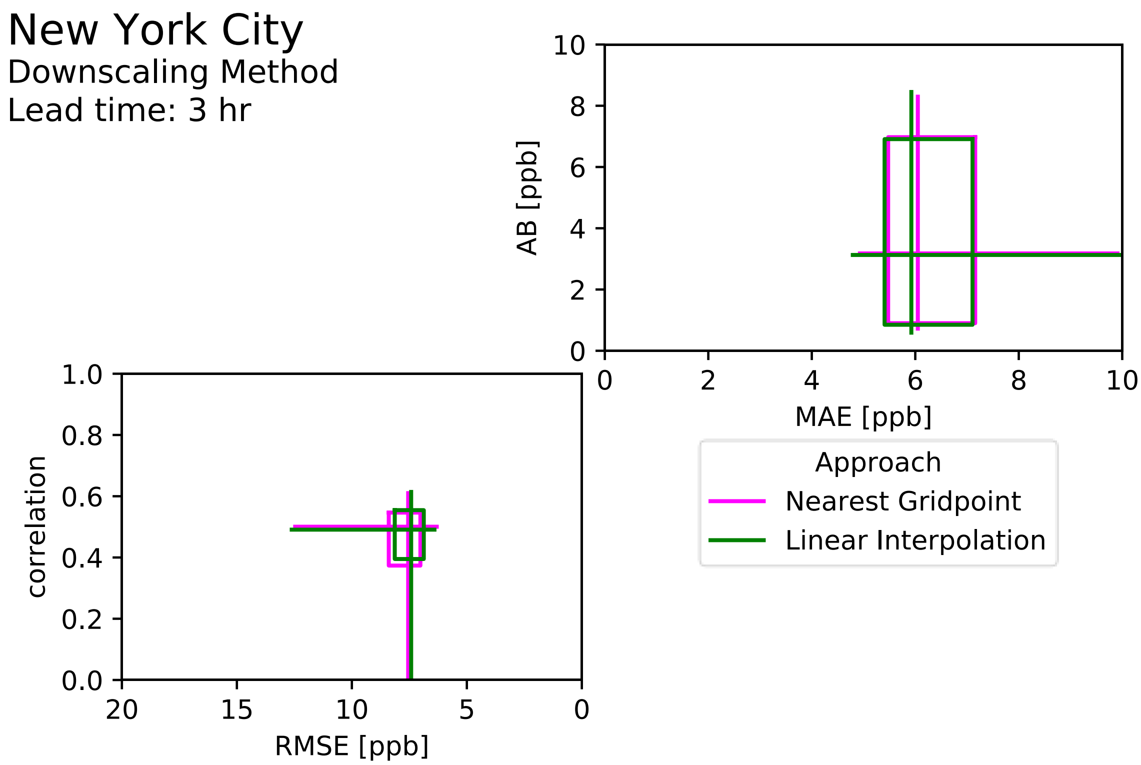

Figure S3. Effects of the downscaling approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, restricted patterns (Equation 3), pattern combination via addition (Equation 6), and decaying periodic regression weighting (Equation 11).

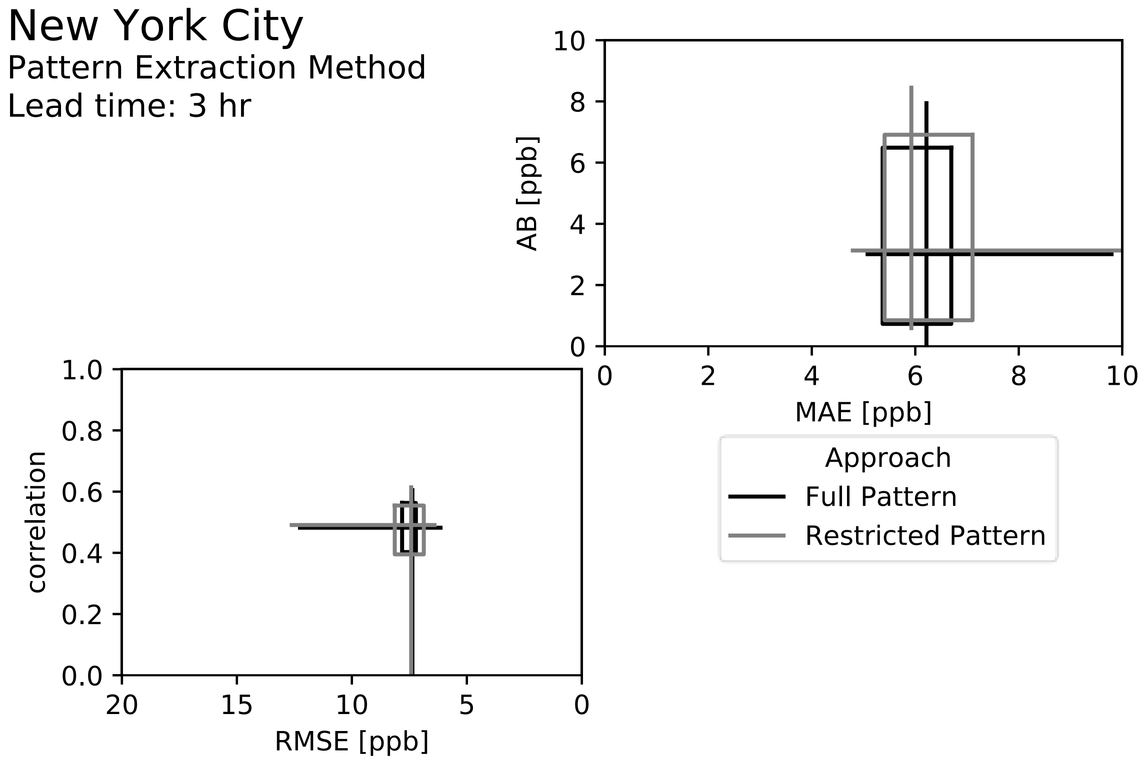

Figure S4. Effects of the pattern extraction approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, pattern combination via addition (Equation 6), and decaying periodic regression weighting (Equation 11).

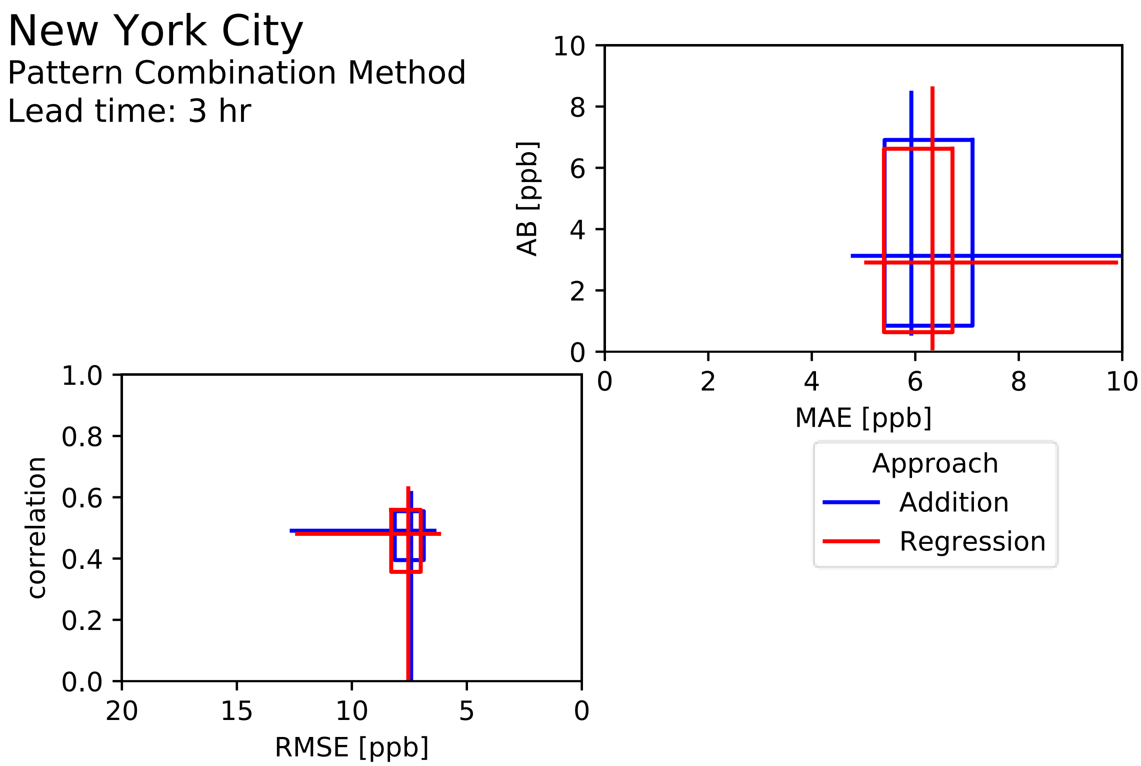

Figure S5. Effects of the pattern combination approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and decaying periodic regression weighting (Equation 11).

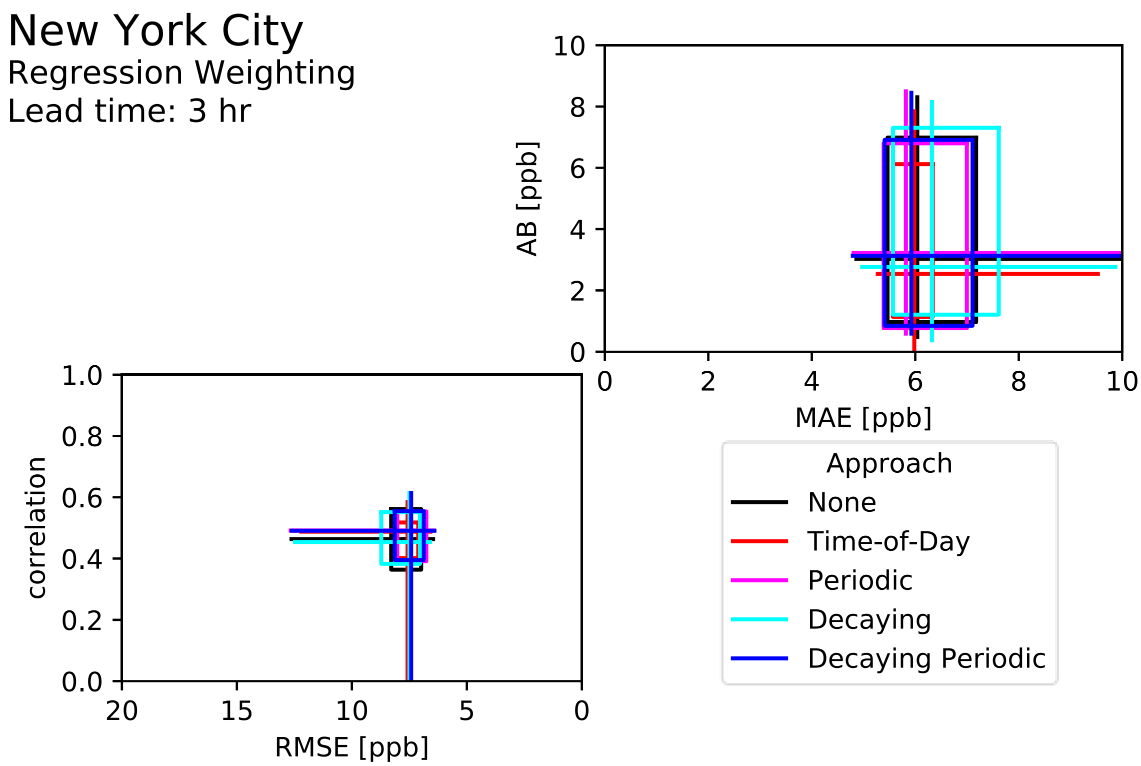

Figure S6. Effects of the regression weighting approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation3), and pattern combination via addition (Equation 6).

S2.1.2 New Orleans

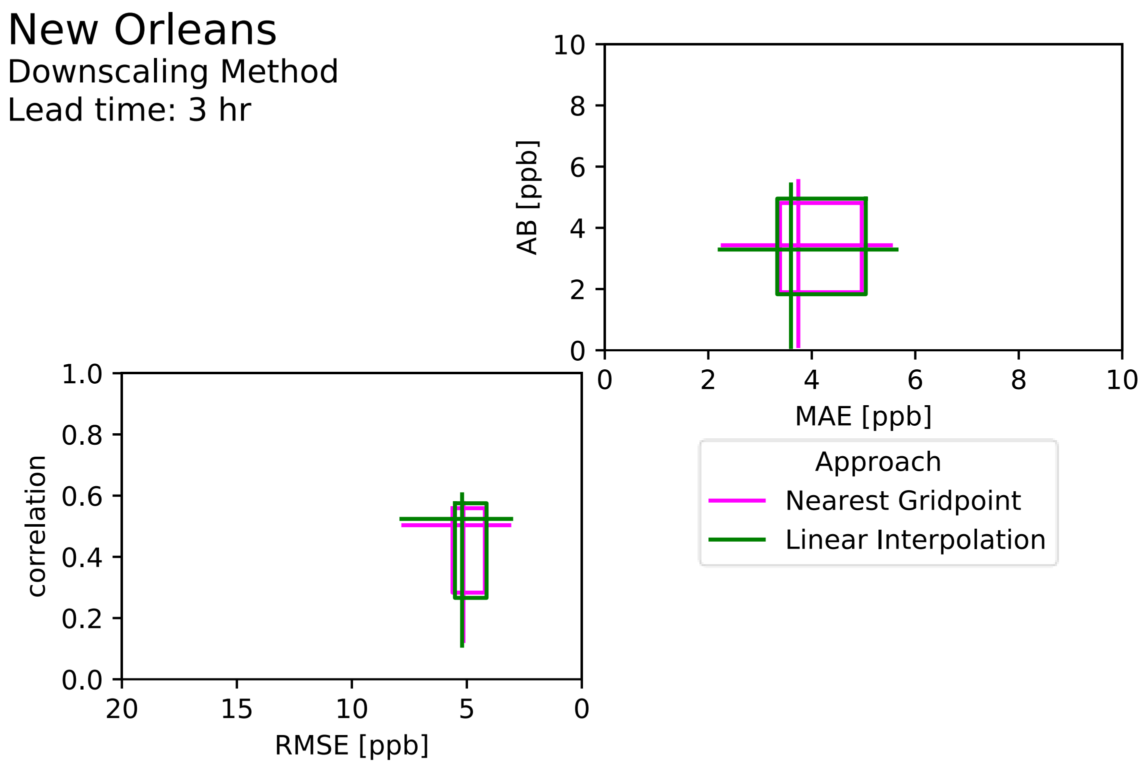

Figure S7. Effects of the downscaling approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, restricted patterns (Equation 3), pattern combination via addition (Equation 6), and decaying periodic regression weighting (Equation 11).

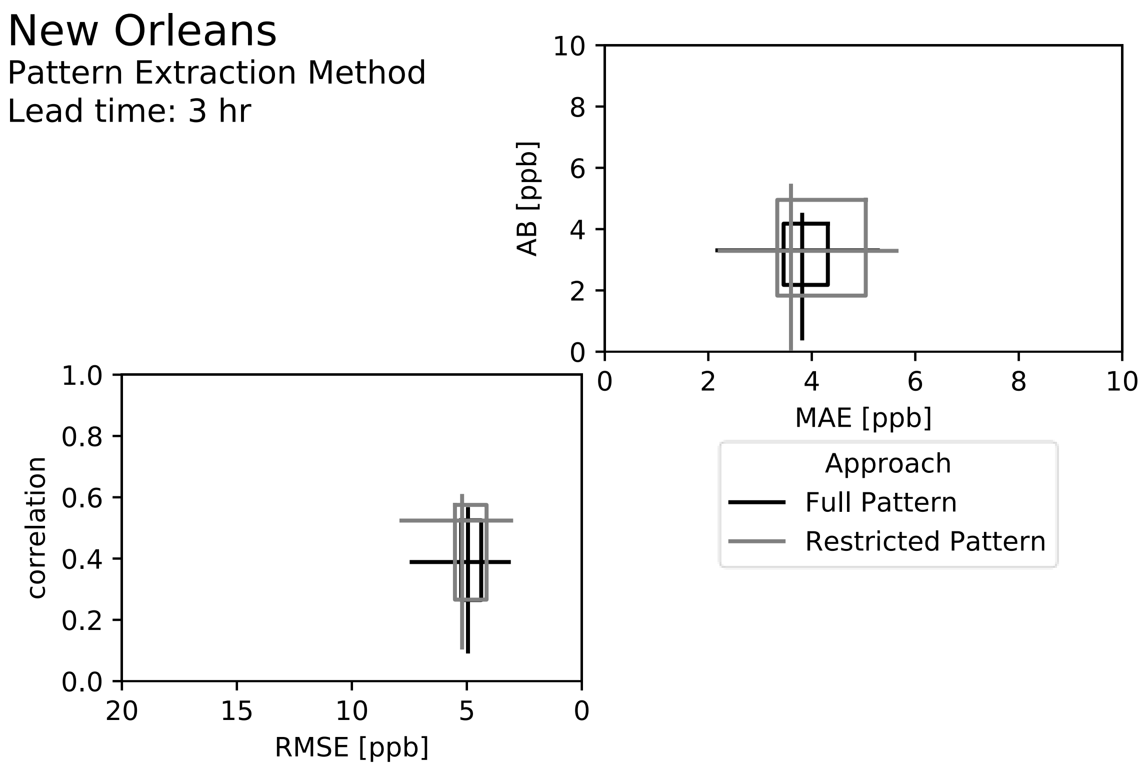

Figure S8. Effects of the pattern extraction approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, pattern combination via addition (Equation 6), and decaying periodic regression weighting (Equation 11).

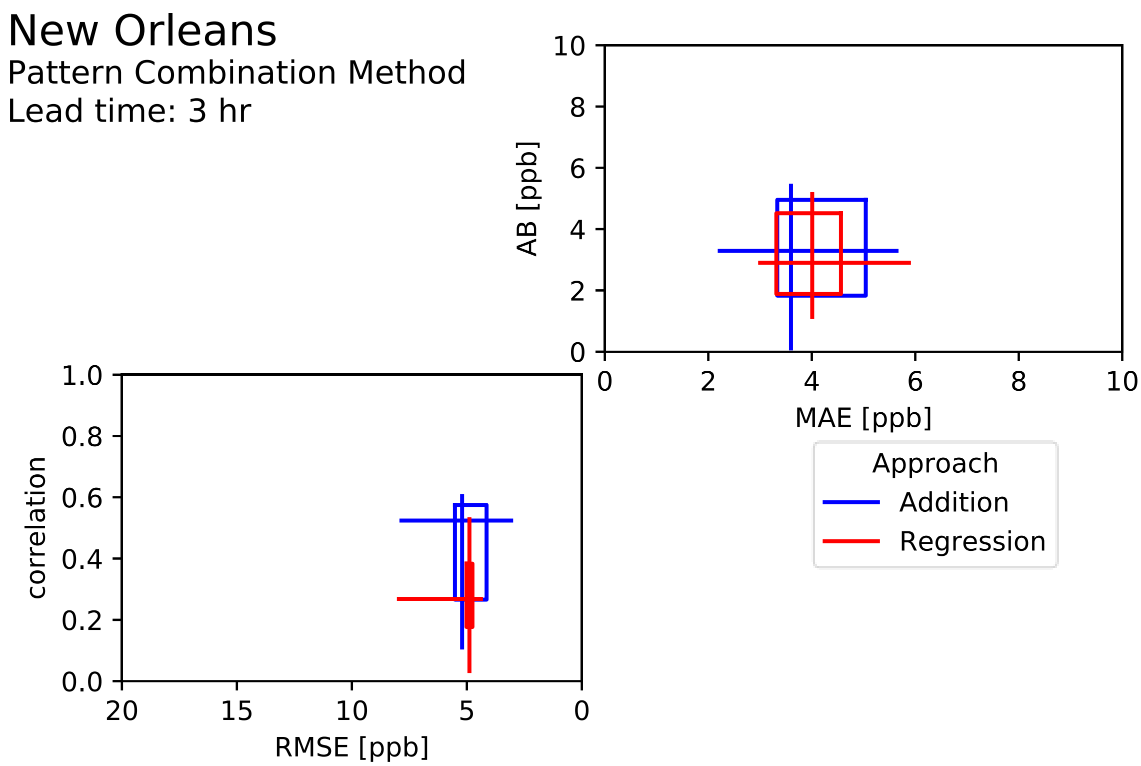

Figure S9. Effects of the pattern combination approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and decaying periodic regression weighting (Equation 11).

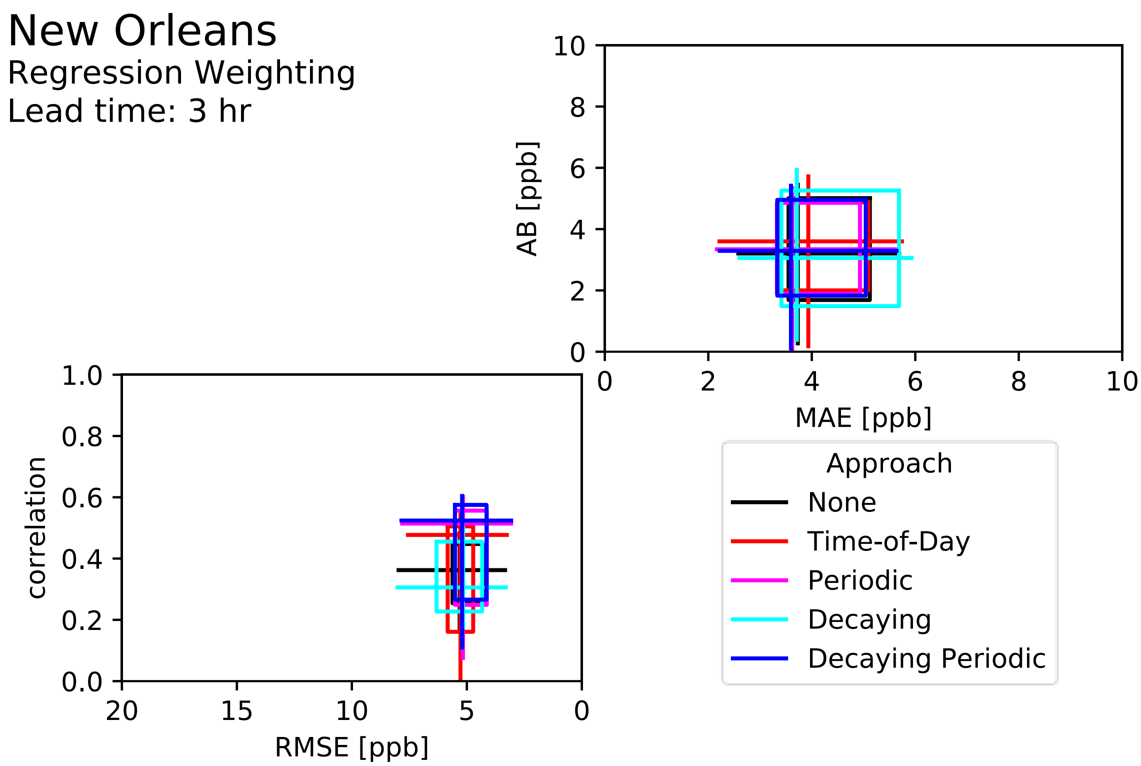

Figure S10. Effects of the regression weighting approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and pattern combination via addition (Equation 6).

S2.1.3 Las Vegas

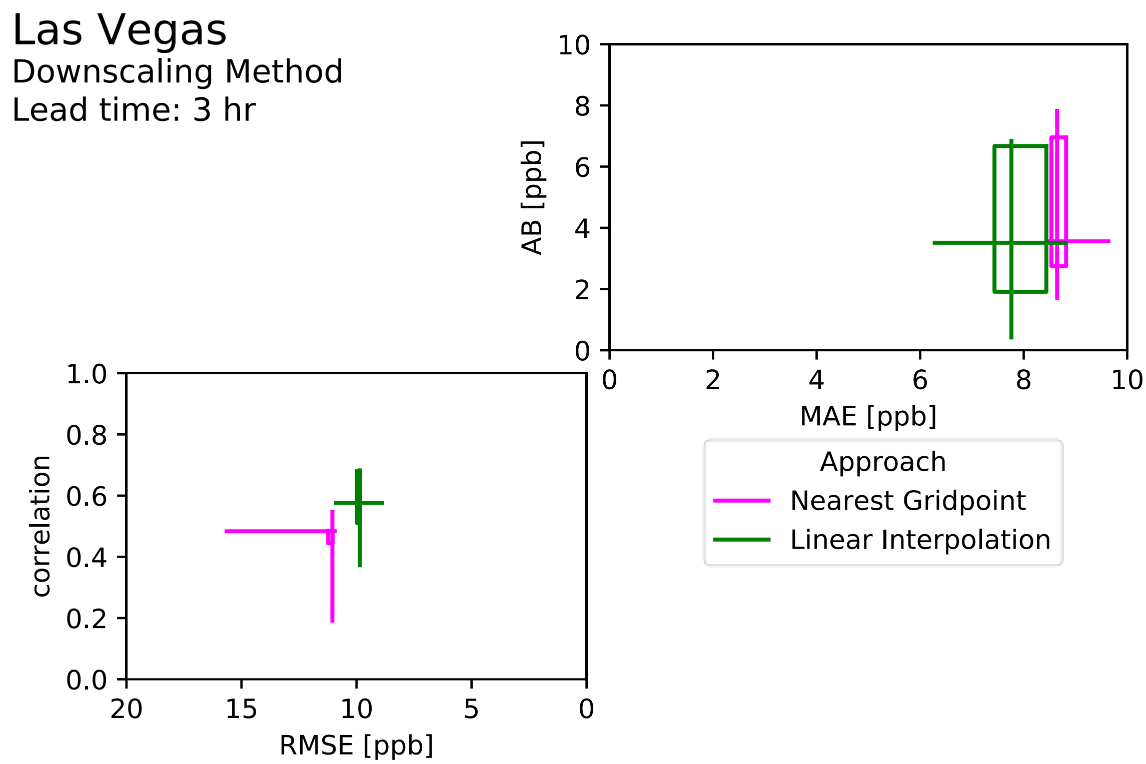

Figure S11. Effects of the downscaling approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, full patterns (Equation 2), pattern combination via addition (Equation 6), and time-of-day regression weighting.

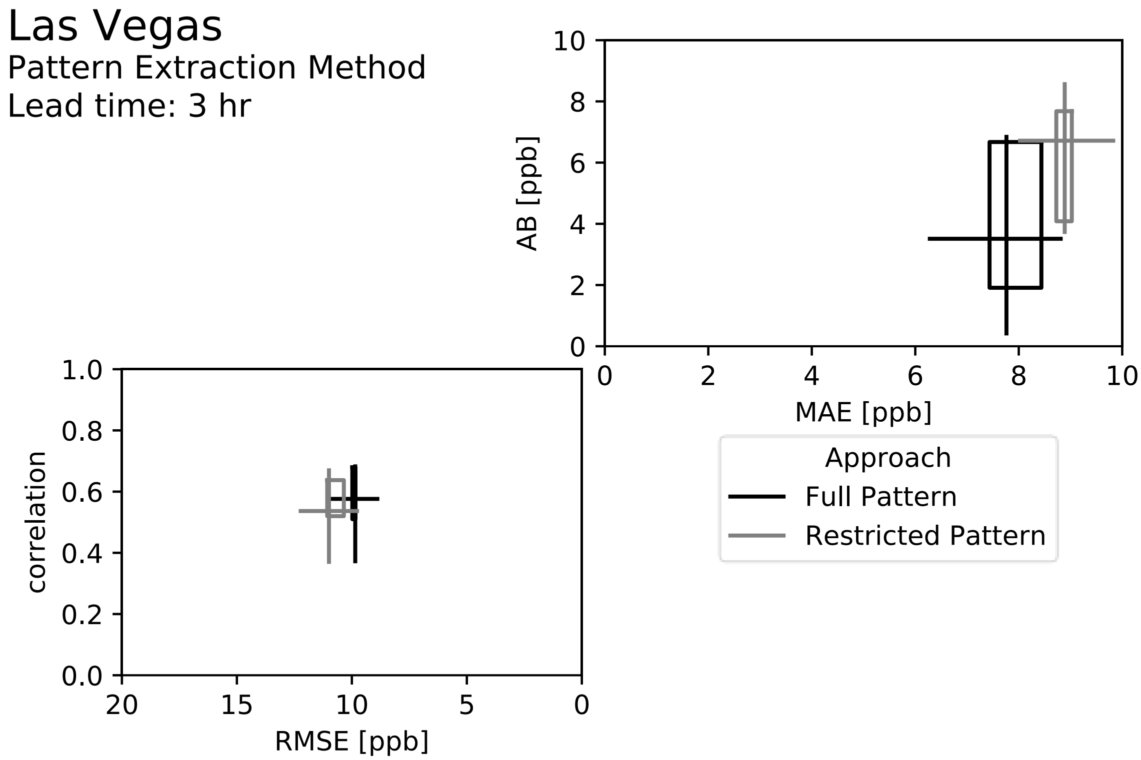

Figure S12. Effects of the pattern extraction approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, pattern combination via addition (Equation 6), and time-of-day regression weighting.

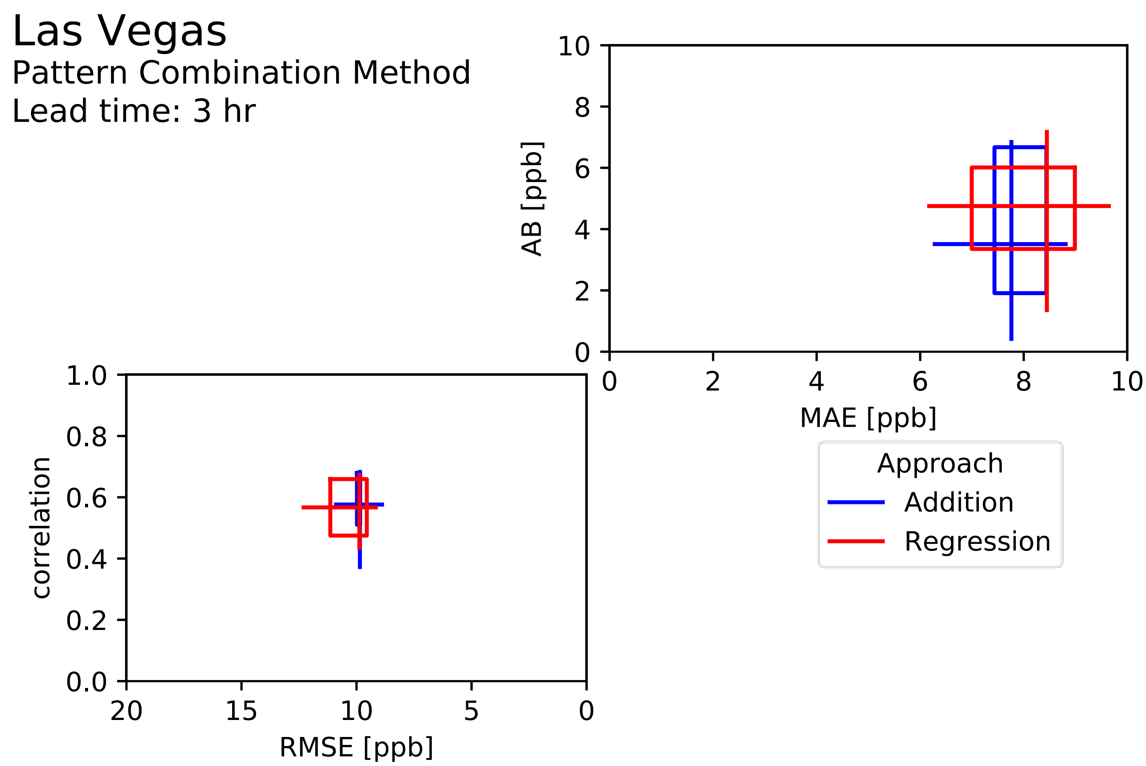

Figure S13. Effects of the pattern combination approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, full patterns (Equation 2), and time-of-day regression weighting.

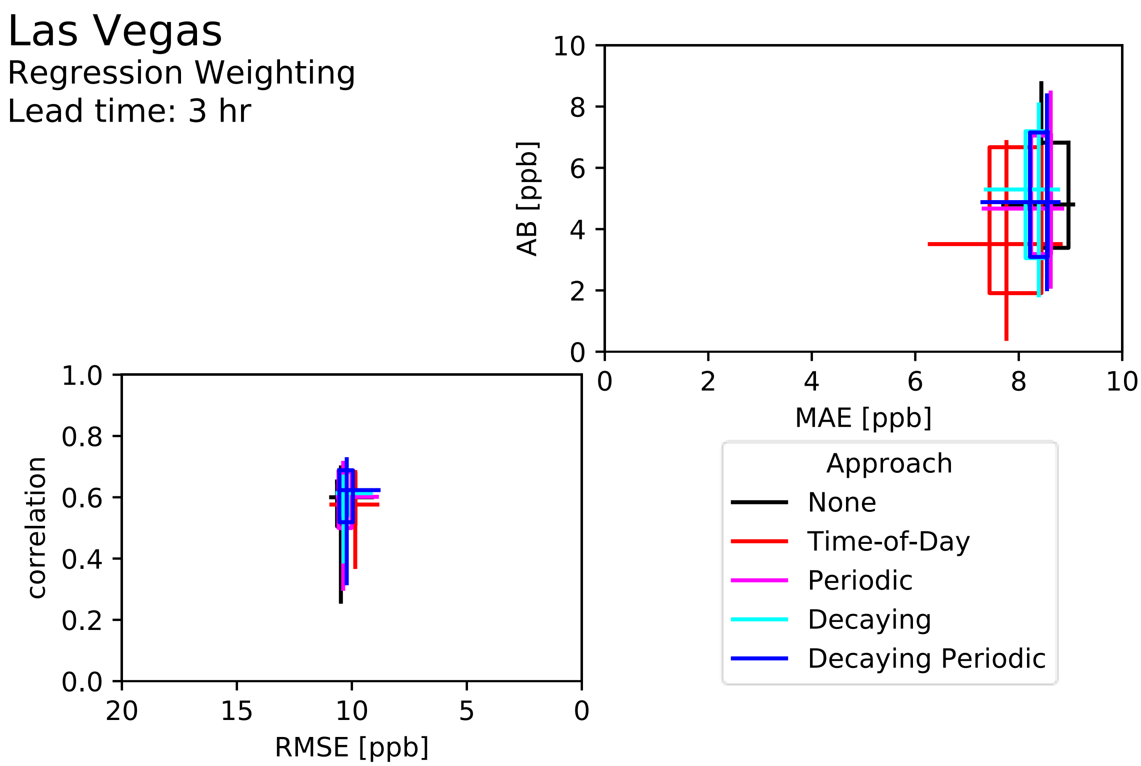

Figure S14. Effects of the regression weighting approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, full patterns (Equation 2), and pattern combination via addition (Equation 6).

S2.1.4 Salt Lake City

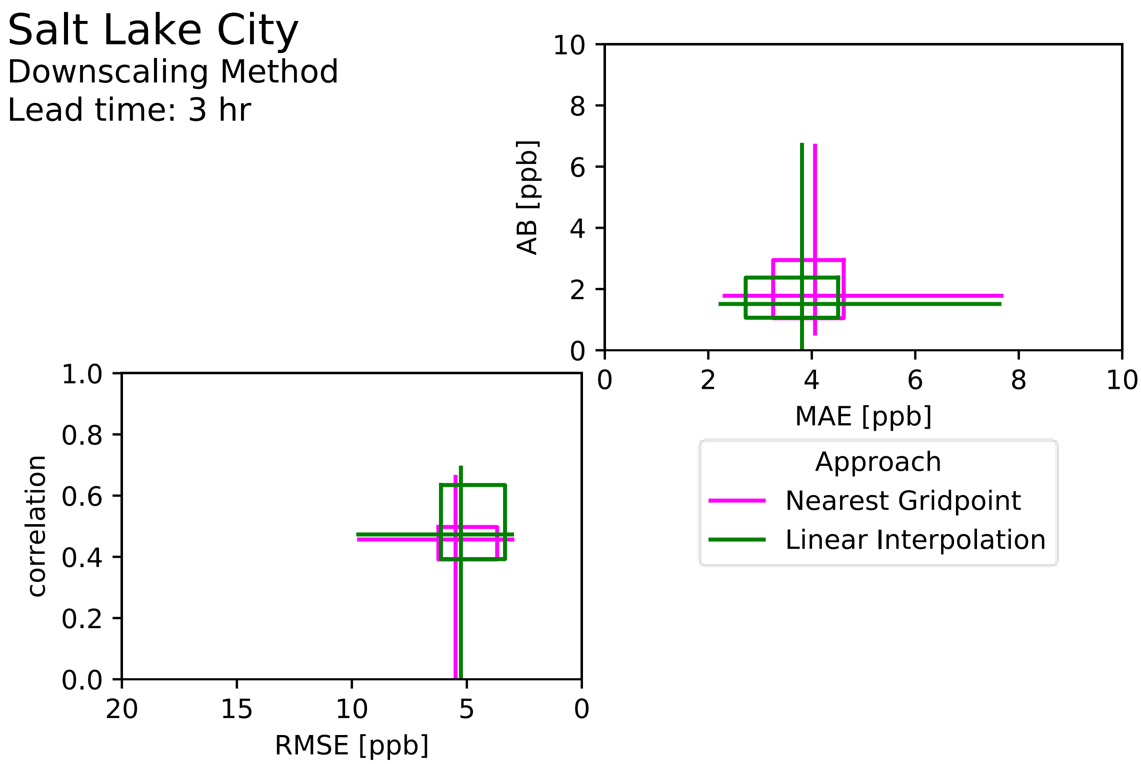

Figure S15. Effects of the downscaling approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, restricted patterns (Equation 3), pattern combination via addition (Equation 6), and time-of-day regression weighting.

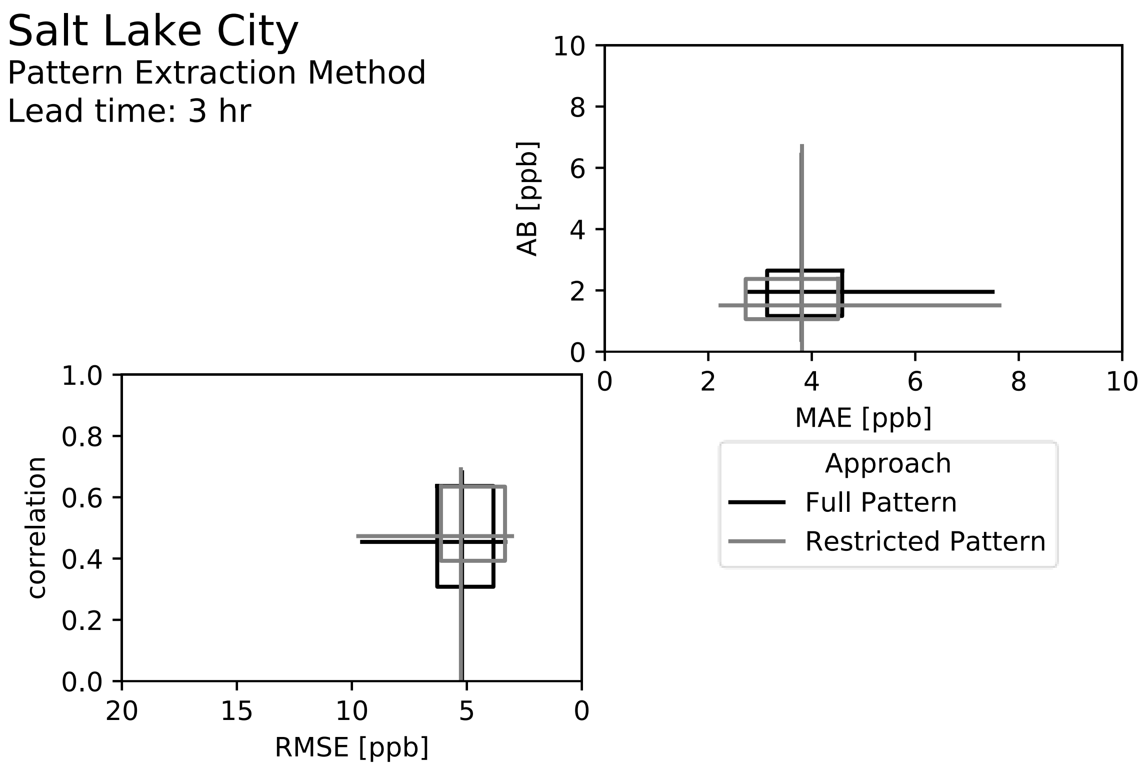

Figure S16. Effects of the pattern extraction approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, pattern combination via addition (Equation 6), and time-of-day regression weighting.

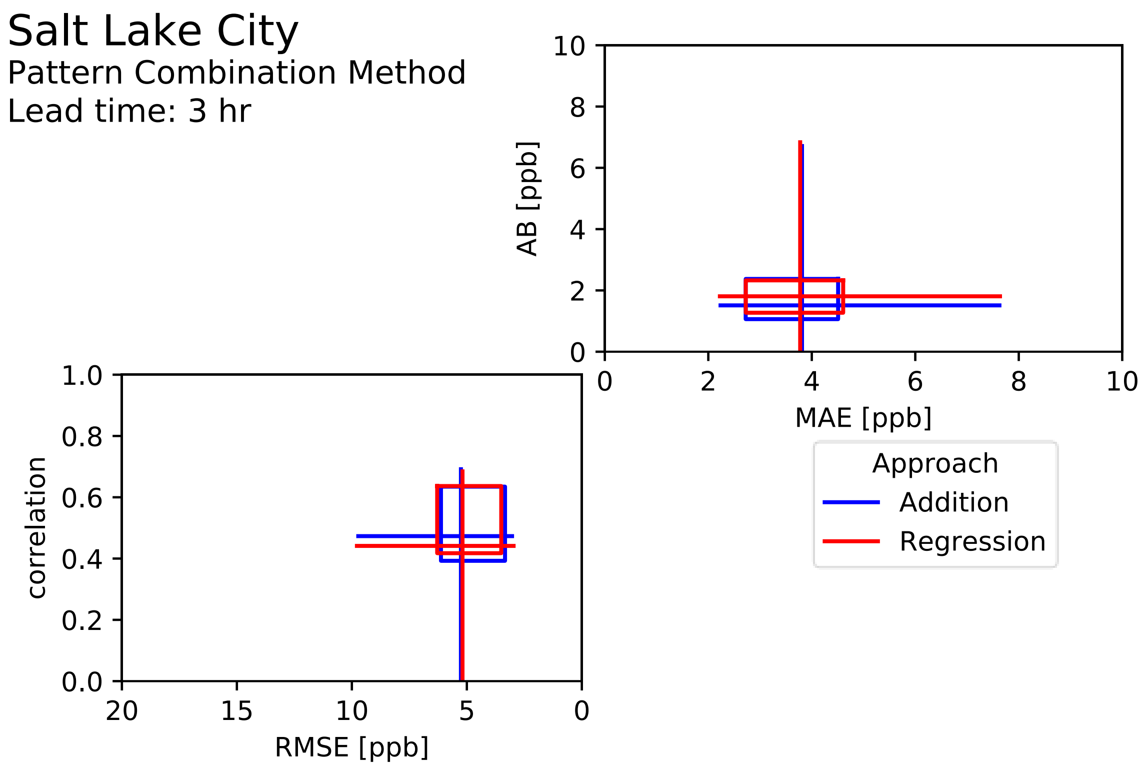

Figure S17. Effects of the pattern combination approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and time-of-day regression weighting.

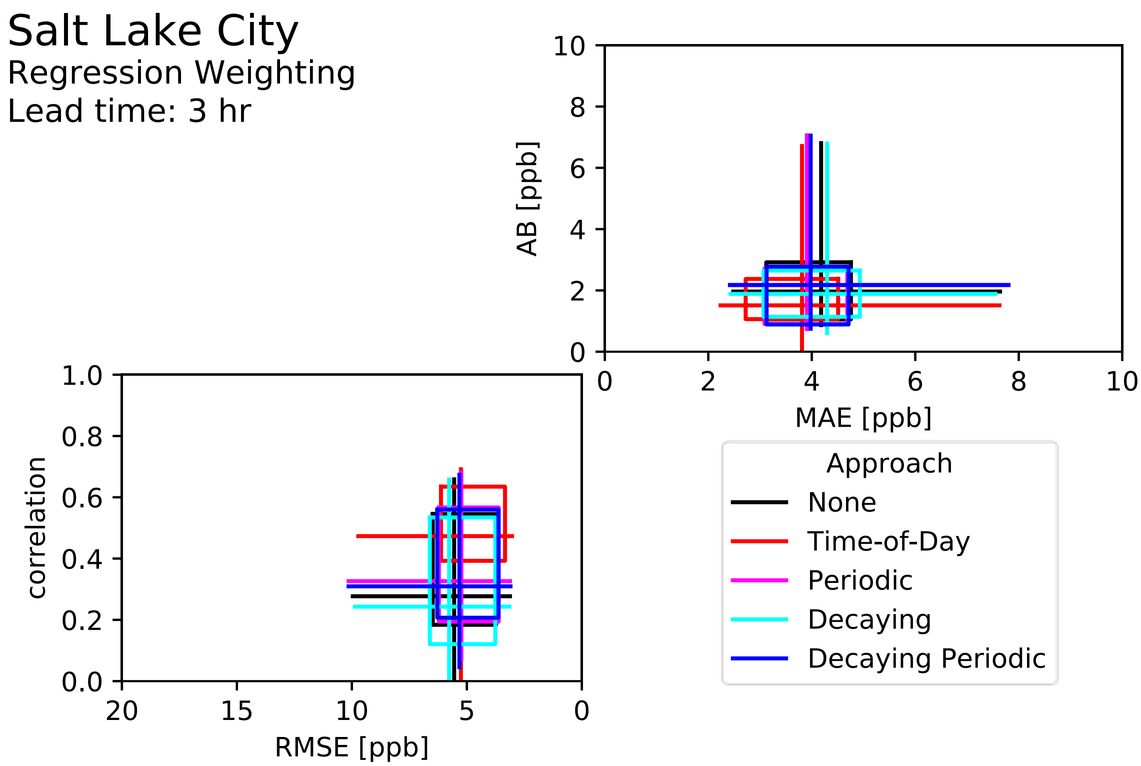

Figure S18. Effects of the regression weighting approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and pattern combination via addition (Equation 6).

S2.1.5 San Francisco

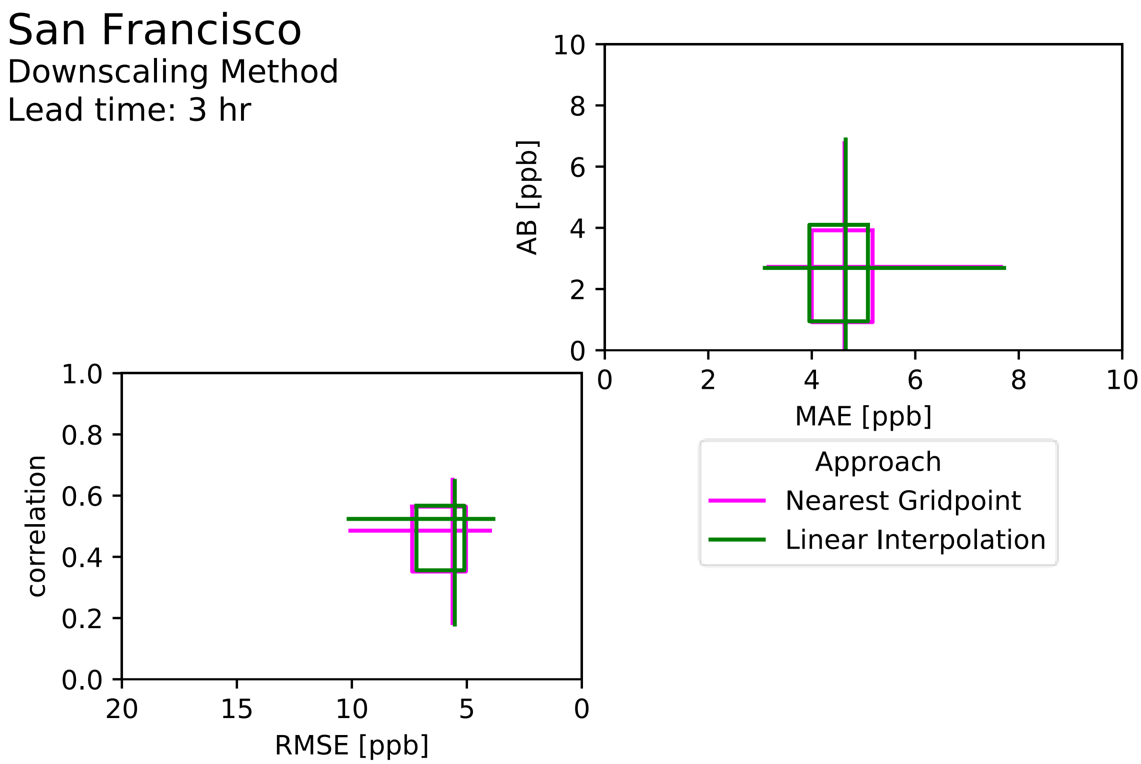

Figure S19. Effects of the downscaling approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, restricted patterns (Equation 3), pattern combination via addition (Equation 6), and decaying periodic regression weighting (Equation 11).

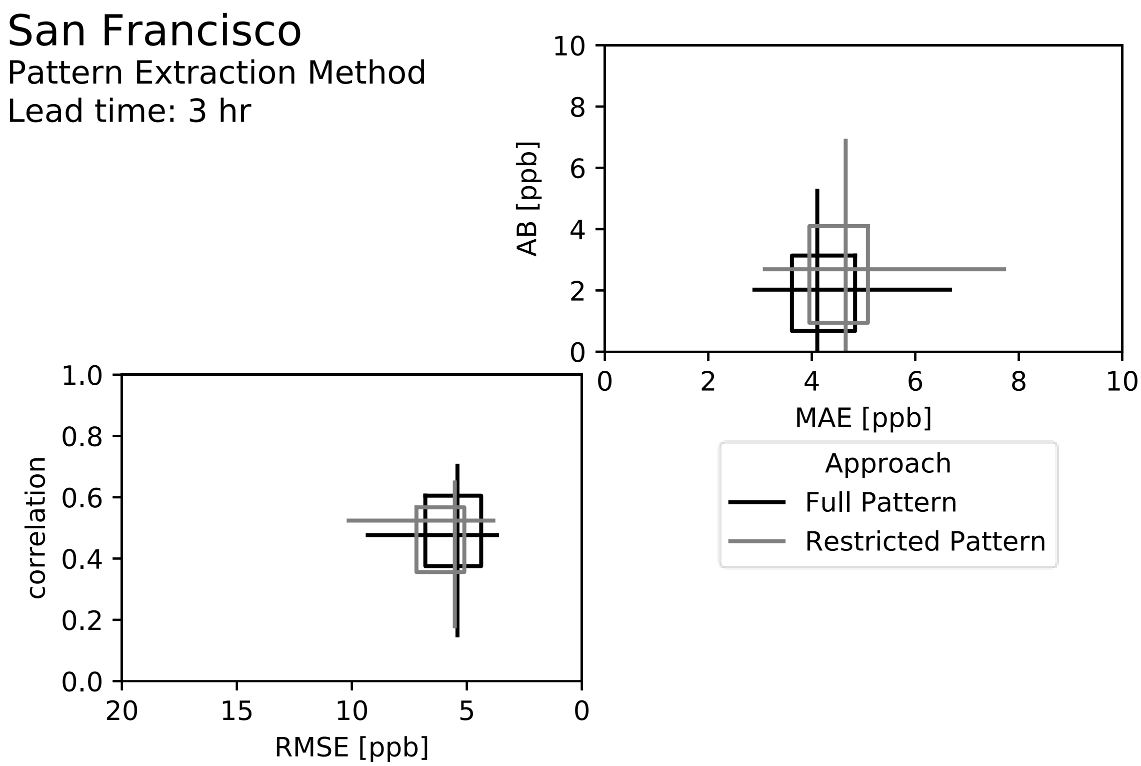

Figure S20. Effects of the pattern extraction approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, pattern combination via addition (Equation 6), and decaying periodic regression weighting (Equation 11).

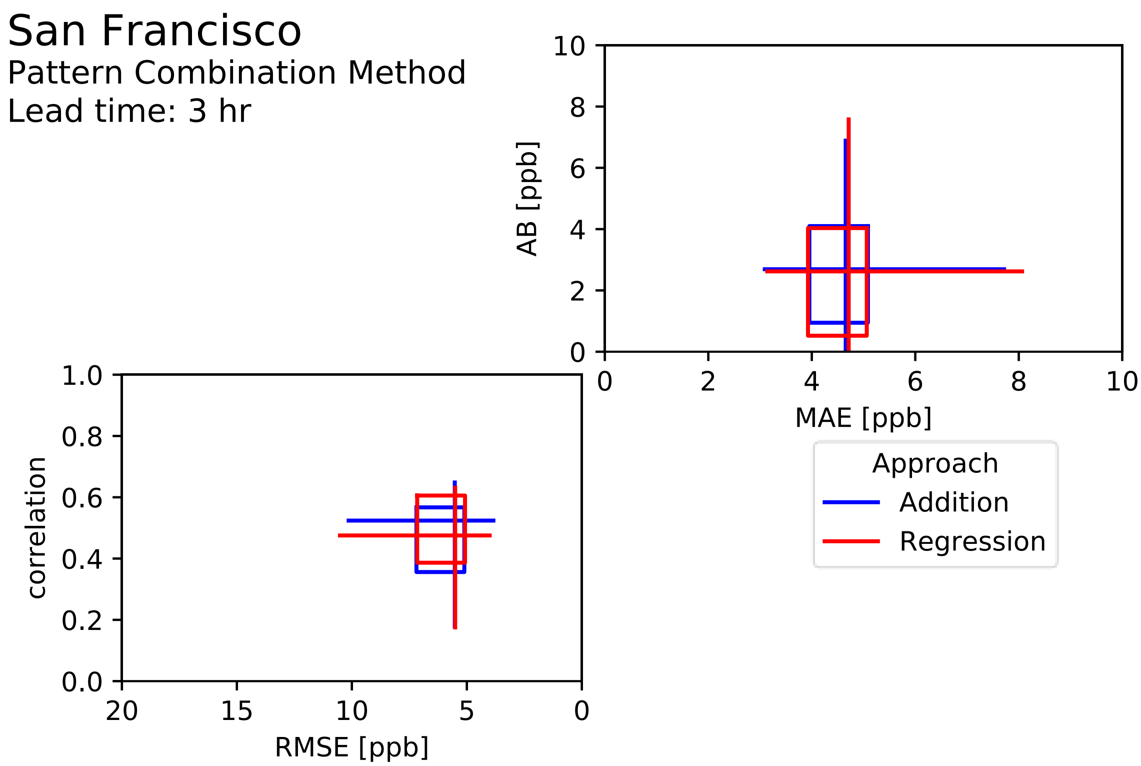

Figure S21. Effects of the pattern combination approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and decaying periodic regression weighting (Equation 11).

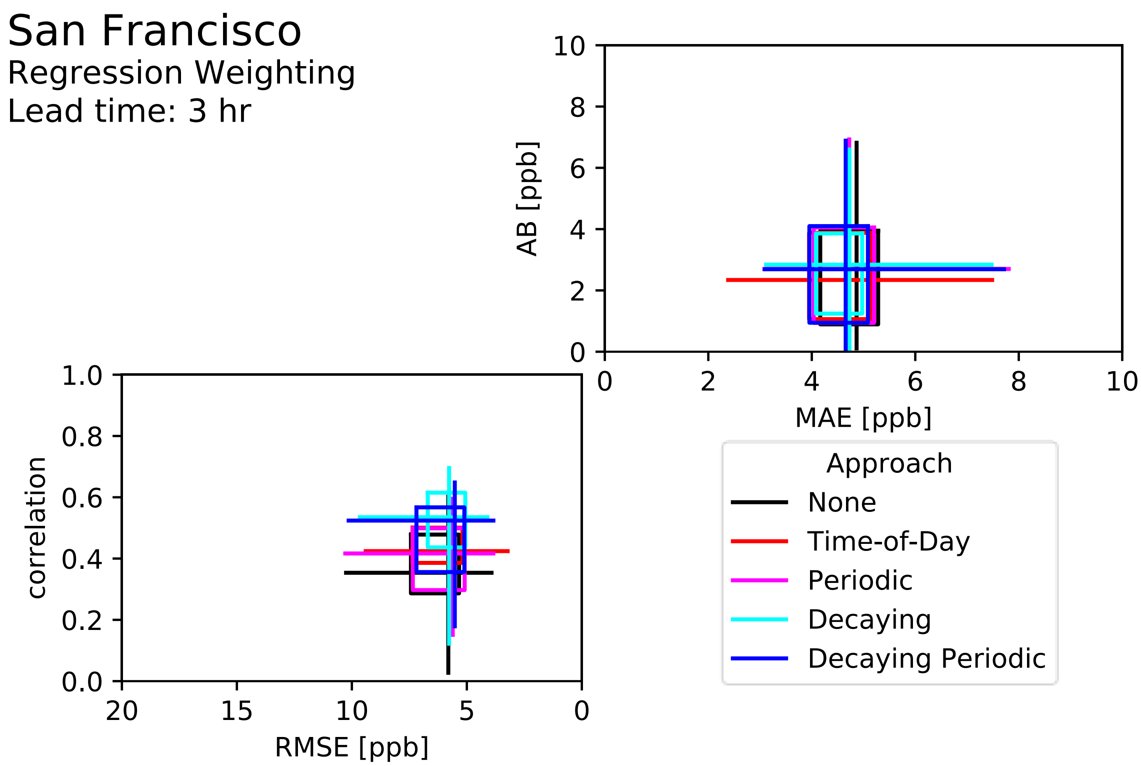

Figure S22. Effects of the regression weighting approach on predictive method performance at 3-hour lead time. The predictive method uses GEOS-CF and TROPOMI information together with ground data, linear interpolation downscaling, restricted patterns (Equation 3), and pattern combination via addition (Equation 6).

S2.2 Performance of methods incorporating different data sources in different areas
S2.2.1 New York City

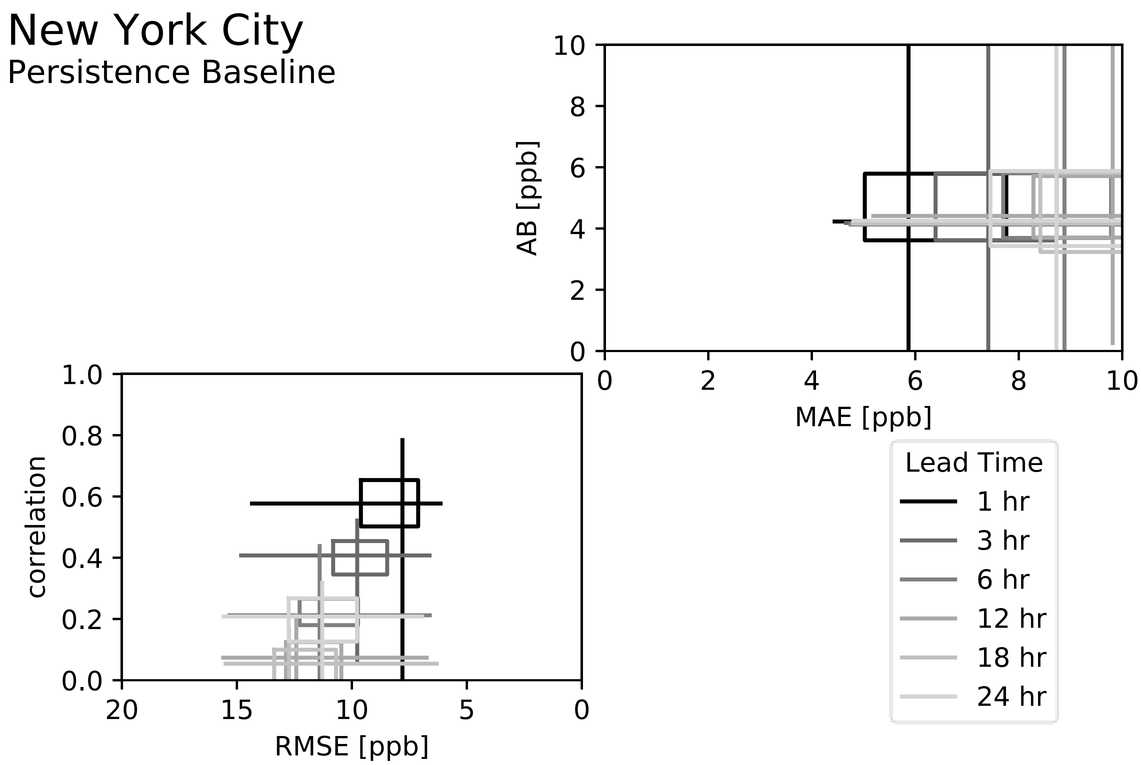

Figure S23. Performance of the persistence baseline method.

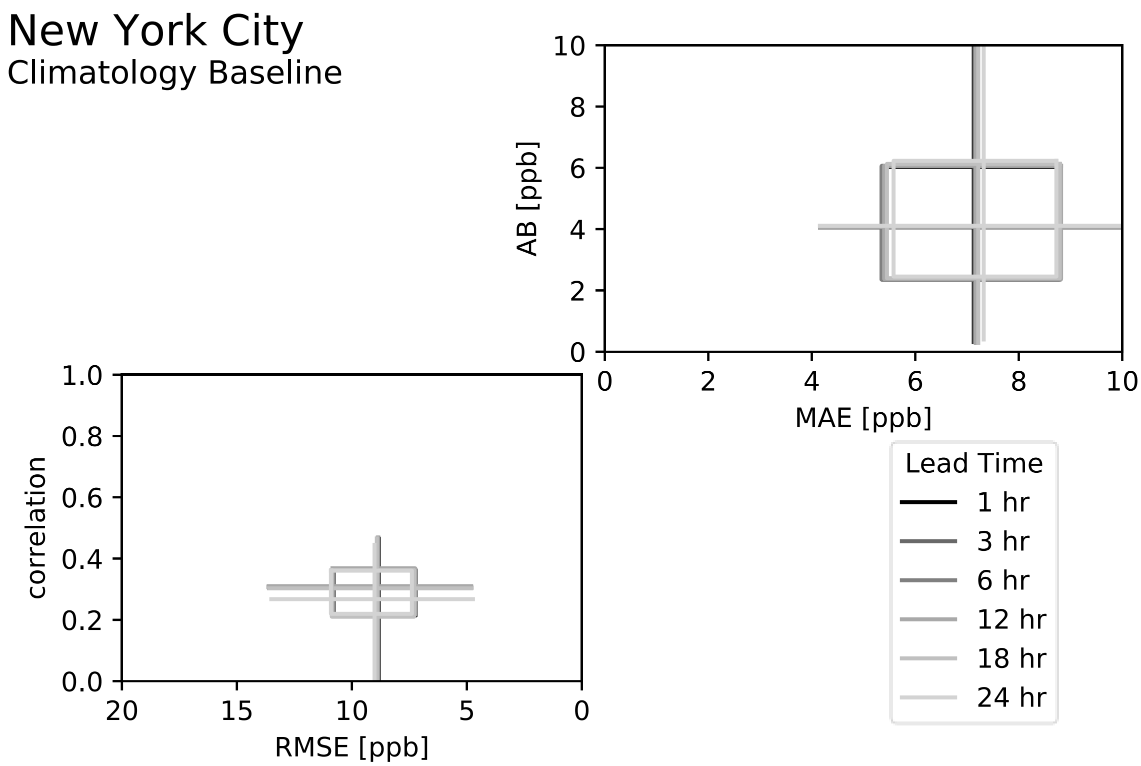

Figure S24. Performance of the climatology baseline method.

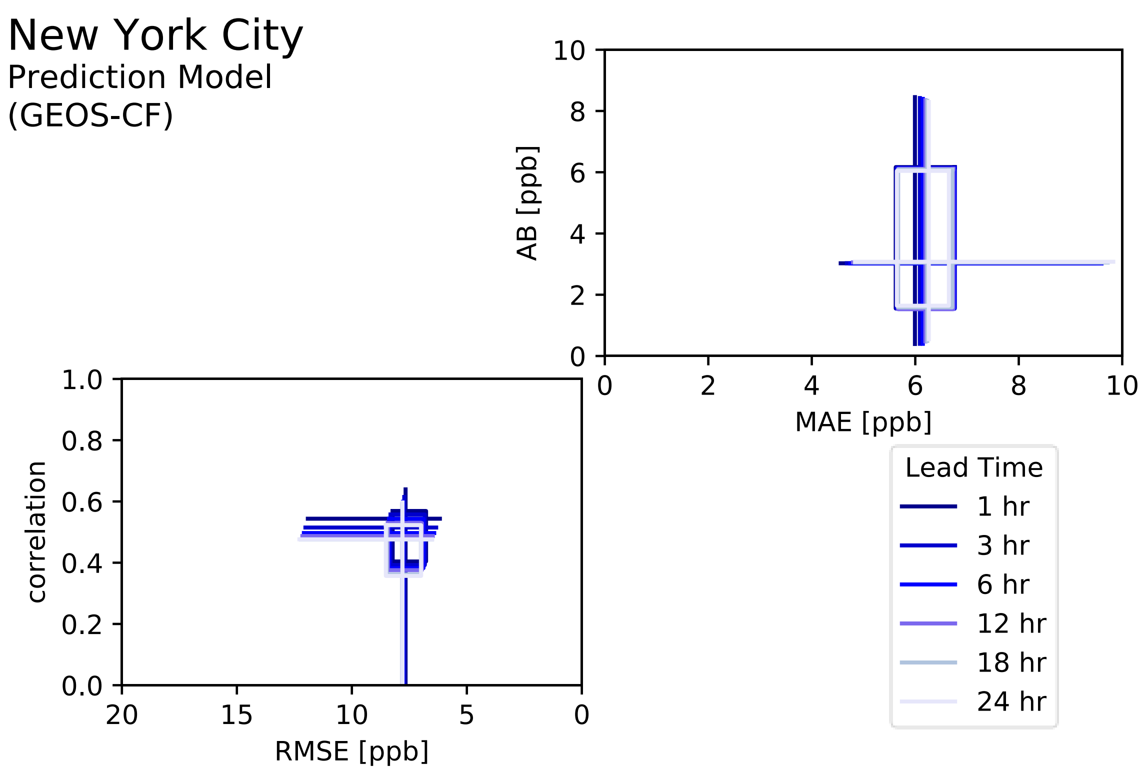

Figure S25. Performance of the proposed method using GEOS-CF and ground information only (downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

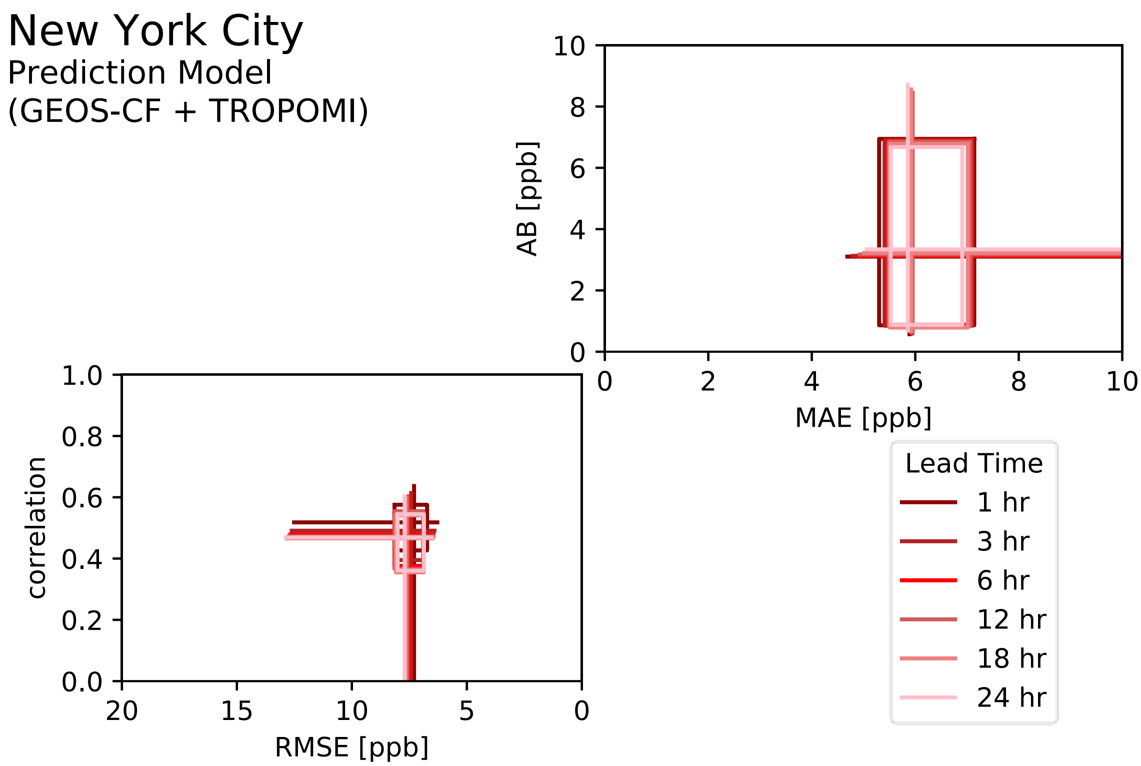

Figure S26. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

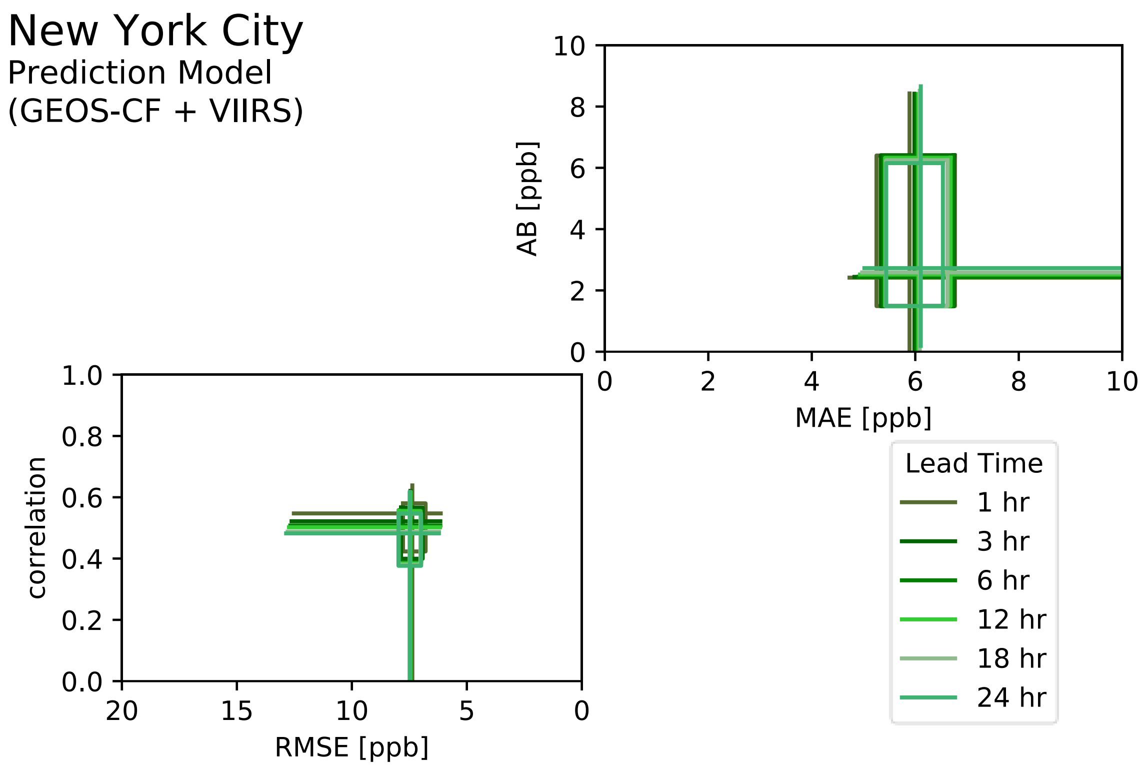

Figure S27. Performance of the proposed method using GEOS-CF and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

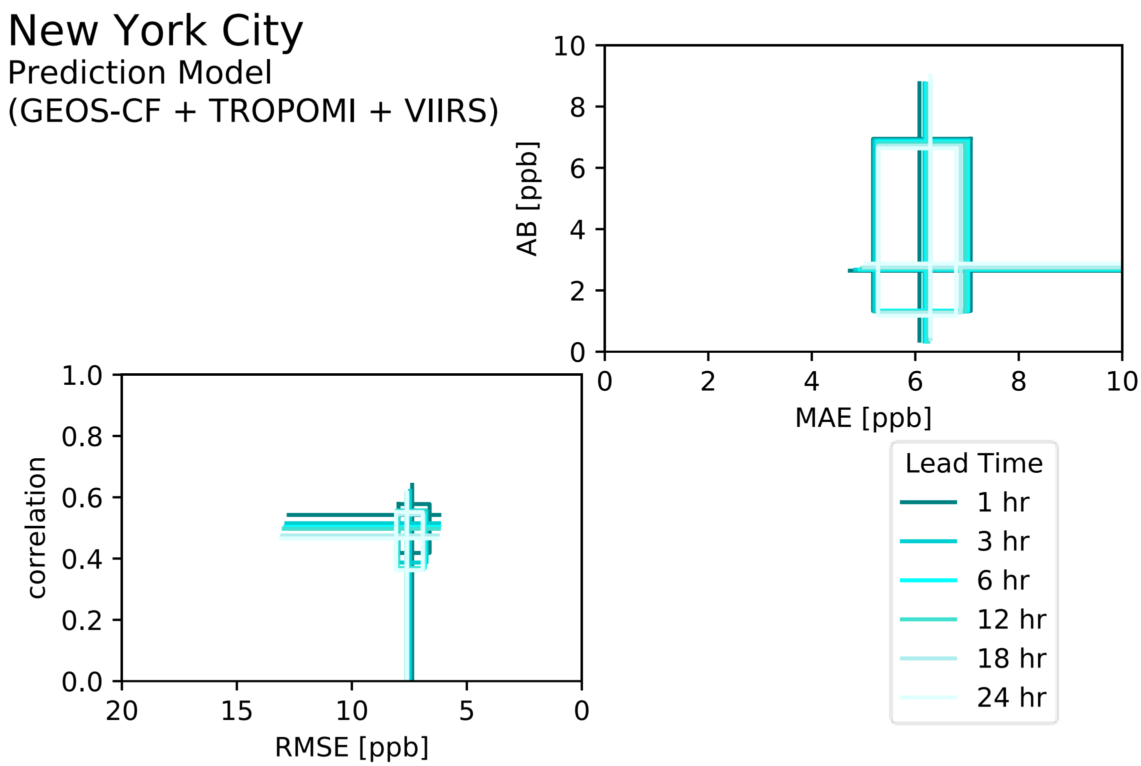

Figure S28. Performance of the proposed method using GEOS-CF, TROPOMI, and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

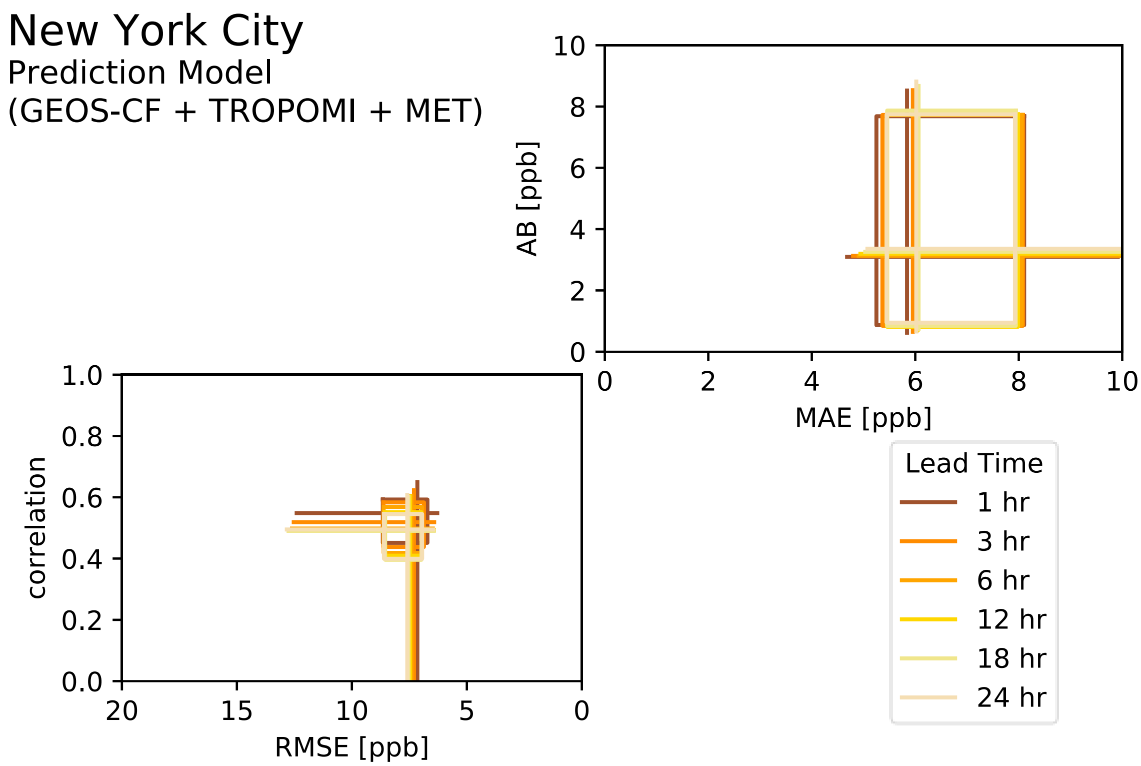

Figure S29. Performance of the proposed method using GEOS-CF, TROPOMI, and meteorological information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

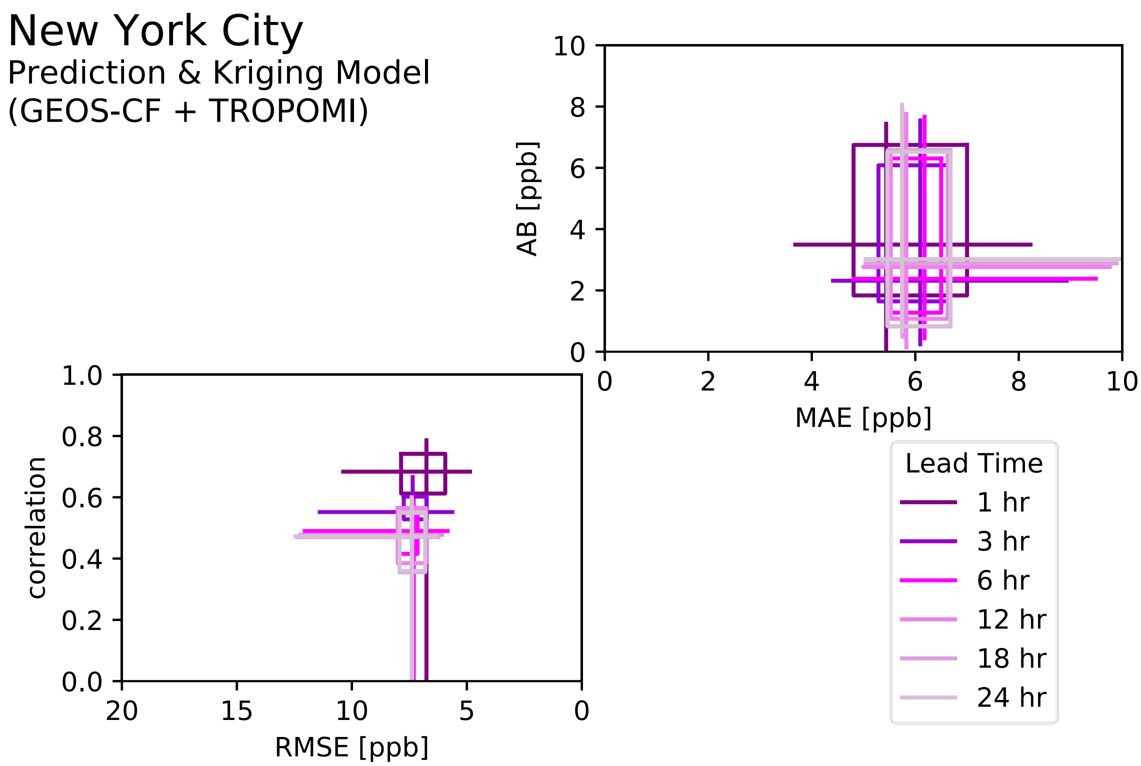

Figure S30. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data, together with a final updating (kriging) based on correlations to the latest available ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

S2.2.2 New Orleans

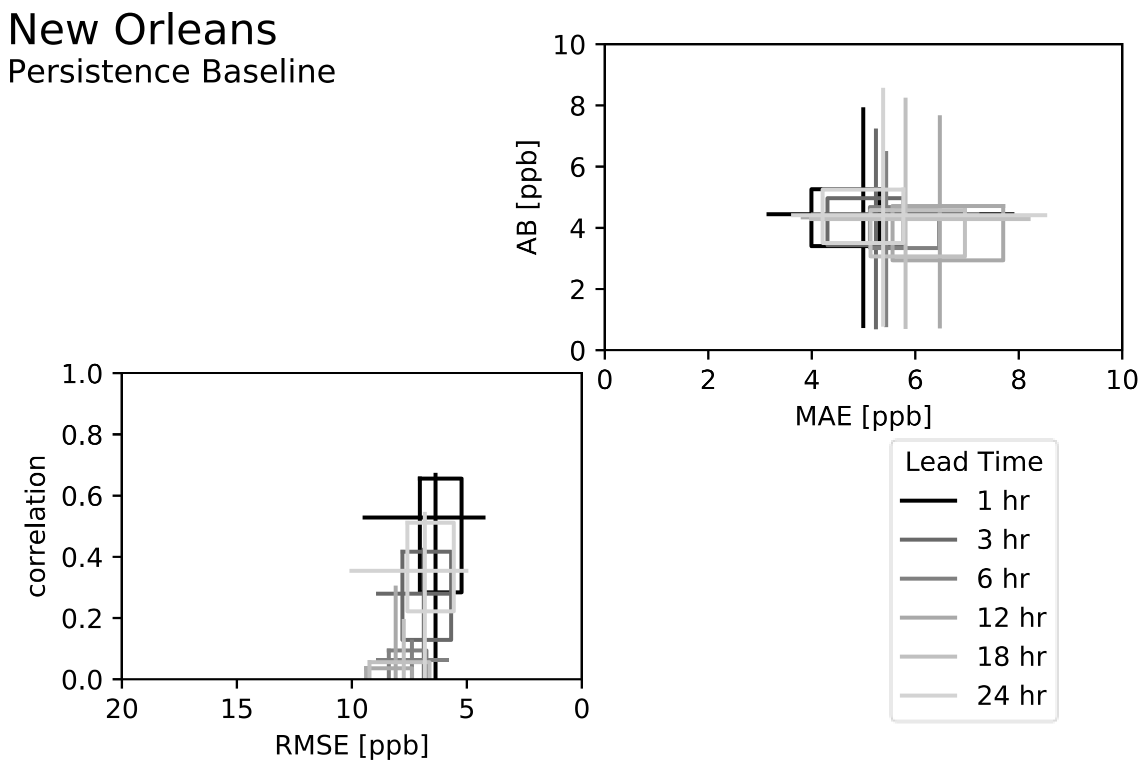

Figure S31. Performance of the persistence baseline method.

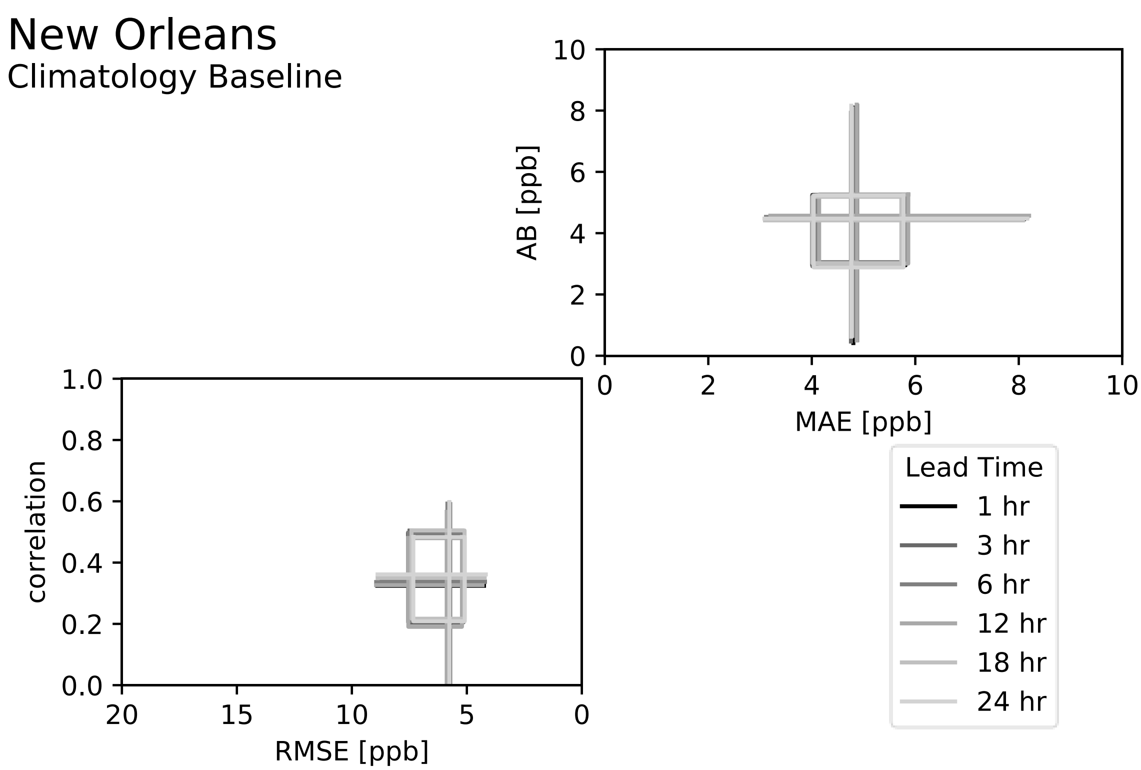

Figure S32. Performance of the climatology baseline method.

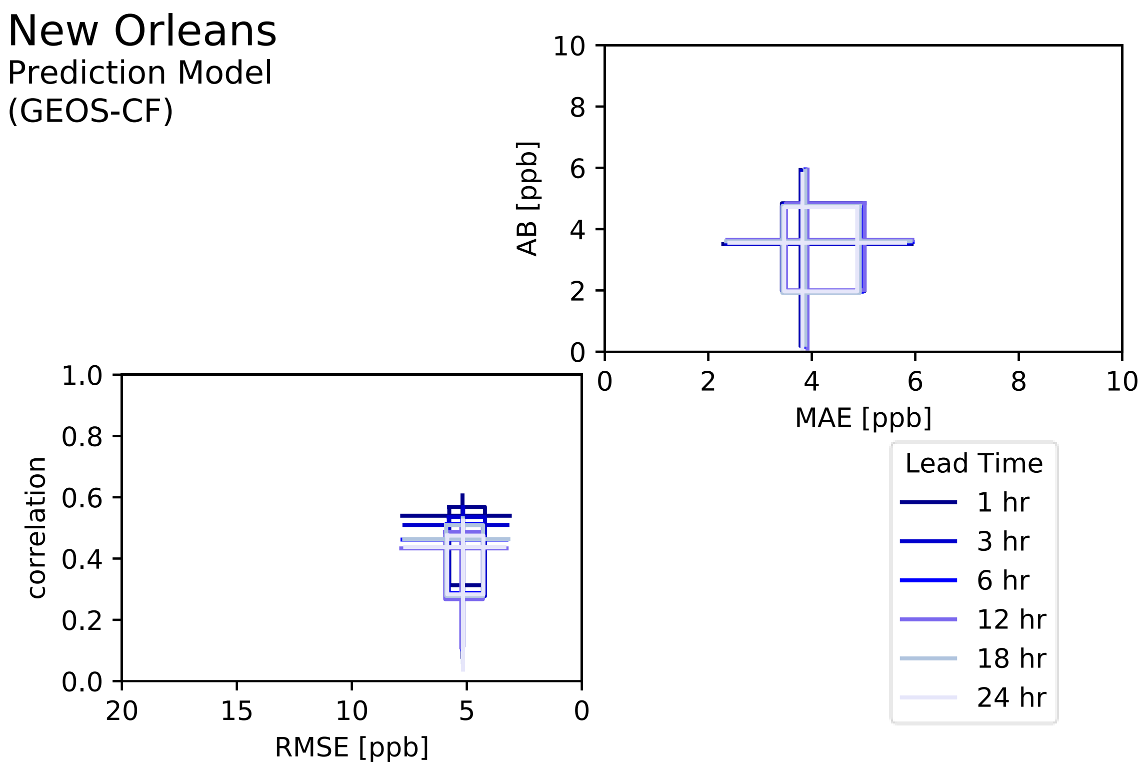

Figure S33. Performance of the proposed method using GEOS-CF and ground information only (downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

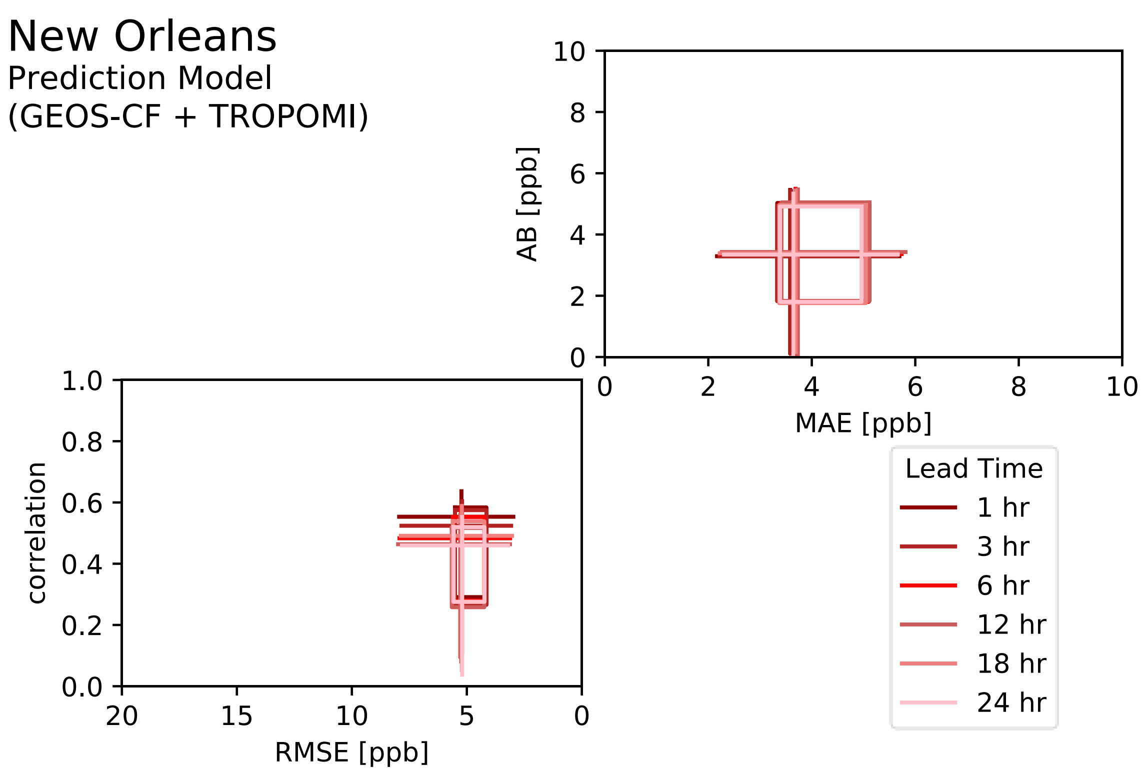

Figure S34. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern

combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

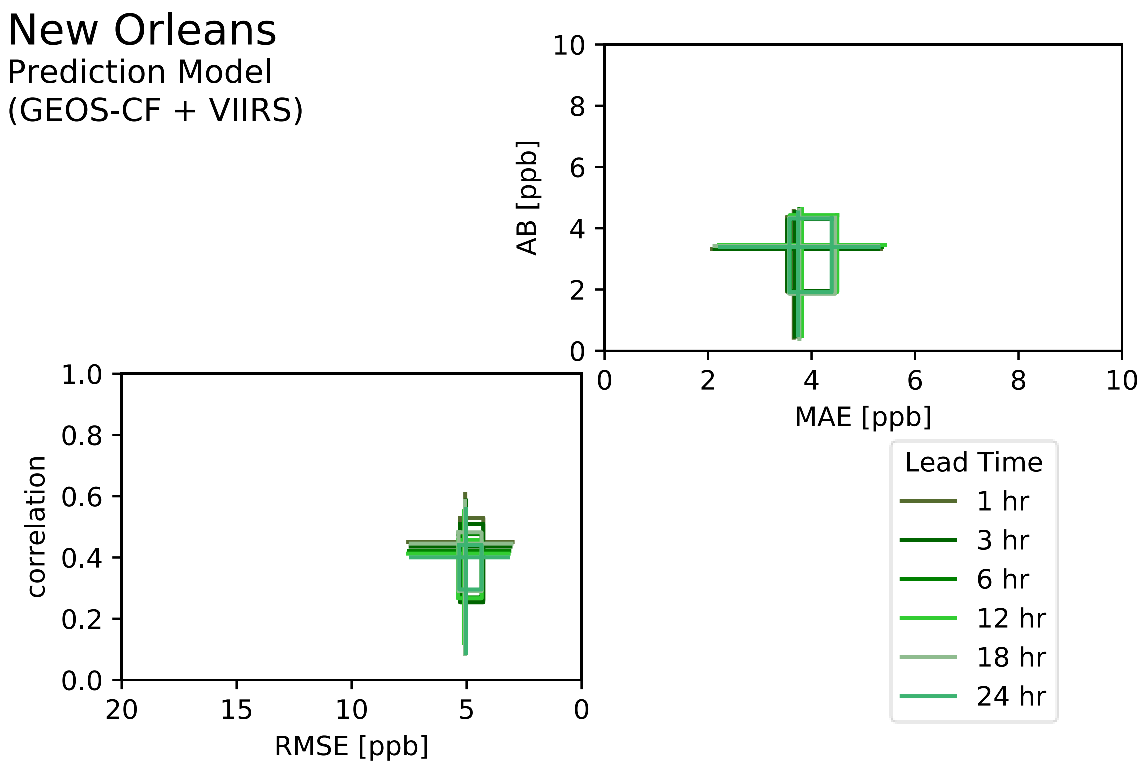

Figure S35. Performance of the proposed method using GEOS-CF and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

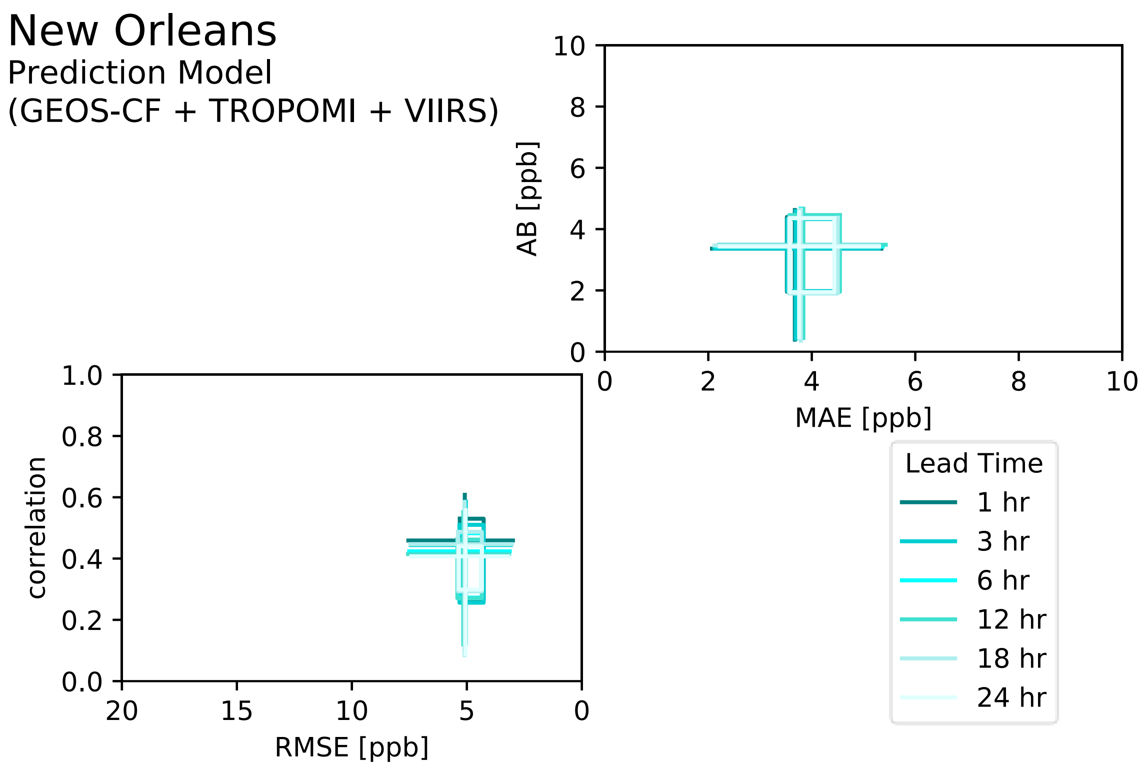

Figure S36. Performance of the proposed method using GEOS-CF, TROPOMI, and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

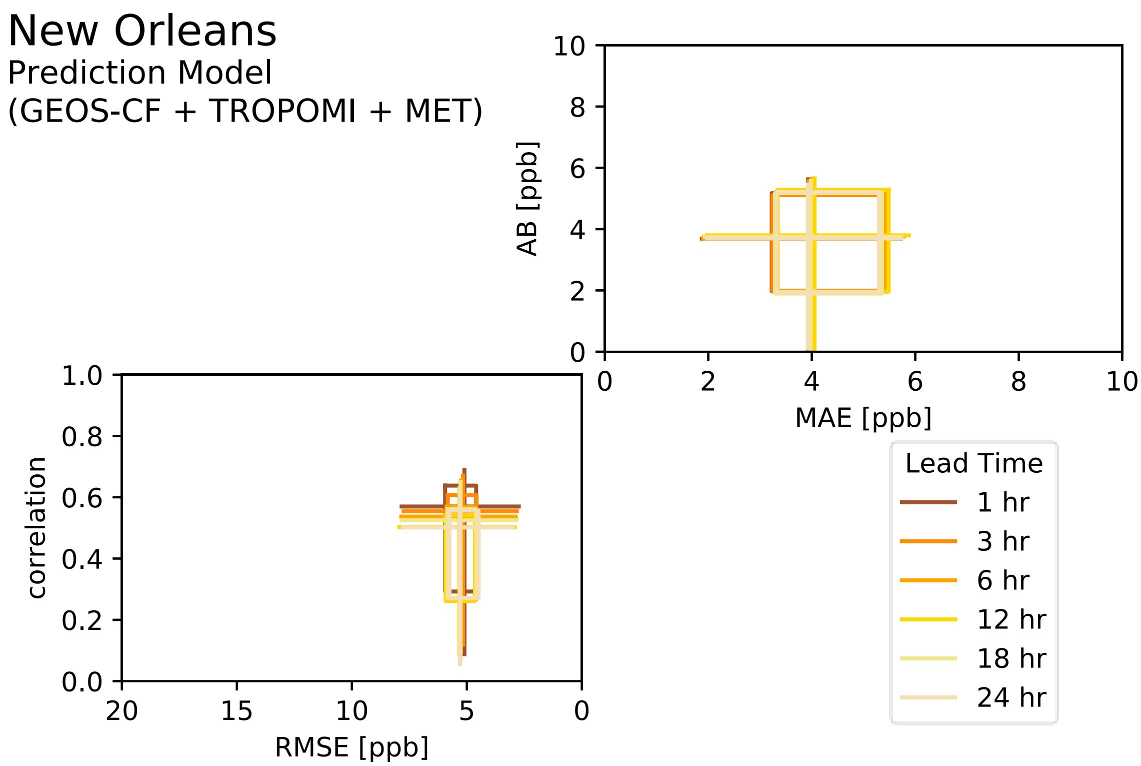

Figure S37. Performance of the proposed method using GEOS-CF, TROPOMI, and meteorological information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

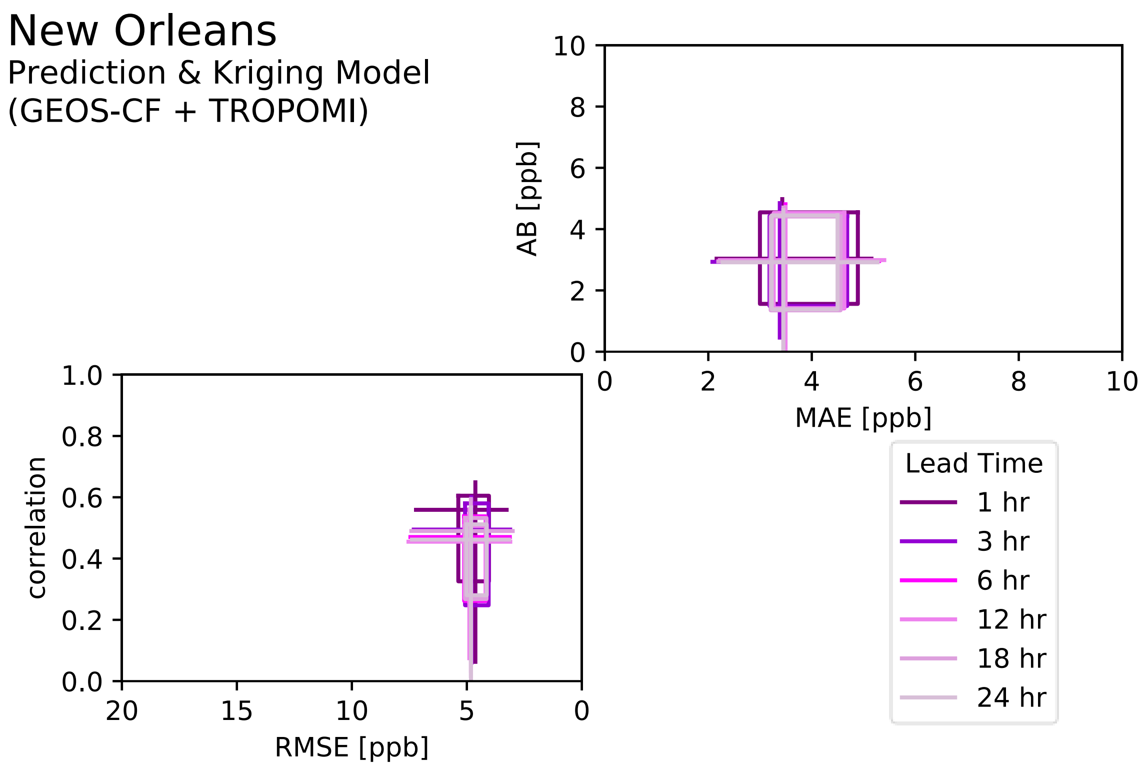

Figure S38. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data, together with a final updating (kriging) based on correlations to the latest available ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

S2.2.3 Las Vegas

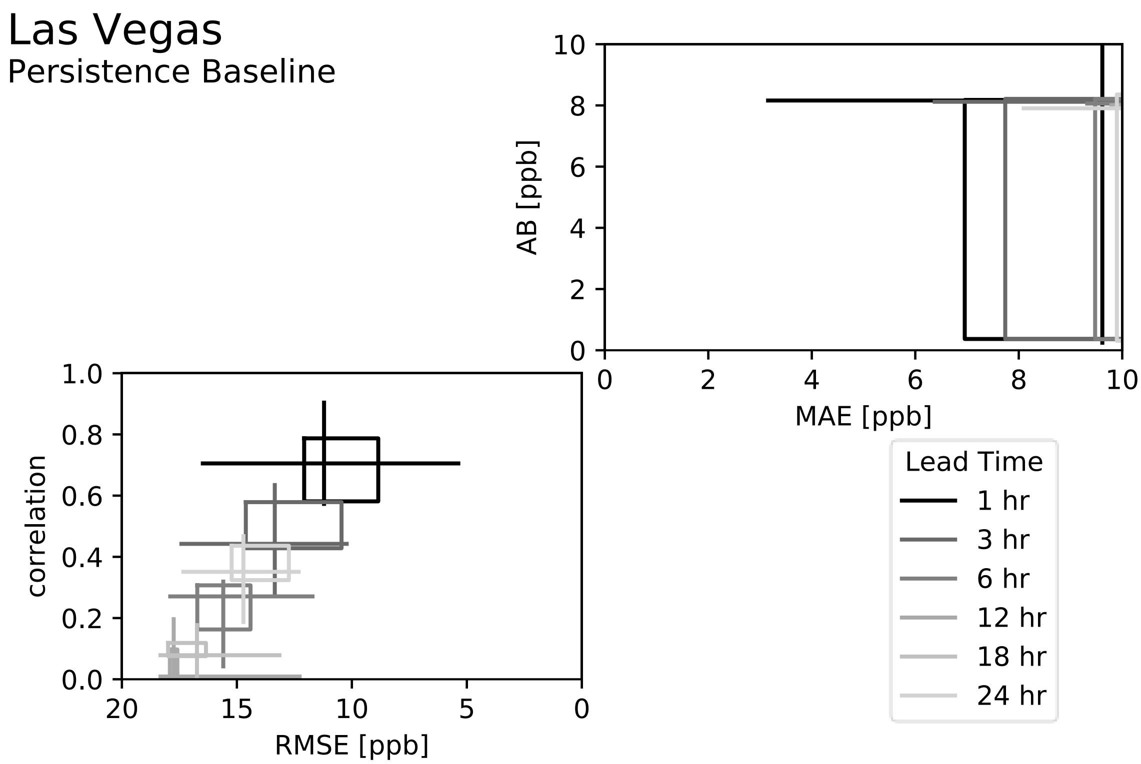

Figure S39. Performance of the persistence baseline method.

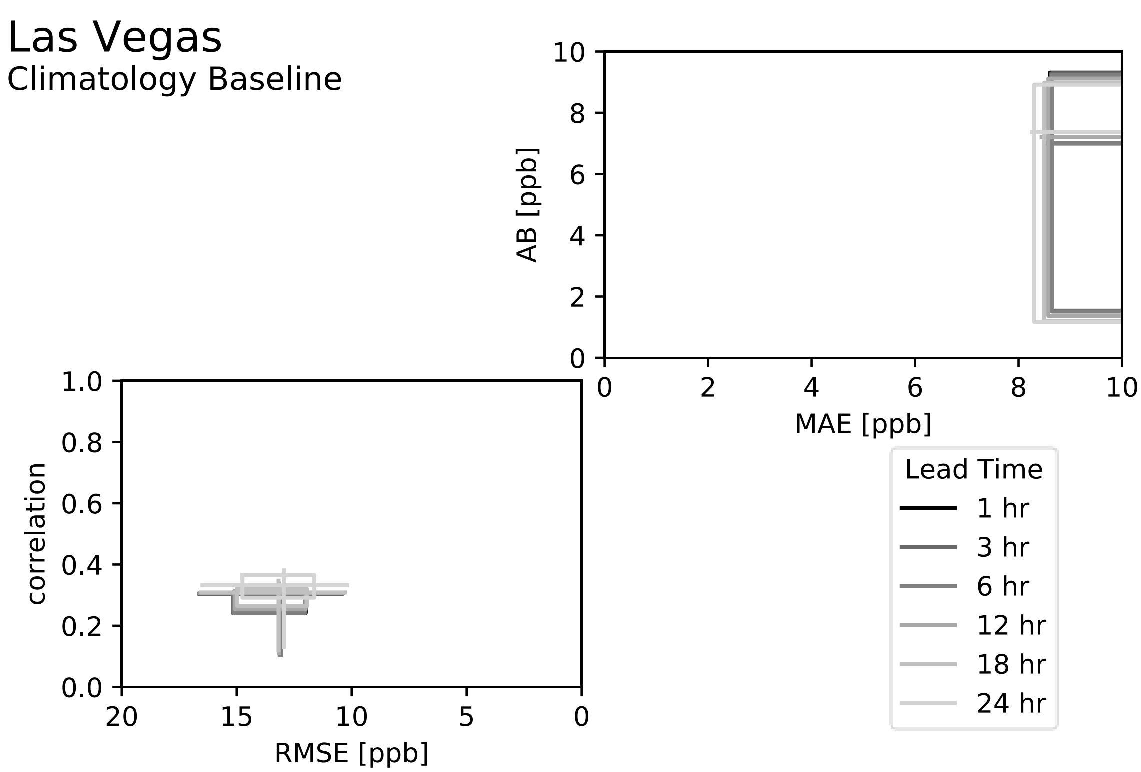

Figure S40. Performance of the climatology baseline method.

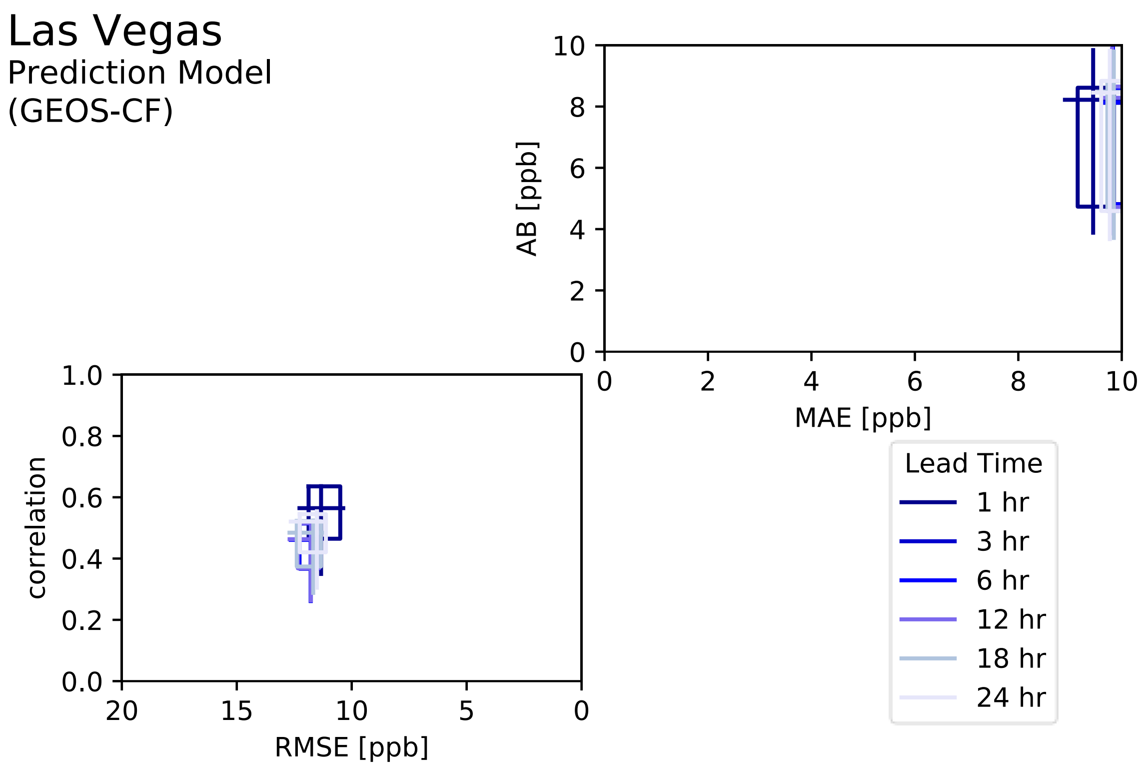

Figure S41. Performance of the proposed method using GEOS-CF and ground information only (downscaling via linear interpolation, time-of-day regression weighting).

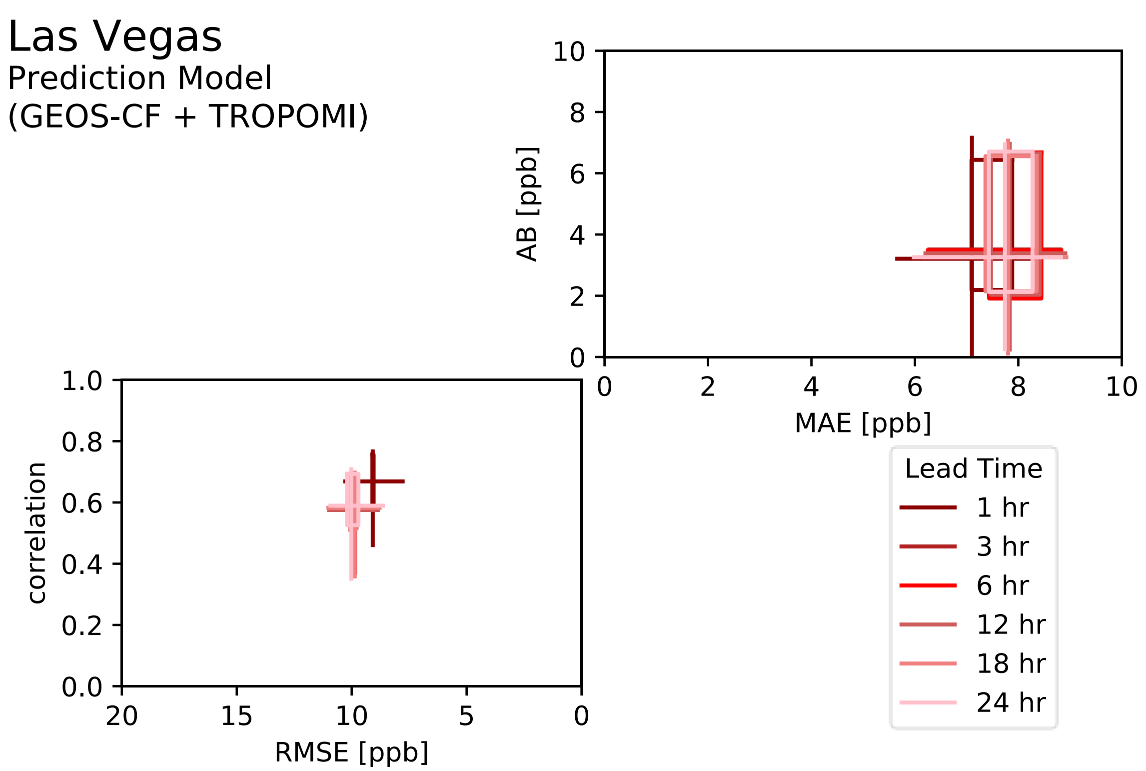

Figure S42. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data (pattern extraction from full calibration period as in Equation 2, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

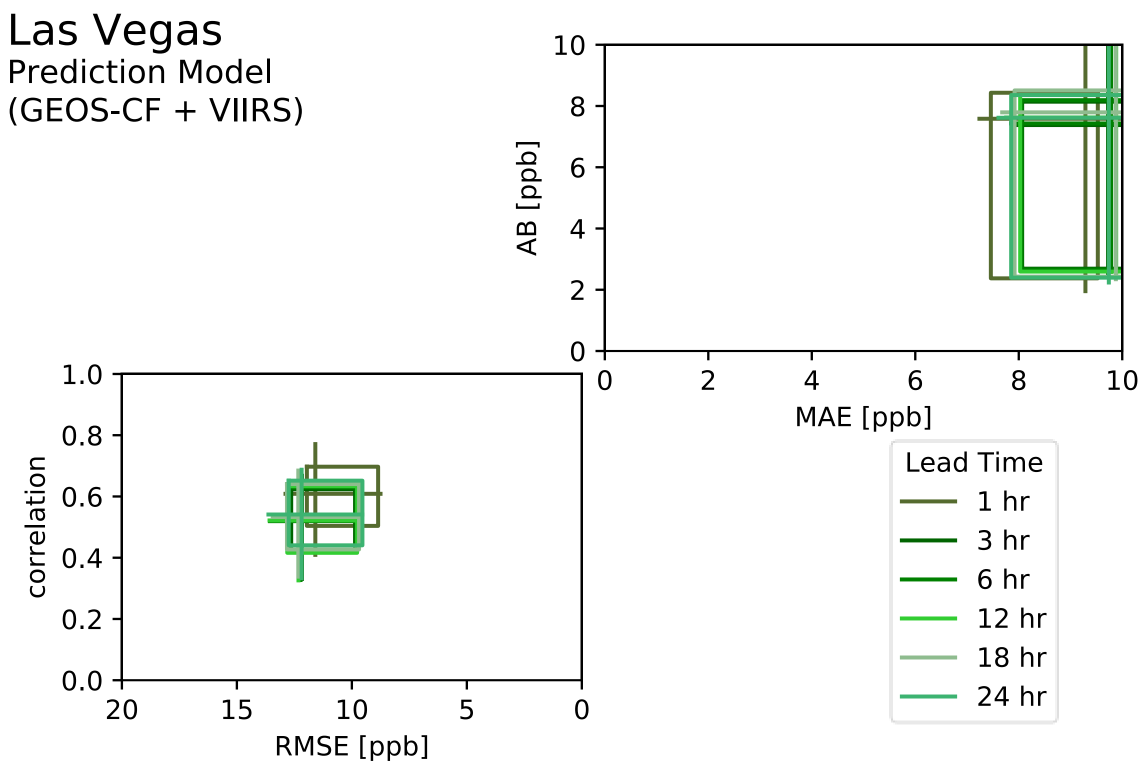

Figure S43. Performance of the proposed method using GEOS-CF and VIIRS information together with ground data (pattern extraction from full calibration period as in Equation 2, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

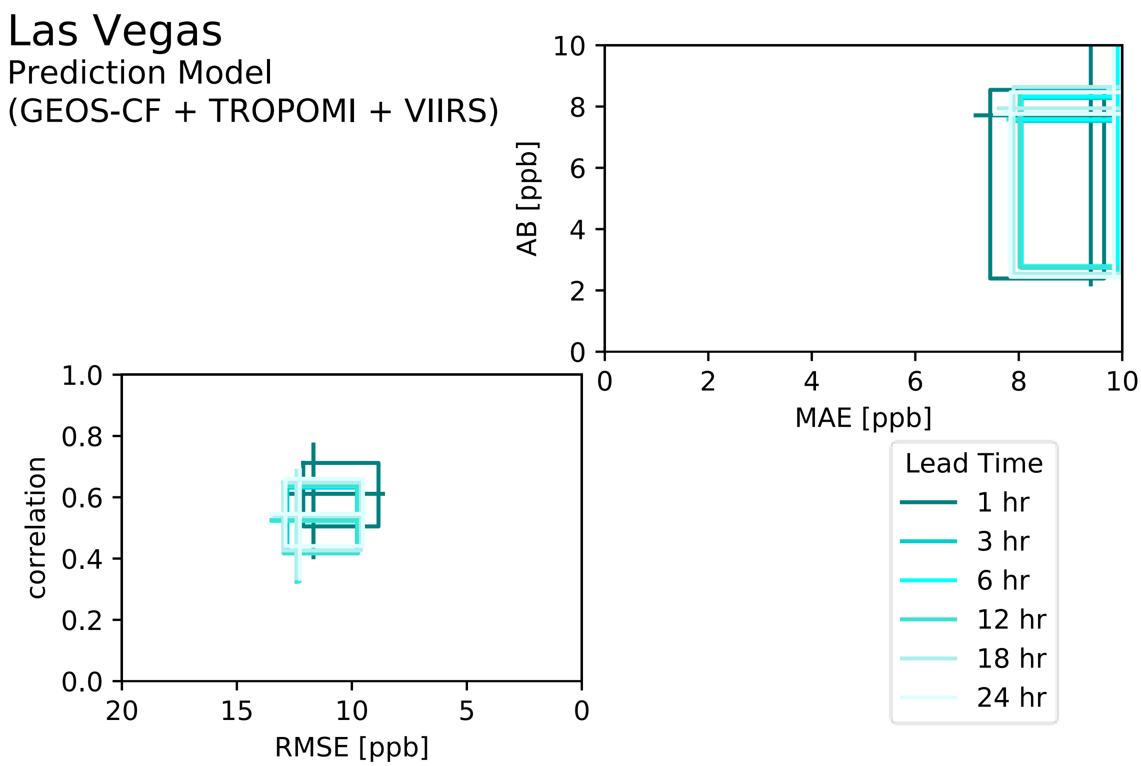

Figure S44. Performance of the proposed method using GEOS-CF, TROPOMI, and VIIRS information together with ground data (pattern extraction from full calibration period as in Equation 2, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

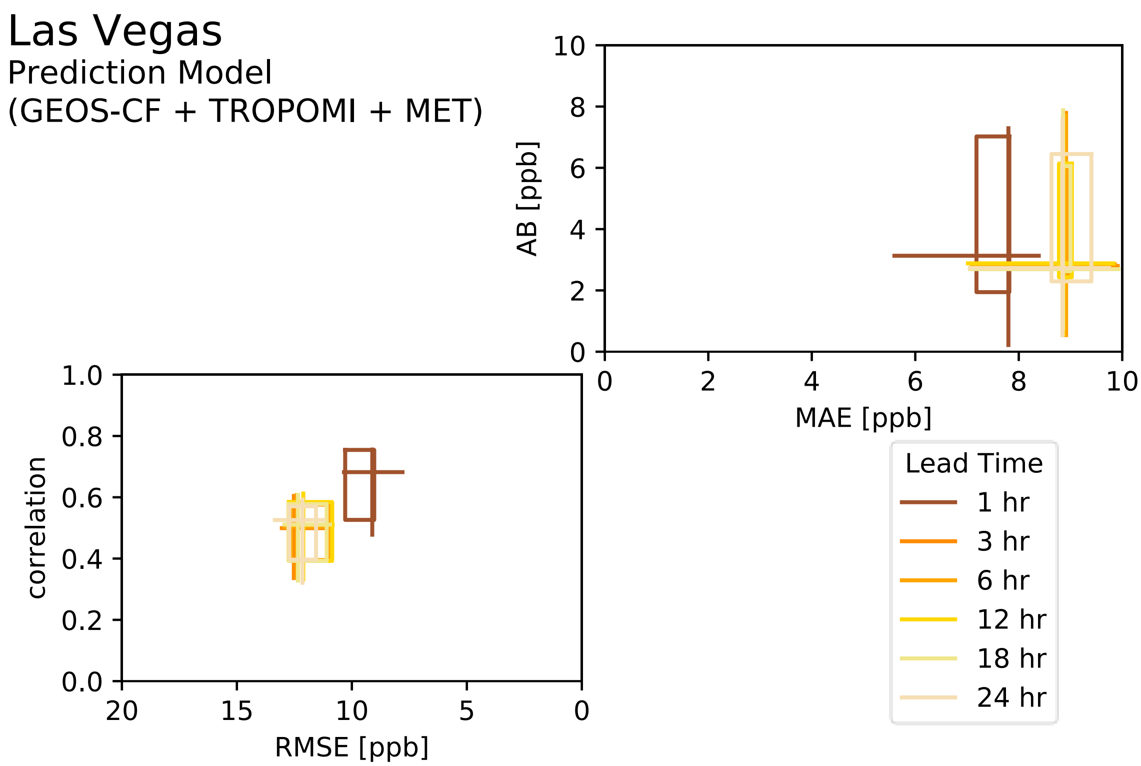

Figure S45. Performance of the proposed method using GEOS-CF, TROPOMI, and meteorological information together with ground data (pattern extraction from full calibration period as in Equation 2, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

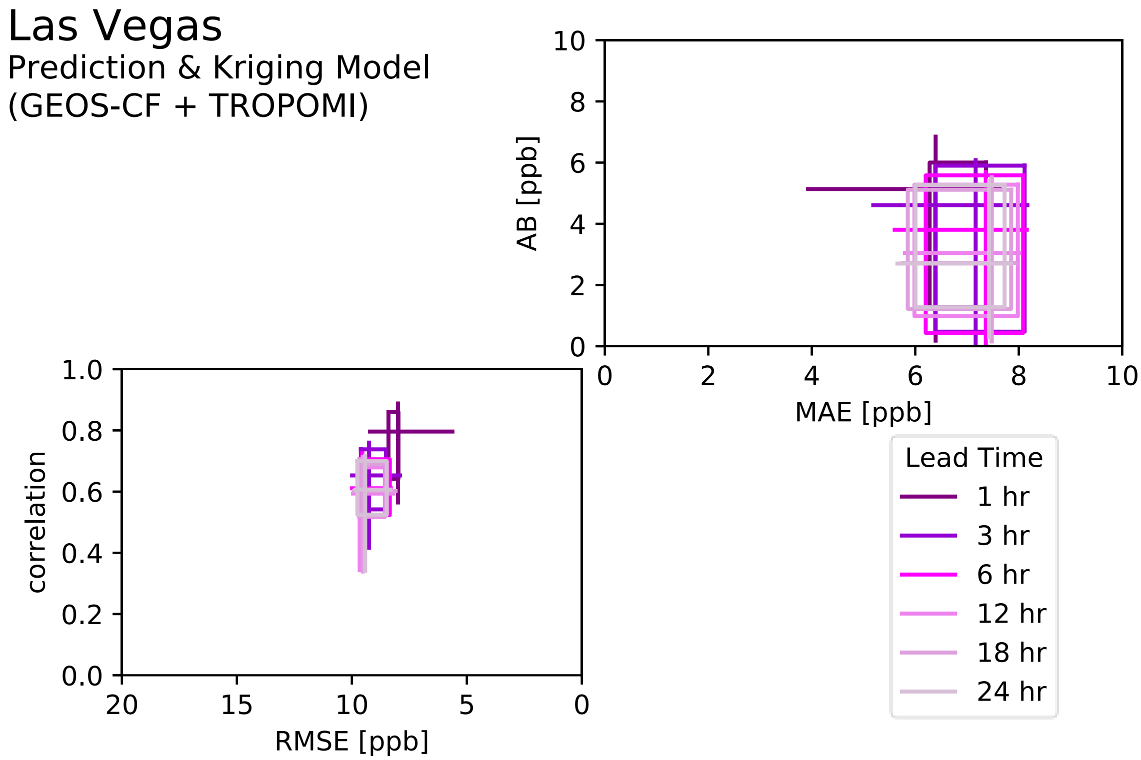

Figure S46. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data, together with a final updating (kriging) based on correlations to the latest available ground data (pattern extraction from full calibration period as in Equation 2, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

S2.2.4 Salt Lake City

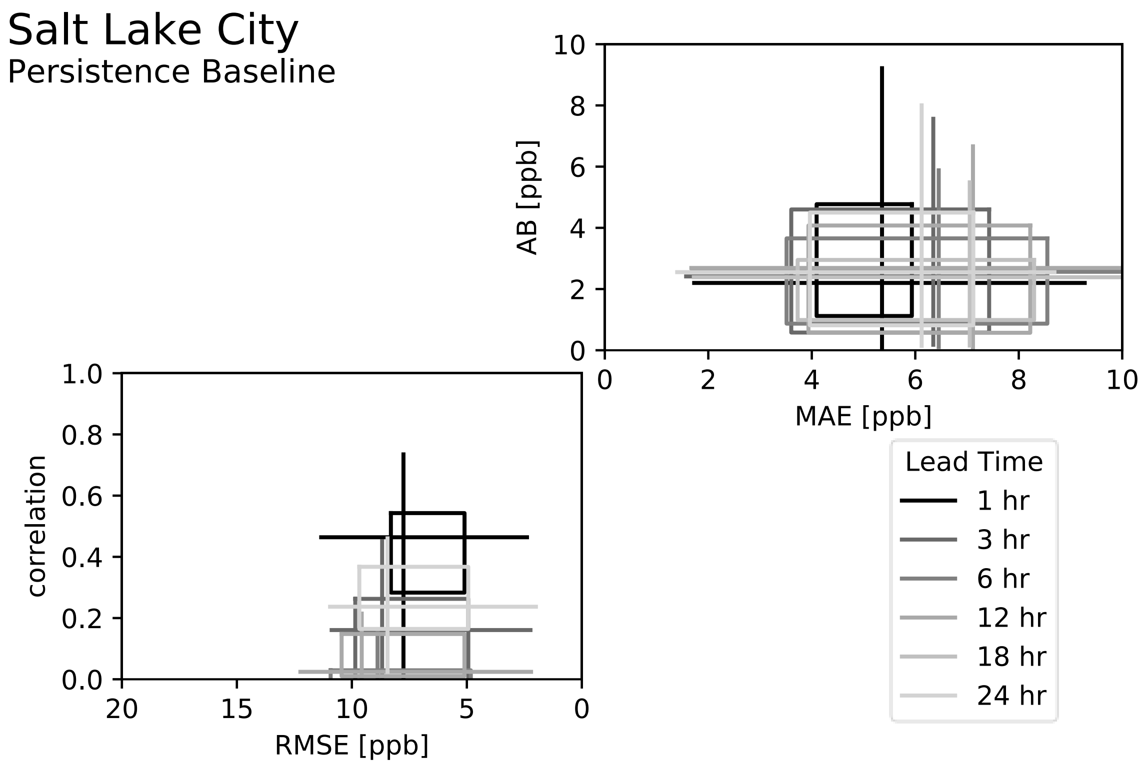

Figure S47. Performance of the persistence baseline method.

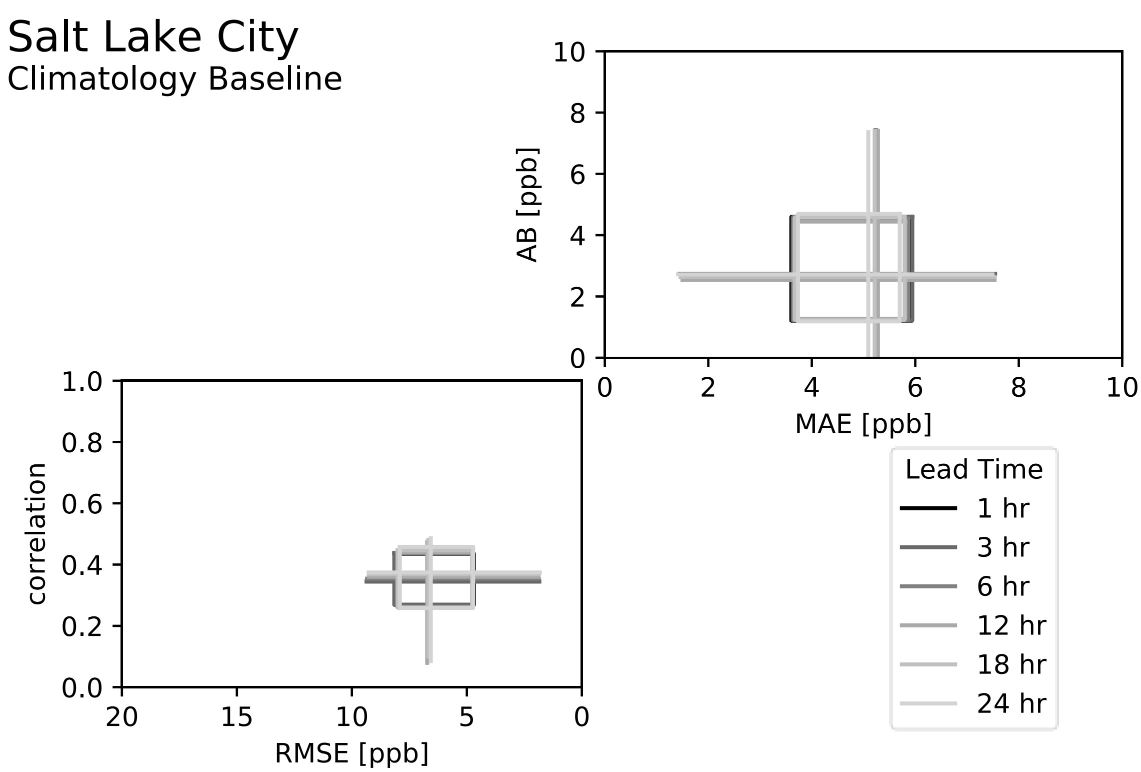

Figure S48. Performance of the climatology baseline method.

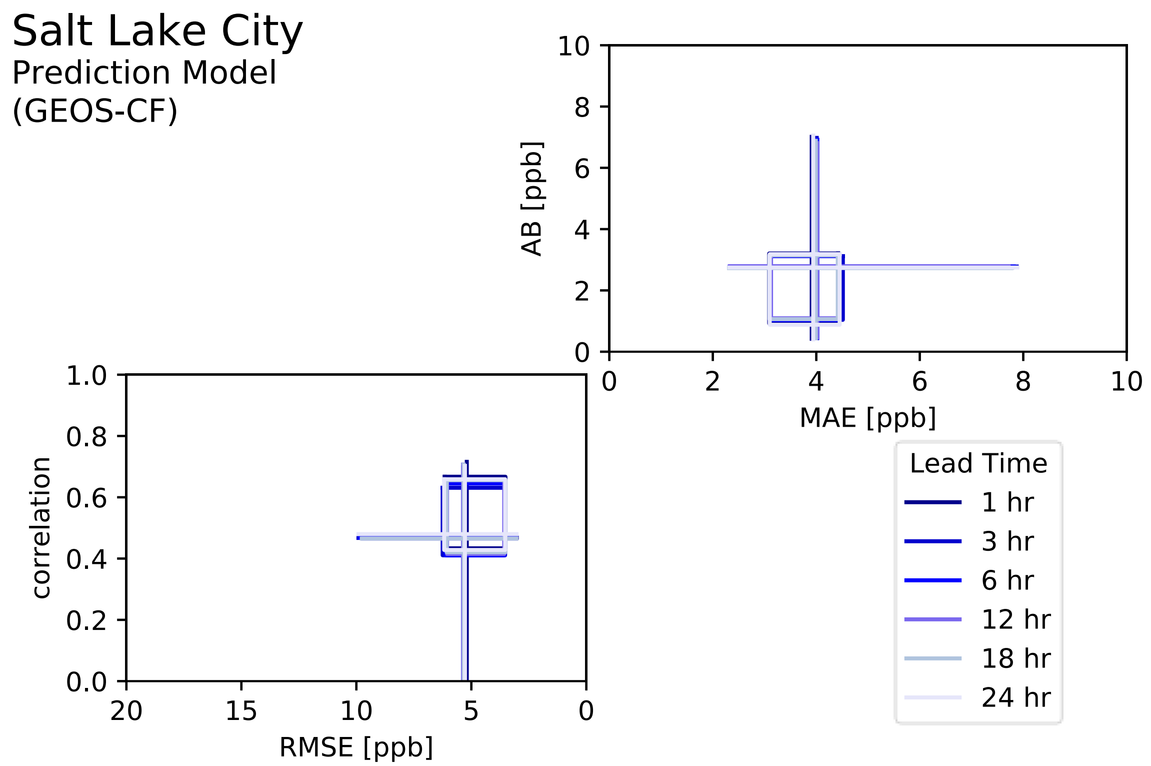

Figure S49. Performance of the proposed method using GEOS-CF and ground information only (downscaling via linear interpolation, time-of-day regression weighting).

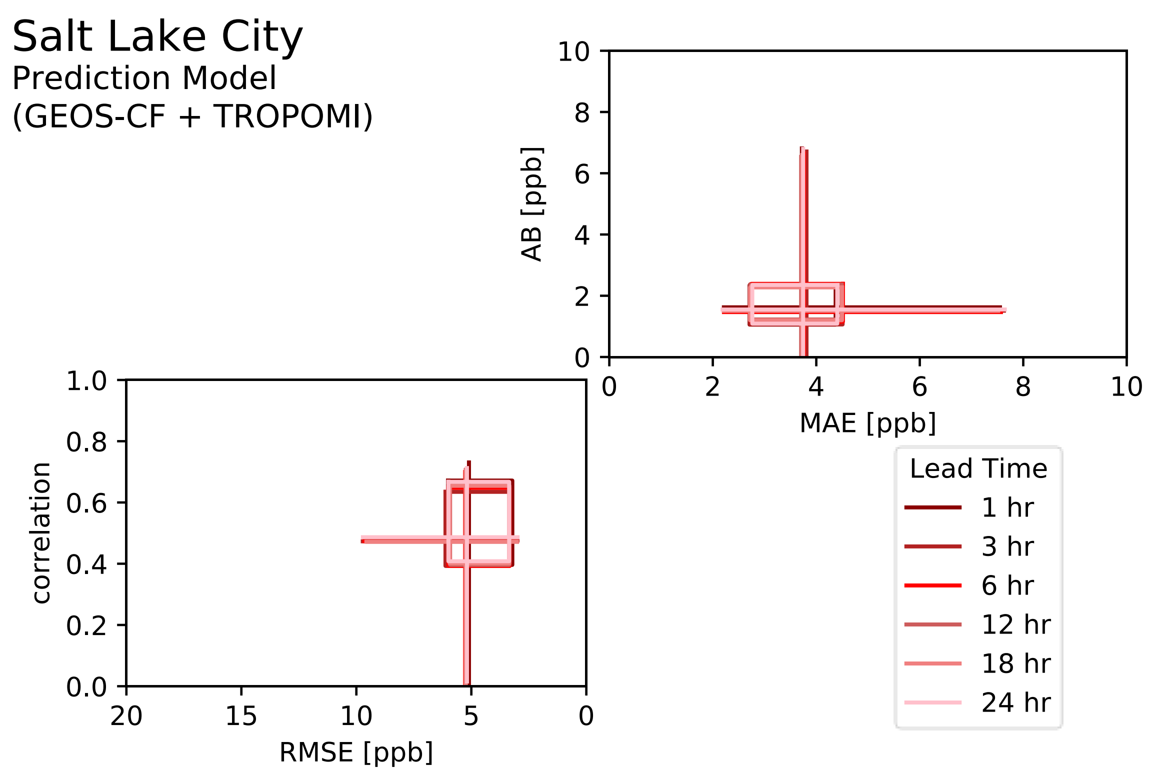

Figure S50. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern

combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

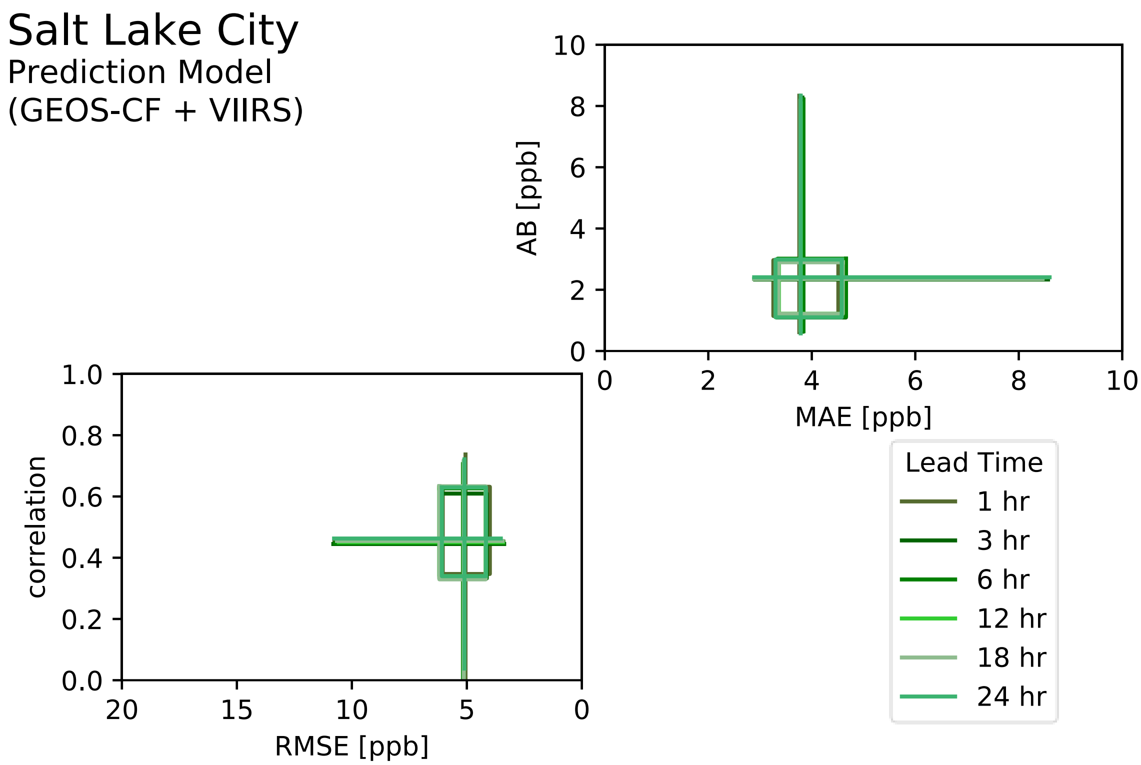

Figure S51. Performance of the proposed method using GEOS-CF and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

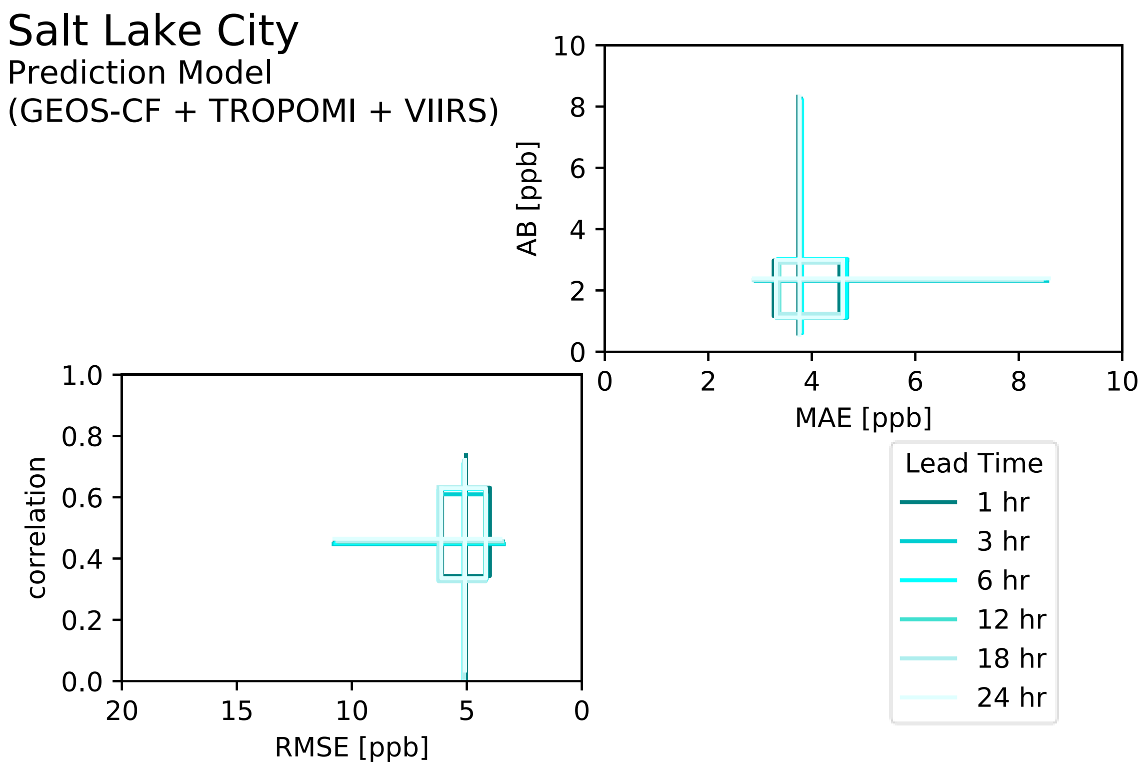

Figure S52. Performance of the proposed method using GEOS-CF, TROPOMI, and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

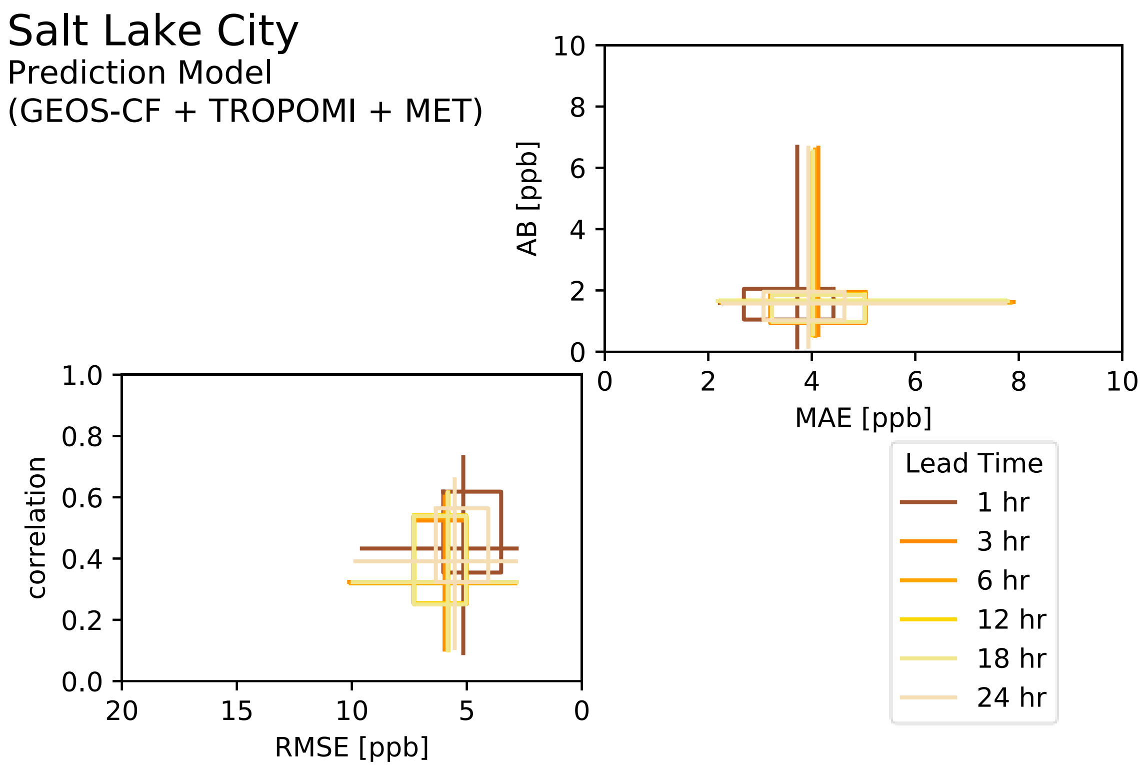

Figure S53. Performance of the proposed method using GEOS-CF, TROPOMI, and meteorological information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

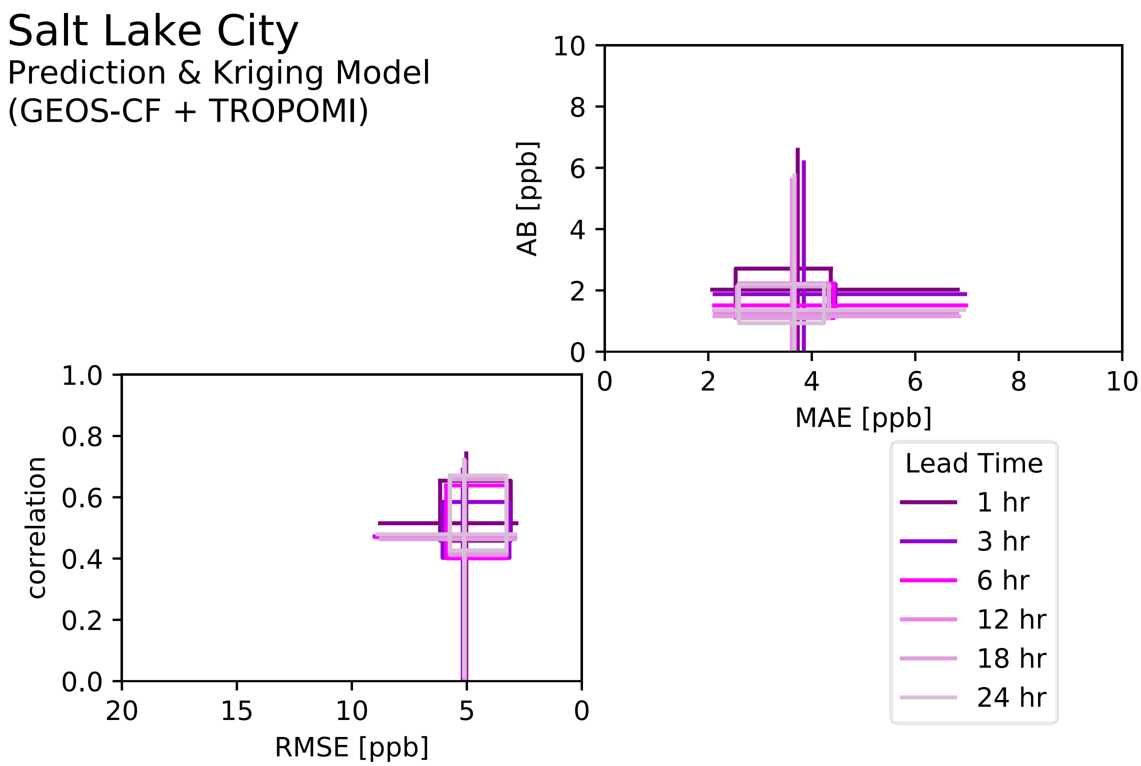

Figure S54. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data, together with a final updating (kriging) based on correlations to the latest available ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, time-of-day regression weighting).

S2.2.5 San Francisco

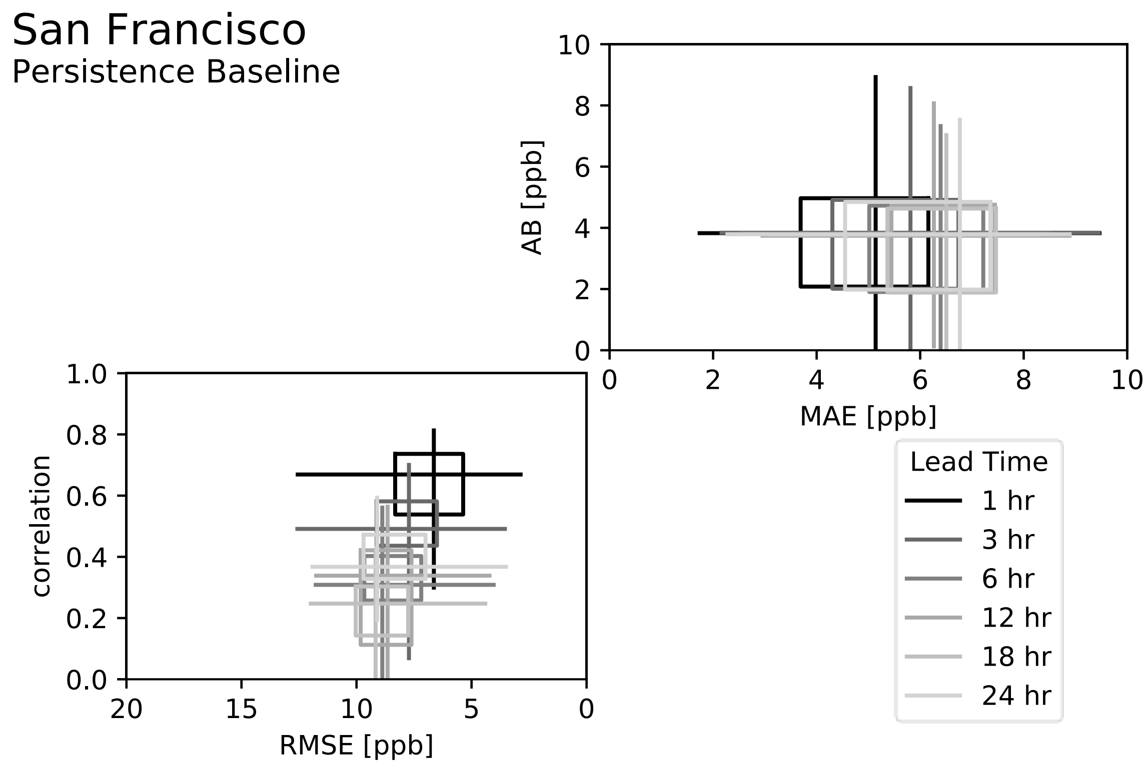

Figure S55. Performance of the persistence baseline method.

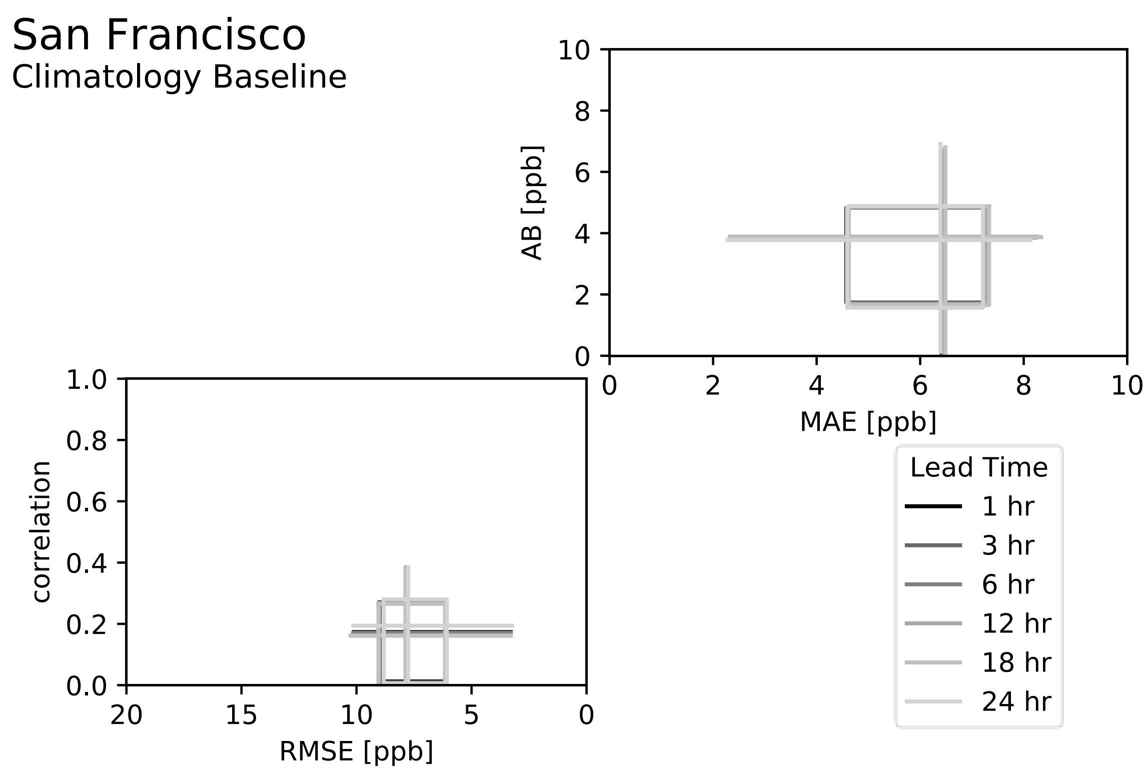

Figure S56. Performance of the climatology baseline method.

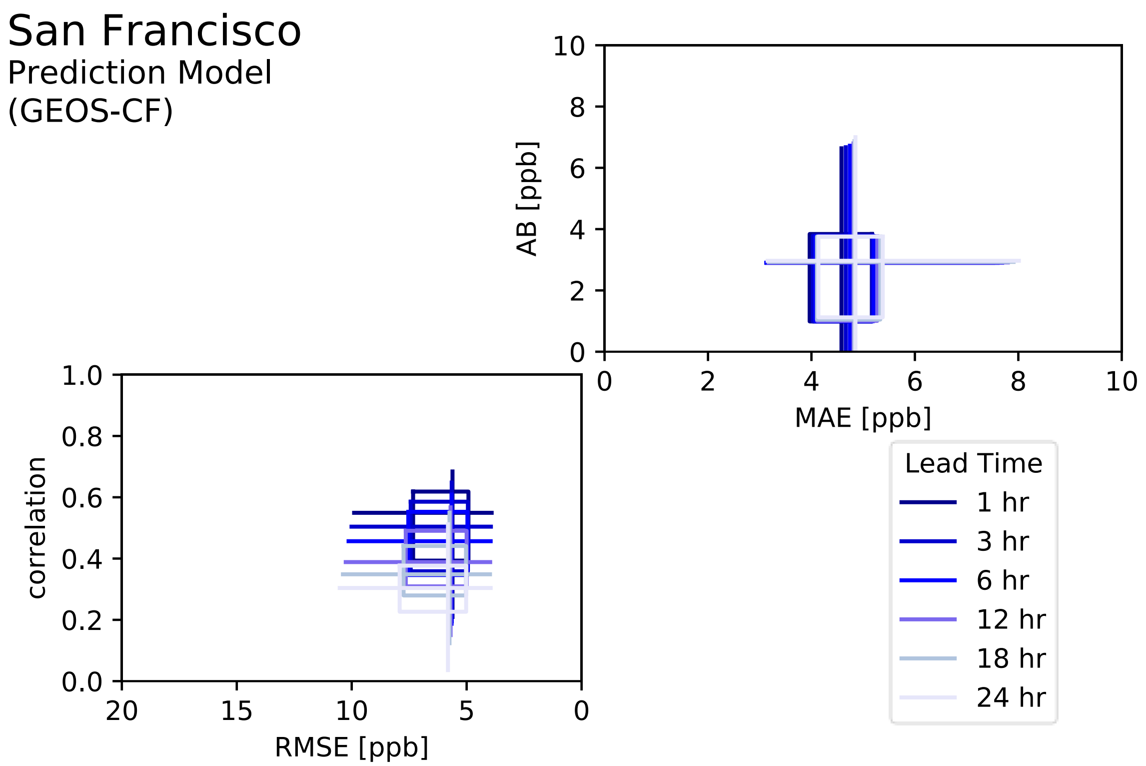

Figure S57. Performance of the proposed method using GEOS-CF and ground information only (downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

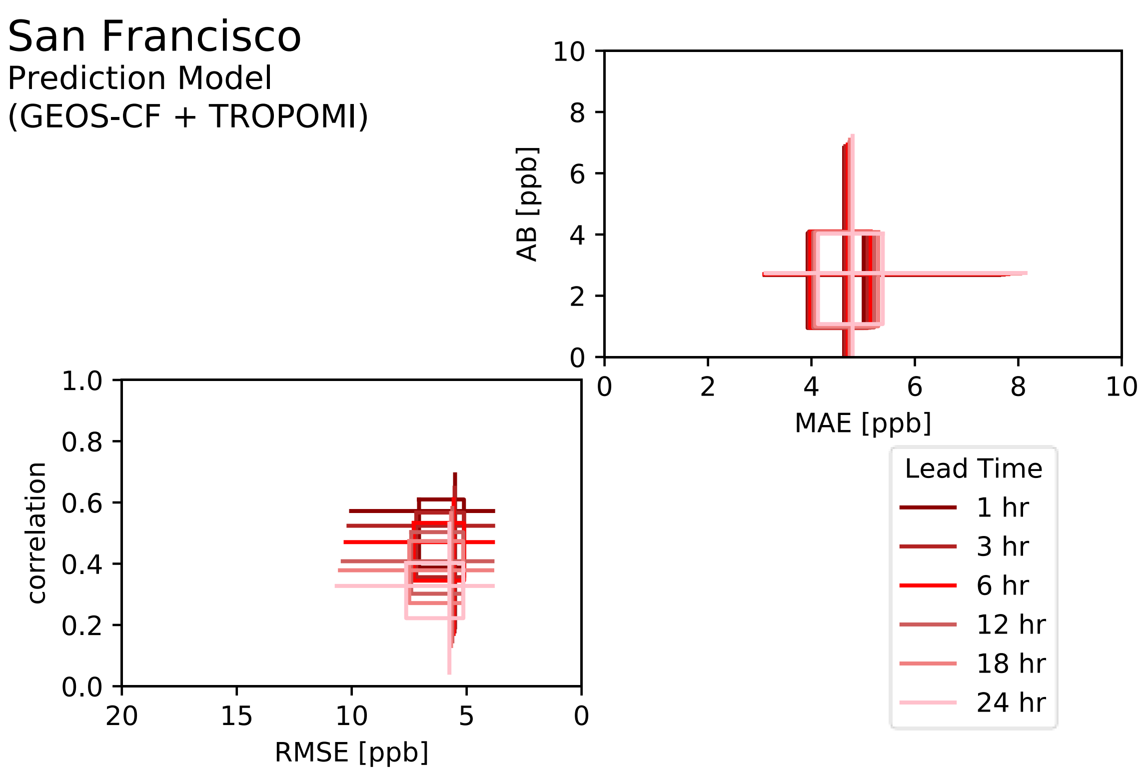

Figure S58. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern

combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

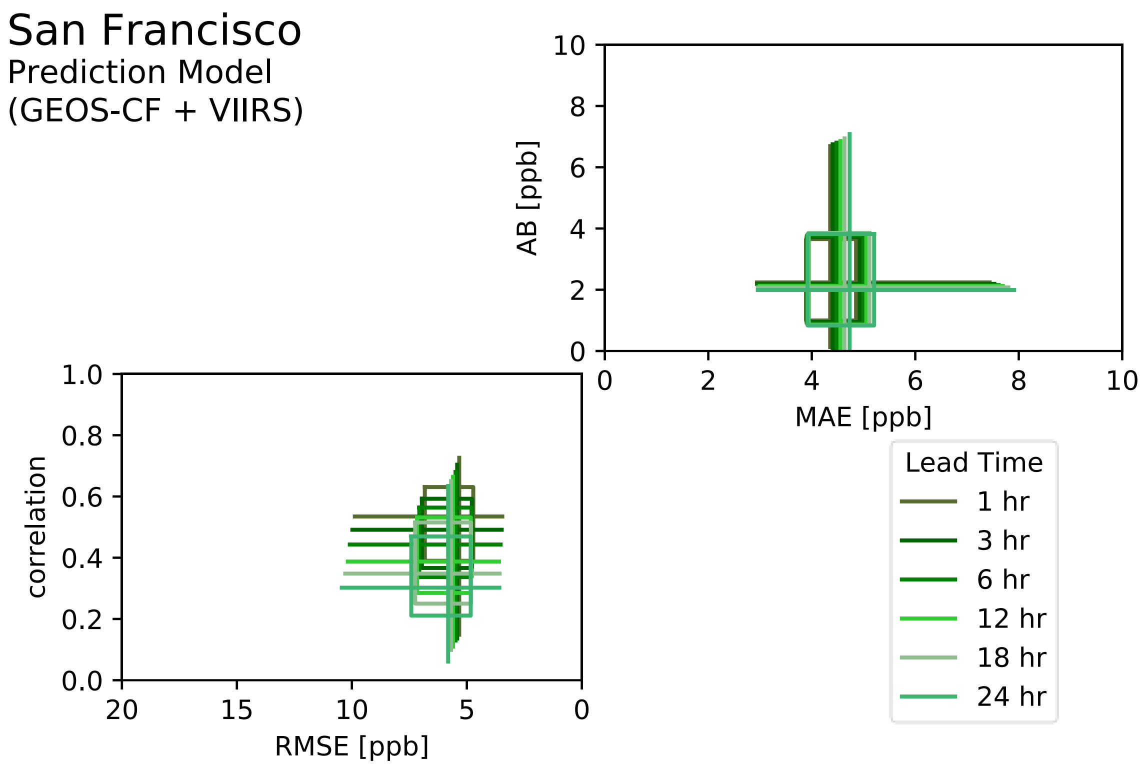

Figure S59. Performance of the proposed method using GEOS-CF and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

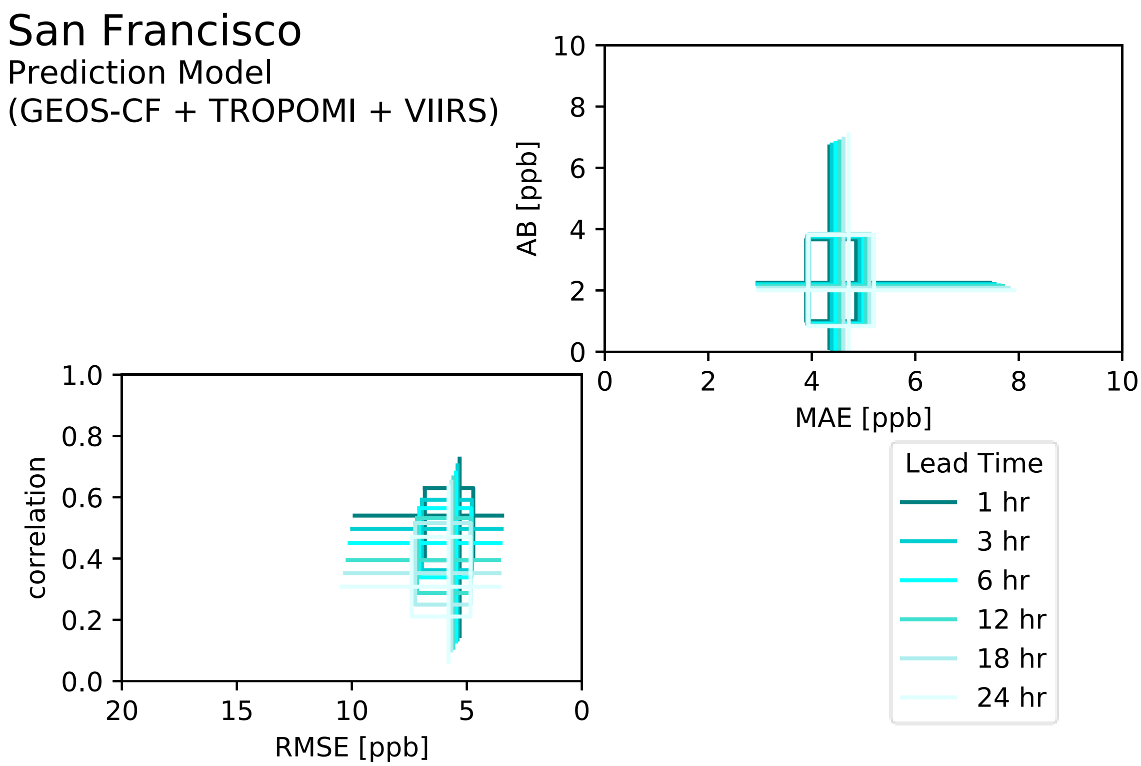

Figure S60. Performance of the proposed method using GEOS-CF, TROPOMI, and VIIRS information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

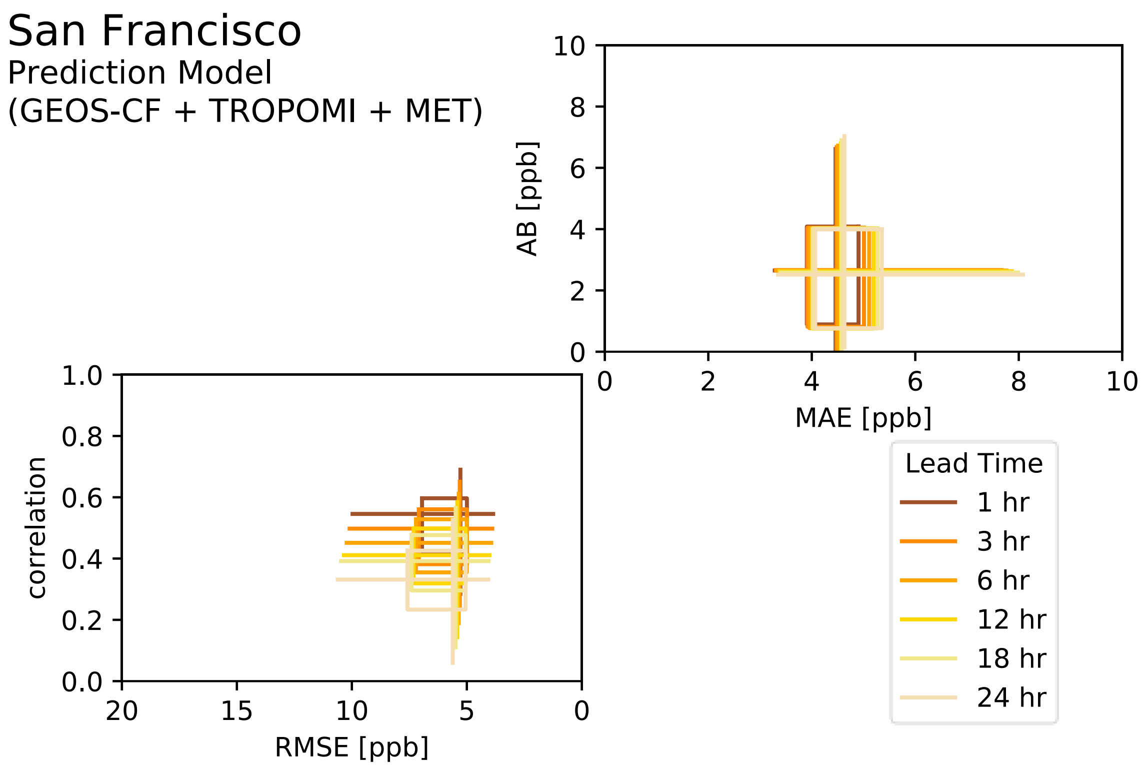

Figure S61. Performance of the proposed method using GEOS-CF, TROPOMI, and meteorological information together with ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).

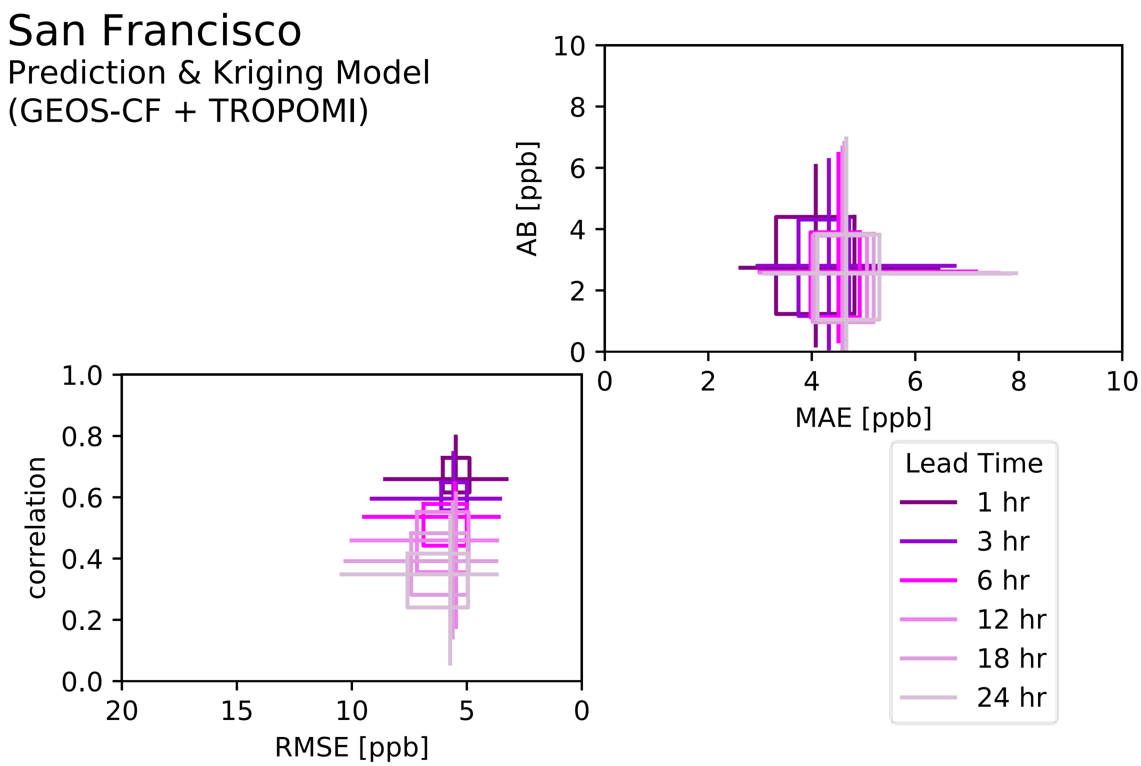

Figure S62. Performance of the proposed method using GEOS-CF and TROPOMI information together with ground data, together with a final updating (kriging) based on correlations to the latest available ground data (pattern extraction at satellite overpass times as in Equation 3, pattern combination via addition as in Equation 6, downscaling via linear interpolation, decaying periodic regression weighting as in Equation 11).