

BAYESIAN MAXIMUM ENTROPY INTEGRATION OF OZONE OBSERVATIONS AND AIR
QUALITY MODEL PREDICTIONS FOR IMPROVED EXPOSURE ESTIMATES

Yadong Xu

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Environmental Sciences and Engineering in the Gillings School of Global Public Health.

Chapel Hill
2016

Approved by:

William Vizuete

Marc Serre

Sarav Arunachalam

Jason West

Karin Yeatts

© 2016
Yadong Xu
ALL RIGHTS RESERVED

ABSTRACT

Yadong Xu. Bayesian Maximum Entropy Integration of Ozone Observations and Air Quality Model Predictions for Improved Exposure Estimates
(Under the direction of William Vizuite and Marc Serre)

To support the Women’s Health Initiative (WHI) Memory Study (WHIMS), a nationwide cohort study, accurate ozone exposure estimates for ambient concentrations needed to be generated at a national scale for years 1993-2010. For this large spatial and temporal coverage we investigated different geo-statistical approaches to generate estimates that integrate routine monitoring from surface ozone observations and episodic chemical transport model (CTM) outputs. The goal is to take advantage of the accuracy of the observational data and the continuous spatial/temporal coverage of CTM model outputs.

In this work, we demonstrate a Bayesian Maximum Entropy (BME) data integration geo-statistical approach for making national scale ozone estimates that models the non-linear and non-homoscedastic relation between air pollution observations and CTM predictions. This is the first application of BME that fully accounts for variability in CTM model performance through our novel Regionalized Air Quality Model Performance (RAMP) approach. A validation analysis was completed using only non-collocated data outside of a validation radius r_v and the error statistics between observations and re-estimated values were obtained. We show that by accounting for the spatial and temporal variability in model performance there is 3-12 fold increase in R^2 (the squared Pearson correlation coefficient) percentage change for the daily ozone concentrations compared to estimates that assume model performance does not change across space and time.

Our second project is to investigate the differences of the predictive capacity for two upscaling methods: USM1 (data aggregation from hourly to daily followed by BME approach estimation) and USM2 (perform BME approach estimation on hourly ozone followed by data aggregation). We found that the less

computationally intensive method USM1 outperforms the method USM2. This highlights the capability of the RAMP approach that was able to capture the spatial temporal variability in CTM model performance at time scale of interest. Thus, we recommend to use upscaling method USM1 to integrate CTM model predictions through RAMP approach because USM1 can achieve higher estimation accuracy and also associated with much lower computational cost.

To my mentors and friends, I couldn't have done this without you.
Thank you for all of your support along this wonderful journey.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Dr. William Vizueté and Dr. Marc Serre for their guidance and mentorship. I greatly appreciate their trust of giving me the opportunity to work on the research project for the Women's Health Initiative (WHI) memory study. Their knowledge and expertise facilitated successful model development and analysis. Without their help, we could not have finished the WHIMS project successfully. I also would like to thank my committee member, Dr. Sarav Arunachalam, Dr. Jason West and Dr. Karin Yeatts for their feedback and helpful suggestions.

I would like to express my deepest gratitude and appreciations to my husband Morris and my son Haoran for their continuous support and encouragement during this long journey. Thank them for believing in me and providing a quiet and pleasant environment at home during long hours of model development and manuscript writing. I am also deeply grateful to my extended family and friends for their support and care.

I am thankful for the support I received from my department's administrative staff, particularly Jack Whaley, who always answered my questions promptly and provided great help in my class registration and academic development.

I am also thankful to my research classmates/colleagues in the CHAQ/MAQ/BME labs for the great learning environment we shared. I am especially grateful to Pradeepa Vennam and Matt Woody for their help in regards to CMAQ/CAMx post analysis tools. I am thankful to Dr. Yasuyuki Akita and Jeanette Reyes from BMElab for sharing their expertise in BMELib, geo-statistics and also MATLAB programming.

I would like to acknowledge the National Institute on Aging (NIA) of the National Institutes of Health (NIH) (award number R01AG033078) and also the National Institute of Environmental Health

Sciences (NIEHS) training program (grant number T32ES007018) of NIH for providing me with financial support.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS AND SYMBOLS	xx
CHAPTER 1 – INTRODUCTION	1
1.1 Ozone estimates for epidemiologic studies	2
1.2 Environmental sources of ozone data	4
1.3 Geo-statistical approaches for integration of environmental data from multiple sources	6
1.4 Thesis Hypothesis and approach	8
CHAPTER 2 – BAYESIAN MAXIMUM ENTROPY INTEGRATION OF OZONE OBSERVATIONS AND MODEL PREDICTIONS: A NATIONAL APPLICATION	10
2.1 Introduction	10
2.2 Data	12
2.2.1 Ozone Monitoring Data	12
2.2.2 Air Quality Model Predictions	13
2.3 Methods	13
2.3.1 BME Estimation Methodology	13
2.3.2 Variability of CTM Model Performance Evaluation across the Continental US	16
2.3.3 The Proposed Regionalized Air Quality Model Performance (RAMP) Evaluation Framework	17
2.3.4 Offset analysis	19
2.3.5 Space-time Covariance Model.....	19
2.3.6 Validation analysis	20

2.4 Results	22
2.4.1 BME Ozone estimates	22
2.4.2 Soft data construction using the RAMP approach.....	24
2.4.3 Validation results	26
2.5 Discussion.....	28
2.6 Acknowledgments	30
 CHAPTER 3 – BME INTEGRATION OF OZONE OBSERVATIONS AND CTM PREDICTIONS AT MULTIPLE TIME SCALES	 31
3.1 Introduction	31
3.2 Data Sources	33
3.3 Methods	34
3.3.1 BME Estimation Methodology	34
3.3.2 Variability of CTM Model Performance Evaluation across the Continental US	35
3.3.3 Regionalized Air Quality Model Performance (RAMP) Analysis for Hourly Ozone	36
3.3.4 Offset analysis	38
3.3.5 Space-time Covariance Model.....	38
3.3.6 Validation analysis	39
3.4 Results	41
3.4.1 BME Ozone estimates	41
3.4.2 Validation results	43
3.5 Discussion.....	47
 CHAPTER 4 – CONCLUSIONS	 50
4.1 Scientific findings.....	53
4.2 Methodological aspects and study limitations	54
4.3 Uncertainties and future research	55

APPENDIX A – SUPPORTING INFORMATION FOR BAYESIAN MAXIMUM ENTROPY INTEGRATION OF OZONE OBSERVATIONS AND MODEL PREDICTIONS: A NATIONAL APPLICATION	57
A.1 Ozone CTM model performance evaluation for daily metrics DM8A and D24A	57
A.1.1 CTM model performance evaluation for the DM8A ozone concentrations	59
A.1.2 CTM model performance evaluation for the D24A ozone concentrations	64
A.2 Parameter selection for the offset analysis and covariance model	68
A.2.1 Maps and figures of the offset analysis for the DM8A ozone concentrations	70
A.2.2 Maps and figures of the offset analysis for the D24A ozone concentrations	73
A.2.3 Covariance model	77
A.3 Additional results of BME O ₃ estimates	80
A.3.1 Additional maps showing the soft data for the DM8A O ₃	80
A.3.2 Additional maps of BME O ₃ estimates for D24A O ₃	81
A.4 Testing the statistical significance of the increase in R^2 for the cross-validation analysis	85
A.5 Additional results for the cross-validation analysis	86
A.6 The cokriging approach with a parametric relationship between the observations and the CTM model predictions	88
A.7 A discussion about the limitation of using the daily averages of the hourly ozone observations as hard data	96
APPENDIX B – SUPPORTING INFORMATION FOR BME INTEGRATION OF OZONE OBSERVATIONS AND CTM PREDICTIONS AT MULTIPLE TIME SCALES	98
B.1 Hourly Ozone CTM model performance evaluation	98
B.2 Parameter selection for the offset analysis and covariance model	104
B.2.1 Offset analysis	104
B.2.2 Covariance model	109
B.3 The additional results of BME ozone estimates	110

B.4 The additional results for the validation analysis.....	112
B.4.1 The additional error statistics for the validation analysis.....	112
B.4.2 The additional validation analyses show the influence of the grid cell resolution on the accuracy of the BME estimates.....	112
APPENDIX C-WHIMS PROJECT REPORT-CODE DOCUMENTATION AND QUALITY ASSURANCE FOR THE ESTIMATION OF DAILY O3 FROM 1993 TO 2010 USING OBSERVATIONS AND CTM	
C.1 Introduction.....	115
C.2 Materials	116
C.3 Location of the study participants.....	117
C.4 Methods	118
C.5 Numerical implementation.....	120
C.6 Results.....	123
C.7 QA/QC.....	123
C.8 Date and version number	125
REFERENCES	126

LIST OF TABLES

Table 2. 1: Validation statistics for BME data integration scenarios OBS, CAMP, and RAMP*	27
Table 3. 1: The list of BME data integration simulations used in the validation analysis	40
Table 3. 2: Validation statistics hourly O3 for BME data integration scenarios OBS and RAMP #	44
Table 3. 3: Compare validation statistics for DM8A and D24A O3 between the upscaling methods USM1 and USM2 ##	45
Table 3. 4: Validation statistics for BME data integration scenarios OBS and RAMP using different soft data (for 888 sites only) for upscaling method USM1	47
Table A. 1s: Spatial and temporal offset parameters and their label	69
Table A. 2s: Parameter values of the 3-structured covariance model for each offset range for DM8A O3	77
Table A. 3s: Parameter values of the 3-structured covariance model for each offset range for D24A O3	77
Table A. 4s: Validation statistics of MNB and MNGE for BME data integration scenarios OBS, CAMP, and RAMP	87
Table A. 5s: Validation statistics for BME data integration scenarios OBS, CAMP, and RAMP	88
Table A. 6s: Error statistics of validation analysis to compare BME with RAMP and Co-kriging with a parametric approach.	96
Table B. 1s: Summary of error statistics for CTM model performance at ozone monitoring sites	104
Table B. 2s. Spatial and temporal offset parameters and their label.....	105
Table B. 3s: Parameter values of the 3-structured covariance model for each set of offset ranges	109
Table B. 4s. Computational costs for generating BME ozone maps USM1 vs USM2#	112
Table B. 5s. Validation statistics for BME data integration scenarios OBS and RAMP	112

LIST OF FIGURES

Figure 2. 1: RAMP analysis conducted specifically at time $t=11\text{-Jul-2005}$ and for sites ID 060372005(left) and 120713002(right). The empty circles show the pairs of observed-modeled values (z_j, z_j) that are within 120 days of t . The vertical lines show the stratification of these pairs in 10 bins. The interpolation lines connecting the filled circles and triangles show how the mean of the observed value in each bin, $\lambda_1(z_b, sn, t)$ (filled circles), and the corresponding standard deviation, $\lambda_2(z_b, sn, t)$ (filled triangles) change as a function of the average modeled value z_b in that bin.....	18
Figure 2. 2. Maps of BME mean estimates (Top) and corresponding standard deviations of BME estimates (Bottom) of the DM8A ozone concentrations (ppb) on day Jul-11-2005 obtained with scenario OBS (Left), CAMP (middle panels) and RAMP (Right). Circles in the maps represent locations of monitoring sites and color match legend for observed concentrations.....	23
Figure 2. 3: The DM8A ozone concentrations in the United States on 11-July-2005 using (Left) the raw CTM Model predictions, (Middle) the bias-corrected expected values $\lambda_1(z_i, p_i)$ for the estimation scenario CAMP, and (Right) the bias-corrected expected values $\lambda_1(z_i, p_i)$ for the estimation scenario RAMP.	25
Figure 3. 1: RAMP analysis conducted specifically at time $t=11\text{-Jul-2005}$ and for sites ID 060372005(left) and 120713002(right). The empty circles show the pairs of hourly observed-modeled values (z_j, z_j) that are within 120 days of t . The vertical lines show the stratification of these pairs in 10 percentile bins. The interpolation lines connecting the filled circles and triangles show how the mean of the observed value in each bin, $\lambda_1(z_b, sn, t)$ (filled circles), and the corresponding standard deviation, $\lambda_2(z_b, sn, t)$ (filled triangles) change as a function of the average modeled value z_b in that bin.....	37
Figure 3. 2. Maps of Jul-11-2005 ozone BME estimates in ppb of the DM8A (Top) and the D24A (Bottom) obtained from upscaling method USM1 (Left) and USM2 (Middle). Also shown are the absolute differences (USM2-USM1) between these two methods (Right).	42
Figure A. 1s: The ozone monitoring sites (circles), CAMx modeling domain with $36 \times 36 \text{ km}^2$ grid cell resolution (dash-dotted line rectangle) and CAMx modeling domain with $12 \times 12 \text{ km}^2$ grid cell resolution (dashed line rectangle).....	58
Figure A. 2s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the DM8A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution.	59

Figure A. 3s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the DM8A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with 36x36km ² and 12X12km ² grid cell resolution.	60
Figure A. 4s: The DM8A O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km ² grid cell resolution.	60
Figure A. 5s: The DM8A O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km ² grid cell resolution.	61
Figure A. 6s: The DM8A O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km ² grid cell resolution.	61
Figure A. 7s: The DM8A O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km ² grid cell resolution.	62
Figure A. 8s: Boxplots of DM8A O3 mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 36x36km ² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	63
Figure A. 9s: Boxplots of DM8A O3 mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 12x12km ² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	63
Figure A. 10s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of D24A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with 36x36km ² and 12X12km ² grid cell resolution.	64
Figure A. 11s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of D24A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with 36x36km ² and 12X12km ² grid cell resolution.	64
Figure A. 12s: D24A O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km ² grid cell resolution.	65
Figure A. 13s: D24A O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km ² grid cell resolution.	65
Figure A. 14s: D24A O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km ² grid cell resolution.	66

Figure A. 15s: D24A O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km ² grid cell resolution.	66
Figure A. 16s: Boxplots of D24A O3 mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 36x36km ² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	67
Figure A. 17s: Boxplots of D24A O3 mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 12x12km ² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	68
Figure A. 18s: The observed DM8A O3 concentrations on the monitoring sites on day 21-July-2005	70
Figure A. 19s: The short range offset for DM8A O3 on the monitoring sites on day 21-July-2005	71
Figure A. 20s: The intermediate range offset for the DM8A O3 on the monitoring sites on day 21-July-2005	71
Figure A. 21s: The long range offset for the DM8A O3 on the monitoring sites on day 21-July-2005	72
Figure A. 22s: The very long range offset for the DM8A O3 on the monitoring sites on day 21-July-2005	72
Figure A. 23s: The time series of the DM8A O3 concentrations and four offsets on a randomly selected site (SiteID:060719004) for year 2005	73
Figure A. 24s: The observed D24A O3 concentrations on the monitoring sites on day 21-July-2005	74
Figure A. 25s: The short range offset for the D24A O3 on the monitoring sites on day 21-July-2005.....	74
Figure A. 26s: The intermediate range offset for the D24A O3 on the monitoring sites on day 21-July-2005	75
Figure A. 27s: The long range offset for the D24A O3 on the monitoring sites on day 21-July-2005.....	75
Figure A. 28s: The very long range offset for the D24A O3 on the monitoring sites on day 21-July-2005	76
Figure A. 29s: The time series of the D24A O3 concentrations and four offsets on a randomly selected site (SiteID:060719004) for year 2005	76
Figure A. 30s: Graphs of the 3-structured exponential/exponential covariance models with respect to the spatial lag when the temporal lag is set to 0 (Left) and with respect to the temporal lag when the spatial lag is set to 0 (Right) for DM8A O3	78

Figure A. 31s: Graphs of the 3-structured exponential/exponential covariance models with respect to the spatial lag when the temporal lag is set to 0 (Left) and with respect to the temporal lag when the spatial lag is set to 0 (Right) for D24A O3.....	78
Figure A. 32s: Dominance plot showing how the variance changes with respect to the spatial range (Left) and with respect to the temporal range (Right) for covariance model corresponding to each offset range for DM8A O3	79
Figure A. 33s: Dominance plot showing how the variance changes with respect to the spatial range (Left) and with respect to the temporal range (Right) for covariance model corresponding to each offset range for D24A O3	79
Figure A. 34s: Maps showing the square root of soft data variance $\lambda_2(z_i, p_i)$ obtained across the continental United States on 11-July-2005 in (left) scenario CAMP and (right) scenario RAMP for DM8A O3.....	80
Figure A. 35s: Maps of BME mean estimates (top panel) and corresponding standard deviations of BME estimates (bottom panel) of the D24A ozone (ppb) on day Jul-21-2005 obtained with scenario OBS (left panels), CAMP (middle panels) and RAMP (right panels). Circles in the maps represent locations of monitoring sites and color match legend for observed concentrations.	82
Figure A. 36s: D24A O3 concentrations in the United States on 21-July-2005 using (left) the raw CTM Model predictions, (middle) the bias-corrected expected values $\lambda_1(z_i, p_i)$ for the estimation scenario CAMP, and (right) the bias-corrected expected values $\lambda_1(z_i, p_i)$ for the estimation scenario RAMP.	84
Figure A. 37s: Maps showing the square root of soft data variance $\lambda_2(z_i, p_i)$ obtained across the continental United States on 21-July-2005 in (left) scenario CAMP and (right) scenario RAMP for D24A O3.	85
Figure A. 38s: Percent change in mean normalized gross error MNGE shown as a function of cross-validation radius R_v between Scenario RAMP and Scenario OBS. Each curve corresponds to the MNGE calculated using only observations above a given cutoff (0 ppb, 10ppb) of all observations values for the DM8A O3 (Left) and the D24A O3 (Right).....	87
Figure A. 39s: Graphs of the 3-structured exponential/exponential covariance functions with respect to the spatial lag (Top panel) and the temporal lag (Bottom panel) for CTM with $36 \times 36 \text{ km}^2$ cell resolution (Left) and CTM with $12 \times 12 \text{ km}^2$ cell resolution (Right) for D24A O3.....	92
Figure A. 40s: Map of the parameter $\beta_{0s, t}$ across the continental United States for day 11-Jul-2005 for the DM8A ozone concentrations.....	94
Figure A. 41s: Map of the parameter $\beta_{1s, t}$ across the continental United States for day 11-Jul-2005 for the DM8A ozone concentrations.....	94

Figure A. 42s: Map of the parameter $\beta_{0s,t}$ across the continental United States for day 21-Jul-2005 for the D24A ozone concentrations	95
Figure A. 43s: Map of the parameter $\beta_{1s,t}$ across the continental United States for day 21-Jul-2005 for the D24A ozone concentrations	95
Figure B. 1s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the hourly ozone concentrations at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution.....	99
Figure B. 2s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the hourly ozone concentrations at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution.....	99
Figure B. 3s: The hourly O ₃ mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $36 \times 36 \text{ km}^2$ grid cell resolution.....	100
Figure B. 4s: The hourly O ₃ standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $36 \times 36 \text{ km}^2$ grid cell resolution	100
Figure B. 5s: The hourly O ₃ mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $12 \times 12 \text{ km}^2$ grid cell resolution.....	101
Figure B. 6s: The hourly O ₃ standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $12 \times 12 \text{ km}^2$ grid cell resolution.....	101
Figure B. 7s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with $36 \times 36 \text{ km}^2$ grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	102
Figure B. 8s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with $12 \times 12 \text{ km}^2$ grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	102
Figure B. 9s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with $36 \times 36 \text{ km}^2$ grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	103

Figure B. 10s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with 12x12km ² grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).	103
Figure B. 11s: The observed hourly O ₃ concentrations on the monitoring sites for UTC hour 20050711T000000	106
Figure B. 12s: The short range offset for hourly O ₃ on the monitoring sites for UTC hour 20050711T000000	106
Figure B. 13s: The intermediate range offset for hourly O ₃ on the monitoring sites for UTC hour 20050711T000000	107
Figure B. 14s: The long range offset for hourly O ₃ on the monitoring sites for UTC hour 20050711T000000	107
Figure B. 15s: The very long range offset for hourly O ₃ on the monitoring sites for UTC hour 20050711T000000	108
Figure B. 16s: The time series of the hourly O ₃ concentrations and four offsets on a randomly selected site (SiteID:481130069) for year 2005	108
Figure B. 17s: Graphs of the 3-structured exponential/exponential/cosine covariance models with respect to the spatial lag when the temporal lag is set to 0 (Left) and with respect to the temporal lag when the spatial lag is set to 0 (Right) for hourly O ₃	109
Figure B. 18s: Dominance plot showing how the variance changes with respect to the spatial range (Left) and with respect to the temporal range (Right) for covariance model corresponding to each offset range for hourly O ₃	110
Figure B. 19s: Maps of scenario OBS BME mean estimates of the DM8A (Top) and the D24A (Bottom) ozone concentrations (ppb) on day Jul-11-2005 obtained from upscaling method USM1 (Left) and USM2 (Middle), and the absolute differences between these two methods (Right).	111
Figure B. 20s: Maps of R ² (Left) and RMSE (Right) changes for validation analysis conducted with two sets of soft data, one from CTM with 36x36km ² and the other from CTM with 12x12km ² grid cell resolution for D24A at validation radius R _v =0km. Red color in the maps means R ² and RMSE increase when change the soft data from 36x36km ² to 12x12km ² , while blue color means R ² and RMSE decrease.....	113
Figure B. 21s: Maps of R ² (Left) and RMSE (Right) changes for validation analysis conducted with two sets of soft data, one from CTM with 36x36km ² and the other from CTM with 12x12km ² grid cell resolution for DM8A at validation radius R _v =0km. Red color in the maps means R ² and RMSE increase when change the soft data from 36x36km ² to 12x12km ² , while blue color	114

Figure C. 1: The time series plot showing the BME mean estimates on simulated ID18 and the hard data in the neighborhood.....	124
---	-----

Figure C. 2: The time series plot showing the BME mean estimates on simulated ID416 and the hard data in the neighborhood.....	124
--	-----

LIST OF ABBREVIATIONS AND SYMBOLS

AQS	Air Quality Systems
BME	Bayesian Maximum Entropy
CAMx	Comprehensive Air Quality Model with extensions
CMAQ	Community Multi-scale Air Quality Modeling System
CTM	Chemical transport models
DM8A	Daily maximum 8-hour average
D24A	Daily 24-hour average
EPA	The U.S. Environmental Protection Agency
G-KB	General knowledge base
MAQSIP	Multiscale Air Quality Simulation Platform
ME	Mean error
MNB	Mean Normalized Bias
MNGE	Mean Normalized Gross Error
NAAQS	National Ambient Air Quality Standards
NO _x	Nitrogen oxides
PCR ²	Percent change in R ²
PDF	Probability Distribution Function
ppb	Parts per billion
O ₃	Ozone
OAQPS	The office of air quality planning and standards
R^2	Squared Pearson Correlation Coefficient
RAMP	Regionalized Air Quality Model Performance

SE	Standard deviation of the prediction error
S-KB	Site specific knowledge base
S/TRF	Space/time random field
km	Kilometer
km ²	Squared kilometer
UTC	Coordinated Universal Time
WHIMS	Women's Health Initiative Memory Study
a_r	Spatial offset kernel smoothing range
a_t	Temporal offset kernel smoothing range
\hat{c}_X	Experimental covariance
c_X	Covariance model
$E[.]$	Stochastic expectation
m_X	Mean trend function
$p=(s,t)$	p is the prediction point; s is the spatial coordinate and t is time
$o_Z(\mathbf{p})$	Offset at the prediction point p
\mathbf{p}_k	Space/time estimation point
\mathbf{p}_m	Space/time soft data points
\mathbf{p}_o	Space/time observation points
f_G	Prior probability distribution function
f_K	BME posterior probability distribution function
$f_S(\mathbf{x}_m)$	Soft data PDF characterizing the offset-removed ozone values \mathbf{x}_m
$\lambda_1(\tilde{z}_i)$	Expected value
$\lambda_2(\tilde{z}_i)$	Variance

r	Spatial lag
r_v	Validation radius
τ	Temporal lag
x_k	Offset-removed ozone values at the estimation points
\mathbf{x}_m	Offset-removed ozone values at the soft data points
\mathbf{x}_o	Offset-removed hard data at the observational points
\tilde{z}_j	Ozone CTM model predictions
\mathbf{z}_o	Ozone observations

CHAPTER 1 – INTRODUCTION

Ozone is one of the six “criteria” pollutants with established standards in the Clean Air Act charged by the U.S. Environmental Protection Agency (EPA). The National Ambient Air Quality Standards (NAAQS) for ozone has been updated a couple times in the history. The current ozone standard requires that annual fourth-highest daily maximum 8-hour ozone concentrations, averaged over 3 years, should be less than 70 ppb (parts per billion). Tropospheric ozone has been associated with a wide range of adverse health outcomes including respiratory effects, cardiovascular effects, central nervous systems effects and mortality[1].

Most of the evidence on health effects of ozone relates to short-term exposure. The accumulated evidences on impacts in populations residing in areas with elevated ozone levels for prolonged periods are more difficult to be detected and are highly uncertain. The Women’s Health Initiative (WHI) memory study (WHIMS), which involved a nationwide, multicenter cohort of older women aged 65 to 80 years old, aims to investigate the neurodegenerative effects of long-term ozone exposures in older women. This large scale cohort study lasted for over 10 years, from year 1996 to year 2006. Our work is to support an exposure model used to estimate personal exposures of these participants, who came from multiple metropolitan or rural areas across the continental United States. Spatial and temporal variability in ozone concentrations vary across different geographical regions and local urban sectors, this has been a major contributor to the uncertainties in air pollution epidemiologic studies. To achieve the goal of understanding the adverse health effects of long-term exposure to ozone, accurate ambient estimates of the spatiotemporal variation of ozone levels at fine space and for long time periods are needed.

1.1 Ozone estimates for epidemiologic studies

Epidemiologic studies investigate the associations between health effects and exposure of human populations to ambient air pollution. These studies fall into several categories, including cross-sectional, cohort, panel and time-series studies. Despite the epidemiologic study design, the investigator usually needs to collect data in regards to air pollution exposure level for the study participants or population and their health outcomes. Exposure measurement error, which is the uncertainty associated with the exposure metrics used to represent exposure of an individual or population, is an important contributor to uncertainty in air pollution epidemiologic study results. Exposure error can influence observed epidemiologic associations between ambient pollutant concentrations and health outcomes by biasing effect estimates toward or away from the true associations and widening confidence intervals. The difference between true and estimated exposure to ambient pollutants has been one of the major components that contribute to exposure measurement error in air pollution epidemiologic studies. Spatial and temporal variability in ozone concentrations can contribute to exposure error in epidemiologic studies, especially for cross-sectional and large-scale cohort studies, if the ambient ozone concentrations measured at the central site monitor is used as an ambient exposure surrogate, which is often different from the actual ambient ozone concentrations outside a participant's or a population's residence. Community exposure using the ambient ozone measurements at nearby monitoring stations may not be well represented when monitors cover large areas with several sub-communities having different emission sources and topographies, such as in Los Angeles, California, where ozone monitors are found to have a much wider range of inter-monitor correlations (-0.06 to 0.97) than the ones in Atlanta, Georgia (0.61 to 0.96) [1].

Ozone epidemiologic studies use different exposure metrics and have different sources of exposure error. For ozone short-term exposure, different studies report different daily metrics, including the maximum 8-hour running average of the hourly concentrations occurring in a 24-hour period (8-hour daily max)[2-4], the maximum hourly concentrations occurring in a 24-hour period (1-hour daily max)[5, 6] and the average of the hourly concentrations occurring in a 24-hour period (24-hour average)[7, 8]. According to the observed ozone concentrations at monitoring sites, the correlations among these common daily metrics vary site by site. Overall, the two daily peak values, daily 1-hour maximum and daily 8-hour maximum, are well correlated, with a median correlation of 0.97 across the AQS sites. The correlation between the 8-hour maximum and 24-hour average are somewhat less well correlated with a median correlation of 0.89[1]. This indicates the influence of the overnight period on the 24-hour average ozone concentrations. In contrast, the 1-hour daily max and 8-hour daily max are more indicative of the daytime ozone concentrations. Little consensus exists as to which metric is the most appropriate. Preferably, epidemiologic studies are recommended to report results using multiple metrics. For ozone long-term exposure, a long-term arithmetic mean, such as monthly, quarterly or yearly averages of the above daily metrics is often computed for the exposure assessment. It is important to recognize the different averaging times to interpret the health effect estimates reported in epidemiologic studies.

Epidemiologic studies use a wide variety of methods to assign exposure. The commonly used exposure assessment methods, from simple indicators to complex models, include exposure indicators, personal monitoring, dispersion modeling, land-use regression modeling and geo-statistical spatial interpolation methods. Each method has its advantages and disadvantages when applied to individual studies. For example, personal monitoring has the advantage of providing relatively accurate individual-level exposure data, but the disadvantage is that it is very costly and time consuming so it is only practical in small scale studies involving a limited number of participants. The major disadvantage of dispersion modeling is that it requires highly specific input data, including specific emission inventories and meteorological information.

Among different geo-statistical methods, the four commonly used methods are spatial averaging, nearest neighbor, inverse distance weighting and kriging. All of these four methods are weighted average methods, with the interpolation process involving the following steps: 1) defining the search area or neighborhood around the point of interest; 2) locating the observed data points within this neighborhood; and 3) assigning appropriate weights to each of the observed data points. The differences are in their choices of sample weights. With spatial averaging, the same fractional weights are assigned to all sampled values within a fixed distance. With nearest neighbor method, only a single sampled value is used and a weight of 1 is assigned. With inverse distance weighting, the closer samples are assigned with larger weights. With kriging, the weights are assigned based on the spatial autocorrelation statistics of the sampled dataset. The common limitation of these interpolation methods is that it relies on the observational data alone, which poses a bigger challenge for those areas where the monitoring stations are very sparse and/or those time periods where ozone monitoring data is missing.

1.2 Environmental sources of ozone data

An important environmental source of ozone data is measurements from routine monitoring networks. In the United States, EPA regulations require state environmental agencies to operate air pollution monitoring stations and report air monitoring data to the Air Quality System (AQS) database, which is a repository of the monitoring data collected across various monitoring networks. The hourly ozone observational data from these monitoring stations are available from year 1993 to the present. The office of air quality planning and standards (OAQPS) rely upon ozone measurements for air quality assessment and attainment/non-attainment designations. By year 2015, there are over 1250 ozone monitors reporting hourly data to AQS. Strict quality assurance and quality control procedures for ozone monitoring have been developed and implemented at the monitoring stations. The hourly ozone concentrations reported to the AQS database can be considered as a reliable and accurate data source. There are, however, some limitations in this data source. The distribution of ozone monitors across urban areas varies between cities because the number and location of required ozone monitors in an urban area

depend on many factors, such as the magnitude of the concentrations and population density. The densest ozone monitoring sites are located in California and the eastern U.S, while relatively scarce across the central U.S. Further, the monitoring durations on the stations are not consistent. Due to the strong seasonality of ozone concentrations, many states limit their ozone monitoring to a certain portion of the year, termed the ozone season, the length of which varies from one area of the country to another. As a result, less than half of the ozone monitoring sites in the U.S. operate year-round. The majority of the sites only operate for summer months. This is why the estimation approaches solely based on observational data in many of the previous epidemiological studies suffer from the missing data issues due to the sparse monitoring network across space and the inconsistent monitoring durations.

Besides ozone monitoring networks, numerical model predictions have become a second source of environmental ozone data. For more than a decade, air quality models such as Community Multi-scale Air Quality Modeling System (CMAQ) and Comprehensive Air Quality Model with extensions (CAMx) have been used as powerful computational tools for air quality management. These models unite three major types of models, including meteorological models, emission models and a chemistry-transport model. They are designed to approach air quality as a whole by including state-of-science capabilities for modeling multiple air quality issues. These models can simulate air pollution concentrations as averaged values of grid cells with continuous spatial and temporal coverage. For the purpose of air quality management and evaluation in the United States, there has been a wide range of modeling simulations completed which cover various model configurations, domains, episodes, chemical mechanisms and aerosol modules[9-14]. The acceptability of these models' performance was judged by comparisons of the model predicted concentrations, usually the daily 8-hour maximum ozone, to the corresponding observed values at monitoring sites. The modeling community has made significant progress in reducing the emission uncertainties and inaccuracy of the chemical mechanisms in the air quality models to reduce the prediction errors. Ozone model performance has been slowly improved as these modeling systems advance. Overall, the daily 8-hour maximum ozone performance at AQS monitoring sites are relatively good, with the Mean Normalized Bias (MNB) within +/- 10% and Mean Normalized Gross Error (MNGE)

less than 20%. Although these models still have inherent uncertainties and weakness, the ozone concentrations predicted by these modeling platforms can closely reflect the corresponding observed concentrations in space and time. Our work is to take what is available and make use of them.

Due to limited computational resources, CTM model applications on national scale usually use a coarser horizontal grid cell resolution, such as 36x36 km² for Continental U.S. or 12x12 km² for the eastern U.S. covering thirty seven eastern states. Model predictions from the 36x36 km² Continental U.S. domain were often used to provide initial and boundary concentrations for simulations in the 12x12 km² domain. For those applications studying air quality at local scale, finer horizontal grid cell resolutions, such as 4x4 km² or 2x2 km² have been used. In theory, higher resolution modeling is expected to yield better predictions because of better resolved model input fields, such as topography, land cover or emissions, and better mathematical characterization of physical and chemical processes. Ozone model performance dependence on grid resolution have been examined [15, 16]. In general, finer grid scales are found to be able to better resolve the local scale spatial variability of ozone concentrations. The newest release CMAQ 5.1 enables improved fine-scale simulations allowing users to simulate air quality at smaller settings like metropolitan areas as fine as 1x1km² grid cell resolution. Improvements in computational efficiency are expected to enable higher resolution in the future release of these modeling system.

1.3 Geo-statistical approaches for integration of environmental data from multiple sources

Geo-statistical approaches provide useful solution to integrate air pollution measurements and other relevant information. Several Bayesian inference approaches [17-19] have been developed to provide a sophisticated statistical framework for the data integration of observations and CTM model predictions to improve ambient air pollution exposure estimates. These approaches share the following characteristics: parameterize the relationship between air pollution observations and predictions, using kriging to obtain air pollution estimates for any given value of the parameters, and use Bayesian inference to obtain air pollution estimates that accounts for parameter uncertainty. These methods, have the

following limitations: they assume that the relationship between air pollution observations and predictions is linear and homoscedastic, they share the linear limitations of the kriging estimator, and require a high computational cost.

One approach is the Bayesian Maximum Entropy (BME) method of modern geo-statistics, a knowledge-processing framework, because of its following advantages. First of all, it can incorporate different kinds of knowledge bases, such as general knowledge derived from physical laws, scientific theories and specific knowledge processed from a given situation. Secondly, there are no assumptions about the shape and distribution of the underlying probability law. Therefore, it can integrate a wide variety of nonlinear, non-Gaussian uncertain datasets in a probabilistic way. Thirdly, it is computationally effective in spatial and temporal domains.

In the past few years, BME has been applied to map criteria pollutant [20, 21]. Using BME to integrate air monitoring observations and numerical model predictions has been proven to be a cost-effective and efficient technique in improving spatial predictions of ozone concentrations. It allows us to take advantage of the strength from both data sources, the accuracy of the observational data and the good spatial/temporal coverage of air quality model outputs without assuming a parametric relationship between these two data sources. In de Nazelle et al. [20], BME framework was used to develop ozone estimates for the state of North Carolina for a short study period, June 19th to June 30th of year 1996. The observational data from the state's ozone monitoring network in combination with model outputs from the Multiscale Air Quality Simulation Platform (MAQSIP) modeling system were integrated. In this study, the BME framework gave preference to measured ozone data, also used MAQSIP model outputs as a function of model performance. It showed that the BME data integration approach improves the accuracy and the precision of ozone estimations across the state of North Carolina when compared to a spatial interpolation of observational data alone.

1.4 Thesis Hypothesis and approach

Our hypothesis is that fully characterizing the spatial and temporal heterogeneity in CTM model performance in our geo-statistical approach can increase estimation accuracy. In de Nazelle's work [20], air quality model performance was assumed to be homogeneous for the study domain, so the bias and uncertainty associated with the model predictions were assumed to be the same across space and time. Therefore, the soft data was processed through pooling all the paired observed and modeled ozone concentrations in the domain at one time. This assumption might be reasonable given the small study domain and short study period, but may not be applicable due to the documented spatial heterogeneity and temporal variability of ozone model performance across the country. Therefore, we need to extend the work of de Nazelle et. al's by developing a new approach that can accommodate the spatial or seasonal variability in the ozone model performance of the CTM.

To test this hypothesis, we describe in Chapter 2 the development of a Regionalized Air Quality Model Performance (RAMP) approach to characterize the ozone model prediction errors that changes across space/time. Instead of making the assumption of air quality model performance homogeneity, we generate soft data as secondary information, to reflect the bias and uncertainty of model predictions changing across space and time. As a result, the RAMP approach is expected to capture geographical and temporal changes in bias and uncertainty associated with air quality model predictions. The soft data generated from RAMP approach is integrated with the ozone observations in our BME model framework to produce ozone estimates. We first compare the RAMP estimation with two other estimation scenarios, one using only ozone observations, and the other is a Constant Air Quality Model Performance (CAMP) scenario assuming that CTM model performance does not change across space and time. We also compare our BME estimation to a cokriging estimation based on a parametric relationship between the observations and the CTM model predictions.

For the WHIMS work, the BME approach was used to interpolate directly the daily ozone concentrations by first aggregating the hourly observations and CTM model predictions. An alternative approach would be to first generate hourly BME estimates then aggregate it into a daily metrics. This alternative approach could be especially useful for those epidemiologic studies that require higher temporal resolution of ambient exposure estimates, such as those exposure models combining microenvironmental concentrations with human activity data to estimate personal exposures. This could be relevant given the known diurnal patterns seen in hourly ozone data. The disadvantage of this alternative approach is the computational intensity, requiring over 200 times more CPU runtime. In Chapter 3, our first task was to investigate the extent of the improvement on the accuracy of the hourly ozone estimates when incorporating CTM hourly model predictions through our RAMP approach. Our second task is to investigate the differences of the predictive capacity between these two choices of generating daily ozone estimates. We conducted a comparison of two upscaling methods: USM1 (data aggregation from hourly to daily followed by BME approach estimation) and USM2 (perform BME approach estimation on hourly ozone followed by data aggregation). A validation analysis using only non-collocated data outside of a validation radius was performed and the error statistics between the observations and re-estimated values for two daily metrics, the daily maximum 8-hour average (DM8A) and the daily 24-hour average (D24A) ozone concentrations, were obtained to investigate the estimation accuracy.

CHAPTER 2 – BAYESIAN MAXIMUM ENTROPY INTEGRATION OF OZONE OBSERVATIONS AND MODEL PREDICTIONS: A NATIONAL APPLICATION¹

2.1 Introduction

According to EPA's newly released Integrated Science Assessment for tropospheric ozone[1], the evidence of public health impacts on populations residing in areas with elevated ozone levels for prolonged periods are still uncertain. A better understanding of the adverse health effects to chronic ozone requires accurate exposure estimates at multiple temporal scales and at fine spatial resolutions. Estimates of ozone concentrations typically rely on environmental data collected from two sources: monitoring networks and air quality chemical transport models (CTM). The first source gives measurement concentrations for a long temporal time, but only at a point where the monitor is located. The CTM provides predictions for all locations, but is an average concentration based on the spatial resolution of the model grid cell. Further, given the intensive resources needed to build a CTM, the numbers of days that are simulated are limited. Several categories of data integration methods, including Kalman filter methods[22], variational methods [23], optimal interpolation [24] and Bayesian methods [17-19] have been developed to integrate these two types of data and rely on their individual strengths to build a more refined air pollution estimate. In this work, we choose the BME method of modern geostatistics, a knowledge-processing framework, because of its advantage of integrating a wide variety of nonlinear, non-Gaussian knowledge bases.

¹ This chapter previously appeared as an article in the Journal of Environmental Science & Technology. The original citation is as follows: Xu, Yadong, Serre, L. Marc, Reyes, Jeanette, Vizuete, William. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: A National Application. Environmental Science & Technology (2016) 50 (8), 4393-4400.

We developed our data integration approach to obtain two metrics of ozone estimates, the DM8A and D24A ozone concentrations. Both of these metrics are commonly used in epidemiology studies [7, 8, 25].

Several Bayesian inference approaches [17-19] provide a sophisticated statistical framework for the data integration of ozone observations and model predictions and production of multiple time averaged estimates. These approaches share the following characteristics: parameterize the relationship between air pollution observations and predictions, use kriging to obtain air pollution estimates for any given value of the parameters, and use Bayesian inference to obtain air pollution estimates that accounts for parameter uncertainty. These methods, however, have the following limitations: they assume that the relationship between air pollution observations and predictions is linear and homoscedastic, they share the linear limitations of the kriging estimator, and have a high numerical cost.

To overcome these limitations de Nazelle et al. [20] introduced an approach based on the nonlinear extension of kriging provided by the Bayesian Maximum Entropy (BME) method of modern spatiotemporal geostatistics[26]. This approach uses a non-parametric methodology that fully accounts for the non-linearity and non-homoscedasticity of the relationship between air pollution observations and predictions. Their application of this approach showed that the BME method provided a numerically efficient data integration framework that combines a wide variety of nonlinear, non-Gaussian knowledge bases that are out of the reach of kriging-based methods. That study applied the BME framework to integrate ozone observations and model predictions simulated by the Multiscale Air Quality Simulation Platform (MAQSIP) in the state of North Carolina and exposure were estimated for a short study period, June 19th to June 30th of year 1996. That study demonstrated that the BME data integration approach, by incorporating the MAQSIP model predictions along with ozone observations, improved both the accuracy and the precision of ozone estimations across North Carolina.

It is clear from the de Nazelle et al.'s work that the authors assumed that the model performance from the air quality model for ozone was homogeneous for the entire state. This was a reasonable assumption given the small study domain and short study period. In our work here, however, we are providing ozone estimates for the entire continental United States (US) for multiple time averages that could include a full year. Thus, de Nazelle et al.'s assumption may not be applicable due to the unknown spatial heterogeneity and temporal variability of ozone model performance across the country. Therefore, we extend the work of de Nazelle et al.'s by developing our new RAMP approach that can accommodate any spatial or seasonal variability in the model performance of the CTM. The refined ozone estimates that we obtain could be applied for health assessments or adapted to generate exposure estimates for other criteria air pollutants.

2.2 Data

2.2.1 Ozone Monitoring Data

The DM8A and D24A ozone concentrations for each monitoring site and day for the year 2005 were constructed based on raw monitoring data from ozone monitoring stations measuring hourly O₃ concentrations using the procedure described here.

We downloaded hourly ozone monitoring data (raw data) sampled from 1179 sites in the Air Quality Systems (AQS) database maintained by the U.S. Environmental Protection Agency (EPA), which is a repository of the monitoring data collected across various monitoring networks. Then we computed the DM8A and D24A of hourly ozone concentrations at each monitoring site to construct a daily ozone concentration database. These daily averages are considered as hard data, an error-free proxy, in our later interpolation analysis (see SI section 7 for details).

2.2.2 Air Quality Model Predictions

The air quality model data consists of hourly ozone concentrations predicted by the Comprehensive Air Quality Model with extensions (CAMx)[27] modeling system on a $36 \times 36 \text{ km}^2$ grid cell resolution domain covering the continental U.S. and a $12 \times 12 \text{ km}^2$ grid cell resolution domain covering the Eastern U.S. as shown in Figure 1s. CAMx is a publicly available Eulerian grid-based model that can address tropospheric ozone, acid deposition, visibility, fine particulates and other air pollutants issues in the context of a “one atmosphere” perspective. The modeling simulations were created by the U.S. EPA as base-case simulations in their analysis of the final Transport Rule. These air quality-modeling simulations used the CAMx version v5.30 with gas-phase chemistry mechanism CB05, and also refined meteorological and emission fields for the year 2005 across the United States. Detailed model configurations and evaluation are discussed elsewhere[28]. The hourly model predictions were used to compute the DM8A and D24A ozone concentrations at each grid cell. These CTM predicted daily ozone concentrations are used to construct the soft data, as secondary information with uncertainties, consisting of the expected values of the daily concentrations and the uncertainties associated with the expected values at each grid cell. The details of soft data construction are described in section 3.3.

2.3 Methods

2.3.1 BME Estimation Methodology

BME is a modern geo-statistical method [26] for spatial-temporal interpolation that incorporates information from many different data sources. The implementation and performance of BME have been detailed in other works[21, 29], and its application to the integration of O₃ observations and model predictions was described by de Nazelle et al[20]. In short, we model the (offset-removed) transform, which is a commonly used deterministic transformation[30], of air pollution as a Space/Time Random Field (S/TRF) $X(\mathbf{p})$ at space/time coordinate $\mathbf{p}=(s,t)$, where s is the spatial coordinate and t is time. Our notation for S/TRFs consists of denoting a single random variable X in capital letters, its realization, x , in lower case; and vectors in bold faces (e.g. $\mathbf{x} = [x_1, \dots]^T$). The general knowledge base (G-KB)

characterizing $X(\mathbf{p})$ consists of the mean function $m_X(\mathbf{p}) = E[X]$, where $E[.]$ is the stochastic expectation, describing its consistent trends, and the covariance function $c_X(\mathbf{p}, \mathbf{p}') = E[(X(\mathbf{p}) - m(\mathbf{p}))(X(\mathbf{p}') - m(\mathbf{p}'))]$ describing its space/time dependencies. Likewise the site specific knowledge base (S-KB) consists of the hard data \mathbf{x}_o at space/time observation points \mathbf{p}_o located at the monitoring stations, and the soft data characterizing the S/TRF values \mathbf{x}_m at the space/time model prediction points \mathbf{p}_m in terms of a site-specific PDF $f_S(\mathbf{x}_m)$. Denoting the G-KB as $G = \{m_X(\mathbf{p}), c_X(\mathbf{p}, \mathbf{p}')\}$ and the S-KB as $S = \{\mathbf{x}_o, f_S(\mathbf{x}_m)\}$, we can summarize the BME steps as 1) using the *Maximum Entropy* principle of information theory to process the G-KB in the form of a prior Probability Distribution Function (PDF) f_G , 2) integrating the S-KB using an epistemic *Bayesian* conditionalization rule to create a BME posterior PDF f_K characterizing the value x_k taken by $X(\mathbf{p})$ at any estimation point \mathbf{p}_k of interest, and 3) computing space/time estimates based on the BME posterior PDF. The BME posterior PDF is given by the BME equation

$$f_K(x_k) = A^{-1} \int d\mathbf{x}_m f_S(\mathbf{x}_m) f_G(\mathbf{x}_{map}) \quad (\text{E2-1})$$

where $\mathbf{x}_{map} = (x_k, \mathbf{x}_o, \mathbf{x}_m)$ is the value of $X(\mathbf{p})$ at points $\mathbf{p}_{map} = (\mathbf{p}_k, \mathbf{p}_o, \mathbf{p}_m)$ and A is a normalization constant.

Let $Z(\mathbf{p}) = Z(\mathbf{s}, t)$ be the Space/Time Random Field (S/TRF) representing daily ozone. In this study we define $Z(\mathbf{s}, t)$ as the sum of a homogenous/stationary S/TRF and a known offset as follows. We first define the transformation of the ozone observational data \mathbf{z}_o at locations \mathbf{p}_o as

$$\mathbf{x}_o = \mathbf{z}_o - o_Z(\mathbf{p}_o) \quad (\text{E2-2})$$

where $o_Z(\mathbf{p})$ may be any deterministic offset that can be mathematically calculated without error as a function of the space/time coordinate \mathbf{p} . We then define $X(\mathbf{p})$ as a homogeneous/stationary S/TRF representing the variability and uncertainty associated with the transformed data \mathbf{x}_o , and we let $Z(\mathbf{p}) = X(\mathbf{p}) + o_Z(\mathbf{p})$ be the S/TRF representing daily O3. We can then calculate \hat{z}_k , the estimated daily O3 at unmonitored location \mathbf{p}_k by obtaining the BME estimate \hat{x}_k for the transformed S/TRF $X(\mathbf{p})$ at the estimation point \mathbf{p}_k , and adding back $o_Z(\mathbf{p}_k)$, the offset calculated at \mathbf{p}_k .

The soft data are described by the PDF $f_S(\mathbf{x}_m)$ characterizing the offset-removed ozone values \mathbf{x}_m at the soft data points \mathbf{p}_m corresponding to the centroids of the n_m CTM computational nodes. The offset-removed ozone model predictions \tilde{x}_i are calculated at these nodes. As a key conceptual aspect of our work, the generation of this soft data PDF requires not only the offset-removed ozone model predictions, but also the observation-prediction pairs where the observed and CTM predicted ozone concentrations are paired across space and time. This PDF is expressed as

$$f_S(\mathbf{x}_m) = \prod_i^{n_m} f(x_i|\tilde{x}_i, \mathbf{p}_i) \quad (\text{E2-3})$$

which essentially characterizes how well each CTM offset-removed ozone value \tilde{x}_i predicts the true offset-removed ozone concentration x_i at the computational prediction point \mathbf{p}_i . Procedurally equation (3) is simply obtained by first calculating $f_S(\mathbf{z}_s) = \prod_i^{n_m} f(z_i|\tilde{z}_i, \mathbf{p}_i)$, where z_i and \tilde{z}_i are observed and CTM predicted ozone values, respectively, and then using the offset relationship $\mathbf{x}_s = \mathbf{z}_s - o_Z(\mathbf{p}_s)$ to obtain $f_S(\mathbf{x}_s)$.

As described by de Nazelle et al [31], the PDF $f(z_i|\tilde{z}_i, \mathbf{p}_i)$ is modeled using a parameterized statistical distribution, chosen to be the normal distribution truncated below zero with an expected value $\lambda_1(\tilde{z}_i)$ and variance $\lambda_2(\tilde{z}_i)$, such that:

$$f(z_i|\tilde{z}_i) = \Phi(z_i; \lambda_1(\tilde{z}_i), \lambda_2(\tilde{z}_i)) \quad (\text{E2-4})$$

In the soft data construction approach implemented by de Nazelle et al [31] the parameters $\lambda_1(\tilde{z}_i)$ and $\lambda_2(\tilde{z}_i)$ vary as a function of the model prediction \tilde{z}_i but are constant with respect to the space/time point \mathbf{p}_i , hence their implementation is based on a Constant Air quality Model Performance (CAMP). The CAMP approach was appropriate since in their application the air quality model performance did not change across their small study geographical domain (North Carolina) and short study duration (<15 days).

Our aim, however, is to extend the BME methodological framework to the national domain by modeling $\lambda_1(\tilde{z}_i, \mathbf{p}_i)$ and $\lambda_2(\tilde{z}_i, \mathbf{p}_i)$ as a function of both \tilde{z}_i and the space and time coordinate \mathbf{p}_i , expressed as below.

$$f(z_i|\tilde{z}_i, \mathbf{p}_i) = \Phi(z_i; \lambda_1(\tilde{z}_i, \mathbf{p}_i), \lambda_2(\tilde{z}_i, \mathbf{p}_i)) \quad (\text{E2-5})$$

Therefore, we need to investigate how the air quality model performance varies across the continental US.

2.3.2 Variability of CTM Model Performance Evaluation across the Continental US

Each observed daily concentration z_j is paired with its corresponding CTM prediction value \tilde{z}_j , and the error for the observation-prediction pair is defined as $e_j = \tilde{z}_j - z_j$. To evaluate the air quality model performance over a given space time region R of interest, we calculate error statistics such as the Mean prediction Error (ME), the Standard deviation of the prediction Error (SE), the mean normalized bias (MNB) and the mean normalized gross error (MNGE) as defined in SI equations A.1s-A.4s.

According to the model performance analysis of this CTM (see Supporting Information (SI) section A.1), for the DM8A O₃, we find that overall the CAMx simulation with $12 \times 12 \text{ km}^2$ grid cell resolution has a substantially lower over-prediction (median ME=+1.4ppb) than that with $36 \times 36 \text{ km}^2$ grid cell resolution (median ME=+4.5 ppb). Furthermore, as summarized in Figures A.2s and A.3s, the ME, SE, MNB and MNGE at individual monitoring sites vary over a wider range for the simulation with $36 \times 36 \text{ km}^2$ grid cell resolution. The variability of these ME and SE values exhibit clear geographical trends (SI figures A.4s-7s for the DM8A and figures A.12s-15s for the D24A): Urban cities located in the east and west coast tend to have higher over-prediction bias (i.e. higher ME) and higher imprecision (i.e. higher SE) than sites located in the central United States. We also found noticeable seasonal differences in the model performance for both CTM simulations (SI Figures A.8s-9s for the DM8A and figures A.16s-17s for the D24A).

The results of this analysis provide strong evidence that the performance of CTM varies considerably across the national domain and over seasons. Therefore, there is a need to extend the implementation of the BME framework to account for this space/time variability in model performance. We use the Regionalized Air Quality Model Performance (RAMP) method to quantify how the expected value $\lambda_1(\tilde{z}_i, \mathbf{p}_i)$ and variance $\lambda_2(\tilde{z}_i, \mathbf{p}_i)$ for the ozone soft data derived from CTM outputs vary as a function of both the CTM prediction \tilde{z}_i and the space/time computational node \mathbf{p}_i for which that prediction was calculated. The goal of the RAMP method is to select the most relevant observation-prediction pairs to most accurately identify the CTM bias associated with the prediction value \tilde{z}_i outputted for any space/time computational node \mathbf{p}_i of interest.

2.3.3 The Proposed Regionalized Air Quality Model Performance (RAMP) Evaluation Framework

In the first stage of the RAMP analysis, we pool for each monitoring site the observation-prediction pairs (z_j, \tilde{z}_j) that are within a time tolerance of $\Delta T = 120$ days of a particular time of interest t . Examples of two selected sites are shown in Figure 1. These pairs are highly relevant to the location \mathbf{s}_n where the monitoring station is sited, and the 120 days time window is chosen to balance the abundance of the pairs and the intention to retain seasonal specificity in the $\tilde{z}_j - z_j$ differences. We stratify the pairs in 10 equal percentile bins of increasing predicted values \tilde{z}_j , and for each bin we calculate the mean and variance of observed values,

$$\hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t) = \frac{1}{n_0(\tilde{z}_b, \mathbf{s}_n, t)} \sum_{j=1}^{n_0(\tilde{z}_b, \mathbf{s}_n, t)} z_j \quad (\text{E2-6})$$

$$\hat{\lambda}_2(\tilde{z}_b, \mathbf{s}_n, t) = \frac{1}{n_0(\tilde{z}_b, \mathbf{s}_n, t)} \sum_{j=1}^{n_0(\tilde{z}_b, \mathbf{s}_n, t)} (z_j - \hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t))^2 \quad (\text{E2-7})$$

where $n_0(\tilde{z}_b, \mathbf{s}_n, t)$ is the number of (z_j, \tilde{z}_j) pairs in the b^{th} bin, z_j is the j^{th} observation value in these $n_0(\tilde{z}_b, \mathbf{s}_n, t)$ pairs, and \tilde{z}_b is the average of the predictions \tilde{z}_j in these $n_0(\tilde{z}_b, \mathbf{s}_n, t)$ pairs.

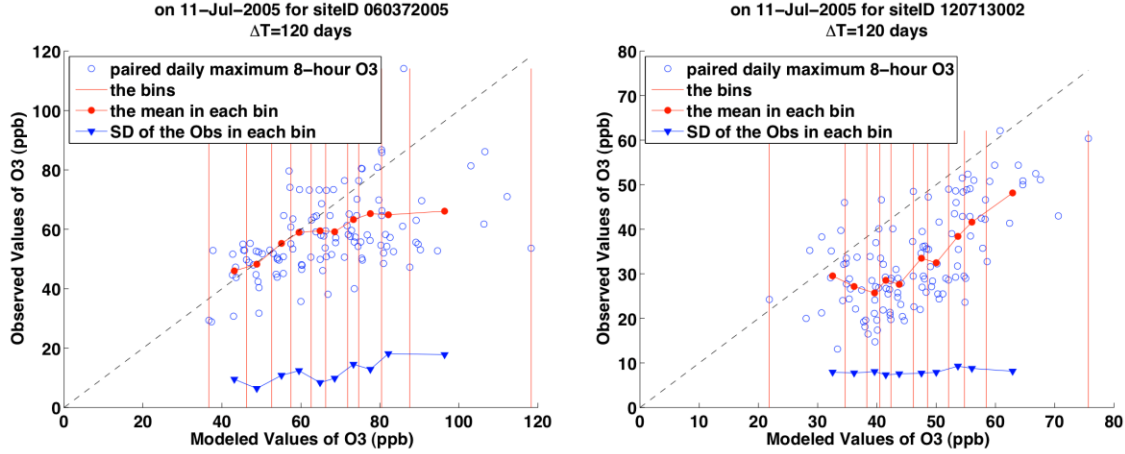


Figure 2. 1: RAMP analysis conducted specifically at time $t=11\text{-Jul-2005}$ and for sites ID 060372005(left) and 120713002(right). The empty circles show the pairs of observed-modeled values (z_j, \tilde{z}_j) that are within 120 days of t . The vertical lines show the stratification of these pairs in 10 bins. The interpolation lines connecting the filled circles and triangles show how the mean of the observed value in each bin,

$\hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t)$ (filled circles), and the corresponding standard deviation, $\sqrt{\hat{\lambda}_2(\tilde{z}_b, \mathbf{s}_n, t)}$ (filled triangles) change as a function of the average modeled value \tilde{z}_b in that bin.

In the second stage of the RAMP analysis we obtain $\lambda_1(\tilde{z}_i, \mathbf{p}_i)$ and $\lambda_2(\tilde{z}_i, \mathbf{p}_i)$ for actual predicted values \tilde{z}_i and space/time grid point $\mathbf{p}_i = (\mathbf{s}_i, t_i)$ as follows. For each monitoring site \mathbf{s}_n we perform a linear interpolation/extrapolation of the $\hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t_i)$ and $\hat{\lambda}_2(\tilde{z}_b, \mathbf{s}_n, t_i)$ values to obtain $\lambda_1(\tilde{z}_i, \mathbf{s}_n, t_i)$ and $\lambda_2(\tilde{z}_i, \mathbf{s}_n, t_i)$ (see interpolation lines in figure 3), and then we do a spatial interpolation of these values to obtain $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ and $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ at $\mathbf{p}_i = (\mathbf{s}_i, t_i)$ using the following formula

$$\hat{\lambda}_{1 \text{ or } 2}(\tilde{z}_i, \mathbf{s}_i, t_i) = \frac{\sum_{n=1}^N w_i(\mathbf{s}_i, \mathbf{s}_n) \hat{\lambda}_{1 \text{ or } 2}(\tilde{z}_i, \mathbf{s}_n, t_i)}{\sum_{n=1}^N w_i(\mathbf{s}_i, \mathbf{s}_n)}, \quad w_i(\mathbf{s}_i, \mathbf{s}_n) = \frac{1}{\text{dist}(\mathbf{s}_i, \mathbf{s}_n)} \quad (\text{E2-8})$$

where $n=1, \dots, N$ refers the N monitoring sites closest to the location of the computational node \mathbf{s}_i of interest, and $w_i(\mathbf{s}_i, \mathbf{s}_n)$ is a weight equal to the inverse of the distance between the computational node \mathbf{s}_i of interest and the n -th neighboring monitoring station.

Stated simply $\tilde{z}_i - \hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ is the bias characterizing systematic errors associated with a CTM prediction value of \tilde{z}_i calculated at space/time grid point $\mathbf{p}_i = (\mathbf{s}_i, t_i)$, and $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ is the variance characterizing the associated imprecision. The strength of the RAMP method is that it does not make any assumption on the relationship between observed and predicted values, and therefore geographical and temporal changes in non-linear and non-homoscedastic relationships are automatically captured in the

calculation of $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ and $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$, which are fully integrated in the BME soft data through equation (E2-4).

2.3.4 Offset analysis

The offset is used to transform the daily O3 data into residual offset-removed data. The ozone offset $o_z(\mathbf{p}_i)$ at an arbitrary location $\mathbf{p}_i = (\mathbf{s}_i, t_i)$ is obtained using an exponential kernel smoothing of the surrounding observed O3 data [30]

$$o_z(\mathbf{s}_i, t_i) = \sum_{j=1}^N w_j z_j / \sum_{j=1}^N w_j \quad (\text{E2-9})$$

where z_j is the observed value at space/time observation point $\mathbf{p}_j = (\mathbf{s}_j, t_j)$ within the neighborhood of the point \mathbf{p}_i of interest, and the kernel smoothing weights are $w_j = \exp(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{a_r} - \frac{|t_i - t_j|}{a_t})$, a_r is the spatial offset kernel smoothing range and a_t is the temporal offset kernel smoothing range.

An optimal offset ($a_r = 50 \text{ km}$ and $a_t = 10 \text{ day}$) was chosen to ensure the transformed data has a low variance so that the geostatistical estimation error variance is minimized, while retaining high autocorrelation to ensure that neighboring data locations are informative at the estimation location (see SI section A.2 for details).

2.3.5 Space-time Covariance Model

The covariance model for the homogeneous/stationary S/TRF $X(\mathbf{p})$ is developed from the experimental covariance of the transformed observational data $\mathbf{x}_o = \mathbf{z}_o - o_z(\mathbf{p}_o)$. The experimental covariance value for a spatial lag r and a temporal lag τ is calculated as

$$\hat{c}_X(r, \tau) = \frac{1}{N(r, \tau)} \sum_{j=1}^{N(r, \tau)} x_{head,j} x_{tail,j} - m_X^2 \quad (\text{E2-10})$$

Where $N(r, \tau)$ is the number of pairs of values $(x_{head,j}, x_{tail,j})$ separated by a spatial lag of r and temporal lag of τ , and m_X is the mean of the \mathbf{x}_o data. In practice $\hat{c}_X(r, 0)$ and $\hat{c}_X(0, \tau)$ are calculated and plotted separately to facilitate the visualization of the space/time covariance models (SI Figures A.30s & 31s). A 3-structured exponential covariance model was chosen for the subsequent BME analysis (see SI section A.2 for details).

The formula of the 3-structured exponential covariance model is given by:

$$c_X(r, \tau) = C_0 \left[\alpha \exp\left(\frac{-3r}{ar_1}\right) \exp\left(\frac{-3\tau}{at_1}\right) + \beta \exp\left(\frac{-3r}{ar_2}\right) \exp\left(\frac{-3\tau}{at_2}\right) + (1 - \alpha - \beta) \exp\left(\frac{-3r}{ar_2}\right) \exp\left(\frac{-3\tau}{at_1}\right) \right] \quad (\text{E2-11})$$

Tables A.2s&3s list the parameters used in the space time covariance model.

2.3.6 Validation analysis

A validation analysis is used to assess the accuracy of different BME estimation approaches. Each observed value z_j at space/time point $= (\mathbf{s}_j, t_j)$ is compared with the corresponding ozone concentration z_j^* re-estimated using only non-located data outside of a radius r_v of \mathbf{s}_j .

The validation error, which is the difference between each re-estimated value $z_j^*(r_v)$ and observed value z_j is defined as $e_j^*(r_v) = z_j^*(r_v) - z_j$. The estimation accuracy is quantified based on statistics of these estimation errors, which is a function of the validation radius r_v . They consist of the Root Mean Square Error RMSE (ppb), the R^2 (unitless), the Mean Normalized Bias MNB (%) and the Mean Normalized Gross Error MNGE (%) between observations and re-estimated values, calculated as a function of r_v as shown in SI equations A.E7s-10s.

Using the validation error statistics $RMSE(r_v)$ and $R^2(r_v)$, we compare the following three BME data fusion scenarios

1. Scenario OBS: Uses only ozone observations in the BME framework. This is the kriging limiting case of the BME data integration framework since kriging is the linear limiting case of BME when only hard data are used.
2. Scenario CAMP: Integrates both observations and CTM predictions in the BME data integration framework using CTM soft data constructed with the CAMP approach, which assumes that CTM model performance is constant across space and time.
3. Scenario RAMP: Integrates both observations and CTM predictions in the BME data integration framework, with CTM soft data obtained through the RAMP approach introduced here to account for the space/time variation in CTM model performance.

We let R^2_{OBS} , R^2_{CAMP} and R^2_{RAMP} be the coefficient of determination of estimation error for scenario OBS, scenario CAMP and RAMP, respectively, and we define the percent change $PCR^2_{OBS \rightarrow CAMP}$ and $PCR^2_{OBS \rightarrow RAMP}$ as follow:

$$PCR^2_{OBS \rightarrow CAMP} = 100 \frac{R^2_{CAMP}(r_v) - R^2_{OBS}(r_v)}{R^2_{OBS}(r_v)} \quad (E2-12)$$

$$PCR^2_{OBS \rightarrow RAMP} = 100 \frac{R^2_{RAMP}(r_v) - R^2_{OBS}(r_v)}{R^2_{OBS}(r_v)} \quad (E2-13)$$

A positive PCR^2 indicates an increase in R^2 , which corresponds to the percent improvement in estimation precision resulting in integrating air quality model predictions in the BME data integration.

2.4 Results

2.4.1 BME Ozone estimates

When incorporating CTM prediction as soft data in the BME data integration framework (scenarios CAMP and RAMP), we use the soft data with the finer grid cell resolution when it is available. That means for the areas where both $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution are available, we incorporate the $12 \times 12 \text{ km}^2$ CTM prediction values.

Figure 2.2 shows for Jul-11-2005 the BME estimates of DM8A ozone concentrations z_k obtained for the three estimation scenarios. This day was chosen because it has the highest standard deviation (at 33.1 ppb) for CTM prediction errors at ozone monitoring sites, which means the CTM model performance has the highest spatial variability among sites. It is clear that on this day the BME mean estimates (in the top panel of Figure 2.2) in the immediate proximity of the monitoring stations (marked in circles) are at very similar levels in the three maps, being in good agreement with the observed data in their local neighborhood. As the estimation location moves away from the monitoring stations, the difference among these three maps becomes more substantial. For example, in scenario OBS we see a wider area of high ozone value, with the area above 70ppb covering $811,296 \text{ km}^2$ across the continental United States. In scenario CAMP the ozone plume above 70ppb only covers a much smaller area ($545,184 \text{ km}^2$) with the peak ozone concentrations in the plume reaching 85ppb. By contrast, in scenario RAMP the size of the ozone plume (the area where ozone levels are above 70ppb) is $570,096 \text{ km}^2$ and the highest peak ozone concentration reaches 90ppb.

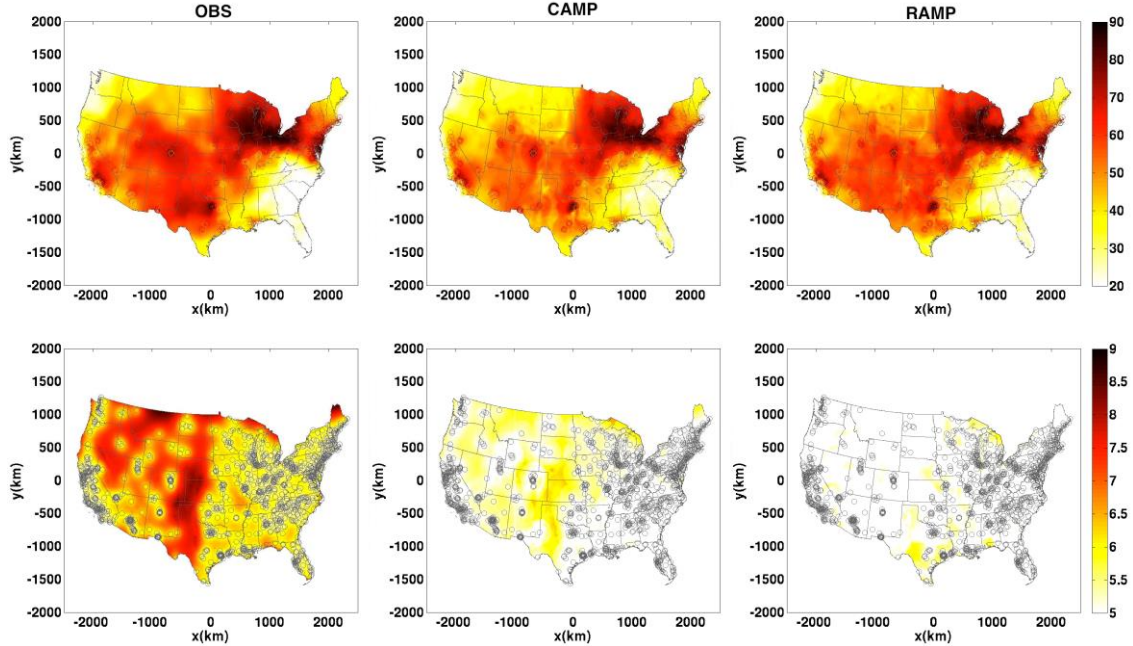


Figure 2. 2. Maps of BME mean estimates (**Top**) and corresponding standard deviations of BME estimates (**Bottom**) of the DM8A ozone concentrations (ppb) on day Jul-11-2005 obtained with scenario OBS (**Left**), CAMP (middle panels) and RAMP (**Right**). Circles in the maps represent locations of monitoring sites and color match legend for observed concentrations.

The uncertainty associated with the BME estimates are quantified by the corresponding BME standard estimation error (bottom panel of Figure 2.2). For estimation scenario OBS there is a higher estimation uncertainty, with the highest BME standard estimation error reaching 8.7ppb for areas far away from any monitoring stations, and with an average standard estimation error of about 6.4ppb across the continental United States. This is in contrast to estimation scenarios CAMP and RAMP, where the BME standard estimation error remains relatively low, with the highest standard estimation error reaching 6.5ppb and 6.3ppb for scenarios CAMP and RAMP. This indicates that integrating both observations and model predictions improved the quality of the ozone estimates, especially for areas far away from any monitoring station. Overall, scenario RAMP has the lowest standard estimation error, with an average standard estimation error of about 4.6ppb across the continental United States.

2.4.2 Soft data construction using the RAMP approach

The construction of the soft data using the proposed RAMP approach can be illustrated by comparing scenario RAMP that accounts for the space/time variability of CTM performance, with scenario CAMP that does not. Two important parameters that differed in these two scenarios are the bias-corrected expected values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ and the corresponding soft data variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$.

Figure 2.3 shows a map of the raw CAMx modeled DM8A average ozone predictions for 11-July-2005. Also shown in the figure are the bias-corrected CTM predicted values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ from scenarios CAMP and RAMP for the same day. Both scenarios CAMP and RAMP corrected the CTM prediction bias to some extent, especially for areas close to the monitoring sites. There are, however, substantial differences of the bias-corrected CTM predicted values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ between scenarios CAMP and RAMP. For scenario CAMP, the map of the bias-corrected CTM predicted values $\hat{\lambda}_1$ shows lower values than scenario RAMP, with the area above 70 ppb covering $221,616 \text{ km}^2$ across the continental United States and the highest bias-corrected CTM predicted value $\hat{\lambda}_1$ at 105 ppb. By contrast, in scenario RAMP the size of the area with bias-corrected CTM ozone levels greater than 70 ppb is $431,856 \text{ km}^2$, with the peak bias-corrected CTM ozone level reaching 111 ppb. This substantial difference is due to the assumed homogeneity of the CTM model performance in scenario CAMP that forces the same correction throughout the study domain. This correction results in an over correction in some local areas such as the area covering the western and southwestern states of Nevada, Idaho, Utah, Wyoming, Colorado, Arizona, New Mexico and California. In contrast, the scenario RAMP is better able to account for regional biases in model performance.

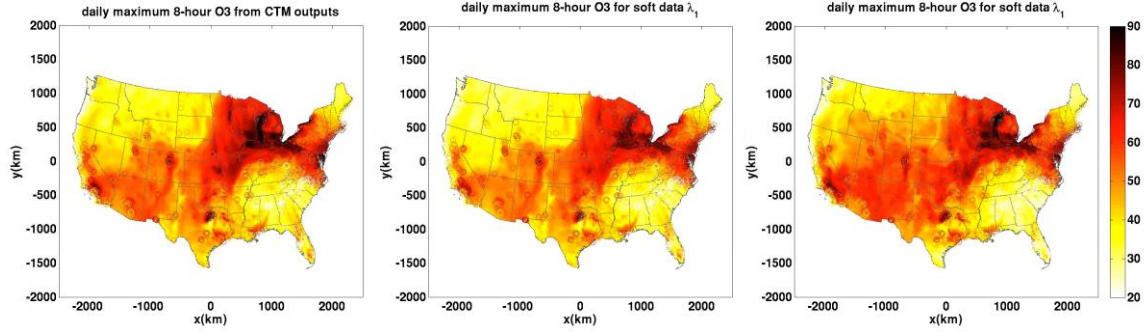


Figure 2. 3: The DM8A ozone concentrations in the United States on 11-July-2005 using **(Left)** the raw CTM Model predictions, **(Middle)** the bias-corrected expected values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ for the estimation scenario CAMP, and **(Right)** the bias-corrected expected values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ for the estimation scenario RAMP.

The maps for the corresponding square root of soft data variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ are shown in SI Figure A.34s. This map characterizes the imprecision associated with the bias-corrected CTM predicted values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$. We find that scenario RAMP has more localized gradients for the variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$, with the square root of $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ spanning from a low value of 2.6 ppb to a high of 20.5 ppb, and averaging about 8.7 ppb across the continental United States. By contrast, this variance has less spatial variability for the scenario CAMP; with a narrower span of $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ values ranging from 9.2 ppb to 14.1 ppb, and a higher average over the continental United States of 10.3 ppb. This illustrates that the proposed RAMP method used in scenario RAMP has a greater ability to characterize regional changes in the precision of bias-corrected CTM predictions. This is an important methodological improvement explaining the improved performance in scenario RAMP in the cross validation analysis described next.

2.4.3 Validation results

The validation analysis was conducted to compare estimation scenarios OBS, CAMP, and RAMP. For those monitoring site locations covered by both CTM domains, soft data with a finer grid cell resolution of $12 \times 12 \text{ km}^2$ are incorporated for scenarios CAMP and RAMP.

Table 2.1 shows the percent change in R^2 as a function of validation radius r_v . As shown in the table, scenario RAMP has the highest R^2 for all radii r_v . Furthermore, the $\text{PCR}^2_{\text{OBS} \rightarrow \text{CAMP}}$, representing the percent change in R^2 from scenario OBS to scenario CAMP, is consistently positive when r_v larger than 0 km , indicating that de Nazelle's approach, even when applied beyond the condition for which it was developed, is still more accurate than relying on observational data alone, and its percent increase in R^2 consistently improves as r_v increases. The $\text{PCR}^2_{\text{OBS} \rightarrow \text{RAMP}}$, representing the percent change in R^2 when comparing scenarios OBS and RAMP, is also consistently positive. More importantly, it is larger than $\text{PCR}^2_{\text{OBS} \rightarrow \text{CAMP}}$, with an overall 0.73% increase for the DM8A and 2.6% increase for the D24A O3 in R^2 for $r_c = 0 \text{ km}$. Furthermore, there is 2.9% increase for the DM8A and 5.9% increase for the D24A in R^2 between scenario OBS and RAMP at locations more than 108 km away from a monitoring station. We also calculated the percent change in R^2 when we aggregate the BME estimates into weekly and monthly estimates (SI Table A.5s). Scenario RAMP still has the highest R^2 increase for these aggregated metrics. More results of the cross-validation analysis are documented in SI section 5.

These results demonstrate that integrating both observations and soft data processed through the RAMP approach improves the capability of estimating both of the DM8A and the D24A ozone concentrations compared to using only observational data and through the CAMP approach. Compared with the CAMP approach, the RAMP approach consistently results in a further improvement in estimation, as evidenced by the fact that the $\text{PCR}^2_{\text{OBS} \rightarrow \text{RAMP}}$ values are over 12 times greater than the $\text{PCR}^2_{\text{OBS} \rightarrow \text{CAMP}}$ values for the DM8A ozone concentrations and over 3.5 times greater than the $\text{PCR}^2_{\text{OBS} \rightarrow \text{CAMP}}$ values for the D24A ozone concentrations.

Table 2. 1: Validation statistics for BME data integration scenarios OBS, CAMP, and RAMP*

Validation radius r_v (km)	DM8A				D24A			
	0	36	72	108	0	36	72	108
$RMSE_{OBS}$ (ppb)	5.536	6.309	6.799	7.041	5.705	6.178	6.303	6.422
$RMSE_{CAMP}$ (ppb)	5.675	6.442	6.966	7.250	5.803	6.222	6.383	6.545
$RMSE_{RAMP}$ (ppb)	5.445	6.109	6.531	6.732	5.487	5.835	5.917	6.004
R^2_{OBS} (unitless)	0.886	0.853	0.829	0.817	0.792	0.757	0.747	0.738
R^2_{CAMP} (unitless)	0.884	0.853	0.831	0.819	0.794	0.765	0.758	0.750
R^2_{RAMP} (unitless)	0.893	0.866	0.849	0.841	0.813	0.789	0.786	0.781
$PCR^2_{OBS \rightarrow CAMP}$ (%)	-0.223	0.014	0.197	0.235	0.230	1.100	1.480	1.617
$PCR^2_{OBS \rightarrow RAMP}$ (%)	0.726	1.602	2.407	2.936	2.642	4.338	5.190	5.898
p-value $OBS \rightarrow CAMP$ (unitless)	1	0.990	0.321	0.0009	0.259	0.00041	<0.00001	<0.00001
p-value $OBS \rightarrow RAMP$ (unitless)	0.980	0.0001	<0.00001	<0.00001	0.0725	<0.00001	<0.00001	<0.00001

* The analysis uses a constant offset and corresponding covariance model. $r_v(km)$ is the validation radius around monitoring stations within which all observation points are excluded in the validation estimations; R^2_{OBS} , R^2_{CAMP} , R^2_{RAMP} are the squared spearman's correlation coefficients between the ozone observations and the BME estimates for the OBS, CAMP and RAMP data integration scenarios, respectively, $RMSE_{OBS}$, $RMSE_{CAMP}$ and $RMSE_{RAMP}$ are the corresponding root mean square errors; $PCR^2_{OBS \rightarrow CAMP}$ is the percent change in R^2 from OBS scenario to CAMP scenario, $PCR^2_{OBS \rightarrow RAMP}$ is the percent change from scenario OBS to scenario RAMP; p-value $_{OBS \rightarrow CAMP}$ is the p-value testing the hypothesis that there is no difference between the R^2 in scenarios OBS and CAMP; p-value $_{OBS \rightarrow RAMP}$ is the p-value testing the hypothesis that there is no difference between the R^2 in scenarios OBS and RAMP.

2.5 Discussion

We have presented an ozone estimation method that is able to integrate observations with predictions from a CTM. These predictions are weighted according to model performance that varies across space and time based on a soft data construction utilizing the newly developed RAMP method. Thus, estimates are produced that put priority on observations and take advantage of air quality model predictions based on how well they reproduce the observed values. Spatial fields generated from this approach provide an observation and CTM informed representation of ozone across space/time that is more accurate and precise than relying only on observation data. This was especially true for locations away from monitoring stations.

We developed the RAMP method by extending the CAMP framework presented by de Nazelle et al[20]. We tested the RAMP method by comparing the percent change in R^2 and found the percent increase achieved by the RAMP method ($PCR^2_{OBS \rightarrow RAMP}$) was four to ten-fold greater than that of the CAMP method ($PCR^2_{OBS \rightarrow CAMP}$). This improvement is attributed to the RAMP ability to account for the spatial and temporal variability in model performance.

Approaches used to model the uncertainty associated with the CTM model predictions can be divided into parametric approaches that parameterize the relationship between the air pollution observations Z and predictions \tilde{Z} [17, 18], and non-parametric approaches such as our RAMP method that directly model air quality performance based on paired observed and predicted values. For example, Fuentes et al [17] assumes that $\tilde{Z}(\mathbf{s}) = \beta_0(\mathbf{s}) + \beta_1(\mathbf{s})Z(\mathbf{s}) + \varepsilon(\mathbf{s})$, while Berrocal et al [18] assumes that $Z(\mathbf{s}, t) = \beta_0(\mathbf{s}, t) + \beta_1(\mathbf{s}, t)\tilde{Z}(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$, where in both cases the relation is linear and homoscedastic since the noise term is assumed to have a constant error variance, i.e. $\varepsilon(\mathbf{s}, t) \sim N(0, \sigma_\varepsilon^2)$. By contrast, our novel RAMP approach is a non-parametric approach that fully accounts for the non-linear, non-homoscedastic relationship between observations Z and predictions \tilde{Z} , and accounts for the spatiotemporally varying nature of that relationship.

To illustrate the difference between parametric and non-parametric approaches, we also compared our non-parametric estimates to the ones generated from a cokriging estimation with a parametric relationship between the observations and the CTM model predictions (SI section A.6). In this analysis, we found that the disadvantage of cokriging is that it is limited by the parametric relationship and the final estimates tend to be heavily influenced by CTM model predictions. Based on the validation results (SI Table A.6s), our approach outperforms the Co-kriging approach in terms of smaller root mean squared error, 5.45 ppb for RAMP and 6.5 ppb for Co-kriging, and higher spearman's R^2 , 0.893 for RAMP and 0.845 for Co-kriging for the DM8A ozone.

To the best of our knowledge, our proposed framework is one of the first to fully account for the spatiotemporal variation of the non-linear, non-homoscedastic relationship between air pollution observations and predictions. Major strengths of our approach are that its numerical implementation is based on a straight forward analysis of paired observations and predictions, which is computationally efficient and trivially implemented on parallel computers, and it reduces the uncertainty of the mapping error by putting more weight on air quality predictions where they reproduce well the observed values. This is particularly useful in large regulatory or health studies that need to incorporate air quality predictions with widely varying model performance across the study domain, such as studies examining the entire continental United States rather than some small portions of it, or studies combining air quality predictions from a variety of air quality model simulations with significantly varying model performance.

2.6 Acknowledgments

This work was supported by National Institute on Aging (NIA) of the National Institutes of Health (NIH) under award number R01AG033078. It was also supported by a grant (T32ES007018) from the National Institute of Environmental Health Sciences (NIEHS) training program of NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The CAMx CTM modeling was performed by the U.S. Environmental Protection Agency (EPA), but this work has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency.

CHAPTER 3 – BEM INTEGRATION OF OZONE OBSERVATIONS AND CTM PREDICTIONS AT MULTIPLE TIME SCALES²

3.1 Introduction

For environmental epidemiologists and exposure scientists to assess the risk to the human population due to ozone exposure at national or regional scales requires accurate ambient ozone estimates at fine spatial and temporal resolutions. Most air pollution epidemiologic investigations, however, rely on ambient ozone estimates generated solely from air quality monitoring networks[4, 8, 25, 32, 33], such as the Air Quality System (AQS) operated by the U.S. Environmental Protection Agency (EPA). The AQS monitoring network has constraints in the geographic and temporal coverage, especially for the ozone-monitoring network. In addition to hourly averages and sparse national coverage the network has less than half of its sites operating year round. Only during the summer are all sites operating. Pollutant predictions from a chemical transport model (CTM) is another useful environmental data source that is used for generating estimates. CTM model predictions have the advantage of having continuous spatial coverage, but sometimes are limited in their temporal coverage. Further there is inherent error and uncertainty in CTM model outputs that can be difficult to quantify. The continued need for ambient air pollution estimates with longer time periods and larger spatial scales has pushed the exposure community to find ways to combine these two data sources.

² This chapter is planned to be submitted as an article to the Journal of Environmental Science & Technology. Xu, Yadong, Serre, L. Marc, Reyes, Jeanette, Vizuite, William. BME Integration of Ozone Observations and CTM predictions at Multiple Time Scales.

We have developed a methodology[34] that attempts to incorporate the spatial and temporal changes in CTM model performance to guide the integration of ozone observations into a BME estimation framework. We have shown that by integrating the dynamic spatial and temporal changes in CTM performance we can improve ambient ozone exposure estimates at a national scale. In our previous study[34], we first aggregated both of the ozone observational data and CTM predictions from a hourly to a daily average and then applied the BME framework to estimate the daily maximum 8-hour average (DM8A) and daily 24-hour average (D24A) ozone concentrations. The use of daily averages in our BME estimation framework substantially reduced the computational effort. If for example, hourly averages were used in the BME estimation framework it would require 200 times more CPU time. The disadvantage is the loss of information that could be found in hourly averages that could be used by the BME framework and potentially improves accuracy of the estimations.

Ideally the exposure estimates should be at time scales that are relevant to the health outcome. These times scales can range from hourly averages for research acute effects [4, 35] to monthly averages to study long-term health effects [32, 36]. In practice, many studies have used long-term arithmetic mean exposure estimates derived from small time scale estimates. Like in our previous study, the hourly observations and CTM model predictions for ozone could first be aggregated into the appropriate time scale then apply an interpolation technique to estimate exposure at the large time scale of interest. An alternative method would be applying interpolation at a finer time scale and then aggregating the BME estimated concentrations into a longer time scale. The benefit of this method is that it keeps the spatiotemporal structure of the original dataset. For ozone it would retain the characteristics of the diurnal pattern and hourly changes in CTM model performance. Having the estimates at a finer time scale provides the freedom to construct other hourly-based exposure metrics such as daily 1-hour maximum, daily 8-hour maximum or daily 24-hour averages depending on the need of the studies. It is unknown whether the substantial increase in computational burden is worth the additional information provided to the BME framework.

Although the cost of implementation of these two up-scaling methods are known, less is known about their impact on estimation accuracy. Yu et al[21] compared estimates of ambient PM10 and ozone using these two upscaling methods in their BME analysis. Their study showed mixed results for ozone estimates that USM1(data aggregation followed by BME spatiotemporal interpolation) had lower estimation error than USM2 (perform BME interpolation at small time scales followed by data aggregation) at weekly and monthly time scales, but USM2 performed better when the time scales increased to larger than three months to one year. These results are constrained by a limited observational data set. The BME analysis relied solely on the AQS observed ozone dataset from 77 ozone-monitoring stations in the Carolinas over year 1995-2002. Their study also did not start with hourly ozone concentrations, instead, daily averages were used and aggregated into longer time scales.

It is still unclear the extent to which keeping the diurnal pattern of ozone present in an hourly dataset would benefit the estimation accuracy when up-scaled to longer time scales. This work quantifies the impact in predictive accuracy when starting with hourly ozone concentrations contrasting it with the increased cost in computational burden. The work relies on a large national scale observational data set consisting of 1179 monitors for the year of 2005. National scale hourly CTM predictions for the entire year of 2005 are also used in a BME estimation framework.

3.2 Data Sources

The ozone observational data are hourly ozone monitoring data sampled from 1179 sites for the year 2005 downloaded from the Air Quality Systems (AQS) database maintained by the U.S. Environmental Protection Agency (EPA). Details on the processing and quality assurance of this data can be found in Xu et al[34].

The CTM model predictions consist of hourly ozone concentrations for the entire year of 2005 predicted by the Comprehensive Air Quality Model with extensions (CAMx) modeling system. Model prediction has a spatial grid cell resolution of $36 \times 36 \text{ km}^2$ covering the continental U.S. and a $12 \times 12 \text{ km}^2$ grid cell resolution domain covering the Eastern U.S. Detailed model configurations and evaluation are documented in the final Transport Rule by the U.S. EPA[28].

3.3 Methods

3.3.1 BME Estimation Methodology

A detailed description of the BME framework used to generate ozone estimates can be found in Xu et al. 2016[34]. In short, first the AQS hourly data was paired with CTM modeled hourly ozone concentrations in space/time. Then, localized bias-corrected CTM data were constructed as the soft data through RAMP approach, where the hourly ozone CTM predictions are weighted according to model performance that varies across space and time. A transformation of these data was used, which consisted in removing from the data an offset obtained using an exponential kernel smoothing of the data. The exponential kernel smoothing was set so that the offset captured the spatial variability of the data over intermediate spatial distances and intermediate time scales. A 3-term exponential/exponential/cosine space/time covariance model was used to characterize the space/time autocorrelation in the offset removed hourly ozone data. Finally, we conduct a BME interpolation that 1) uses the Maximum Entropy principle of information theory to process the general knowledge base G-KB consisting of the mean function and the covariance function of ozone 2) integrates the site specific knowledge base S-KB using an epistemic Bayesian conditionalization rule to create a BME posterior PDF f_K characterizing the value x_k taken by $X(\mathbf{p})$ at any estimation point \mathbf{p}_k of interest, and 3) computes space/time estimates based on the BME posterior PDF. The S-KB consisted of hourly observations treated as hard data and localized bias-corrected hourly CTM data treated as soft data.

3.3.2 Variability of CTM Model Performance Evaluation across the Continental US

An important feature of this methodology is that it fully evaluates the spatial and temporal variability in the CTM performance. For this evaluation each observed hourly concentration z_j is paired with its corresponding CTM prediction value \tilde{z}_j , and the error for the observation-prediction pair is defined as $e_j = \tilde{z}_j - z_j$. To evaluate the CTM model performance over a given space time region R of interest, we calculate error statistics such as the Mean prediction Error (ME) and the Standard deviation of the prediction error (SE), the mean normalized bias (MNB) and the mean normalized gross error (MNGE) as defined in SI equations 1s-4s.

Detailed results of the nationwide model performance analysis of this CTM are provided in SI section 1. In brief, we find that overall the CAMx simulation with $12 \times 12 \text{ km}$ grid cell resolution has a substantially lower over-prediction (median ME=+3.4ppb) than that with $36 \times 36 \text{ km}$ grid cell resolution (median ME=+6.1 ppb). There are clear geographical trends in the variability of these error statistics: Urban cities located in the east and west coast tend to have higher over-prediction bias (i.e. higher ME) and higher imprecision (i.e. higher SE) than sites located in the central United States. The seasonal difference in model performance was also analyzed by recalculating the error statistics at each site separately for the summer (May, June, July and August) and winter (November, December, January and February). The over-prediction bias is noticeably higher in the summer, with the median ME for the summer equal to 4.1 ppb compared to 2.3 ppb for the winter, while the median MNB (237%) and MNGE (267%) values for the winter are much higher than for the summer with the median MNB at 129% and the median MNGE at 143%, indicating the CTM model's difficulty at capturing lower ozone concentrations in the winter.

Generally, the CTM reproduces hourly ozone concentrations with much higher prediction errors than the daily DM8A and D24A ozone values. For example, in the CAMx simulation with 12x12km² grid cell resolution, the median SE for hourly ozone is 13.2 ppb, which is much higher than the ones for DM8A at 9.4 ppb and for D24A at 7.9 ppb. The differences in MNB and MNGE are even greater between the hourly and daily metrics, with the median MNB at 145% for hourly ozone compared to the ones for DM8A at 9.3% and for D24A at 19.8% (See more detailed error statistics in SI table B.1s). This can be explained by the fact that CTM models tend to severely over-predict nighttime hourly ozone concentrations at very low values and under-predict peak hourly ozone concentrations.

3.3.3 Regionalized Air Quality Model Performance (RAMP) Analysis for Hourly Ozone

The RAMP evaluation analysis is conducted in two stages. In the first stage we analyze CTM performance at specific monitoring stations, and in the second stage we perform an inverse weighted distance interpolation to assess CTM performance away from monitoring stations.

In the first stage of the RAMP analysis, we focus specifically on each monitoring site, and for each monitoring site we pool the observation-prediction pairs (z_j, \tilde{z}_j) that are within a time tolerance ΔT of a particular time of interest t . These pairs are highly relevant to the location \mathbf{s}_n where the monitoring station is located. To be consistent with our previous study, we keep the same $\Delta T = 120$ days for the hourly ozone. Examples of two selected sites are shown in Figure 3.1. We stratify the pairs in 10 percentile bins of increasing predicted values \tilde{z}_j , and for each bin we calculate the mean $\hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t)$ and variance $\hat{\lambda}_2(\tilde{z}_b, \mathbf{s}_n, t)$ of observed values (see detailed equations in Xu et al).

In the second stage of the RAMP analysis we obtain $\lambda_1(\tilde{z}_i, \mathbf{p}_i)$ and $\lambda_2(\tilde{z}_i, \mathbf{p}_i)$ for actual predicted hourly values \tilde{z}_i at space/time grid point $\mathbf{p}_i = (\mathbf{s}_i, t_i)$. We first perform a linear interpolation/extrapolation of the $\hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t_i)$ and $\hat{\lambda}_2(\tilde{z}_b, \mathbf{s}_n, t_i)$ values at each monitoring site to obtain $\lambda_1(\tilde{z}_i, \mathbf{s}_n, t_i)$ and $\hat{\lambda}_2(\tilde{z}_i, \mathbf{s}_n, t_i)$ at the predicted values. Then we use inverse weighted distance to do a spatial interpolation of these values from the neighboring monitoring stations to obtain $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ and $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ at $\mathbf{p}_i = (\mathbf{s}_i, t_i)$.

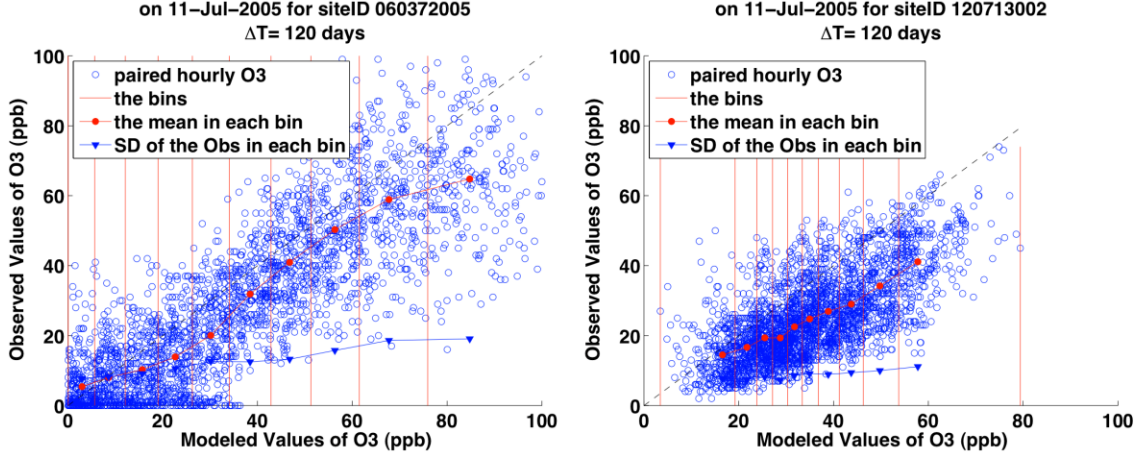


Figure 3. 1: RAMP analysis conducted specifically at time $t=11\text{-Jul-2005}$ and for sites ID 060372005(left) and 120713002(right). The empty circles show the pairs of hourly observed-modeled values (z_j, \tilde{z}_j) that are within 120 days of t . The vertical lines show the stratification of these pairs in 10 percentile bins. The interpolation lines connecting the filled circles and triangles show how the mean of the observed value in each bin, $\hat{\lambda}_1(\tilde{z}_b, \mathbf{s}_n, t)$ (filled circles), and the corresponding standard deviation, $\sqrt{\hat{\lambda}_2(\tilde{z}_b, \mathbf{s}_n, t)}$ (filled triangles) change as a function of the average modeled value \tilde{z}_b in that bin.

3.3.4 Offset analysis

The offset is used to transform the hourly ozone data into residual offset-removed data. The ozone offset $o_Z(\mathbf{p}_i)$ at an arbitrary location $\mathbf{p}_i = (\mathbf{s}_i, t_i)$ is obtained using an exponential kernel smoothing[30] of the surrounding observed ozone data. Several sets of kernel smoothing ranges, including the spatial range a_r and the temporal range a_t , are investigated. An optimal set of kernel smoothing ranges ($a_r = 50 \text{ km}$ and $a_t = 10 \text{ day}$) were chosen to ensure the transformed data has a low variance so that the geo-statistical estimation error variance is minimized, while retaining high autocorrelation to ensure that neighboring data locations are informative at the estimation location. These offset ranges are also consistent with what we used for the daily metrics. (See SI section 2 for details).

3.3.5 Space-time Covariance Model

The covariance model for the homogeneous/stationary S/TRF $X(\mathbf{p})$ is developed from the experimental covariance of the transformed observational data $\mathbf{x}_o = \mathbf{z}_o - o_Z(\mathbf{p}_o)$. The experimental covariance value for a spatial lag r and a temporal lag τ is calculated and plotted separately to facilitate the visualization of the space/time covariance models. Several covariance models were attempted and evaluated. A 3-structured exponential/exponential/cosine covariance model was chosen for the subsequent BME analysis (see SI section B.2 for details).

The formula of the 3-structured exponential/cosine covariance model is given by:

$$c_X(r, \tau) = C_0[\alpha \exp\left(\frac{-3r}{ar_1}\right) \exp\left(\frac{-3\tau}{at_1}\right) + \beta \exp\left(\frac{-3r}{ar_2}\right) \exp\left(\frac{-3\tau}{at_1}\right) + (1 - \alpha - \beta) \exp\left(\frac{-3r}{ar_2}\right) \cos\left(\pi \frac{\tau}{at_2}\right)] \quad (\text{E3-1})$$

Where C_0 is the sill (variance), a_{r1} and a_{r2} are the spatial ranges and a_{t1} and a_{t2} are the temporal ranges, and α , β and $(1-\alpha-\beta)$ are the proportions of variability contributed by the first, second and third covariance structure, respectively. The parameters α , β , a_{r1} , a_{r2} , a_{t1} , and a_{t2} obtained by least square fitting for each offset are shown in SI table B.3s. Compared with the covariance models for the daily metrics, the major difference of the covariance model for the hourly ozone is the cosine function in the third structure of equation 1. This function represents the diurnal pattern occurring in the observational ozone data, which was altered by the aggregation process for the daily metrics.

3.3.6 Validation analysis

A validation analysis is used to assess the accuracy of two BME estimation upscaling methods USM1 and USM2. Each observed value z_j at space/time point $\mathbf{p}_j = (s_j, t_j)$ is compared with the corresponding ozone concentration z_j^* re-estimated using only non-collocated data outside of a radius r_v of s_j . The validation error, which is the difference between each re-estimated value $z_j^*(r_v)$ and observed value z_j is defined as $e_j^*(r_v) = z_j^*(r_v) - z_j$. The estimation accuracy is quantified based on statistics of these estimation errors, which is a function of the validation radius r_v . They consist of the Root Mean Square Error RMSE (ppb), the R^2 (unitless), the Mean Normalized Bias MNB (%) and the Mean Normalized Gross Error MNGE (%) between observations and re-estimated values, calculated as a function of r_v shown as below.

$$RMSE(r_v) = \sqrt{\frac{1}{N_o} \sum_{j=1}^{N_o} (z_j^*(r_v) - z_j)^2} \quad (E3-2)$$

$$R^2(r_v) = \left(\frac{\sum_{j=1}^{N_o} (z_j^*(r_v) - \bar{z}^*(r_v))(z_j - \bar{z})}{\sqrt{\sum_{j=1}^{N_o} (z_j^*(r_v) - \bar{z}^*(r_v))^2} * \sqrt{\sum_{j=1}^{N_o} (z_j - \bar{z})^2}} \right)^2 \quad (E3-3)$$

$$MNB(r_v) = \frac{1}{N} \sum_{j=1}^{N_o} 100\% * (z_j^*(r_v) - z_j) / z_j \quad (E3-4)$$

$$MNGE(r_v) = \frac{1}{N} \sum_{j=1}^{N_o} 100\% * (|z_j^*(r_v) - z_j| / z_j) \quad (E3-5)$$

We first calculate the validation error statistics for the hourly ozone estimates to evaluate the improvement in predictive capacity between two estimation scenarios, with the OBS scenario using ozone observations only and the RAMP scenario incorporating the CTM model predictions through the RAMP approach.

Our next step is to compare the following data fusion simulations (4 simulations for each daily metrics) as in list in Table 1 to investigate the differences in the predictive capacity between USM1 and USM2.

Table 3. 1: The list of BME data integration simulations used in the validation analysis

To simulate DM8A			To simulate D24A		
Simulation name	Upscaling method	Soft data Scenario[#]	Simulation name	Upscaling method	Soft data Scenario[#]
DM8A-1	USM1	OBS	D24A-1	USM1	OBS
DM8A-2	USM2	OBS	D24A-2	USM2	OBS
DM8A-3	USM1	RAMP	D24A-3	USM1	RAMP
DM8A-4	USM2	RAMP	D24A-4	USM2	RAMP

[#] Scenario OBS: only use ozone observations in BME data integration; Scenario RAMP: use both of ozone observations and soft data processed through RAMP approach

To investigate the influence of the CTM grid cell resolution on the accuracy of the BME estimates, we also conduct our validation analysis by using two sets of soft data, one processed from CAMx outputs with 36x36km² grid cell resolution and the other from CAMx outputs with 12x12km² grid cell resolution, for those monitoring sites located both in 36x36km² and 12x12km² modeling domains.

3.4 Results

3.4.1 BME Ozone estimates

To visualize the differences of BME ozone estimates from two up-scaling methods USM1 and USM2, we generate the maps for a selected day (11-July-2005), this day is chosen because it has the highest standard deviation (at 33.1ppb) for CTM prediction errors for DM8A at ozone monitoring sites. To create a map using up-scaling method USM1, we directly conduct BME interpolation with daily DM8A and D24A observations and the corresponding soft data generated through the RAMP approach. To create a map using up-scaling method USM2, We first do a BME interpolation of hourly ozone for all the hours within this selected day. Then we obtain the maps of two daily metrics of ozone concentrations, DM8A and D24A, by aggregating hourly ozone maps with considering the time shift in different time zones in the continental United States. The computational costs and efforts of generating the maps using USM2 is substantially higher than the ones using USM1, with the average CPU time for USM2 requiring 3403 hours compared to the CPU time of 13.4 hours for USM1 when generating a map with 36x36km² grid cell resolution for one selected day (See SI table B.4s).

Figure 3.2 shows for Jul-11-2005 the BME estimates of daily ozone concentrations and their difference from methods USM1 and USM2. The maps are quite similar at capturing the high ozone plume (DM8A over 70ppb) in Wisconsin and Michigan and also the low ozone values (DM8A less than 30ppb) in Georgia and Florida. In fact differences in ozone concentrations were below 5ppb for most of the continental U.S. For DM8A, the averaged difference across the continental U.S is 4.8ppb with the largest difference of 32.5 ppb occurring in Kentucky. For the D24A, the averaged difference is 3.3 ppb with the largest difference of 29.6 ppb. Additional maps for estimation scenario OBS demonstrating the differences between USM1 and USM2 are provided in SI Section B.3.

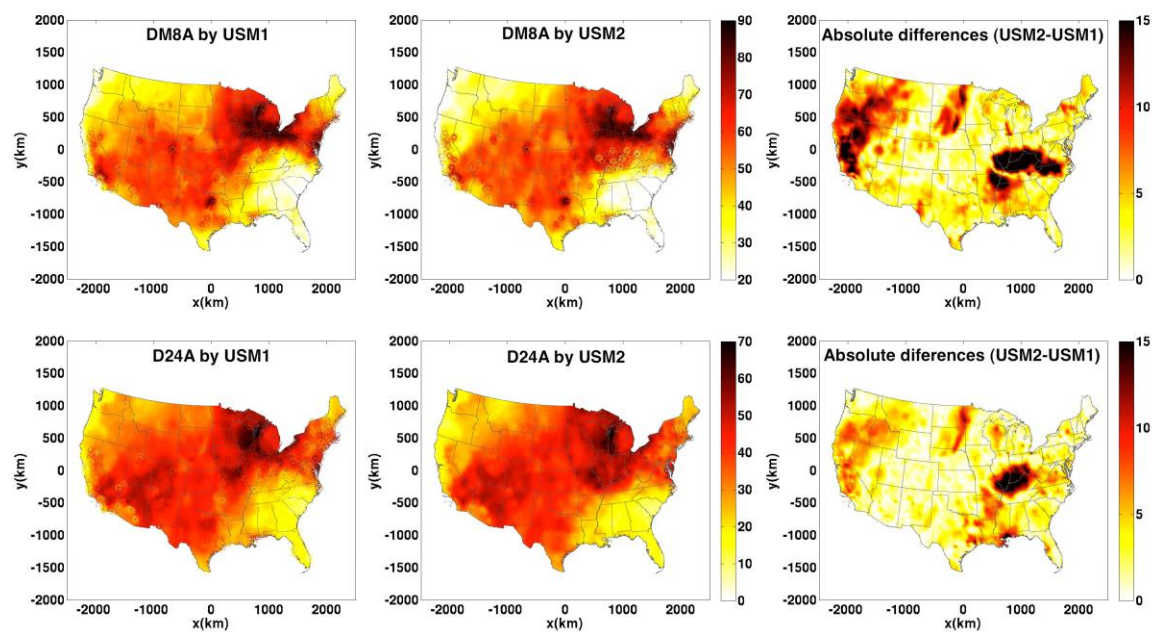


Figure 3. 2. Maps of Jul-11-2005 ozone BME estimates in ppb of the DM8A (**Top**) and the D24A (**Bottom**) obtained from upscaling method USM1 (**Left**) and USM2 (**Middle**). Also shown are the absolute differences (USM2-USM1) between these two methods (**Right**).

3.4.2 Validation results

Table 3.2 Column 1 shows error statistics of hourly ozone concentrations as a function of validation radius r_v between all the observed and re-estimated data points using USM2. We see a R^2 increase and $RMSE$ decrease in scenario RAMP within each validation radius. As shown in the table, the $PCR^2_{OBS \rightarrow RAMP}$, representing the percent change in R^2 when comparing scenarios OBS and RAMP, is consistently positive. These results demonstrate that integrating both observations and soft data processed through RAMP approach improves the capability of estimating hourly ozone concentration values compared to using only observational data.

To investigate seasonal differences in BME estimate performance for hourly ozone concentrations we calculated error statistics by separating validation datasets into summer (May-August) and winter months (January, February, November and December). The results are summarized in Table 1. The R^2 for summer months are much better than the winter months, with R^2 at 0.807 for summer and the corresponding R^2 at 0.693 for the winter when the validation radius R_v at 0km for Scenario RAMP. This is due to increased number of ozone monitoring in the summer. The denser observational data in the summer provides more informative knowledge in regards to the ozone concentrations in the local neighborhood of the estimated space/time location during the BME interpolation process.

There are larger improvements for scenario RAMP over Scenario OBS during the winter months at a validation radius of 0km, 36km and 72km. This is not the case, however, for the larger validation radius of 108km. There are two major factors contribute to this mixed results. First, the soft data constructed from the CTM model predictions have a larger influence on the BME estimates in winter months due to the sparseness of the observational data. Secondly, the CTM ozone model performance in winter is worse compared with the one in the summer as we described in Section 3.3. This has the result of increasing the uncertainty of the soft data during winter months and reduces the accuracy of the BME estimates when the validation radius increases.

Table 3. 2. Validation statistics hourly O3 for BME data integration scenarios OBS and RAMP [#]

Validation radius r_v (km)	Overall				Summer (May/June/July/August)				Winter (Jan/Feb/Nov/Dec)			
	0	36	72	108	0	36	72	108	0	36	72	108
$RMSE_{OBS}$ (ppb)	8.612	9.520	10.00	10.29	8.841	9.876	10.59	10.96	7.757	8.609	8.726	8.850
$RMSE_{RAMP}$ (ppb)	8.525	9.531	9.963	10.15	8.784	10.00	10.57	10.80	7.599	8.412	8.626	8.721
R^2_{OBS} (unitless)	0.794	0.748	0.723	0.707	0.803	0.754	0.718	0.698	0.679	0.605	0.595	0.583
R^2_{RAMP} (unitless)	0.800	0.759	0.737	0.728	0.807	0.762	0.736	0.725	0.693	0.634	0.615	0.605
$PCR^2_{OBS \rightarrow RAMP}$ (%)	0.753	1.424	2.050	2.989	0.519	1.003	2.480	3.822	2.011	4.697	3.266	3.642

[#] The analysis uses a constant offset and corresponding covariance model. $r_v(km)$ is the validation radius around monitoring stations within which all observation points are excluded in the validation estimations; R^2_{OBS} , R^2_{RAMP} are the squared spearman's correlation coefficients between the ozone observations and the BME estimates for the OBS and RAMP data integration scenarios, respectively, $RMSE_{OBS}$ and $RMSE_{RAMP}$ are the corresponding root mean square errors; $PCR^2_{OBS \rightarrow RAMP}$ is the percent change from scenario OBS to scenario RAMP.

To compare the predictive capacity between the upscaling methods USM1 and USM2, we summarize the error statistics of the estimation error in Table 3.3 for two daily ozone metrics, DM8A and D24A. For the DM8A, the USM1 performs better than USM2 yielding lower estimation error for both the OBS and RAMP scenarios. In the case of D24A, for OBS scenario, the values of RMSE and R^2 for USM1 and USM2 are very similar, with less than 0.3% of the absolute differences. For RAMP scenario, the USM1 is generally better than USM2 with slightly higher R^2 (0.813 vs 0.809). Using USM1 also results a bigger improvement, in terms of the R^2 percentage change, for RAMP scenario over the OBS scenario. This can be explained by the fact that the predictive accuracy of USM2 need to rely on the hourly soft data processed from CTM model predictions. As we have seen in SI table B.1, the uncertainty associated with the hourly ozone model performance is over 40% higher, in terms of the median SE, compared with the one for DM8A. This higher uncertainty in hourly model predictions was transferred into less informative soft data, which cause the difficulty of final hourly BME estimates to accurately predict hourly ozone concentrations. This ultimately provides less chance to capture the daily values of DM8A in the later aggregation process.

Table 3. 3. Compare validation statistics for DM8A and D24A O3 between the upscaling methods USM1 and USM2 ^{##}

Daily metrics	Validation radius r_v (km)	USM1				USM2			
		0	36	72	108	0	36	72	108
DM8A	$RMSE_{OBS}$ (ppb)	5.536	6.309	6.799	7.041	5.683	6.566	7.455	7.909
	$RMSE_{RAMP}$ (ppb)	5.445	6.109	6.531	6.732	6.206	7.420	8.127	8.416
	R^2_{OBS} (unitless)	0.886	0.853	0.829	0.817	0.884	0.845	0.803	0.781
	R^2_{RAMP} (unitless)	0.893	0.866	0.849	0.841	0.882	0.846	0.816	0.805
	$PCR^2_{OBS \rightarrow RAMP}$ (%)	0.726	1.602	2.407	2.936	-0.226	0.110	1.539	3.070
D24A	$RMSE_{OBS}$ (ppb)	5.705	6.178	6.303	6.422	5.722	6.170	6.292	6.455
	$RMSE_{RAMP}$ (ppb)	5.487	5.835	5.917	6.004	5.486	5.819	5.997	6.100
	R^2_{OBS} (unitless)	0.792	0.757	0.747	0.738	0.792	0.759	0.749	0.738
	R^2_{RAMP} (unitless)	0.813	0.789	0.786	0.781	0.809	0.789	0.777	0.770
	$PCR^2_{OBS \rightarrow RAMP}$ (%)	2.642	4.338	5.190	5.898	2.121	3.949	3.804	4.354

^{##} R^2_{OBS} , R^2_{RAMP} are the squared spearman's correlation coefficients between the ozone observations and the BME estimates for the OBS and RAMP data integration scenarios, respectively, $RMSE_{OBS}$ and $RMSE_{RAMP}$ are the corresponding root mean square errors; $PCR^2_{OBS \rightarrow RAMP}$ is the percent change from scenario OBS to scenario RAMP.

Based on these results we can conclude the extent that the RAMP approach can make better estimates depends on the accuracy of CTM predictions. This hypothesis was tested by investigating the differences in estimation accuracy of the daily ozone BME estimates when using two sets of soft data with different uncertainty levels. The two sets of soft data were processed from CTM model predictions with different grid cell resolutions. The ozone model performance with a finer grid cell resolution (12x12km²) is more accurate than with the coarser grid cell resolution (36x36km²). We created two more estimation scenarios using USM1. For all 888 monitoring sites located within both of the CTM modeling domains two sets of soft data were created for BME estimation. One set of data was processed through CTM model predictions with 12x12km² grid cell resolution and the other through the ones with 36x36km² grid cell resolution.

Based on the cross-validation results summarized in Table 3.4, we see that the BME interpolation with finer grid cell resolution ($12 \times 12 \text{ km}^2$) resulted in an overall lower estimation error and a higher R^2 . We also calculated the R^2 and RMSE on each individual monitoring site to quantify the improvements contributed by using the finer grid cell resolution soft data. We found that, for the daily metric D24A, the majority of the sites (539 out of 888 sites) have a R^2 increase when using finer grid cell resolution soft data, while 676 out of 888 sites have a RMSE decrease when using soft data with $12 \times 12 \text{ km}^2$ grid cell resolution compared with the one with $36 \times 36 \text{ km}^2$ grid cell resolution. For the daily metric DM8A, we also see a significant improvement at a majority of the sites when using soft data with finer grid cell resolution (See SI figure B.21s&22s). The better ozone CTM model performance in the simulation with finer grid cell resolution can explain this improvement. As we mentioned in the previous section, the CTM model predictions with $12 \times 12 \text{ km}^2$ grid cell resolution have a smaller prediction error and less uncertainty, which lead to a set of soft data with more informative secondary information.

Table 3. 4. Validation statistics for BME data integration scenarios OBS and RAMP using different soft data (for 888 sites only) for upscaling method USM1

Daily metrics	Validation radius r_v (km)	BME with 36x36km soft data				BME with 12x12km soft data			
		0	36	72	108	0	36	72	108
DM8A	$RMSE_{OBS}$ (ppb)	5.174	5.835	6.270	6.525	5.174	5.835	6.270	6.525
	$RMSE_{RAMP}$ (ppb)	5.202	5.820	6.238	6.499	5.113	5.683	6.088	6.340
	R^2_{OBS} (unitless)	0.900	0.873	0.853	0.841	0.900	0.873	0.853	0.841
	R^2_{RAMP} (unitless)	0.901	0.876	0.860	0.849	0.905	0.884	0.868	0.858
	$PCR^2_{OBS \rightarrow RAMP}$ (%)	0.073	0.411	0.765	0.917	0.565	1.246	1.734	1.959
D24A	$RMSE_{OBS}$ (ppb)	5.256	5.596	5.693	5.863	5.256	5.596	5.693	5.863
	$RMSE_{RAMP}$ (ppb)	5.261	5.485	5.590	5.744	5.037	5.279	5.373	5.512
	R^2_{OBS} (unitless)	0.809	0.784	0.777	0.763	0.809	0.784	0.777	0.763
	R^2_{RAMP} (unitless)	0.815	0.798	0.793	0.782	0.830	0.816	0.811	0.802
	$PCR^2_{OBS \rightarrow RAMP}$ (%)	0.776	1.781	2.026	2.408	2.658	4.045	4.445	5.060

3.5 Discussion

We have applied the novel RAMP approach to integrate ozone observations with CTM model predictions to generate ambient hourly ozone estimates across the United States. We see a small improvement, with PCR^2 at 0.753% when the validation radius R_v at 0km, of Scenario RAMP over Scenario OBS in the hourly ozone estimates. This can be explained by the higher uncertainty in hourly CTM model predictions, with the median MNB of 145% for hourly ozone in contrast to the one of 9.3% for DM8A (more details shown in SI Table B.1s). The higher uncertainty in hourly CTM model predictions was captured into the soft data constructed through the RAMP approach, which led to less informative secondary information integrated with the observations.

We then compared the predictive capacity of two upscaling methods USM1 and USM2. The validation analysis demonstrated unexpected results that the USM1 is generally better than USM2 in terms of predictive accuracy when incorporating the soft data processed from CTM model predictions, while nearly no predictive differences between these two methods were found when using the observations only. This result confirms that our RAMP approach can accurately capture the variability in CTM model performance therefore was able to provide secondary information according to the CTM model performance. When there is high uncertainty among CTM model performance, like the one in hourly ozone concentrations, RAMP was able to capture the magnitude of the uncertainty and transferred it into less informative secondary information, this is the reason we saw a less improvement of Scenario RAMP over Scenario OBS for the upscaling method USM2.

On the other hand, we can also reasonably expect that RAMP approach will be able to make a larger improvement if we integrate the observations with a CTM having better model performance. To test this we integrated the same observations with two sets of soft data processed from CTM model predictions with different grid cell resolutions. The ozone model performance with a finer grid cell resolution (12x12km²) was known to be better than the one with coarser grid cell resolution (36x36km²). The validation analysis confirmed again that the RAMP approach was able to capture the better CTM model performance with finer grid cell resolution. Scenario RAMP made a larger improvement over Scenario OBS with over 2-7 times of R² increase percentage when compared with the ones with coarser grid cell resolution.

Our findings are particularly useful for large-scale (large spatial coverage and long time span) epidemiologic studies like WHIMS. These studies involve many participants from different metropolitan areas, it is important to capture the spatial temporal characteristics of local air pollution in the ambient exposure estimates. Due to the long time span (over 10 years for the WHIMS) of the study design, the selection of the most appropriate modeling approach so that the exposure assessment can be completed in a computational efficient manner is also a key point to consider. Our work provides a clear recommendation that using our BME framework to directly integrate daily ozone observations with CTM model predictions through RAMP approach, which can achieve higher improved accuracy and also associated with much lower

computational cost. It also calls the attention of using better CTM model predictions to generate more accurate air pollution estimates.

CHAPTER 4 – CONCLUSIONS

To improve the accuracy of national scale ambient ozone exposure estimates, we focus on developing a BME modeling framework to integrate ozone observations and model predictions from a CTM. BME's flexibility allowed the assimilation of the general knowledge base of G-KB and the specific knowledge base of S-KB with different formats. We developed an approach (RAMP) that was able to account for the spatial and temporal variability in ozone CTM model performance (Chapter 2). We generated ozone estimates of two commonly used daily metrics, DM8A and D24A, for the continental U.S. for an entire year. The ozone estimates that were produced prioritized observations, but also took advantage of CTM model predictions accounting for how well they reproduce the observed values. The validation analysis indicated that the RAMP approach can improve the accuracy and precision with a 12 fold increase in R^2 percentage change for DM8A and over 3.5 times for D24A compared to estimates that assume CTM model performance does not change across space and time. The improvement is attributed to RAMP being able to characterize the spatial heterogeneity and temporal variability of ozone model performance across the country. Thus, CTM model predictions were weighted according to model performance that varies across space and time. This confirmed our speculation that de Nazelle et al's homogeneous model performance assumption is not applicable for our study domain. Spatial fields generated from our BME model provide an observation and CTM informed representation of ozone across space/time that is more accurate and precise than relying only on observation data or using soft data that does not account for the spatial and temporal variability in CTM model performance. This was especially true for locations away from monitoring stations when observational data is not available and our BME estimates heavily rely on the bias-corrected soft information to the extent that CTM model predictions are deemed accurate with respect to observed values in the local neighborhood.

Approaches used to model the uncertainty associated with the CTM model predictions can be divided into parametric approaches that parameterize the relationship between the air pollution observations and CTM model predictions and non-parametric approaches such as our RAMP method that directly model air quality performance based on paired observed and predicted values. The parametric approaches rely on the assumption that the relation between air pollution observations is linear and homoscedastic. By contrast, our novel RAMP approach is a non-parametric approach that fully accounts for the non-linear, non-homoscedastic relationship. A validation analysis show that our model also outperforms the cokriging approach in terms of a smaller RMSE (5.4 ppb vs 6.5ppb) and a higher R^2 (0.893 vs 0.845). This is attributed to the disadvantage of the assumed parametric prediction-observation relationship in cokriging which makes the final estimates heavily influenced by inaccurate CTM model predictions.

In our work to support the Women's Health Initiative (WHI) Memory Study, the BME approach was used to interpolate directly the daily ozone concentrations by first aggregating the hourly observations and CTM model predictions. An alternative approach would be to first generate hourly BME estimates then aggregate it into a daily metrics. This alternative approach could be especially useful for those epidemiologic studies that require higher temporal resolution of ambient exposure estimates, such as those exposure models combining microenvironmental concentrations with human activity data to estimate personal exposures. This could be relevant given the known diurnal patterns seen in hourly ozone data, but it requires over 200 times more CPU runtime than our previous work. In Chapter 3, we investigated the extent of the improvement on the accuracy of the hourly ozone estimates when incorporating CTM hourly model predictions through our RAMP approach. Our validation analysis indicates that incorporating hourly soft data through RAMP approach achieved a small improvement, with PCR^2 at 0.753% when the validation radius R_v at 0km, for hourly ambient ozone estimates compared to using hourly observations only. This can be explained by the inaccuracies found in the hourly CTM model predictions. When compared with observations the hourly CTM model predictions had a median MNB of 145% while it was 9.3% for DM8A. The higher uncertainty in hourly CTM model predictions is well

characterized by our RAMP approach and determined the extent this data is used in generating BME estimates.

We also investigated the differences of the predictive capacity between two upscaling methods: USM1 (data aggregation from hourly to daily followed by BME approach estimation) and USM2 (perform BME approach estimation on hourly ozone followed by data aggregation). A validation analysis was completed using only non-collocated data outside of a validation radius r_v and the R^2 between observations and re-estimated daily ozone concentration using the two upscaling methods. The validation analysis produced unexpected results that the less computationally intensive USM1 is generally better than USM2 in terms of predictive accuracy when incorporating the soft data processed from CTM model predictions. We also found that there were no predictive differences between these two methods when using the observations only. This result confirms that only with our RAMP approach can we accurately capture the variability in CTM model performance and therefore provide secondary information for a BME estimate. When there is high uncertainty among CTM model performance, like for hourly ozone concentrations, RAMP was able to capture the magnitude of the uncertainty and transferred it into less informative secondary information, this is the reason we saw a less improvement of Scenario RAMP over Scenario OBS for the upscaling method USM2.

Based on these results we can conclude the extent that the RAMP approach can make better estimates depends on the accuracy of CTM predictions. This hypothesis was tested by investigating the differences in estimation accuracy of the daily ozone BME estimates when using two sets of soft data with different uncertainty levels. The two sets of soft data were processed from CTM model predictions with different grid cell resolutions. The ozone model performance with a finer grid cell resolution ($12 \times 12 \text{ km}^2$) is more accurate than with the coarser grid cell resolution ($36 \times 36 \text{ km}^2$). The validation analysis confirmed again that the RAMP approach was able to capture the better CTM model performance with finer grid cell resolution. Scenario RAMP made a larger improvement over just using observations with over 2-7 times of R^2 increase percentage when compared with the ones with coarser grid cell resolution.

4.1 Scientific findings

To the best of our knowledge, our proposed framework is one of the first to fully account for the spatiotemporal variation of the non-linear, non-homoscedastic relationship between air pollution observations and predictions. Major strengths of our approach are that its numerical implementation is based on a straight forward analysis of paired observations and predictions, which is computationally efficient and trivially implemented on parallel computers, and it reduces the uncertainty of the mapping error by putting more weight on air quality predictions where they reproduce well the observed values. Our findings are particularly useful for large-scale (large spatial coverage and long time span) epidemiologic studies like WHIMS. These studies involve many participants from different metropolitan areas, it is important to capture the spatial temporal characteristics of local air pollution in the ambient exposure estimates. Due to the 10 year time span of the study design, the selection of the most appropriate modeling approach so that the exposure assessment can be completed in a computational efficient manner is also a key point to consider. Our work provides a clear recommendation that using our BME framework to directly integrate daily ozone observations with CTM model predictions through RAMP approach, which can achieve higher improved accuracy and also associated with much lower computational cost. It also calls the attention of using better CTM model predictions to generate more accurate air pollution estimates.

4.2 Methodological aspects and study limitations

In our work, as one of the limitations of our approach, we treated the daily ozone observations, processed from hourly-observed ozone concentrations, as hard data for computational efficiency. First, we assumed that there is no difference between the true mean of the hourly concentrations and the sample mean calculated from a finite sample size. This assumption may be investigated in the future work. One of the alternative approaches is to use a finite population correction factor to estimate the uncertainty associated with the difference between the true mean of hourly ozone concentrations and the sample mean calculated from an incomplete sample size. Secondly, we ignored those days when the completeness of ozone hourly observations do not meet the 75% criteria proposed by U.S. EPA. An alternative approach would be to incorporate these ignored data as soft site-specific data since the BME method can always express incomplete information with probability functions with arbitrary shapes.

Another limitation of our current application is the grid resolution of $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ of the air quality model, might be too low to assess exposures near local sources. For example, to study ozone impacts near roadways and high traffic densities may require much higher resolution or more sophisticated techniques such as variable-grid resolutions or adaptive grid resolution in CTM models. According to the sensitivity analysis of the influence of the grid cell resolution on the accuracy of the BME estimates (Chapter 3), we found that BME interpolation with finer grid cell resolution $12 \times 12 \text{ km}^2$ can achieve an overall lower estimation error and a higher R^2 compared with the one with $36 \times 36 \text{ km}^2$ grid cell resolution. This improvement can be attributed to the better ozone CTM model performance in the simulation with finer grid cell resolution, which led to a set of soft data with more informative secondary information. This finding support the idea that we should incorporate CTM model predictions with higher grid cell resolution to obtain more spatial temporal resolved air pollution estimates in the future.

The third methodological limitation is that there are multiple ways to generate soft data from a CTM. Our current application generated the soft data by only examining how ozone is reproduced in a CTM model outputs without incorporating other variables, which is computationally efficient and straightforward. However, there are other more sophisticated methods might be worth being explored. For example, other information, such as land-use covariates, population density and meteorological information could be incorporated into the soft data construction process to provide more informative correction of the CTM model performance.

4.3 Uncertainties and future research

There are several sources of uncertainty in our BME estimates. First of all, BME interpolation process for both projects put priority on ozone hourly or daily observed concentrations. Therefore, the estimation quality of our current application is heavily impacted by the completeness of the observational data. As we can see in our validation analysis for hourly ozone estimates, we have a much worse (with R^2 at 0.693 for winter vs 0.807 for summer) model performance during the winter months because of the scarceness of the winter ozone measurements. Secondly, there are inherent uncertainties in CTM modeling, which can be attributed primarily to the uncertainties in chemical mechanism or uncertainties associated with meteorology and emission inputs. The modeling community has continued to reduce model uncertainties through improved physical parameterizations, refined emission inventory, integrated weather observations and data assimilation advancements.

Although no technique will completely eliminate uncertainty, with the available CTM model simulations, there is still a lot of space for improvement. As we mentioned in the previous sections, BME is a knowledge-processing framework. The better we understand the sensitivity of CTM model predictions to variable spatial temporal conditions, the better we can extract the site-specific knowledge from a CTM. Our future research can explore new approaches to generate more informative soft data from a CTM. For example, it is noted that clouds or shortwave radiation differences in meteorological models can dominate daytime ozone variability[37], which result a higher uncertainty in the simulated ozone concentrations than

clear sky days. We could incorporate cloud cover as an additional variable into our soft data construction process. Thus, the paired observation-prediction dataset can be further examined based on different meteorological conditions.

APPENDIX A – SUPPORTING INFORMATION FOR BAYESIAN MAXIMUM ENTROPY INTEGRATION OF OZONE OBSERVATIONS AND MODEL PREDICTIONS: A NATIONAL APPLICATION³

This supporting information provides descriptions of (A.1) ozone CTM model performance evaluation for daily metrics DM8A and D24A, (A.2) the parameter selection of the offset analysis and the covariance model for ozone DM8A and D24A (A.3) the additional maps of BME O₃ estimates, (A.4) the statistical test for the significant increase in R^2 in the validation analysis, (A.5) the additional results of cross-validation analysis and (A.6) the cokriging approach using a parametric relationship between observations and CTM model predictions, (A.7) a discussion about the limitation of using the daily averages of the hourly ozone observations as hard data

A.1 Ozone CTM model performance evaluation for daily metrics DM8A and D24A

In our research, the air quality model predictions were used to construct the soft data, consisted of the expected values and the uncertainties associated with the expected values at each grid cell, based on how well they reproduce the observed values. The ozone monitoring sites along with the CAMx 36x36km² and 12x12km² modeling domains are shown in Figure A1s.

³ This appendix previously appeared as the supporting information for an article in the Journal of Environmental Science & Technology. The original citation is as follows: Xu, Yadong, Serre, L. Marc, Reyes, Jeanette, Vizuete, William. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: A National Application. Environmental Science & Technology (2016) 50 (8), 4393-4400.

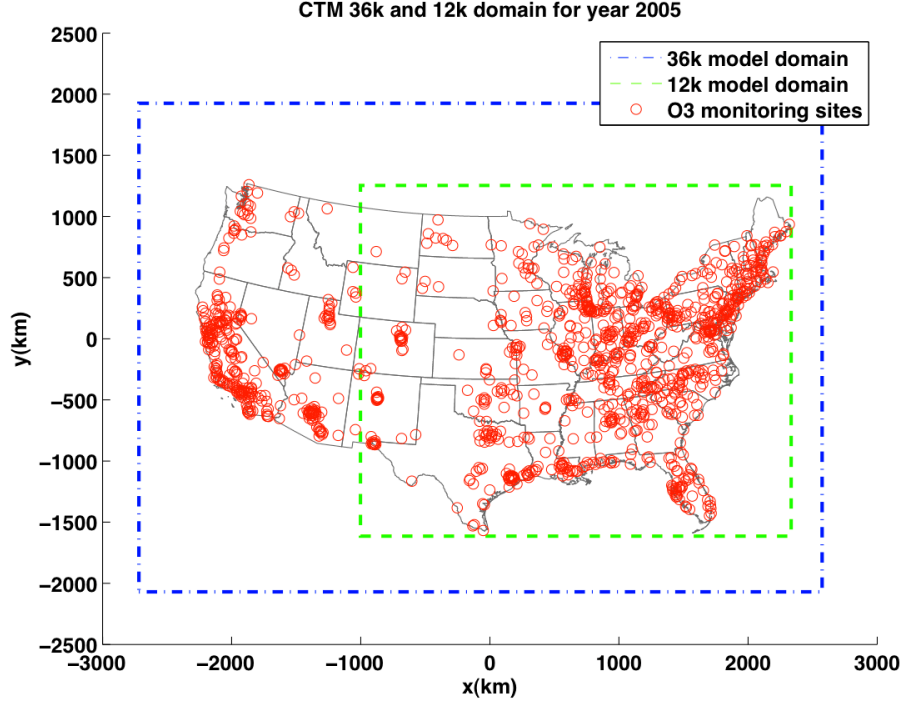


Figure A. 1s: The ozone monitoring sites (circles), CAMx modeling domain with 36x36km² grid cell resolution (dash-dotted line rectangle) and CAMx modeling domain with 12x12km² grid cell resolution (dashed line rectangle).

To develop a meaningful way to characterize model performance we need to investigate how the air quality model performance changes across space and time. To evaluate the overall air quality model performance for daily O₃ concentrations, the modeled daily concentrations are paired with the observational data in space and time at each monitoring site. We calculated error statistics such as mean prediction error (ME), the standard deviation of the prediction error (SE), the mean normalized bias (MNB) and the mean normalized gross error (MNGE) for the DM8A and D24A ozone concentrations on each site.

$$ME(R) = \frac{1}{N(R)} \sum_{j=1}^{N(R)} (\tilde{z}_j - z_j) \quad (A1s)$$

$$SE(R) = \sqrt{\frac{\sum_{j=1}^{N(R)} ((\tilde{z}_j - z_j) - ME(R))^2}{N(R) - 1}} \quad (A2s)$$

$$MNB(R) = \frac{1}{N(R)} \sum_{j=1}^{N(R)} 100\% * (\tilde{z}_j - z_j) / z_j \quad (A3s)$$

$$MNGE(R) = \frac{1}{N(R)} \sum_{j=1}^{N(R)} 100\% * (|\tilde{z}_j - z_j|/z_j) \quad (A4s)$$

Where z_j is the observed daily concentration, \tilde{z}_j is its corresponding CTM prediction value and $N(R)$ is the number of observation-prediction pairs in region R .

The model performance evaluation results for both metrics are summarized in the following sections.

A.1.1 CTM model performance evaluation for the DM8A ozone concentrations

To compare the CAMx model performance for those sites (888 sites in total) covered by both of the domains, we summarize the model prediction error statistics of DM8A, including ME, SDE, MNB and MNGE, in the boxplots shown in Figure A2s and A3s. The variability of ME and SE values across individual sites is shown in the maps of figures A4s - A7s.

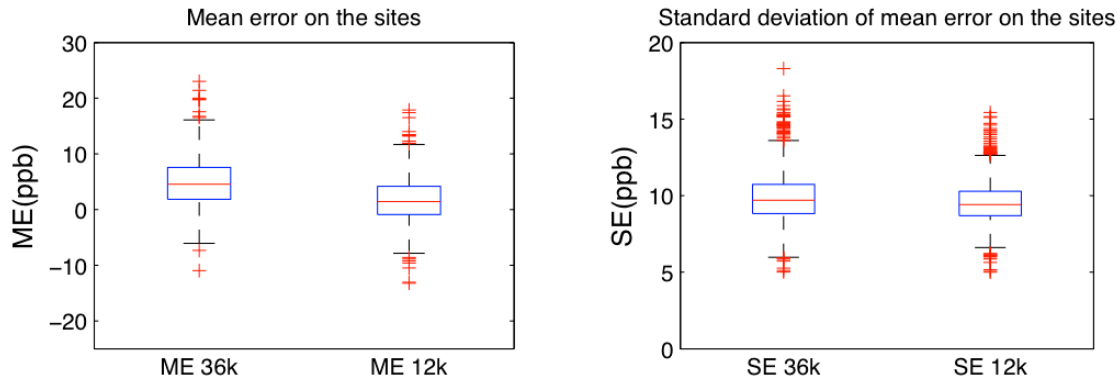


Figure A. 2s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the DM8A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with 36x36km² and 12X12km² grid cell resolution.

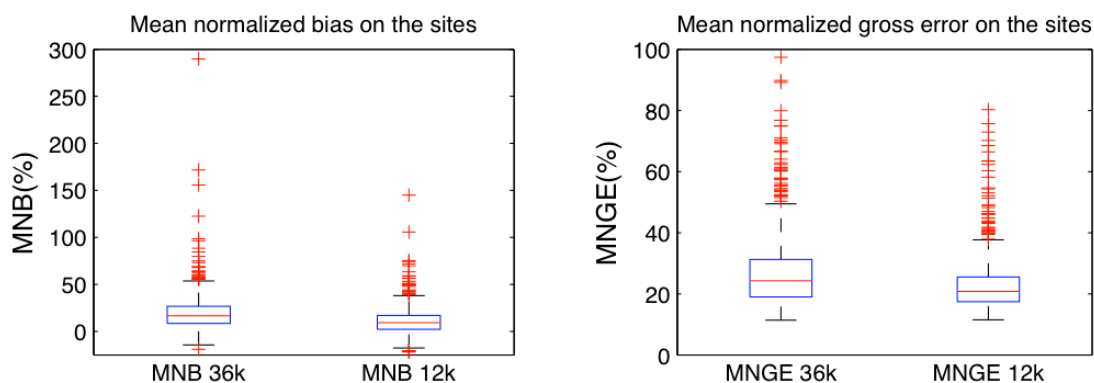


Figure A. 3s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the DM8A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution.

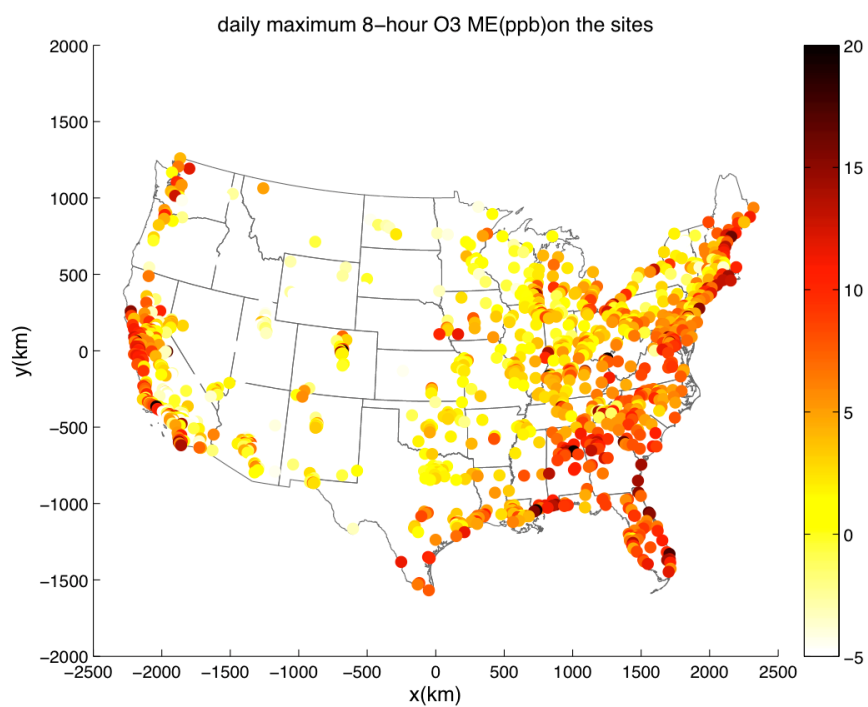


Figure A. 4s: The DM8A O₃ mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $36 \times 36 \text{ km}^2$ grid cell resolution.

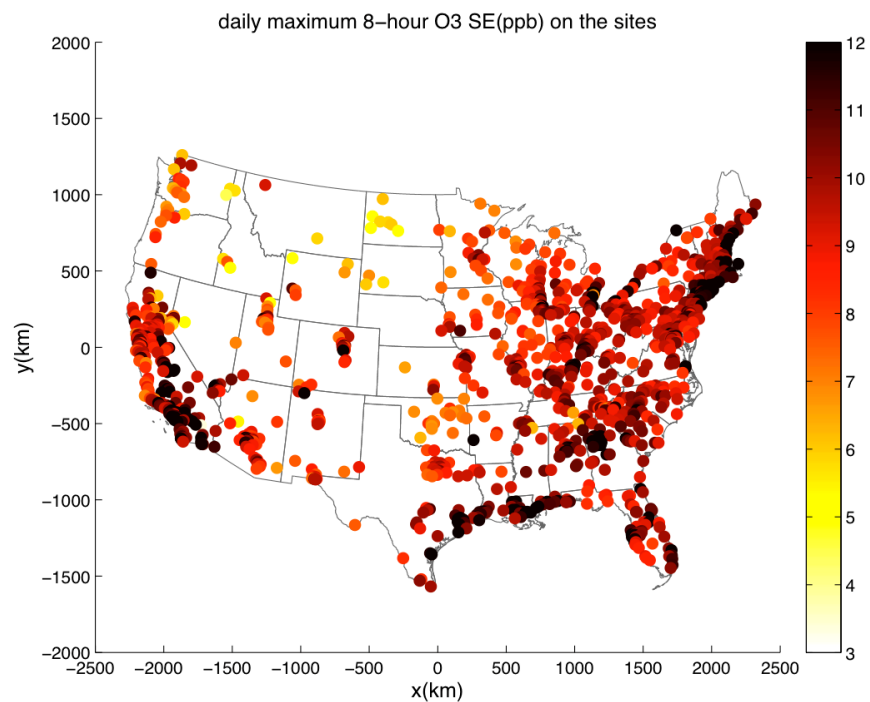


Figure A. 5s: The DM8A O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km² grid cell resolution.

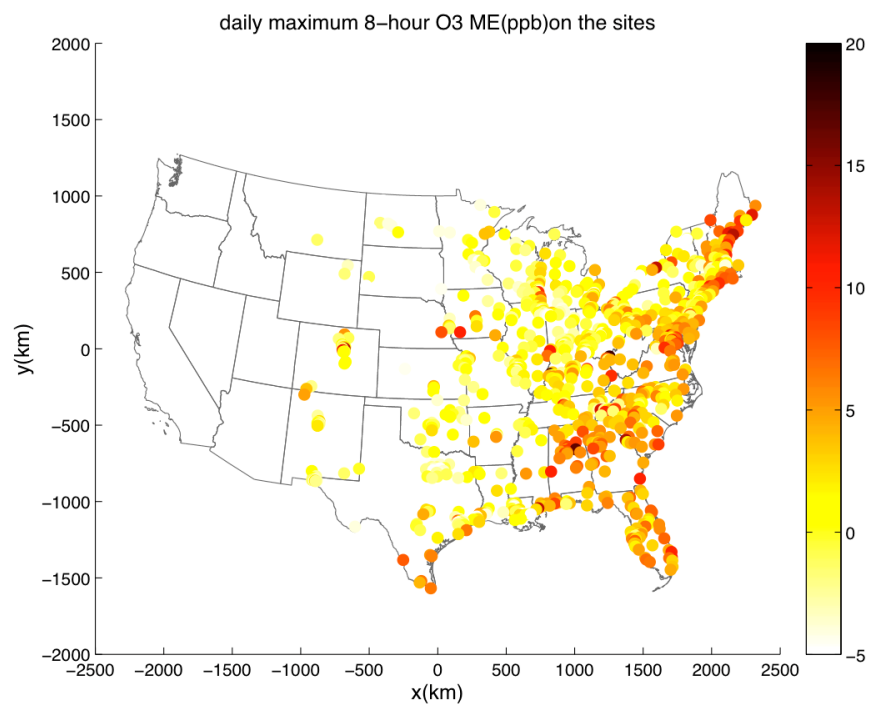


Figure A. 6s: The DM8A O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km² grid cell resolution.

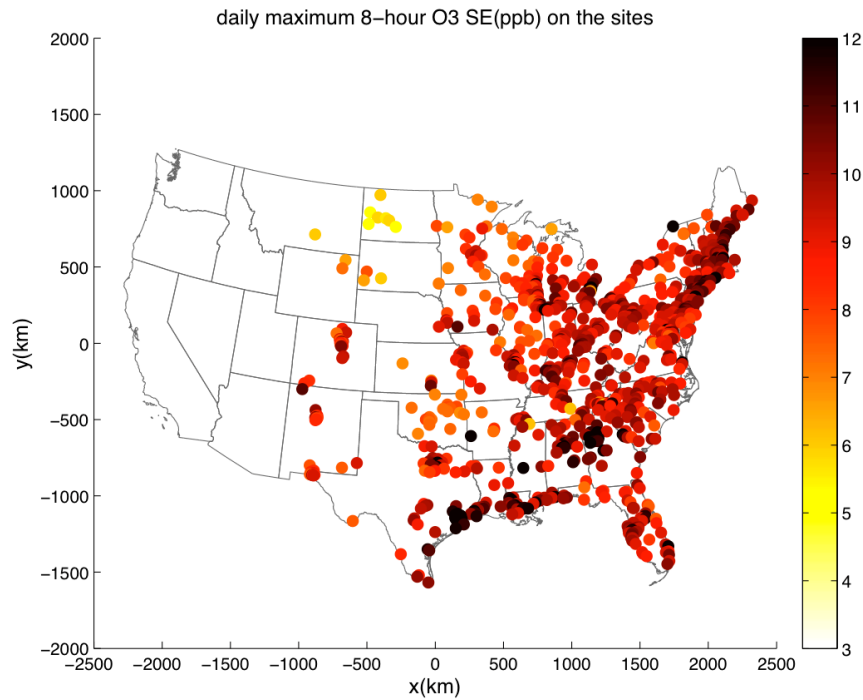


Figure A. 7s: The DM8A O₃ standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km² grid cell resolution.

The seasonal difference in model performance was also analyzed by recalculating the ME and SE at each site separately for the summer (May, June, July and August) and winter (November, December, January and February). For 36x36km² simulation, the over-prediction bias is noticeably higher in the summer, with the median ME for the summer equal to 4.1 ppb compared to 2.3 ppb for the winter. The Interquartile range (IQR) of the SE is 6.05 ppb in the winter, while it is only 1.5 ppb in the summer, indicating that the spatial variability of the prediction imprecision is much larger in the winter.

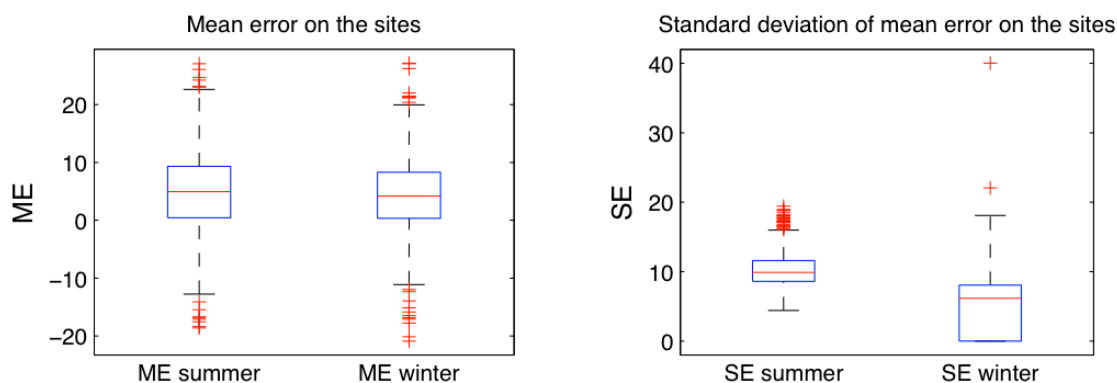


Figure A. 8s: Boxplots of DM8A O₃ mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 36x36km² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

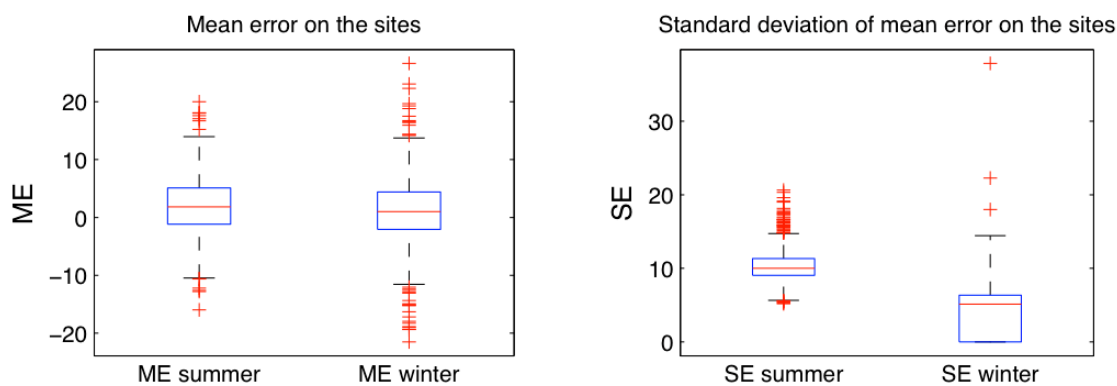


Figure A. 9s: Boxplots of DM8A O₃ mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 12x12km² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

A.1.2 CTM model performance evaluation for the D24A ozone concentrations

To compare the CAMx model performance for those sites (888 sites in total) covered by both of the domains, we summarize the model prediction error statistics of D24A ozone, including ME, SDE, MNB and MNGE, in the boxplots shown in Figure A10s and A11s. The variability of ME and SE values across individual sites is shown in the maps of figures A12s - A15s.

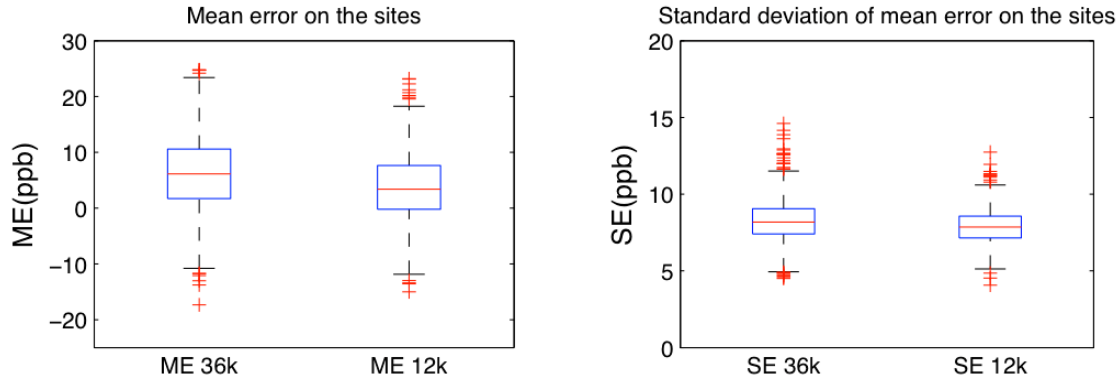


Figure A. 10s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of D24A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution.

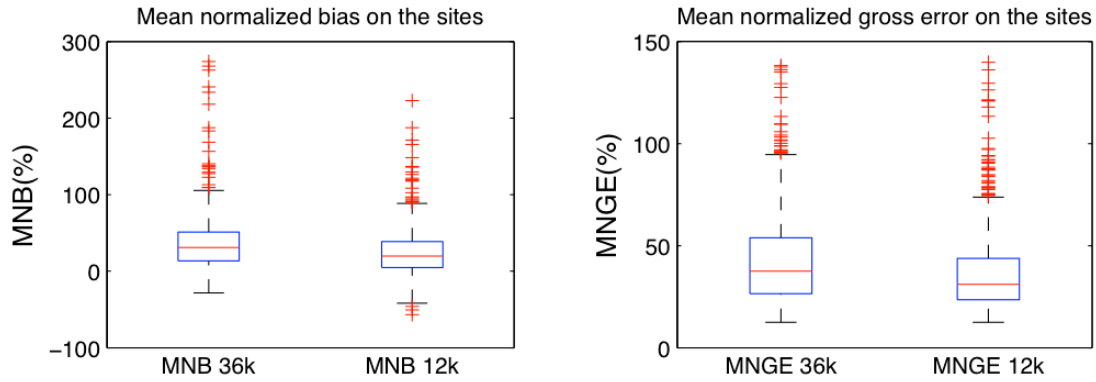


Figure A. 11s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of D24A at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with $36 \times 36 \text{ km}^2$ and $12 \times 12 \text{ km}^2$ grid cell resolution.

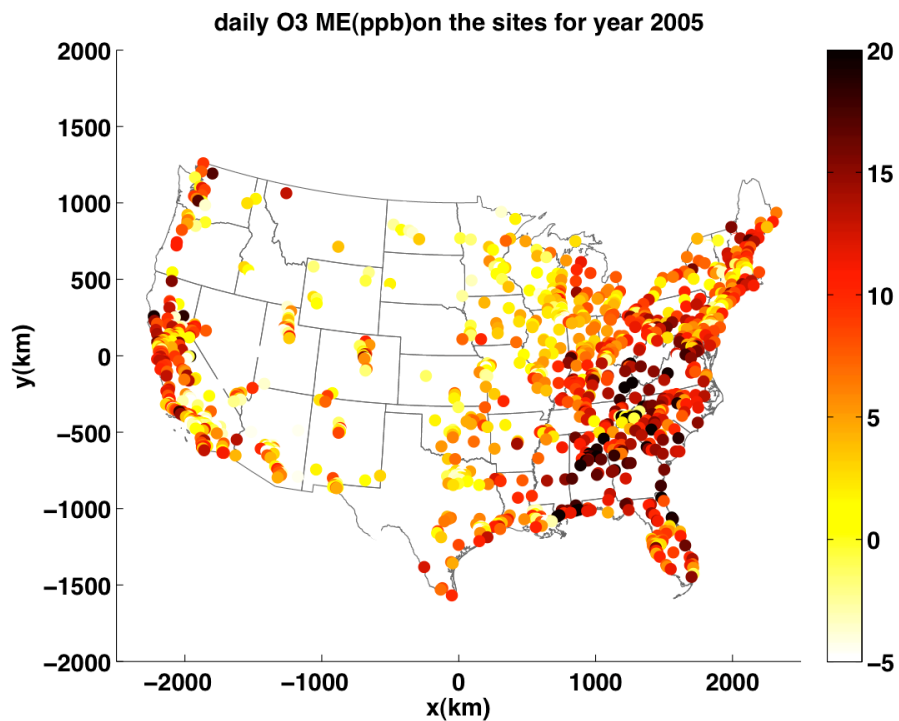


Figure A. 12s: D24A O₃ mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km² grid cell resolution.

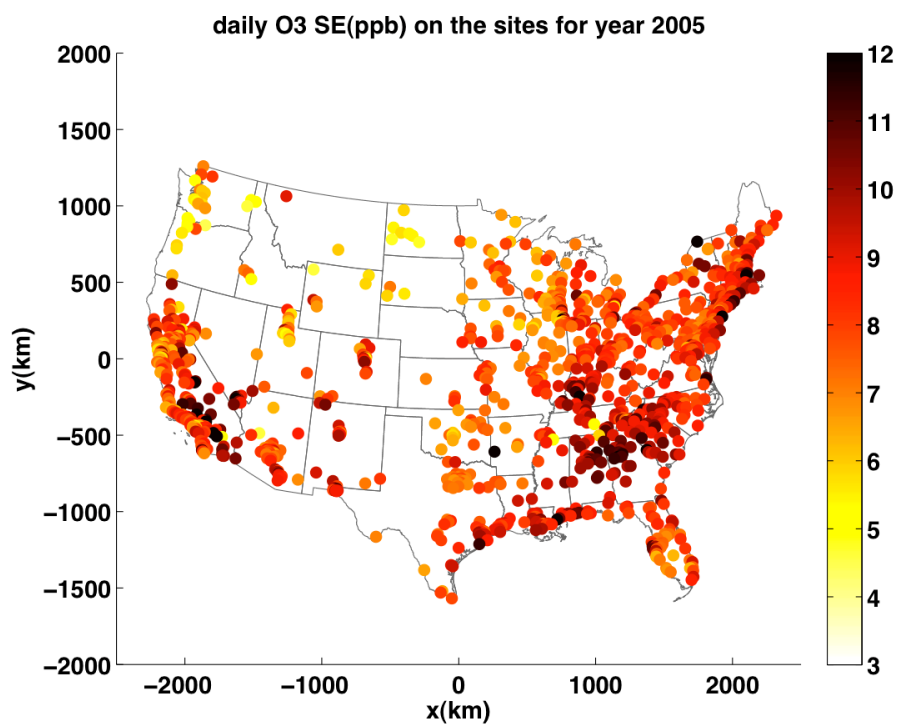


Figure A. 13s: D24A O₃ standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 36x36km² grid cell resolution.

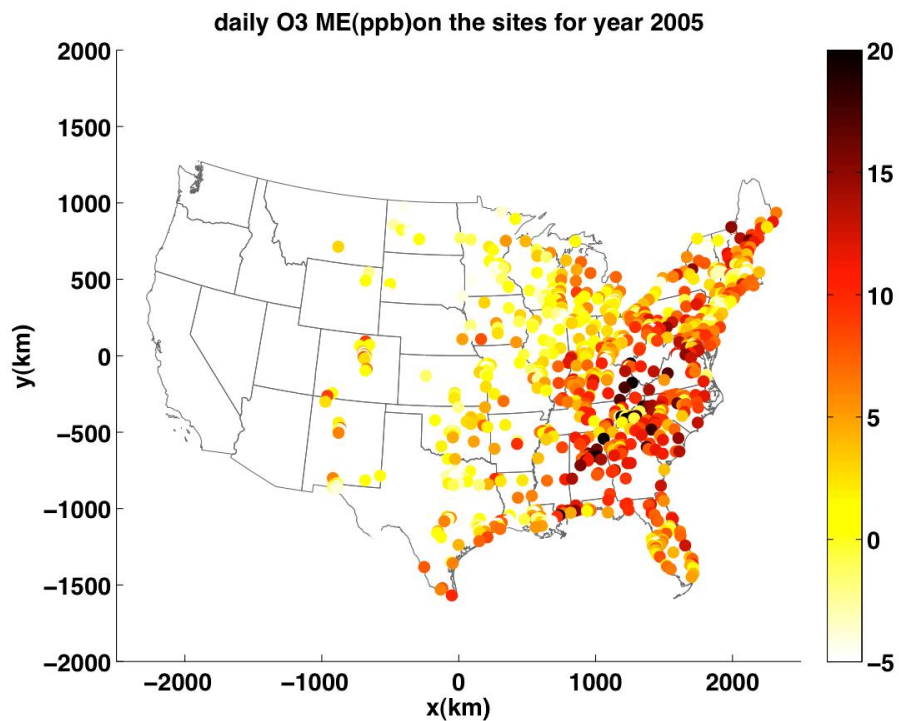


Figure A. 14s: D24A O₃ mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km² grid cell resolution.

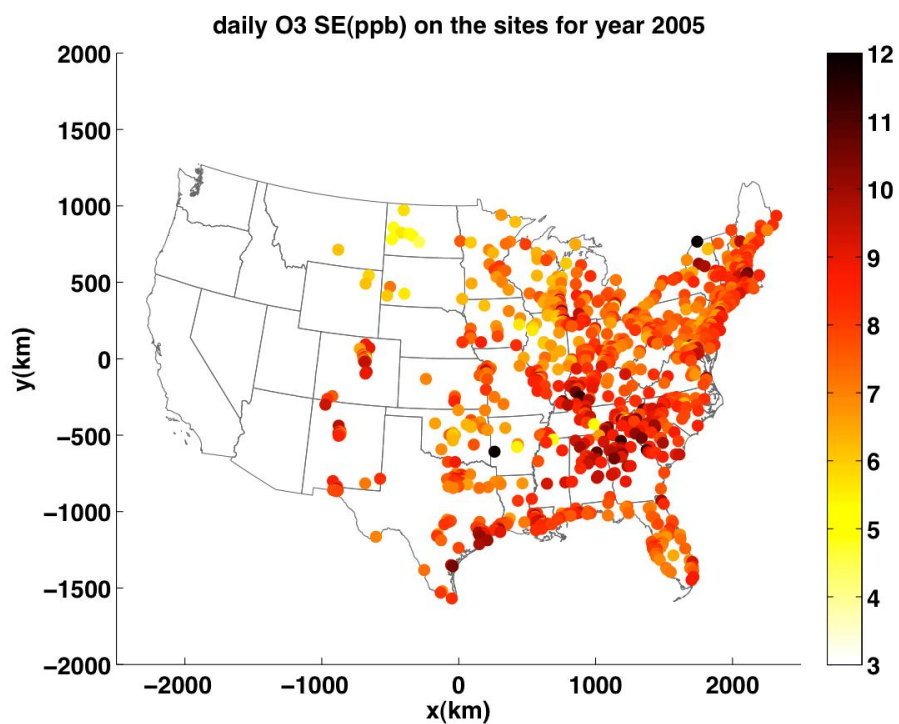


Figure A. 15s: D24A O₃ standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km² grid cell resolution.

The seasonal difference in model performance was also analyzed by recalculating the ME and SE at each site separately for the summer (May, June, July and August) and winter (November, December, January and February). For 36x36km² simulation, the over-prediction bias is noticeably higher in the summer, with the median ME for the summer equal to 5.0 ppb compared to 4.1 ppb for the winter. The Interquartile range (IQR) of the SE is 8.06 ppb in the winter, while it is only 3.0 ppb in the summer, indicating that the spatial variability of the prediction imprecision is much larger in the winter.

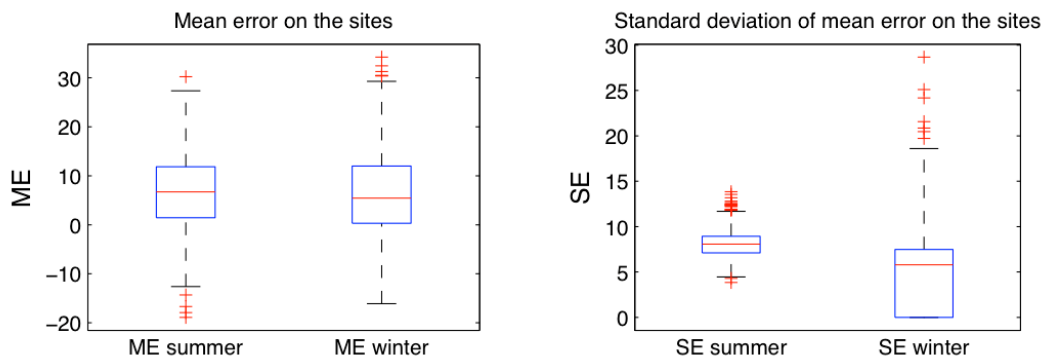


Figure A. 16s: Boxplots of D24A O₃ mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 36x36km² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

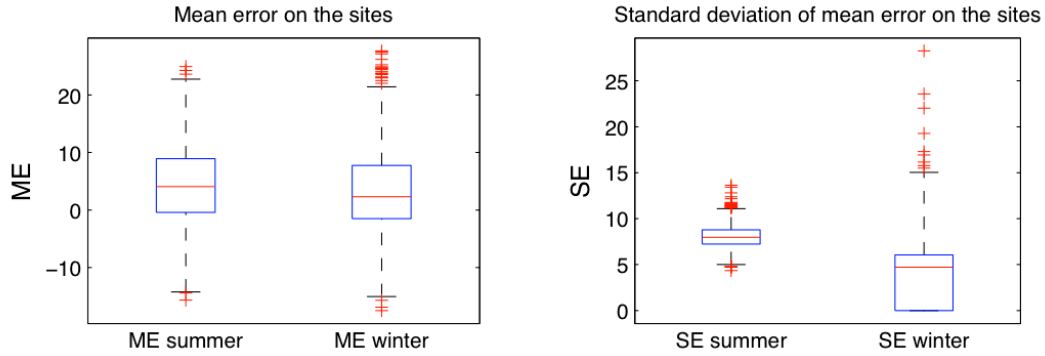


Figure A. 17s: Boxplots of D24A O₃ mean prediction error (ME) and standard deviation of these prediction error (SE) (in ppb) at AQS sites simulated for year 2005 using the CAMx model simulation with 12x12km² grid cell resolution, and separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

A.2 Parameter selection for the offset analysis and covariance model

The offset is considered a deterministic function of space and time that can be mathematically calculated for any space/time point \mathbf{p} without error. Intuitively there is an inverse relationship between the amount variability in the offset and the remaining variability of the transformed data. An offset with shorter ranges (i.e. with small a_r and a_t values) can capture more variability; therefore the transformed data has less variability. Conversely an offset with longer ranges (i.e. with longer a_r and a_t values) can describe less variability; therefore the resulting transformed data retains larger variability of the original data and thus has larger variance. Four sets of offset parameters (Table A1s) were investigated. Each set of offset parameters can be assessed visually through maps (Figures A18s-A22s) and time series plot (SI Figure A23s). The final parameters used in this study need to be selected in conjunction with the autocorrelation of the space-time covariance model processed through the offset-removed transformed data, which is described next.

To investigate the effect of different offset parameters on the variance and autocorrelation of the transformed data, we constructed 4 sets of offset parameters, describing ranges at short, intermediate, long and very long scales, respectively, with a_r and a_t values shown in Table A1s.

Table A. 1s: Spatial and temporal offset parameters and their label

Offset name	a_r (km)	a_t (days)
short	5	2
intermediate	50	10
long	300	50
very long	1,000	100

Each set of offset parameters can be assessed visually through maps and time series. The maps and time series plots for the offset calculated from these sets of parameters are shown in the following sections: section 2.1 for DM8A ozone concentrations and section 2.2 for D24A ozone concentrations.

A.2.1 Maps and figures of the offset analysis for the DM8A ozone concentrations

The map in figure A18s shows DM8A observed O₃ concentrations on day 21-July-2005, while the following four maps in figures A19s, A20s, A21s and A22s show the short, intermediate, long and very long offsets. As can be seen from these figures, the short offset captures variability at a finer scale, which is fairly close to the variability in the original observational daily ozone. As the ranges get longer, the offset at the monitoring sites starts to smooth out the data. In the map (Figure A22s) with the very long range offset, the offsets at all the monitoring sites have nearly the same values.

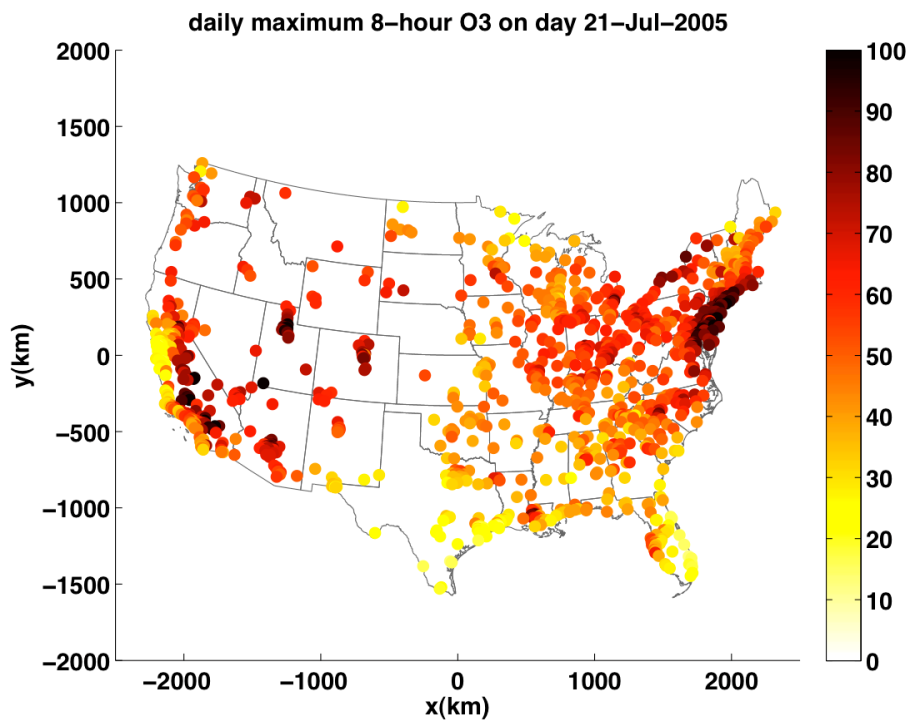


Figure A. 18s: The observed DM8A O₃ concentrations on the monitoring sites on day 21-July-2005

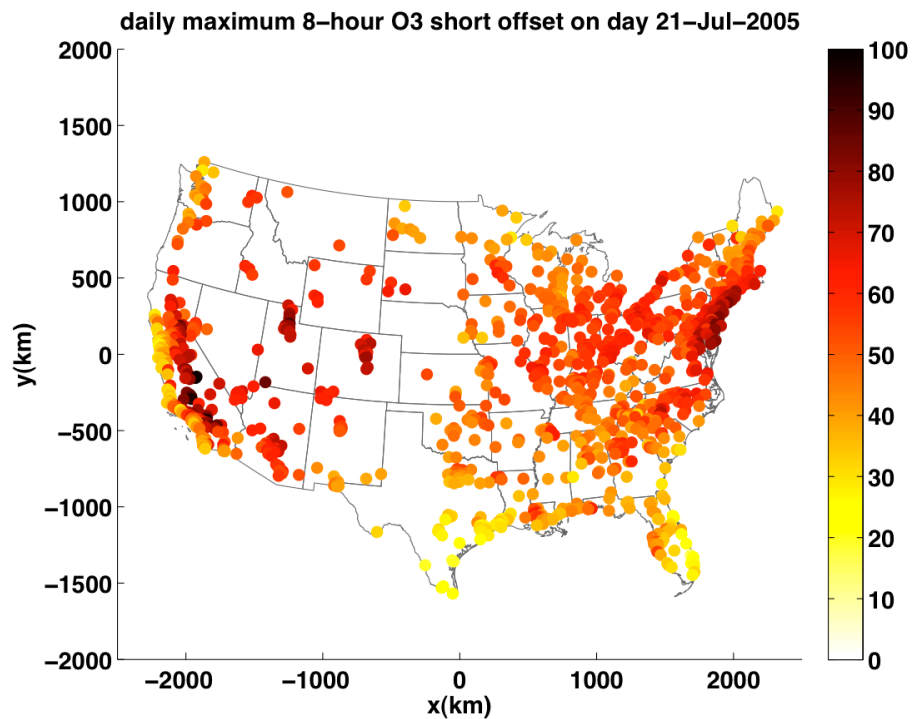


Figure A. 19s: The short range offset for DM8A O3 on the monitoring sites on day 21-July-2005

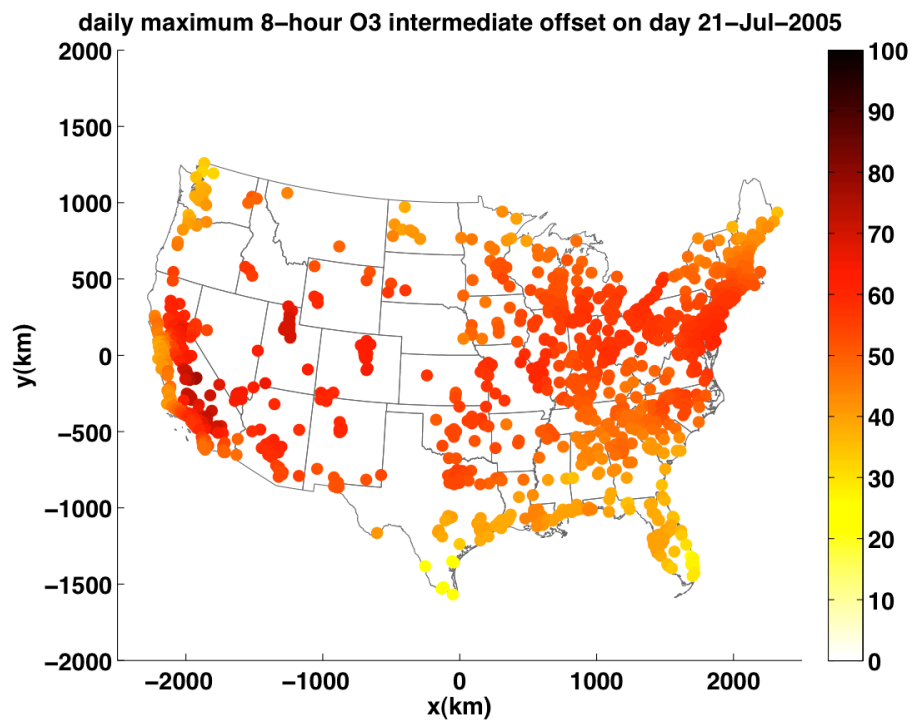


Figure A. 20s: The intermediate range offset for the DM8A O3 on the monitoring sites on day 21-July-2005

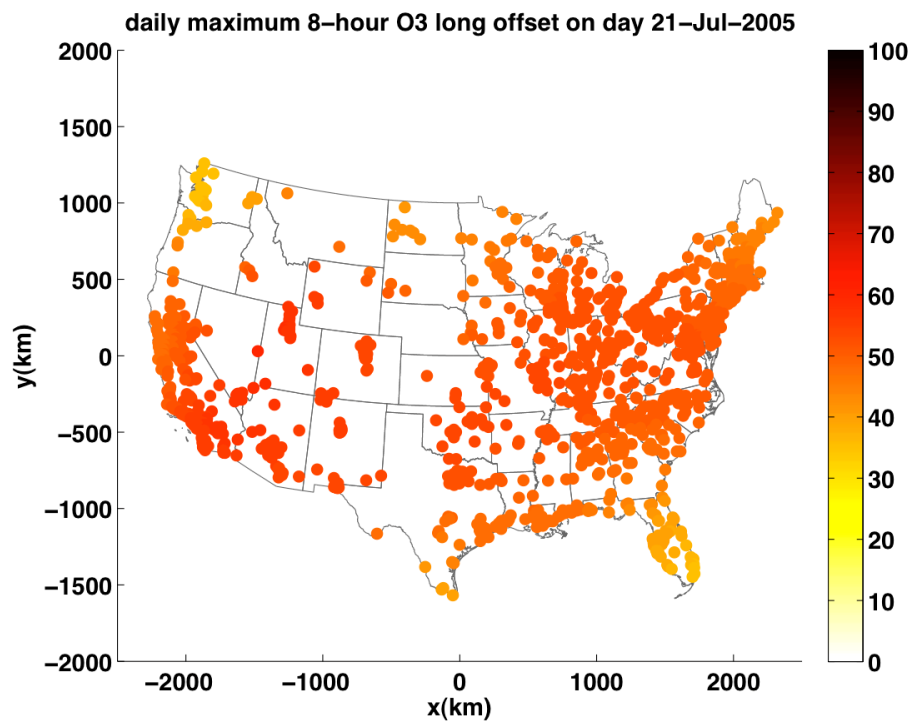


Figure A. 21s: The long range offset for the DM8A O3 on the monitoring sites on day 21-July-2005

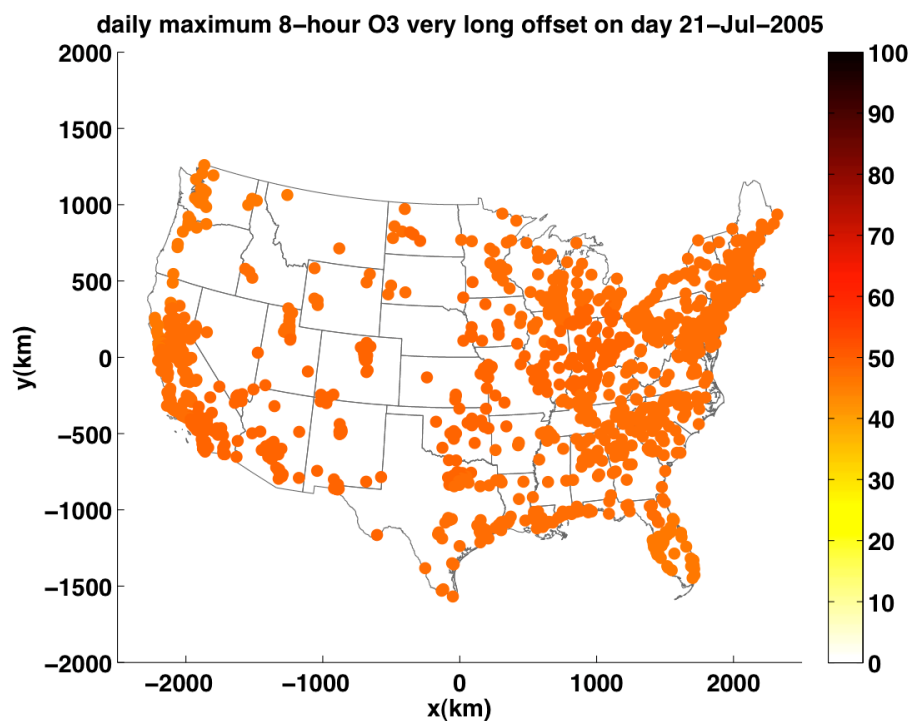


Figure A. 22s: The very long range offset for the DM8A O3 on the monitoring sites on day 21-July-2005

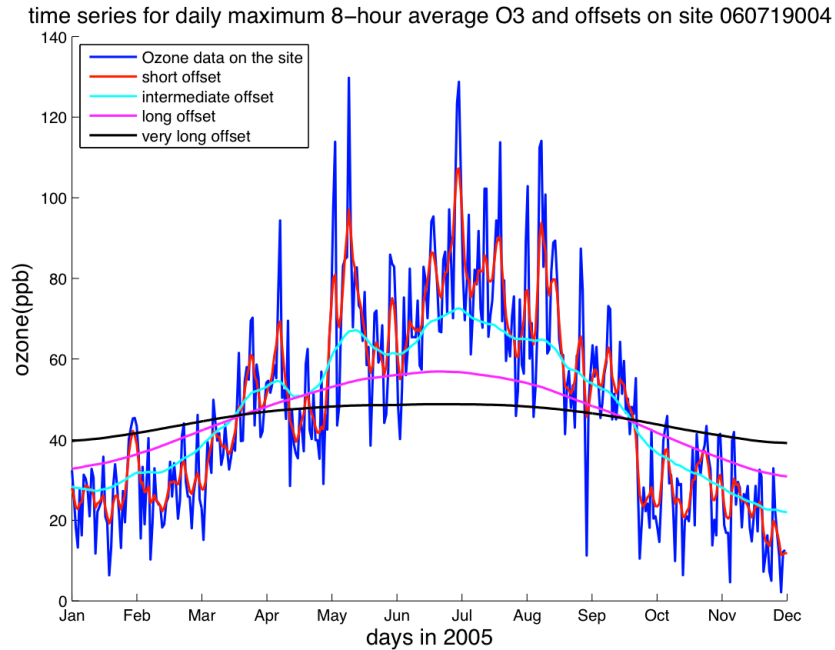


Figure A. 23s: The time series of the DM8A O₃ concentrations and four offsets on a randomly selected site (SiteID:060719004) for year 2005

A.2.2 Maps and figures of the offset analysis for the D24A ozone concentrations

The map in figure A24s shows the observed D24A O₃ concentrations on day 21-July-2005, while the following four maps in figures A25s, A26s, A27s and A28s show the short, intermediate, long and very long offsets. As can be seen from these figures, the short offset captures variability at a finer scale, which is fairly close to the variability in the original observational daily ozone. As the ranges get longer, the offset at the monitoring sites starts to smooth out the data. In the map (Figure A28s) with the very long range offset, the offsets at all the monitoring sites have nearly the same values.

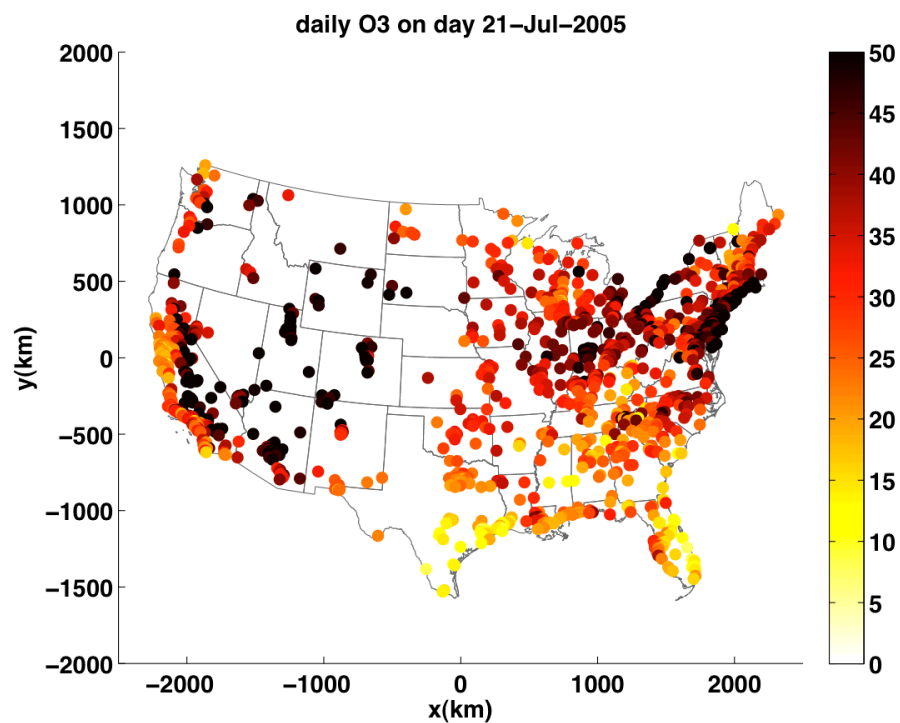


Figure A. 24s: The observed D24A O3 concentrations on the monitoring sites on day 21-July-2005

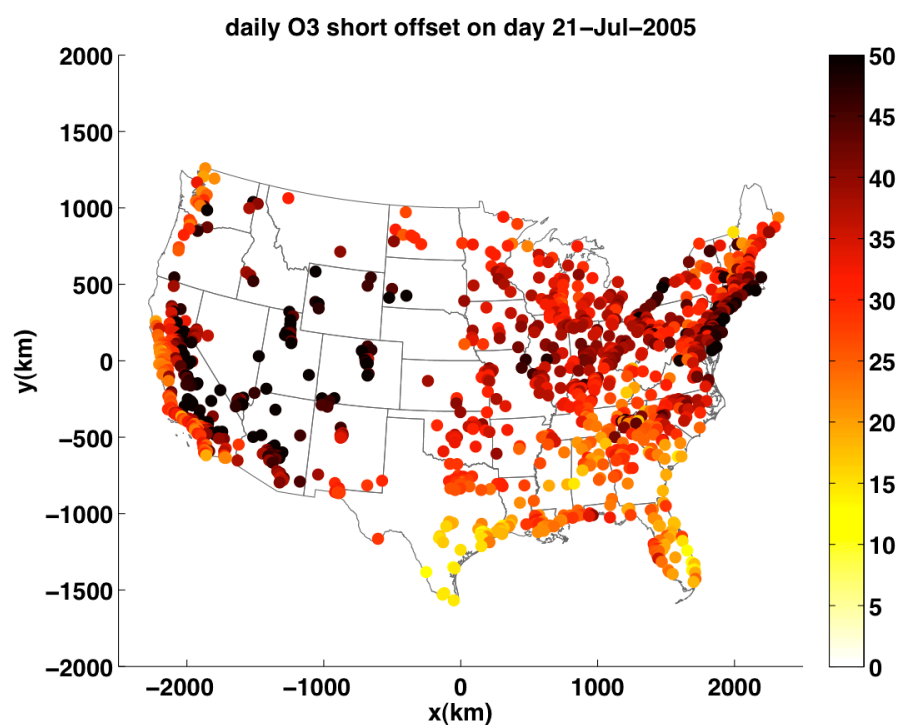


Figure A. 25s: The short range offset for the D24A O3 on the monitoring sites on day 21-July-2005

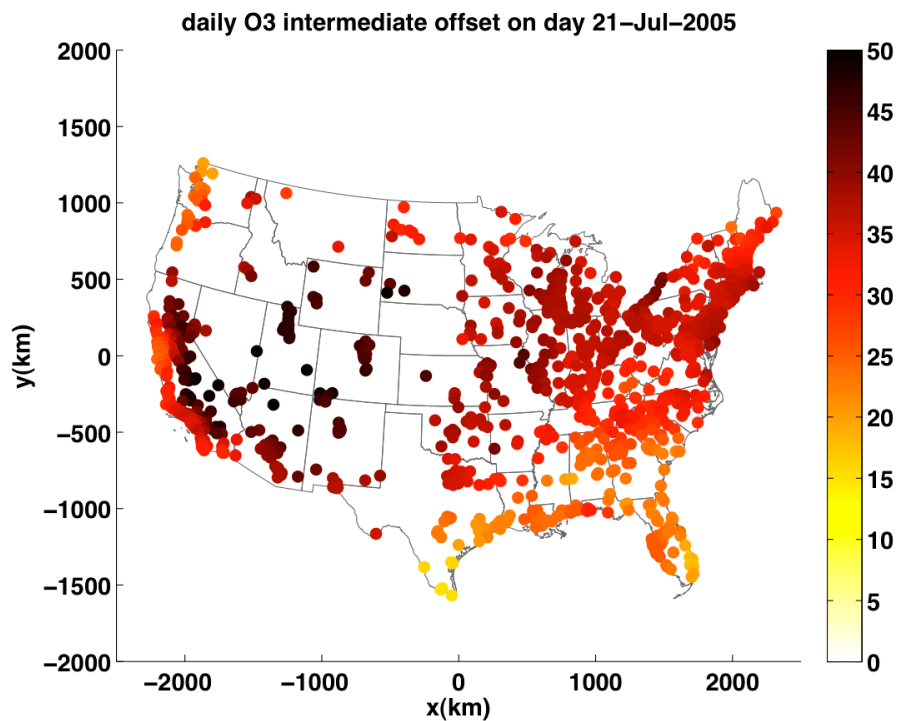


Figure A. 26s: The intermediate range offset for the D24A O3 on the monitoring sites on day 21-July-2005

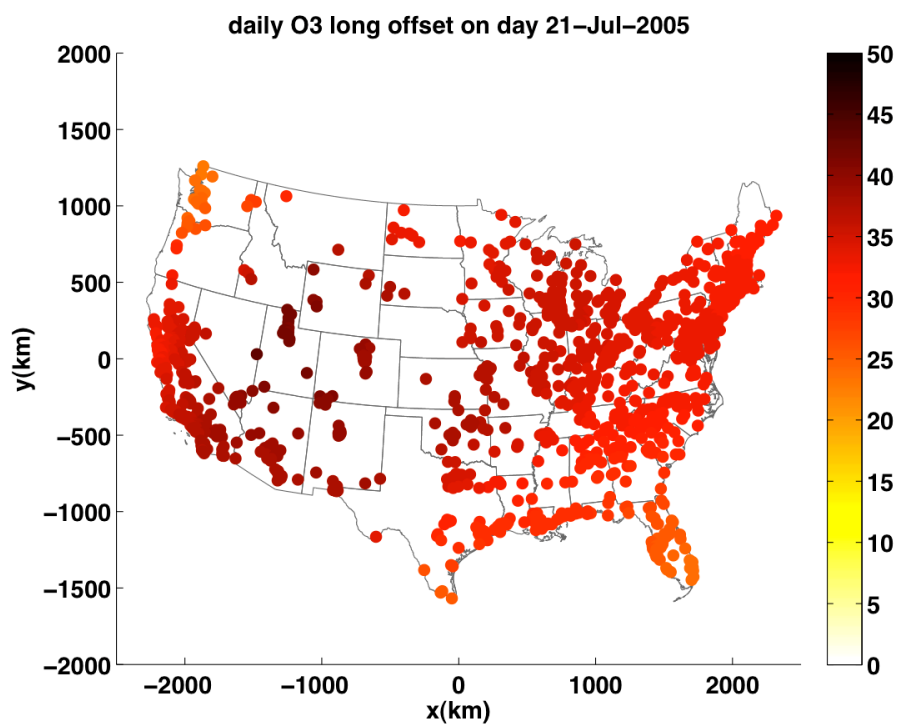


Figure A. 27s: The long range offset for the D24A O3 on the monitoring sites on day 21-July-2005

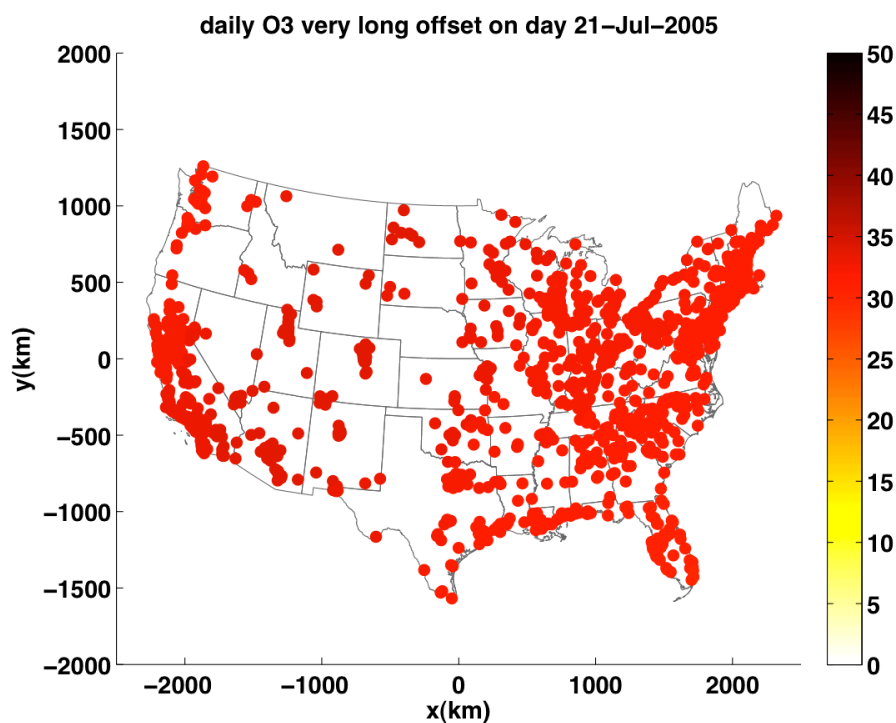


Figure A. 28s: The very long range offset for the D24A O3 on the monitoring sites on day 21-July-2005

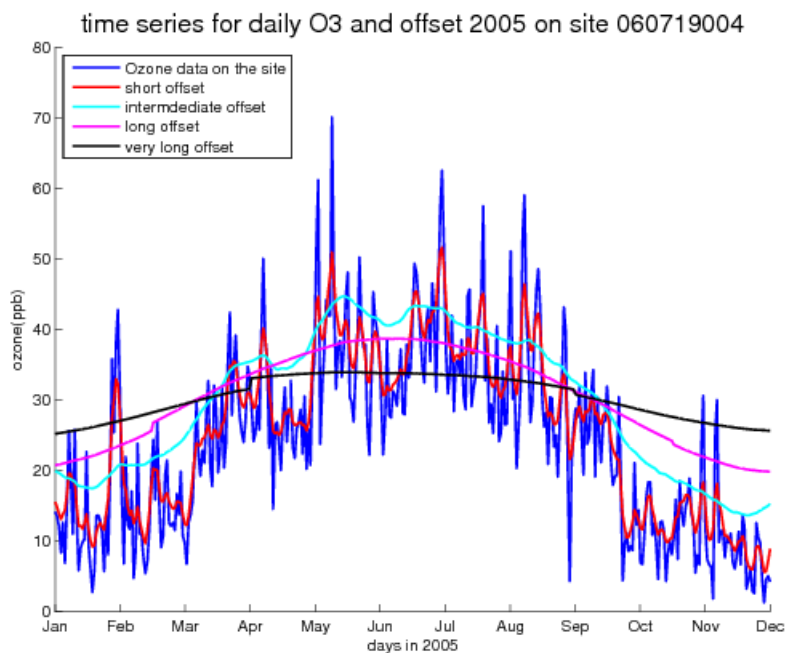


Figure A. 29s: The time series of the D24A O3 concentrations and four offsets on a randomly selected site (SiteID:060719004) for year 2005

A.2.3 Covariance model

Several other covariance models were attempted and evaluated; however, a 3-structured exponential model was selected that provided the best overall performance with respect to least square fit and model interpretability. The equation of the 3-structured exponential covariance model is given by

$$c_X(r, \tau) = C_0[\alpha \exp\left(\frac{-3r}{a_{r1}}\right) \exp\left(\frac{-3\tau}{a_{t1}}\right) + \beta \exp\left(\frac{-3r}{a_{r2}}\right) \exp\left(\frac{-3\tau}{a_{t2}}\right) + (1 - \alpha - \beta) \exp\left(\frac{-3r}{a_{r2}}\right) \exp\left(\frac{-3\tau}{a_{t1}}\right)] \quad (\text{A5s})$$

Where C_0 is the sill (variance), a_{r1} and a_{r2} are the spatial ranges and a_{t1} and a_{t2} are the temporal ranges, and α , β and $(1-\alpha-\beta)$ are the proportions of variability contributed by the first, second and third covariance structure, respectively. The parameters α , β , a_{r1} , a_{r2} , a_{t1} , and a_{t2} obtained by least square fitting for each offset are shown in table A2s&A3s. The corresponding models are graphed in figures A30s and A31s.

Table A. 2s: Parameter values of the 3-structured covariance model for each offset range for DM8A O3

Offset range	C_0 (ppb ²)	α	β	a_{r1} (km)	a_{r2} (km)	a_{t1} (days)	a_{t2} (days)
Short	44.69	0.10	0.25	0.059	1351	0.16	0.05
Intermediate	125.11	0.17	0.77	0.167	2026	3.41	2000
Long	179.40	0.12	0.64	0.198	1942	4.63	230.7
Very long	229.2	0.08	0.49	0.041	1988	4.19	186.55

Table A. 3s: Parameter values of the 3-structured covariance model for each offset range for D24A O3

Offset range	C_0 (ppb ²)	α	β	a_{r1} (km)	a_{r2} (km)	a_{t1} (days)	a_{t2} (days)
Short	23.12	0.28	0	0.034	1289	0.82	66039
Intermediate	75.63	0.41	0.24	0.012	1377	3.127	2000
Long	109.83	0.36	0.41	0.069	1250	4.04	638.89
Very long	136.64	0.30	0.54	0.013	1528	3.71	437.17

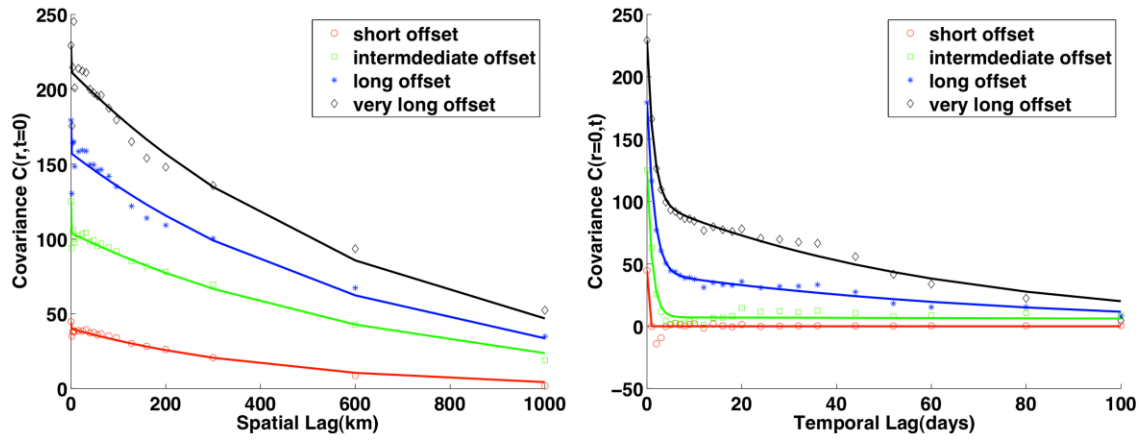


Figure A. 30s: Graphs of the 3-structured exponential/exponential covariance models with respect to the spatial lag when the temporal lag is set to 0 (**Left**) and with respect to the temporal lag when the spatial lag is set to 0 (**Right**) for DM8A O3

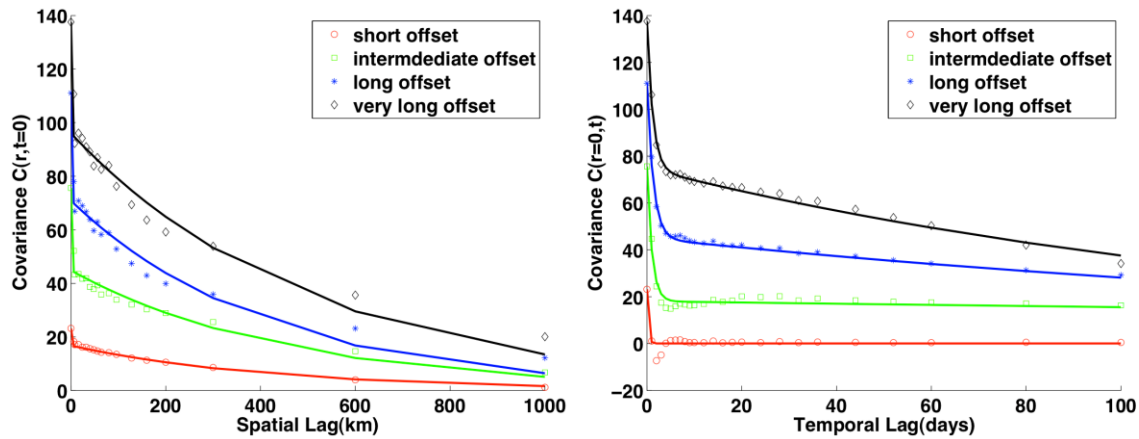


Figure A. 31s: Graphs of the 3-structured exponential/exponential covariance models with respect to the spatial lag when the temporal lag is set to 0 (**Left**) and with respect to the temporal lag when the spatial lag is set to 0 (**Right**) for D24A O3

There are four sets of offsets and therefore there are four resulting covariance models corresponding to each of the transformed data. To assess which of the 4 offsets should be selected to transform DM8A or D24A ozone concentrations for the BME simulation, dominance plots are created as shown below (figure A32s & A33s). According to the dominance plots, the intermediate offset ($a_r = 50 \text{ km}$ and $a_t = 10 \text{ day}$) was chosen because it contains a larger temporal autocorrelation than other

offsets, while only sacrificing a small amount of variance. Thus, the intermediate offset offers an attractive tradeoff of lowering variance while maintaining autocorrelation in the transformed data. As stated above, the offset selected will have a combination of the lowest variance and the highest autocorrelation.

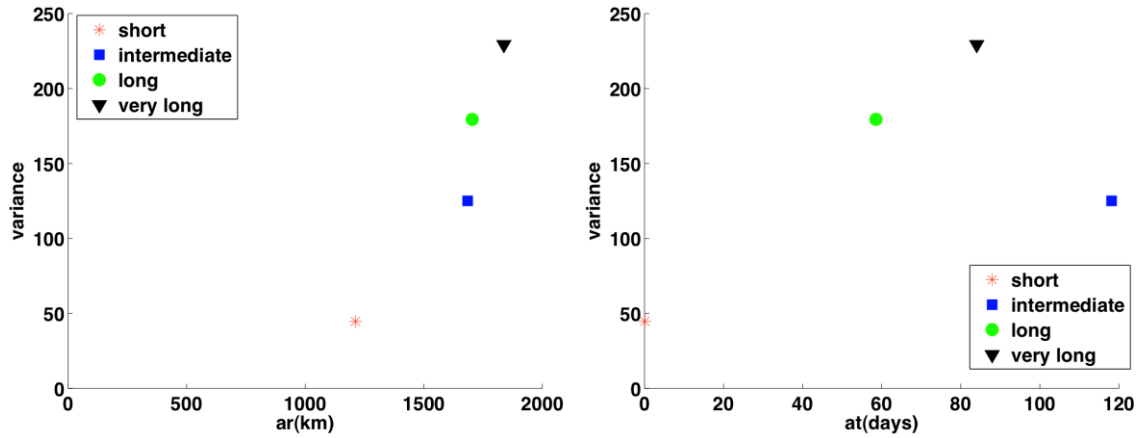


Figure A. 32s: Dominance plot showing how the variance changes with respect to the spatial range (**Left**) and with respect to the temporal range (**Right**) for covariance model corresponding to each offset range for DM8A O3

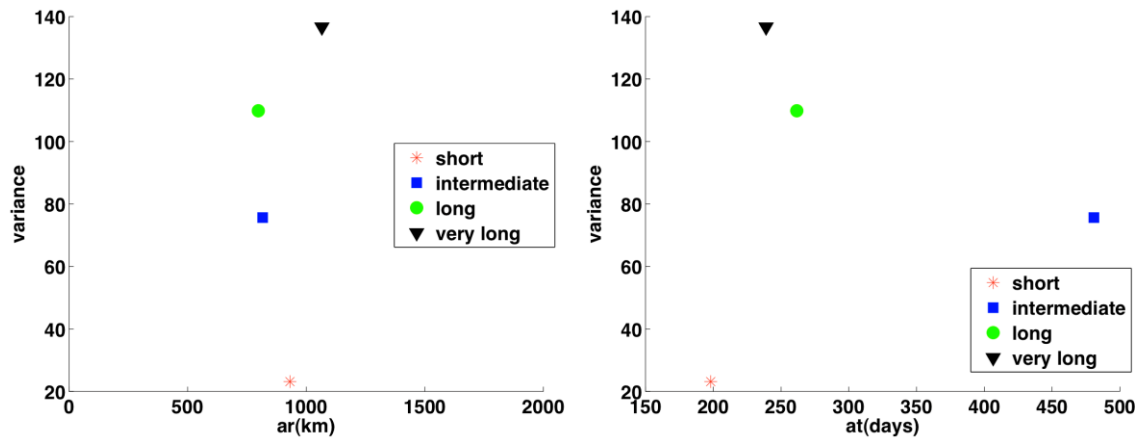


Figure A. 33s: Dominance plot showing how the variance changes with respect to the spatial range (**Left**) and with respect to the temporal range (**Right**) for covariance model corresponding to each offset range for D24A O3

As expected the dominance plots show that the variance of the transformed data increases progressively as the transformation offset range increases. This simply shows that as the offset range increases, the offset becomes smoother and less variable across space and time, and therefore the transformed data retains more variability. The Figure 32s&33s show that the spatial covariance range does not change much among the four offset ranges. The temporal covariance range changes dramatically from the short to intermediate offset. Since our aim is to select a combination of low variance and high autocorrelation, the intermediate offset was chosen because it creates transformed data with a larger temporal autocorrelation than that created by other offsets, while only sacrificing a small amount of variance. Thus, the intermediate offset offers an attractive tradeoff of lowering variance while maintaining autocorrelation in the transformed data. The intermediate offset and its corresponding covariance model are used in the subsequent BME simulation.

A.3 Additional results of BME O3 estimates

A.3.1 Additional maps showing the soft data for the DM8A O3

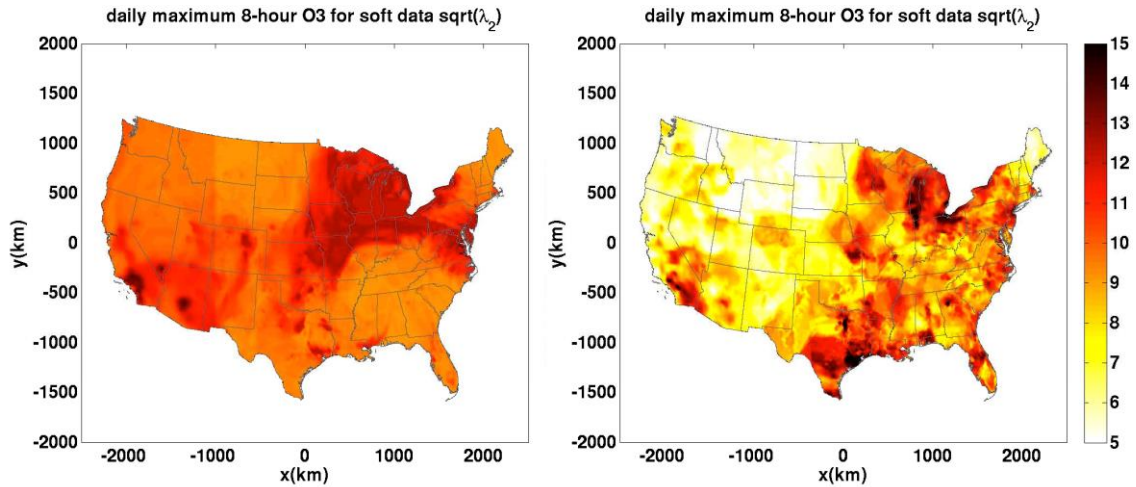


Figure A. 34s: Maps showing the square root of soft data variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ obtained across the continental United States on 11-July-2005 in (left) scenario CAMP and (right) scenario RAMP for DM8A O3.

A.3.2 Additional maps of BME O3 estimates for D24A O3

The main manuscript shows the results of BME O3 estimates for the DM8A on a selected day. Here we provide additional results to demonstrate the BME estimates of the D24A.

Figure A35s shows for Jul-21-2005 the BME estimates of the D24A ozone concentrations z_k obtained for the three estimation scenarios. This day was chosen because it has the highest standard deviation (at 12.8 ppb) for CTM prediction errors at ozone monitoring sites, which means the CTM model performance has the highest spatial variability among sites. It is clear that on this day the BME mean estimates (in the top panel of Figure A34s) in the immediate proximity of the monitoring stations (marked in circles) are at very similar levels in the three maps, being in good agreement with the observed data in their local neighborhood. As the estimation location moves away from the monitoring stations, the difference among these three maps becomes more substantial. For example, in scenario OBS we see a wider area of high ozone value, with the area above 50ppb covering 931,824 km^2 of the western and southwestern states of Nevada, Idaho, Utah, Wyoming, Colorado, Arizona, New Mexico and California. In scenario CAMP the ozone plume above 50ppb only covers a much smaller area (279,648 km^2) with the peak ozone concentrations in the plume reaching 61ppb. By contrast, in scenario RAMP the size of the ozone plume (the area where ozone levels are above 50ppb) is 658,368 km^2 and the highest peak ozone concentration reaches 69.5ppb.

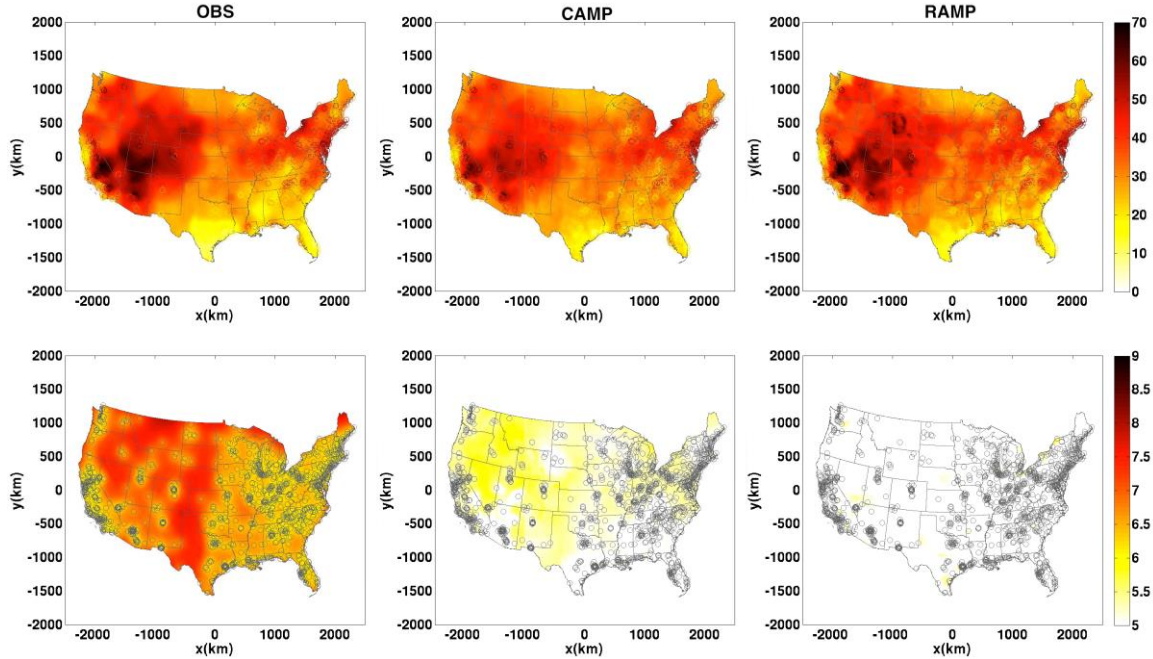


Figure A. 35s. Maps of BME mean estimates (top panel) and corresponding standard deviations of BME estimates (bottom panel) of the D24A ozone (ppb) on day Jul-21-2005 obtained with scenario OBS (left panels), CAMP (middle panels) and RAMP (right panels). Circles in the maps represent locations of monitoring sites and color match legend for observed concentrations.

The uncertainty associated with the BME estimates are quantified by the corresponding BME standard estimation error (bottom panel of Figure A35s). For estimation scenario OBS there is a higher estimation uncertainty, with the highest BME standard estimation error reaching 7.9ppb for areas far away from any monitoring stations, and with an average standard estimation error of about 7.2ppb across the continental United States. This is in contrast to estimation scenarios CAMP and RAMP, where the BME standard estimation error remains relatively low, with the highest standard estimation error reaching 6.7ppb and 6.3ppb for scenarios CAMP and RAMP. This indicates that integrating both observations and model predictions improved the quality of the ozone estimates, especially for areas far away from any monitoring station. Overall, scenario RAMP has the lowest standard estimation error, with an average standard estimation error of about 4.5ppb across the continental United States.

The construction of the soft data for D24A O3 using the proposed RAMP approach can be illustrated by comparing scenario RAMP that accounts for the space/time variability of CTM performance, with scenario CAMP that does not. Two important parameters that differed in these two scenarios are the bias-corrected expected values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ and the corresponding soft data variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$.

Figure A36s shows a map of the raw CAMx modeled D24A ozone predictions for 21-July-2005. Also shown in the figure are the bias-corrected CTM predicted values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ from scenarios CAMP and RAMP for the same day. Both scenarios CAMP and RAMP corrected the CTM prediction bias to some extent, especially for areas close to the monitoring sites. There are, however, substantial differences of the bias-corrected CTM predicted values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ between scenarios CAMP and RAMP. For scenario CAMP, the map of the bias-corrected CTM predicted values $\hat{\lambda}_1$ shows lower values than scenario RAMP. This is especially true for the western and southwestern states of Nevada, Idaho, Utah, Wyoming, Colorado, Arizona, New Mexico and California, where there is no area with bias-corrected CTM ozone levels greater than 50 ppb. By contrast, in scenario RAMP the size of the area with bias-corrected CTM ozone levels greater than 50 ppb is 649,872 km², with the peak bias-corrected CTM ozone level in that area reaching 76.3 ppb. This substantial difference is due to the assumed homogeneity of the CTM model performance in scenario CAMP that forces the same correction throughout the study domain. This correction results in an over correction in some local areas such as the area covering the western and southwestern states of Nevada, Idaho, Utah, Wyoming, Colorado, Arizona, New Mexico and California.. In contrast, the scenario RAMP is better able to account for regional biases in model performance.

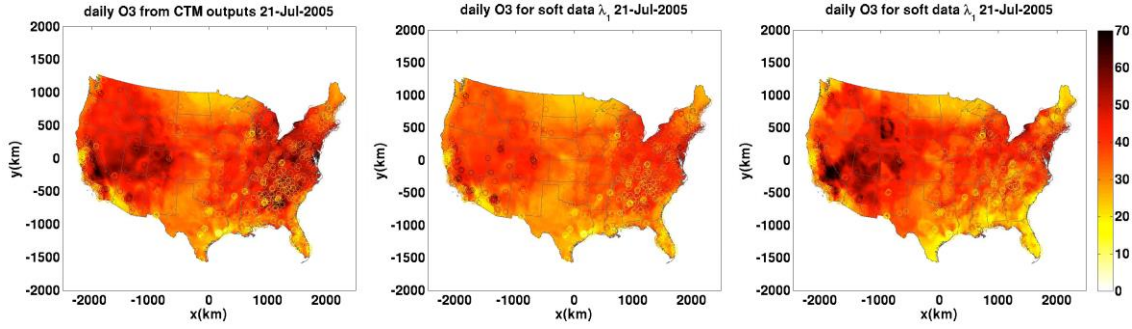


Figure A. 36s: D24A O₃ concentrations in the United States on 21-July-2005 using (left) the raw CTM Model predictions, (middle) the bias-corrected expected values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ for the estimation scenario CAMP, and (right) the bias-corrected expected values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$ for the estimation scenario RAMP.

The maps for the corresponding square root of soft data variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ are shown in figure A37s. This map characterizes the imprecision associated with the bias-corrected CTM predicted values $\hat{\lambda}_1(\tilde{z}_i, \mathbf{p}_i)$. We find that scenario RAMP has more localized gradients for the variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$, with the square root of $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ spanning from a low value of 3.0 ppb to a high of 13.8 ppb, and averaging about 6.9 ppb across the continental United States. By contrast, this variance has less spatial variability for the scenario CAMP; with a narrower span of $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ values ranging from 8.2 ppb to 11.7ppb, and a higher average over the continental United States of 9.2ppb. This illustrates again that the proposed RAMP method used in scenario RAMP has a greater ability to characterize regional changes in the precision of bias-corrected CTM predictions. This is an important methodological improvement explaining the improved performance in scenario RAMP in the cross validation analysis described next.

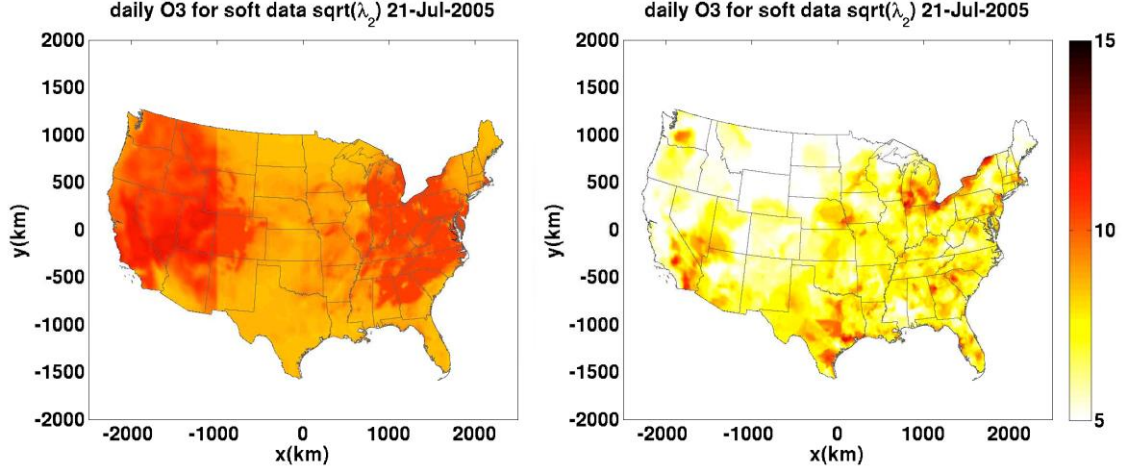


Figure A. 37s: Maps showing the square root of soft data variance $\hat{\lambda}_2(\tilde{z}_i, \mathbf{p}_i)$ obtained across the continental United States on 21-July-2005 in (left) scenario CAMP and (right) scenario RAMP for D24A O3.

A.4 Testing the statistical significance of the increase in R^2 for the cross-validation analysis

We may test whether any findings that the RAMP (or CAMP) analysis scenario did lead to an increase in R^2 is statistically significant, i.e. whether any findings that $R^2_{RAMP} - R^2_{OBS}$ is found to be positive can be said to be a statistically significant result. We apply the test for the difference $D = R^2_{RAMP} - R^2_{OBS}$, where R^2_{RAMP} and R^2_{OBS} are the correlation coefficients for the validation analysis in scenarios RAMP and OBS, respectively.

Under the null hypothesis we assume that D is normally distributed with mean zero and variance var_D . We test the alternative hypothesis that $D > 0$ by using the one-tail p -value formulae to calculate the probability of obtaining a specific $R^2_{RAMP} - R^2_{OBS}$ if the null hypothesis were true,

$$p = \int_{R^2_{RAMP} - R^2_{OBS}}^1 du \Phi(u; 0, var_D) \quad (A6s)$$

Where $\Phi(u; 0, var_D)$ is the Gaussian PDF with mean 0 and variance var_D , and $R^2_{RAMP} - R^2_{OBS}$ is the difference in R^2 obtained in the validation analysis. Traditionally a p -value less than 0.05 indicates that the null hypothesis may be rejected, so we can conclude that the increase in R^2 obtained in validation is statistically significant.

A.5 Additional results for the cross-validation analysis

We used commonly used error statistics to quantify the estimation accuracy, which is a function of the validation radius r_v . They consist of the Root Mean Square Error RMSE (ppb), the R^2 (unitless), the Mean Normalized Bias MNB (%) and the Mean Normalized Gross Error MNGE (%) between observations z_j and re-estimated values z_j^* , calculated as a function of r_v as shown in the equations A7s -A10s:

$$RMSE(r_v) = \sqrt{\frac{1}{N_o} \sum_{j=1}^{N_o} (z_j^*(r_v) - z_j)^2} \quad (A7s)$$

$$R^2(r_v) = \left(\frac{\sum_{j=1}^{N_o} (z_j^*(r_v) - \bar{z}^*(r_v))(z_j - \bar{z})}{\sqrt{\sum_{j=1}^{N_o} (z_j^*(r_v) - \bar{z}^*(r_v))^2} * \sqrt{\sum_{j=1}^{N_o} (z_j - \bar{z})^2}} \right)^2 \quad (A8s)$$

$$MNB(r_v) = \frac{1}{N} \sum_{j=1}^{N_o} 100\% * (z_j^*(r_v) - z_j) / z_j \quad (A9s)$$

$$MNGE(r_v) = \frac{1}{N} \sum_{j=1}^{N_o} 100\% * (|z_j^*(r_v) - z_j| / z_j) \quad (A10s)$$

Where N_o is the number of observed ozone values z_j throughout the study domain, and \bar{z} and $\bar{z}^*(r_v)$ are the averages of observed and re-estimated values, respectively. $RMSE(r_v)$ and $R^2(r_v)$ change as a function of the validation radius r_v considered.

Table A4s show the MNB and MNGE error statistics as a function of cross-validation radius r_v for all the observations in the cross-validation dataset. When comparing Scenario RAMP with Scenario OBS, we can see consistent improvements in MNB and MNGE as R_v increases. As we can see from equations 9s and 10s, the values of MNB and MNGE can be heavily influenced by the low observational values.

Figure A38s show the percent change in MNGE as a function of cross-validation radius r_v between Scenario RAMP and Scenario OBS, considering two curves. One is labeled as “no cutoff”, which includes all observations in the cross-validation data set. The other is labeled as “cutoff at 10ppb”, which excludes the observations below 10ppb. As can be seen, the percent reductions in MNGE are more substantial when we use the cutoff at 10 ppb by excluding those low values in observations.

Table A. 4s: Validation statistics of MNB and MNGE for BME data integration scenarios OBS, CAMP, and RAMP

Validation radius r_v (km)	DM8A				D24A			
	0	36	72	108	0	36	72	108
MNB_{OBS} (%)	3.472	5.298	5.739	5.963	7.595	10.809	11.286	11.576
MNB_{CAMP} (%)	3.990	5.645	6.035	6.220	8.925	11.843	12.361	12.675
MNB_{RAMP} (%)	3.820	5.320	5.580	5.650	8.202	10.785	11.072	11.171
$MNGE_{OBS}$ (%)	11.754	13.874	14.920	15.562	19.278	22.257	22.923	23.553
$MNGE_{CAMP}$ (%)	12.638	14.863	16.117	16.901	20.668	23.641	24.618	25.456
$MNGE_{RAMP}$ (%)	12.020	13.940	14.920	15.470	19.329	21.786	22.373	22.858

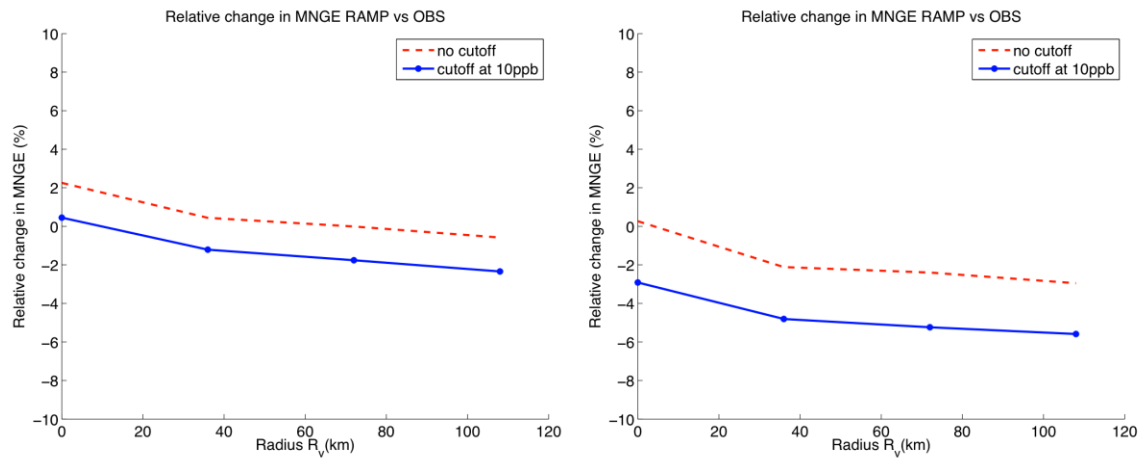


Figure A. 38s. Percent change in mean normalized gross error MNGE shown as a function of cross-validation radius R_v between Scenario RAMP and Scenario OBS. Each curve corresponds to the MNGE calculated using only observations above a given cutoff (0 ppb, 10ppb) of all observations values for the DM8A O3 (**Left**) and the D24A O3 (**Right**).

Table A. 5s: Validation statistics for BME data integration scenarios OBS, CAMP, and RAMP when the daily BME estimates aggregated into weekly and monthly averages

Validation radius r_v (km)	DM8A		D24A	
	Weekly	Monthly	Weekly	Monthly
$RMSE_{OBS}$ (ppb)	4.140	3.696	4.696	4.307
$RMSE_{CAMP}$ (ppb)	4.263	3.794	4.777	4.371
$RMSE_{RAMP}$ (ppb)	3.953	3.446	4.387	3.947
R^2_{OBS} (unitless)	0.903	0.901	0.805	0.799
R^2_{CAMP} (unitless)	0.901	0.898	0.806	0.801
R^2_{RAMP} (unitless)	0.914	0.915	0.833	0.834
$PCR^2_{OBS \rightarrow CAMP}$ (%)	-0.289	-0.262	0.188	0.246
$PCR^2_{OBS \rightarrow RAMP}$ (%)	1.153	1.607	3.546	4.402

A.6 The cokriging approach with a parametric relationship between the observations and the CTM model predictions

A.6.1 Cokriging framework

We let $Z(s, t)$ denote the Space/time Random Field (S/TRF) representing the observed daily ozone concentration at a location \mathbf{s} at time t , and we let $\tilde{Z}(\mathbf{s}, t)$ denote the S/TRF representing the corresponding CTM prediction at (\mathbf{s}, t) . Since the CTM prediction at location \mathbf{s} is given in terms of the average over a grid cell $\mathbf{B}(\mathbf{s})$ containing \mathbf{s} , an alternative notation for the CTM prediction S/TRF is given by $\tilde{Z}(\mathbf{B}, t)$, where \mathbf{B} is a shorthand notation for $\mathbf{B}(\mathbf{s})$.

We assume that the $\tilde{Z}(\mathbf{B}, t)$ S/TRF is Gaussian distributed with mean

$$E[\tilde{Z}(\mathbf{B}, t)] = \mu_{\tilde{Z}}(\mathbf{s}, t) \quad (\text{A11s})$$

and a covariance given by

$$\text{cov}(\tilde{Z}(\mathbf{B}, t), \tilde{Z}(\mathbf{B}', t')) = c_{\tilde{Z}}(r, \tau) \quad (\text{A12s})$$

Where the covariance function $C_{\tilde{Z}}(r, \tau)$ is a function of the spatial lag $r = \|\mathbf{s}' - \mathbf{s}\|$ and temporal lag $\tau = |t' - t|$, i.e. $\tilde{Z}(\mathbf{B}, t) \sim N(\mu_{\tilde{Z}}(\mathbf{s}, t), c_{\tilde{Z}}(r, \tau))$. We then assume that the relationship between the observed and predicted ozone values is given by a linear homoscedastic equation, i.e.

$$Z(\mathbf{s}, t) = \beta_0(\mathbf{s}, t) + \beta_1(\mathbf{s}, t)\tilde{Z}(\mathbf{B}, t) + \varepsilon(\mathbf{s}, t) \quad (\text{A13s})$$

where the error is Gaussian distributed with mean 0 and constant variance, i.e. $\varepsilon(s, t) \sim N(0, \sigma_\varepsilon^2)$, and it is not correlated with the prediction, i.e.

$$\text{cov}\left(\tilde{Z}(\mathbf{B}, t), \varepsilon(\mathbf{s}', t')\right) = 0. \quad (\text{A14s})$$

Given our model assumptions (equations A11s-A13s) it follows that the $Z(\mathbf{s}, t)$ is Gaussian distributed with mean equal to

$$\mu_Z(\mathbf{s}, t) = \beta_0(\mathbf{s}, t) + \beta_1(\mathbf{s}, t)\mu_{\tilde{Z}}(\mathbf{s}, t), \quad (\text{A15s})$$

and that the covariance between $Z(\mathbf{s}, t)$ and $Z(\mathbf{s}', t')$ is equal to

$$\text{cov}(Z(\mathbf{s}, t), Z(\mathbf{s}', t')) = \beta_1(\mathbf{s}, t)\beta_1(\mathbf{s}', t')C_{\tilde{Z}}(r, \tau) + \delta(r, \tau)\sigma_\varepsilon^2, \quad (\text{A16s})$$

while the covariance between $Z(\mathbf{s}, t)$ and $\tilde{Z}(\mathbf{s}', t')$ is equal to

$$\text{cov}\left(Z(\mathbf{s}, t), \tilde{Z}(\mathbf{s}', t')\right) = \beta_1(\mathbf{s}, t)C_{\tilde{Z}}(r, \tau) \quad (\text{A17s})$$

The cokriging approach consists in using the CTM predictions to model their mean $\mu_{\tilde{Z}}(\mathbf{s}, t)$ and covariance function $c_{\tilde{Z}}(r, \tau)$ (see section A6.2), using paired observations and model predictions to model $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$, and then using observations and predictions to estimate ozone at any un-sampled space/time estimation point $\mathbf{p}_k=(s_k, t_k)$ of interest. Let \mathbf{p}_o and \mathbf{p}_m be the space/time coordinates where ozone is Observed and Modeled, respectively, in the space/time neighborhood of the estimation point \mathbf{p}_k . In that localized neighborhood $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$ are approximately constant, $\beta_0(\mathbf{s}, t) \approx \beta_{0k}$ and $\beta_1(\mathbf{s}, t) \approx \beta_{1k}$, where $\beta_{0k} = \beta_0(\mathbf{p}_k)$ and $\beta_{1k} = \beta_1(\mathbf{p}_k)$, and we can re-write equations (A16s) and (A17s) as

$$\text{cov}(Z(\mathbf{s}, t), Z(\mathbf{s}', t')) \approx \beta_{1k}^2 c_{\tilde{Z}}(r, \tau) + \delta(r, \tau)\sigma_\varepsilon^2, \quad (\text{A18s})$$

and

$$\text{cov}\left(Z(\mathbf{s}, t), \tilde{Z}(\mathbf{s}', t')\right) \approx \beta_{1k} c_{\tilde{Z}}(r, \tau) \quad (\text{A19s})$$

Our notation for random variables consists of denoting a single random variable Z in capital letters, its realization, z , in lower case; and vectors in bold faces (e.g. $\mathbf{Z} = [Z_1, \dots]^T$, $\mathbf{z} = [z_1, \dots]^T$). Let Z_k and \mathbf{Z}_o be a random variable and vector of random variables representing ozone concentration at \mathbf{p}_k and \mathbf{p}_o , respectively, and let \mathbf{Z}_m be a vector of random variables representing modeled ozone concentrations at \mathbf{p}_m . Let $\mathbf{Z}_d = \begin{bmatrix} \mathbf{Z}_o \\ \tilde{\mathbf{Z}}_m \end{bmatrix}$ be the vector of random variables at the data points \mathbf{p}_o and \mathbf{p}_m , and let $\mathbf{z}_d = \begin{bmatrix} \mathbf{z}_o \\ \tilde{\mathbf{z}}_m \end{bmatrix}$ be the corresponding cokriging data, consisting of observed and modeled ozone values available in the neighborhood of the estimation point. From our model specification (equations A11s-A19s) it follows that Z_k given \mathbf{z}_d is normally distributed, i.e. $Z_k|\mathbf{z}_d \sim N(\mu_{Z_k|\mathbf{z}_d}, \sigma_{Z_k|\mathbf{z}_d}^2)$, with a mean and variance given by

$$\mu_{Z_k|\mathbf{z}_d} = \mu_{Z_k} + C_{Z_k, \mathbf{Z}_d} C_{\mathbf{Z}_d, \mathbf{Z}_d}^{-1} (\mathbf{z}_d - \mu_{\mathbf{Z}_d}), \quad (\text{A20s})$$

$$\sigma_{Z_k|\mathbf{z}_d}^2 = \sigma_{Z_k}^2 - C_{Z_k, \mathbf{Z}_d} C_{\mathbf{Z}_d, \mathbf{Z}_d}^{-1} C_{\mathbf{Z}_d, Z_k}, \quad (\text{A21s})$$

where

$$\mu_{Z_k} = \beta_{0k} + \beta_{1k} \mu_{\tilde{\mathbf{Z}}_k}, \quad (\text{A22s})$$

$$\mu_{\mathbf{Z}_d} = \beta_{0k} + \beta_{1k} \mu_{\tilde{\mathbf{Z}}_d}, \quad (\text{A23s})$$

$$\sigma_{Z_k}^2 = \beta_{1k}^2 \sigma_{\tilde{\mathbf{Z}}_k}^2 + \sigma_\epsilon^2, \quad (\text{A24s})$$

$$C_{\mathbf{Z}_d, \mathbf{Z}_d} = \begin{bmatrix} C_{\mathbf{Z}_o, \mathbf{Z}_o} & C_{\mathbf{Z}_o, \tilde{\mathbf{Z}}_m} \\ C_{\tilde{\mathbf{Z}}_m, \mathbf{Z}_o} & C_{\tilde{\mathbf{Z}}_m, \tilde{\mathbf{Z}}_m} \end{bmatrix}, \quad (\text{A25s})$$

$$C_{\mathbf{Z}_o, \mathbf{Z}_o} = \beta_{1k}^2 C_{\tilde{\mathbf{Z}}_o, \tilde{\mathbf{Z}}_o} + I * \sigma_\epsilon^2, \quad (\text{A26s})$$

$$C_{\mathbf{Z}_o, \tilde{\mathbf{Z}}_m} = \beta_{1k} C_{\tilde{\mathbf{Z}}_o, \tilde{\mathbf{Z}}_m}, \quad (\text{A27s})$$

and I is the identity matrix, $\mu_{\tilde{\mathbf{Z}}_k}$ and $\mu_{\tilde{\mathbf{Z}}_d}$ are specified by the mean $\mu_{\tilde{\mathbf{Z}}}(\mathbf{s}, t)$ (see section A6.2), $C_{\tilde{\mathbf{Z}}_o, \tilde{\mathbf{Z}}_o}$,

$C_{\tilde{\mathbf{Z}}_o, \tilde{\mathbf{Z}}_m}$ and $C_{\tilde{\mathbf{Z}}_m, \tilde{\mathbf{Z}}_m}$ are specified by the covariance function $c_{\tilde{\mathbf{Z}}}(r, \tau)$ (see section A6.2), and $\beta_{0k} = \beta_0(\mathbf{p}_k)$

and $\beta_{1k} = \beta_1(\mathbf{p}_k)$ are calculated from the models for $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$ (see section A6.3).

Equations A20s and A21s provide the cokriging estimate and its corresponding estimation error variance. The cokriging approach provide estimates of ozone that integrate the ozone observed values \mathbf{z}_o measured at monitoring sites, and the CTM predictions $\tilde{\mathbf{z}}_m$ calculated for the centroid of CTM computational grid. The cokriging approach falls in the category of data integration approaches that (a) parameterize the relationship between air pollution observations and predictions, and (b) use kriging to obtain air pollution estimates for a given value of the parameters. The limitation of these approaches are that they share the limitations of the linear kriging estimator, and they assume that the relationship between air pollution observations and predictions is linear and homoscedastic.

A.6.2 Modeling the mean $\mu_{\tilde{\mathbf{z}}}(s, t)$ and covariance $C_{\tilde{\mathbf{z}}}(r, \tau)$ of the CTM predictions

In this study, a time varying offset $o_{\tilde{\mathbf{z}}}(t)$ is used to transform the daily ozone CTM predictions into transformed offset-removed CTM data. This offset is obtained by calculating the average ozone CTM predicted concentrations over the whole domain for each day. We then define the transformation of the ozone CTM predictions $\tilde{\mathbf{z}}_m$ at space/time grid points $\mathbf{p}_m=(s_m, t_m)$ as

$$\tilde{\mathbf{x}}_m = \tilde{\mathbf{z}}_m - o_{\tilde{\mathbf{z}}}(\mathbf{t}_m). \quad (\text{A28s})$$

We define $\tilde{X}(\mathbf{p})$ as a homogeneous/stationary S/TRF representing the variability and uncertainty associated with the transformed data $\tilde{\mathbf{x}}_m$, and we let $\tilde{Z}(\mathbf{p}) = \tilde{X}(\mathbf{p}) + o_{\tilde{\mathbf{z}}}(t)$ be the S/TRF representing daily ozone CTM predictions.

The covariance model for the homogeneous/stationary S/TRF $\tilde{X}(\mathbf{p})$ is developed from the experimental covariance of the transformed data $\tilde{\mathbf{x}}_m$. The experimental covariance value for a spatial lag r and a temporal lag τ is calculated as

$$\hat{c}_{\tilde{X}}(r, \tau) = \frac{1}{N(r, \tau)} \sum_{i=1}^{N(r, \tau)} \tilde{x}_{head, i} \tilde{x}_{tail, i} - m_{\tilde{X}}^2. \quad (\text{A29s})$$

where $N(r, \tau)$ is the number of pairs of values $(\tilde{x}_{head,i}, \tilde{x}_{tail,i})$ separated by a spatial lag of r and temporal lag of τ , and the mean $m_{\tilde{x}}$ the \tilde{x}_m data is zero. In practice $\hat{c}_{\tilde{x}}(r, 0)$ and $\hat{c}_{\tilde{x}}(0, \tau)$ are calculated and plotted separately to facilitate the visualization of the space/time covariance models (figure A39s).

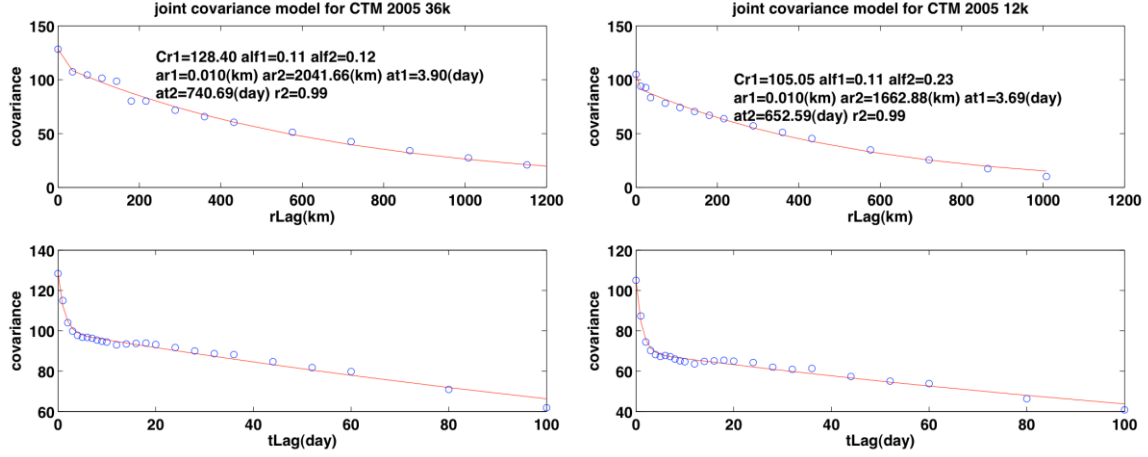


Figure A. 39s: Graphs of the 3-structured exponential/exponential covariance functions with respect to the spatial lag (Top panel) and the temporal lag (Bottom panel) for CTM with 36x36km² cell resolution (Left) and CTM with 12x12km² cell resolution (Right) for D24A O3.

A 3-structured exponential covariance model is fit to the experimental covariance values. The model obtained using least square fitting has the following equation

$$c_{\tilde{x}}(r, \tau) = C_0 \left[\alpha \exp\left(\frac{-3r}{ar_1}\right) \exp\left(\frac{-3\tau}{at_1}\right) + \beta \exp\left(\frac{-3r}{ar_2}\right) \exp\left(\frac{-3\tau}{at_2}\right) + (1 - \alpha - \beta) \exp\left(\frac{-3r}{ar_2}\right) \exp\left(\frac{-3\tau}{at_1}\right) \right] \quad (A30s)$$

Where $C_0 = 128.4$; $\alpha = 0.11$; $\beta = 0.12$, $ar_1 = 0.01\text{km}$, $ar_2 = 2042\text{km}$, $at_1 = 3.90$ days and $at_2 = 740.69$ days for CTM with 36x36km² cell resolution; And $C_0 = 105.05$; $\alpha = 0.11$; $\beta = 0.23$, $ar_1 = 0.01\text{km}$, $ar_2 = 1663\text{km}$, $at_1 = 3.69$ days and $at_2 = 652.59$ days for CTM with 12x12km² cell resolution.

Finally we have $\tilde{Z}(\mathbf{B}, t) \sim N(\mu_{\tilde{Z}}(\mathbf{s}, t), c_{\tilde{Z}}(r, \tau))$ with $\mu_{\tilde{Z}}(\mathbf{s}, t) = o_{\tilde{Z}}(t)$ and $c_{\tilde{Z}}(r, \tau) = c_{\tilde{x}}(r, \tau)$ given by equation A30s.

A.6.3 Modeling the linear homoscedastic relationship between ozone observations and predictions

The linear homoscedastic relationship between ozone prediction $\tilde{Z}(\mathbf{B}, t)$ and its corresponding observation $Z(\mathbf{s}, t)$ is given by equation A15s, $Z(\mathbf{s}, t) = \beta_0(\mathbf{s}, t) + \beta_1(\mathbf{s}, t)\tilde{Z}(\mathbf{B}, t) + \varepsilon(\mathbf{s}, t)$. The parameters $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$ are estimated through pooling the paired modeled and observed data together from the three closest ozone monitoring sites to \mathbf{s} and within 120 days of t . Then a linear regression is done with these selected pair data points to obtain $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$.

Figures A40s&A41s show maps of these estimated parameters $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$ for \mathbf{s} varying across the continental United States and for $t=11\text{-Jul-2005}$ for the DM8A ozone concentrations. And Figures A42s&A43s show maps of these estimated parameters $\beta_0(\mathbf{s}, t)$ and $\beta_1(\mathbf{s}, t)$ for \mathbf{s} varying across the continental United States and for $t=21\text{-Jul-2005}$ for D24A ozone concentrations. We can see that the relationship between the ozone observations and the CTM predictions characterizes the spatial heterogeneity and temporal variability of ozone model performance. Hence the cokriging approach developed here is comparable to our proposed RAMP approach, however its limitations are that it (a) assumes that the observation-prediction relationship is linear and homogeneous, and (b) the estimation relies on the limited linear kriging estimator as opposed to the more general BME framework.

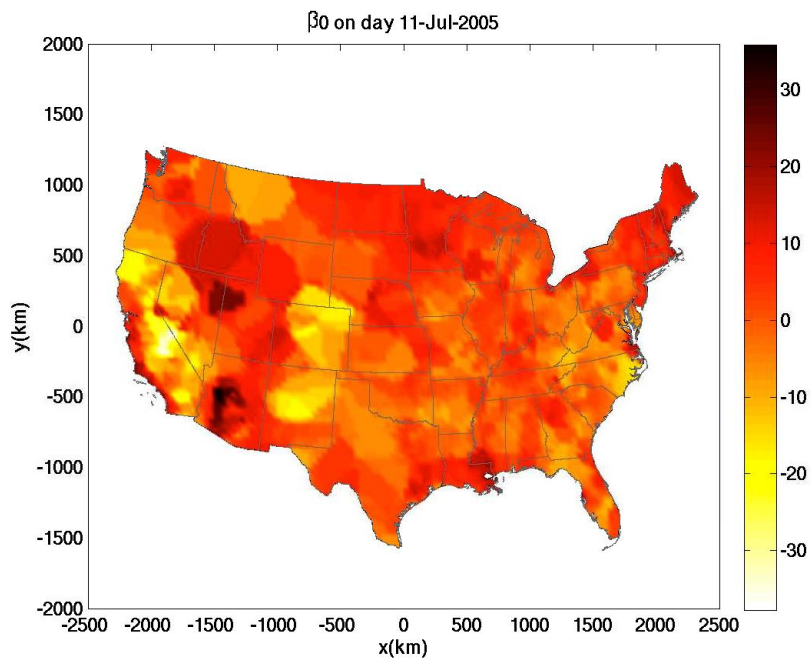


Figure A. 40s: Map of the parameter $\beta_0(s, t)$ across the continental United States for day 11-Jul-2005 for the DM8A ozone concentrations

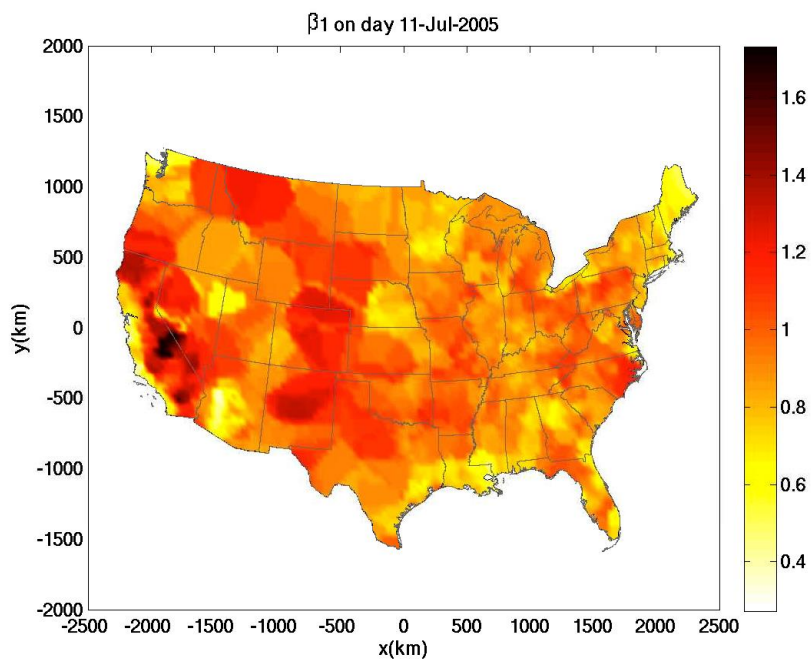


Figure A. 41s: Map of the parameter $\beta_1(s, t)$ across the continental United States for day 11-Jul-2005 for the DM8A ozone concentrations

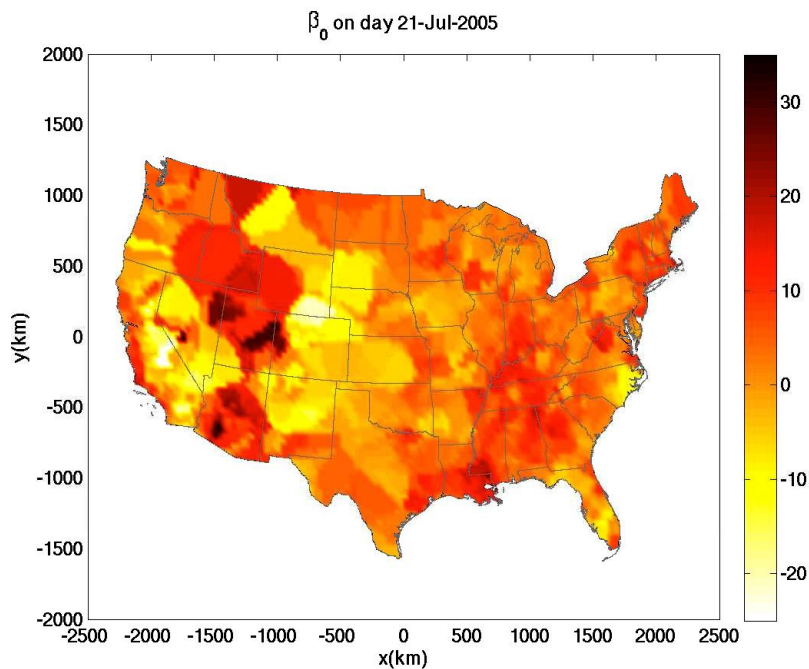


Figure A. 42s: Map of the parameter $\beta_0(s, t)$ across the continental United States for day 21-Jul-2005 for the D24A ozone concentrations

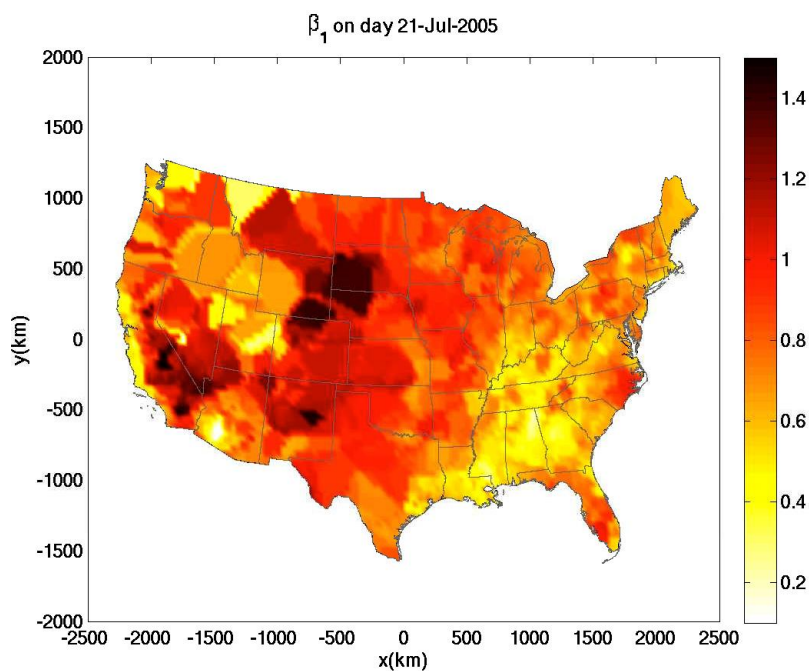


Figure A. 43s: Map of the parameter $\beta_1(s, t)$ across the continental United States for day 21-Jul-2005 for the D24A ozone concentrations

A.6.4. Comparison of the validation results between the RAMP and Co-kriging approaches

A validation analysis was also conducted to assess the accuracy of the co-kriging estimation approaches. Each observed value z_j at space/time point $= (s_j, t_j)$ is compared with the corresponding ozone concentration z_j^* re-estimated using only non-located data outside of a radius $r_v = 0km$ of s_j .

We compare the validation error, including the Root Mean Square Error RMSE (ppb) and the R^2 (unitless) between RAMP and Co-kriging approach. The validation results are shown in Table A6s.

Table A. 6s: Error statistics of validation analysis to compare BME with RAMP and Co-kriging with a parametric approach.

Statistics	DM8A		D24A	
	BME with RAMP	Cokriging	BME with RAMP	Cokriging
RMSE (ppb)	5.445	6.538	5.487	6.560
R^2 (unitless)	0.893	0.845	0.813	0.727

A.7. A discussion about the limitation of using the daily averages of the hourly ozone observations as hard data

In our work, as one of the limitations of our approach, we treated the daily averages processed from the hourly ozone observations as hard data for computational efficiency. We assumed that there is no difference between the true mean of the hourly concentrations and the sample mean calculated from a finite sample size. This limitation may be investigated in the future work. One of the alternative approaches is to use a finite population correction factor to estimate the uncertainty associated with the difference between the true mean of hourly ozone concentrations and the sample mean calculated from an incomplete sample size.

In this case, the daily ozone concentrations are treated as soft data, characterized by a PDF. We may assume that there are $n_{s,t}$ hourly observations $z_{i,j}$ at monitoring site s and time t and an approximation for the PDF is a normal distribution with the mean $\mu_{s,t}$ and the standard deviation $\sigma_{s,t}$ truncated below zero, then the standard deviation can be calculated as below:

$$\sigma_{s,t} = \sqrt{\frac{\sum_{j=1}^{n_{s,t}} (z_{i,j} - \mu_{s,t})^2 / (n_{s,t} - 1)}{n_{s,t}}} X \sqrt{\frac{n^* - n_{s,t}}{n^*}} \quad (\text{A31s})$$

Where the first term is the standard deviation of the sample mean and the second term is a finite population correction factor to account for incompleteness of the finite population size n^* , which is 24 for the D24A ozone concentrations.

APPENDIX B - SUPPORTING INFORMATION FOR BME INTEGRATION OF OZONE OBSERVATIONS AND CTM PREDICTIONS AT MULTIPLE TIME SCALES⁴

B.1. Hourly Ozone CTM model performance evaluation

In our research, the air quality model predictions were used to construct the soft data, consisted of the expected values and the uncertainties associated with the expected values at each grid cell, based on how well they reproduce the observed values. To develop a meaningful way to characterize model performance we need to investigate how the air quality model performance changes across space and time. To evaluate the overall air quality model performance for hourly ozone concentrations, the modeled concentrations are paired with the observational data in space and time at each monitoring site. We calculated error statistics such as mean prediction error (ME), the standard deviation of the prediction error (SE), the mean normalized bias (MNB) and the mean normalized gross error (MNGE) for the hourly ozone concentrations on each site.

$$ME(R) = \frac{1}{N(R)} \sum_{j=1}^{N(R)} (\tilde{z}_j - z_j) \quad (B1s)$$

$$SE(R) = \sqrt{\frac{\sum_{j=1}^{N(R)} ((\tilde{z}_j - z_j) - ME(R))^2}{N(R) - 1}} \quad (B2s)$$

$$MNB(R) = \frac{1}{N(R)} \sum_{j=1}^{N(R)} 100\% * (\tilde{z}_j - z_j) / z_j \quad (B3s)$$

$$MNGE(R) = \frac{1}{N(R)} \sum_{j=1}^{N(R)} 100\% * (|\tilde{z}_j - z_j| / z_j) \quad (B4s)$$

Where z_j is the observed hourly concentration, \tilde{z}_j is its corresponding CTM prediction value and $N(R)$ is the number of observation-prediction pairs in region R .

⁴ This appendix is planned to be submitted as the supporting information of an article to the Journal of Environmental Science & Technology. Xu, Yadong, Serre, L. Marc, Reyes, Jeanette, Vizuite, William. Supporting Information for BME Integration of Ozone Observations and CTM Predictions at Multiple Time Scales.

The hourly ozone model performance error statistics, including ME, SDE, MNB and MNGE, are summarized in the boxplots as below (Figure B.1s & 2s). The variability of ME and SE values across individual sites are shown in the maps of figures B.3s-6s. The seasonal differences in CTM model performance are shown in figures 8s-11s. Table B.1s summarizes the differences of CTM model performance among hourly ozone and daily metrics.

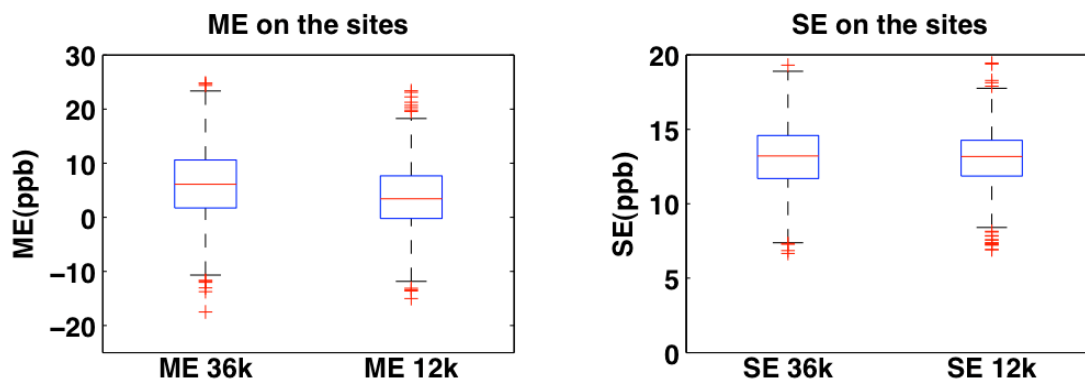


Figure B. 1s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the hourly ozone concentrations at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with 36x36km² and 12X12km² grid cell resolution.

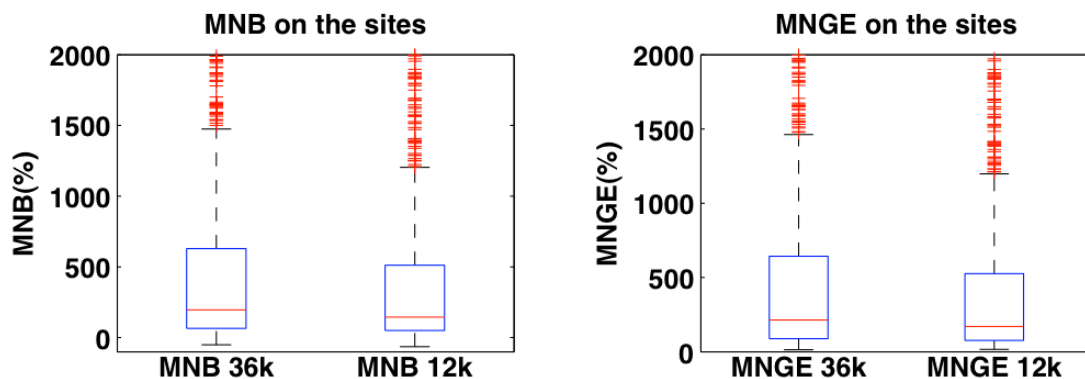


Figure B. 2s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the hourly ozone concentrations at ozone monitoring sites (888 sites in total) covered by both domains of the CAMx model simulations with 36x36km² and 12X12km² grid cell resolution.

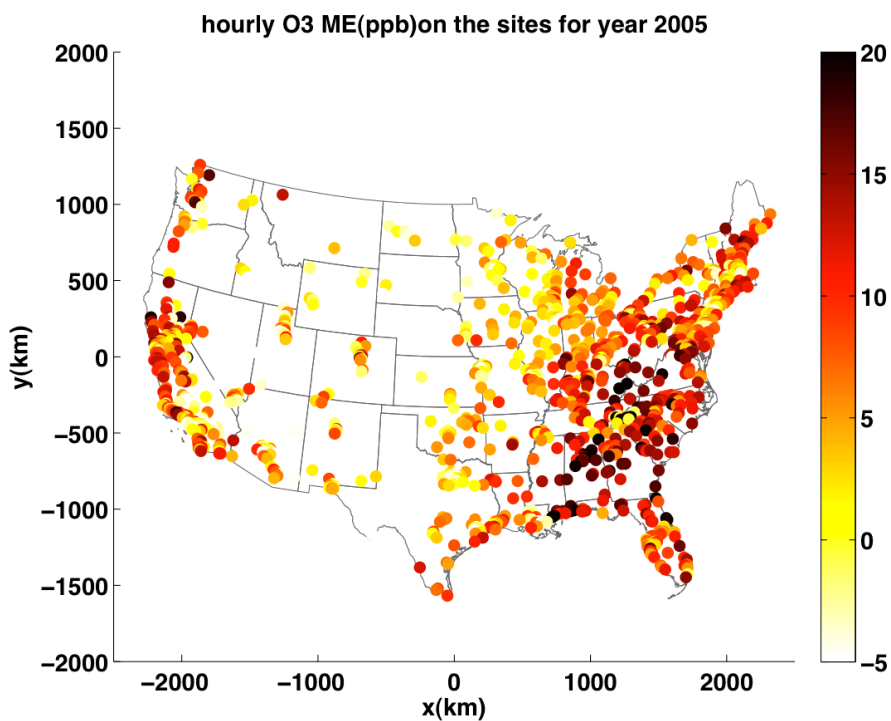


Figure B. 3s: The hourly O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $36 \times 36 \text{ km}^2$ grid cell resolution.

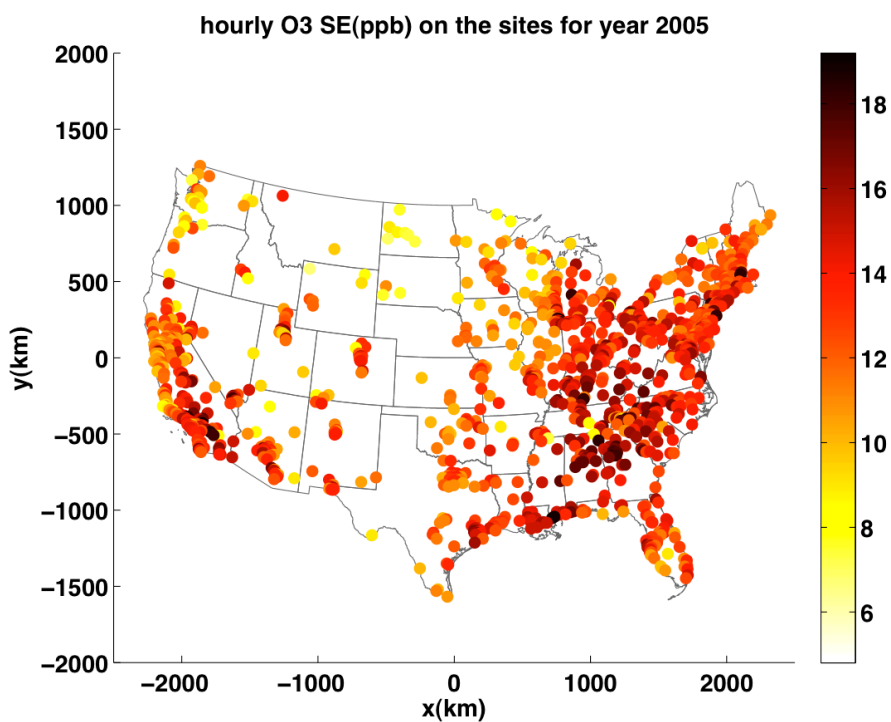


Figure B. 4s: The hourly O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a $36 \times 36 \text{ km}^2$ grid cell resolution

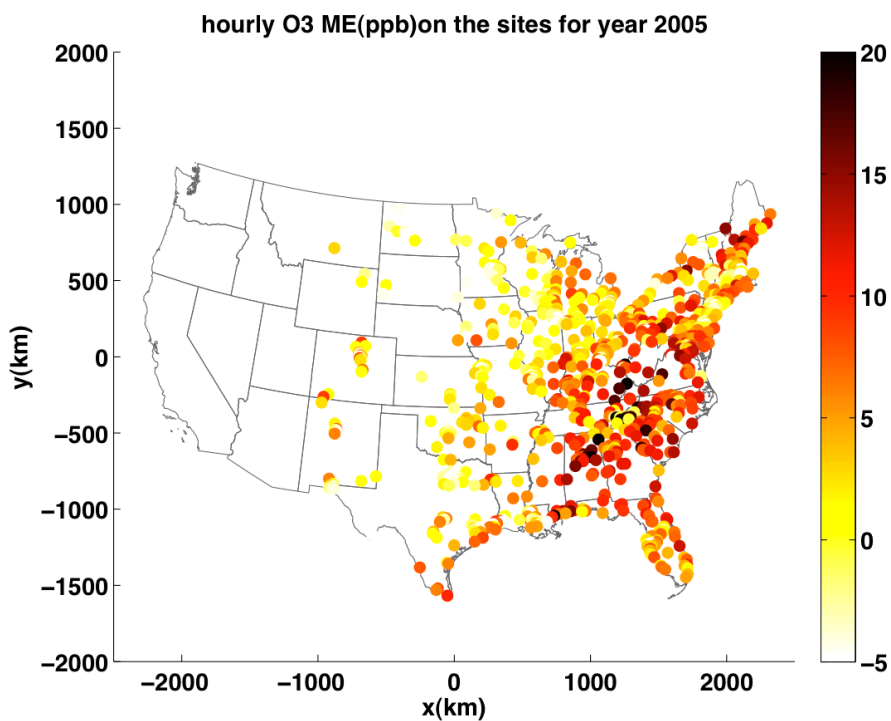


Figure B. 5s: The hourly O3 mean prediction error (ME) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km² grid cell resolution.

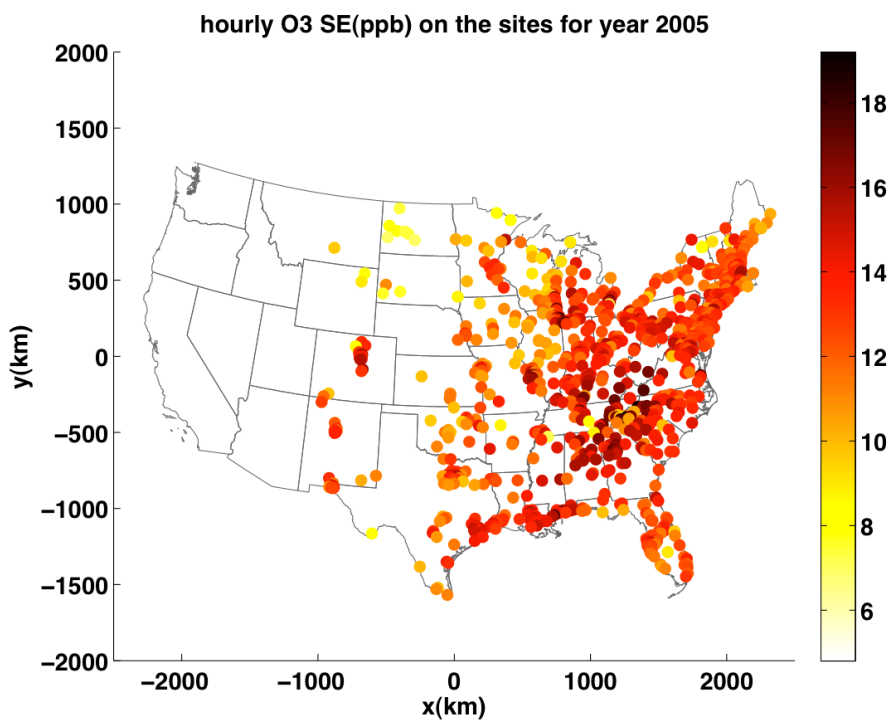


Figure B. 6s: The hourly O3 standard deviation of the prediction error (SE) (in ppb) at each AQS sites for the CAMx simulation of year 2005 using a 12x12km² grid cell resolution.

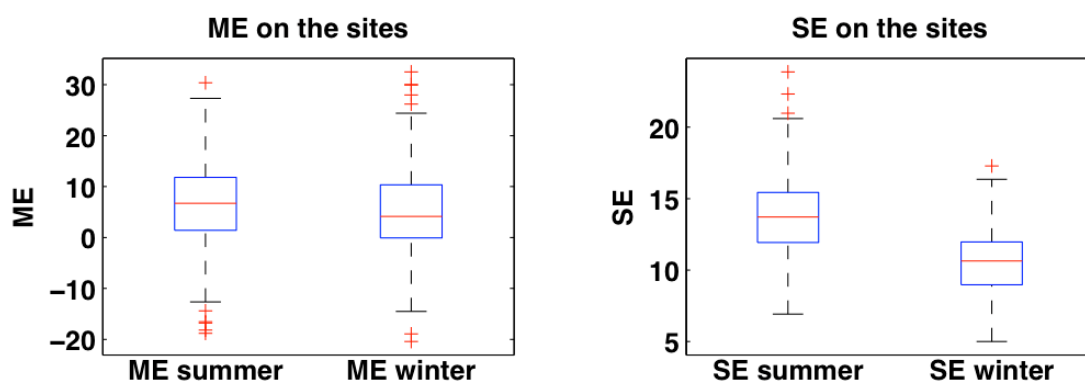


Figure B. 7s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with 36x36km² grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

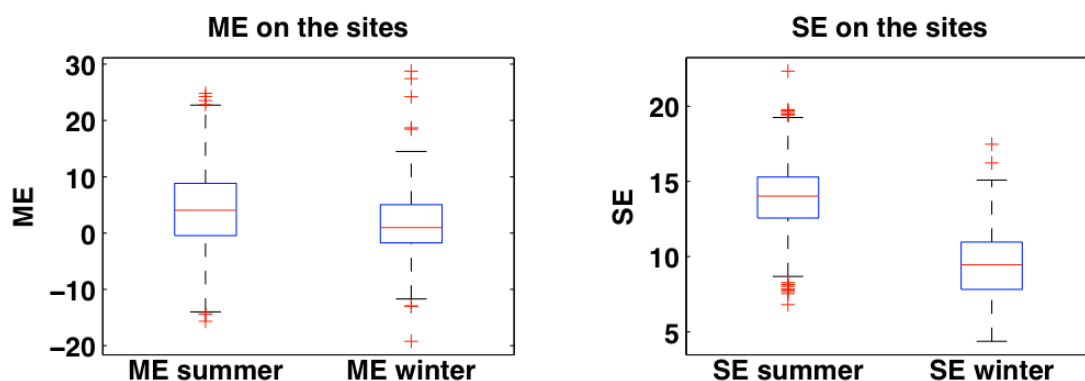


Figure B. 8s: Boxplots for the mean prediction errors (ME) and the standard deviation of these prediction errors (SE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with 12x12km² grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

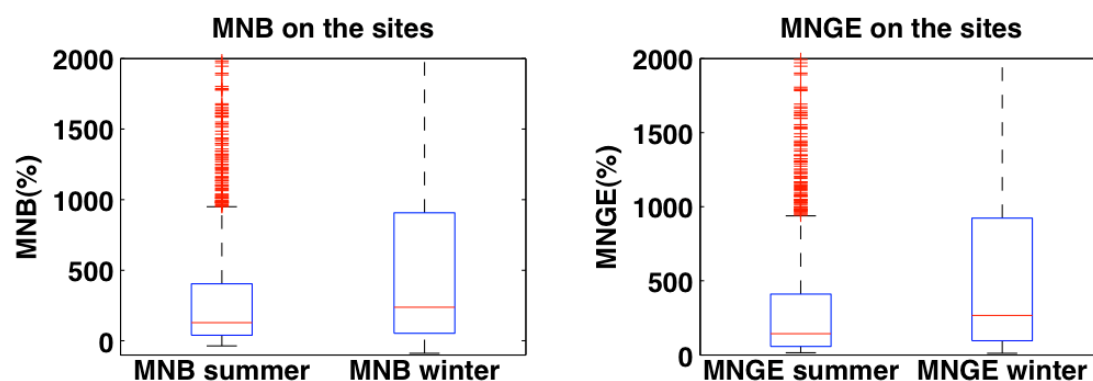


Figure B. 9s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with 36x36km² grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

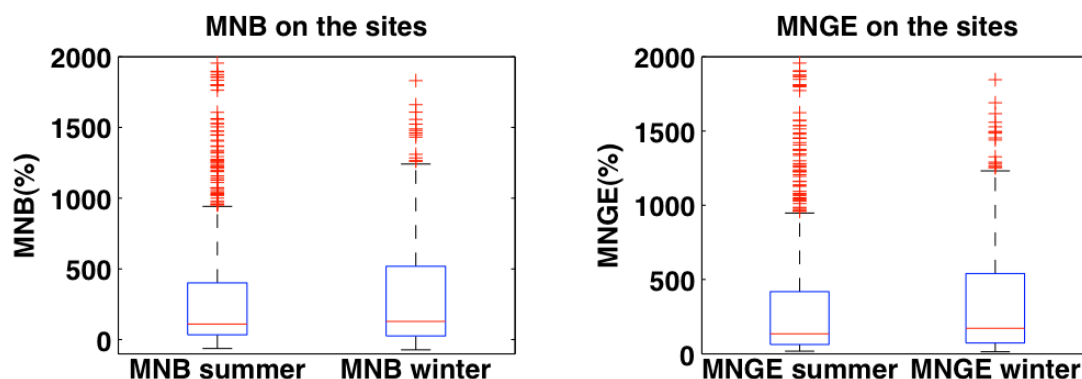


Figure B. 10s: Boxplots for the mean normalized bias (MNB) and the mean normalized gross error (MNGE) of the hourly ozone concentrations at ozone monitoring sites for the CAMx model simulations with 12x12km² grid cell resolution, separated by summer months (May, June, July and August) versus winter months (November, December, January and February).

Table B. 1s: Summary of error statistics for CTM model performance at ozone monitoring sites for three ozone metrics: hourly ozone, DM8A and D24A

Statistics	Sites in CTM domain	Hourly				DM8A				D24A			
		mean	p25	p50	p75	mean	p25	p50	p75	mean	p25	p50	p75
ME (ppb)	36x36	6.1	1.7	6.1	10.6	4.3	1.2	4.2	7.5	6.1	1.7	6.1	10.6
	12x12	3.7	-0.2	3.4	7.7	1.7	-0.9	1.4	4.2	3.7	-0.2	3.4	7.6
SE (ppb)	36x36	13.1	11.7	13.2	14.6	9.8	8.7	9.7	10.8	8.3	7.4	8.2	9.1
	12x12	13.0	11.9	13.2	14.3	9.5	8.7	9.4	10.3	7.9	7.1	7.9	8.6
MNB (%)	36x36	555.8	65.6	196.6	629.5	23.8	7.2	16.5	27.7	43.3	11.8	31.2	55
	12x12	433.2	49.9	145.1	512.3	13.9	2.3	9.3	17	25.8	4.8	19.8	38.6
MNGE (%)	36x36	573.8	87.5	213.4	643.1	32.8	18.5	24.7	33.2	52.9	25.9	39	58.4
	12x12	456.2	77	170.2	526.6	26	17.4	20.8	25.5	39.3	23.5	31.1	43.8

B.2. Parameter selection for the offset analysis and covariance model

B.2.1 Offset analysis

The offset is considered a deterministic function of space and time that can be mathematically calculated for any space/time point \mathbf{p} without error. Intuitively there is an inverse relationship between the amount variability in the offset and the remaining variability of the transformed data. An offset with shorter ranges (i.e. with small a_r and a_t values) can capture more variability; therefore the transformed data has less variability. Conversely an offset with longer ranges (i.e. with longer a_r and a_t values) can describe less variability; therefore the resulting transformed data retains larger variability of the original data and thus has larger variance. To investigate the effect of different offset parameters on the variance and autocorrelation of the transformed data, we constructed 4 sets of offset parameters, describing ranges at short, intermediate, long and very long scales, respectively, with a_r and a_t values shown in Table B.2s.

Table B. 2s. Spatial and temporal offset parameters and their label

Offset name	a_r (km)	a_t (days)
short	5	2
Intermediate	50	10
long	300	50
Very long	1000	100

Each set of offset parameters can be assessed visually through maps (Figures B12s-16s) and time series plot (Figure B17s). The map in figure 12s shows the observed hourly ozone concentrations at UTC hour 20050711T00, while the following four maps in figures B13s, B14s, B15s and B16s show the short, intermediate, long and very long offsets. As can be seen from these figures, the short offset captures variability at a finer scale, which is fairly close to the variability in the original observational ozone. As the ranges get longer, the offset at the monitoring sites starts to smooth out the data. In the map (Figure B17s) with the very long range offset, the offsets at all the monitoring sites have nearly the same values.

The final parameters used in this study need to be selected in conjunction with the autocorrelation of the space-time covariance model processed through the offset-removed transformed data, which is described next.

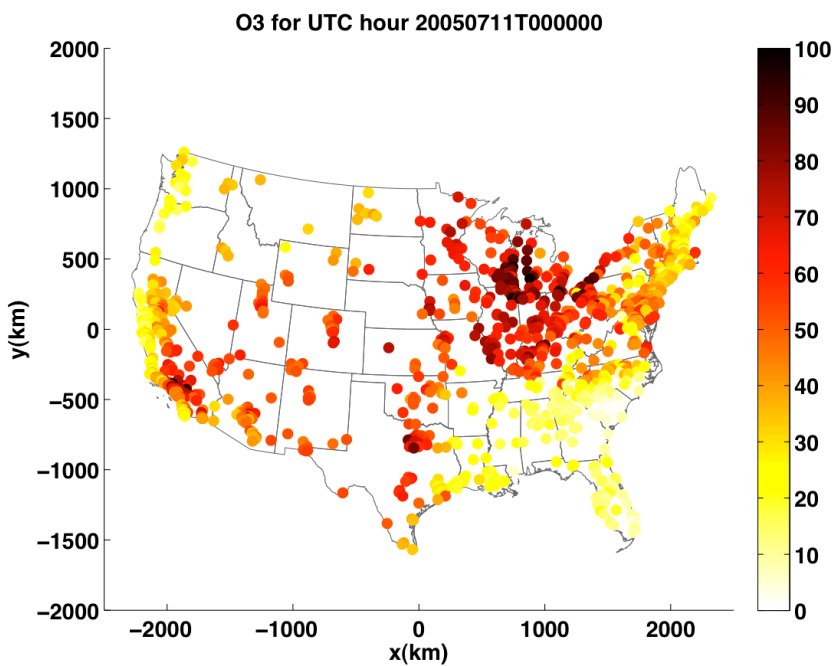


Figure B. 11s: The observed hourly O3 concentrations on the monitoring sites for UTC hour 20050711T000000

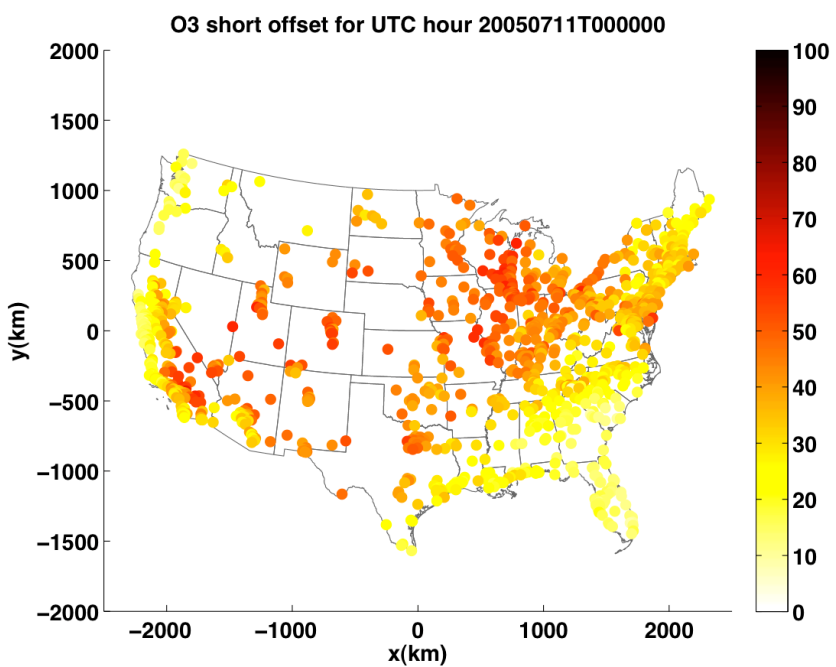


Figure B. 12s: The short range offset for hourly O3 on the monitoring sites for UTC hour 20050711T000000

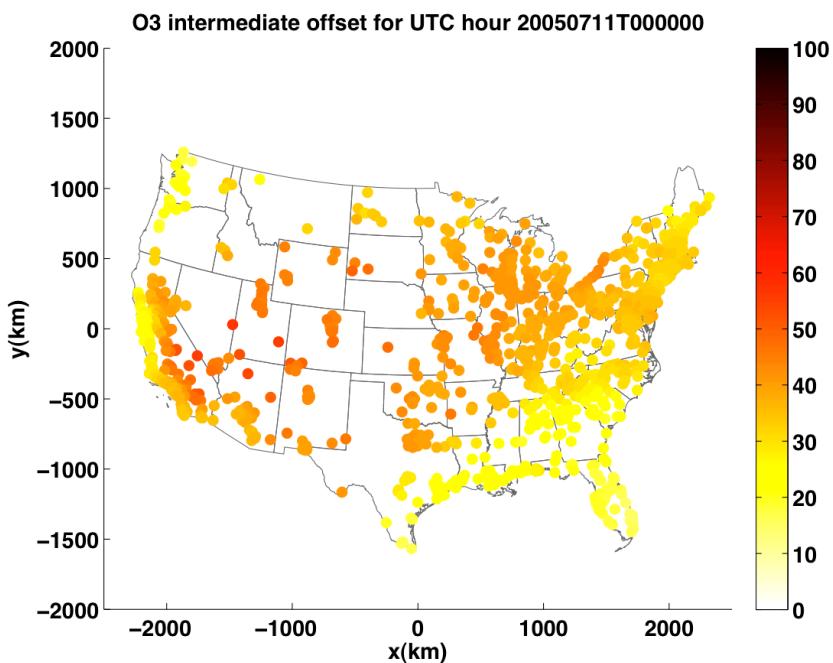


Figure B. 13s: The intermediate range offset for hourly O₃ on the monitoring sites for UTC hour 20050711T000000

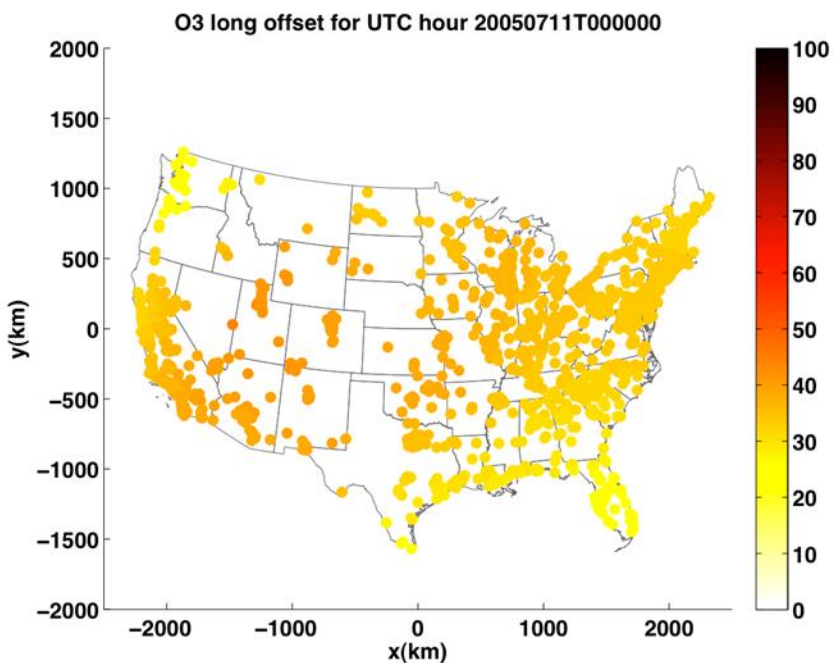


Figure B. 14s: The long range offset for hourly O₃ on the monitoring sites for UTC hour 20050711T000000

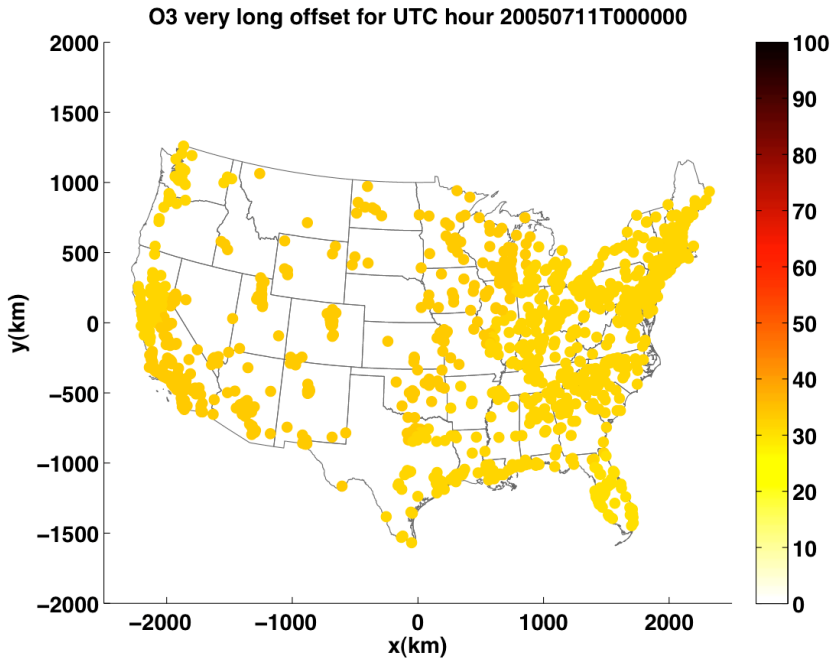


Figure B. 15s: The very long range offset for hourly O₃ on the monitoring sites for UTC hour 20050711T000000

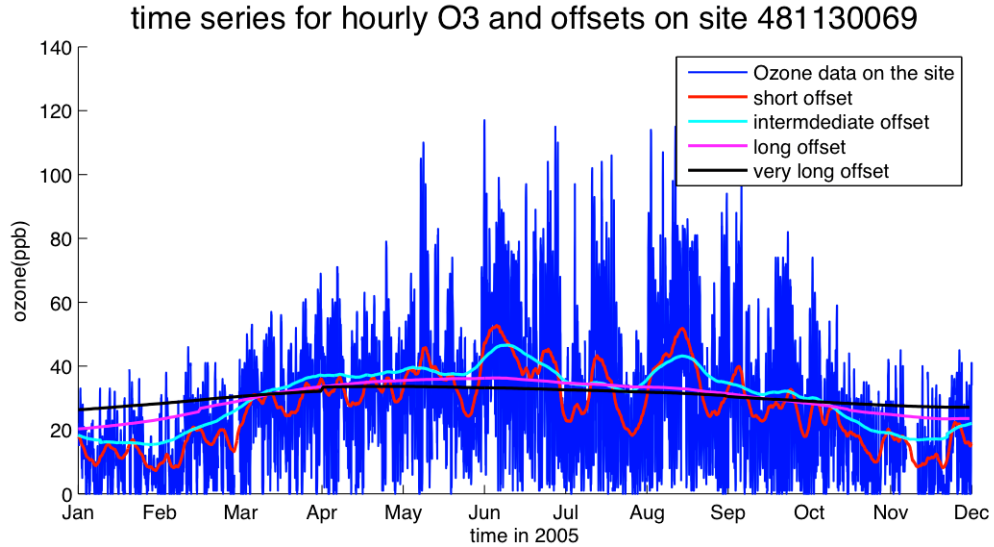


Figure B. 16s: The time series of the hourly O₃ concentrations and four offsets on a randomly selected site (SiteID:481130069) for year 2005

B.2.2 Covariance model

Several other covariance models were attempted and evaluated; however, a 3-structured exponential/cosine model was selected that provided the best overall performance with respect to least square fit and model interpretability. The equation of the 3-structured covariance model is given by the equation 1 in the main manuscript. Table B.3s listed the parameter values of the 3-structured model for each set of offset ranges. The corresponding models are graphed in figures 18s.

Table B. 3s: Parameter values of the 3-structured covariance model for each set of offset ranges

Offset range	C_0	α	β	a_{r1} (km)	a_{r2} (km)	a_{t1} (days)	a_{t2} (days)
short	222.6	0.22	0.24	1.76×10^2	7.83×10^3	0.44	0.5
intermediate	273.6	0.25	0.28	1.05×10^2	8.27×10^3	2.1	0.5
long	334.3	0.28	0.28	8.957×10	6.695×10^3	4.9	0.5
very long	358.3	0.28	0.30	1.024×10^2	6.367×10^3	6.7	0.5

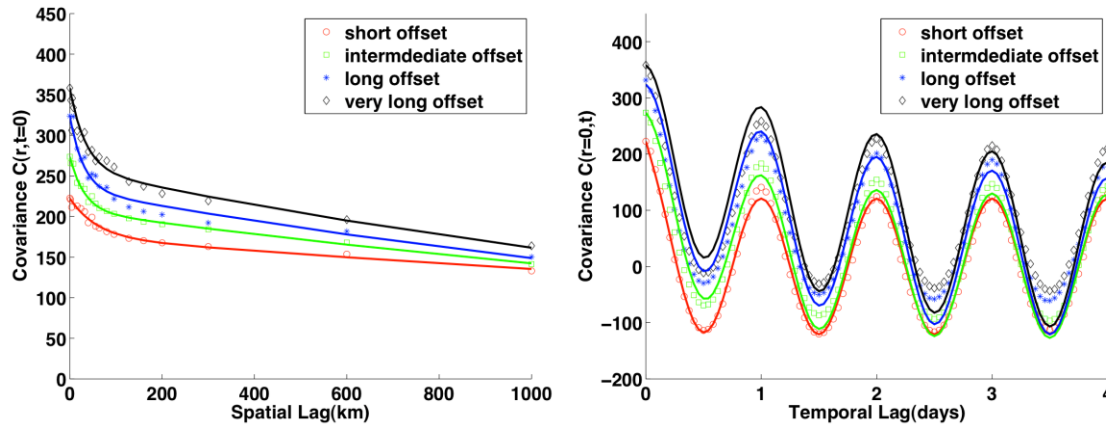


Figure B. 17s: Graphs of the 3-structured exponential/exponential/cosine covariance models with respect to the spatial lag when the temporal lag is set to 0 (**Left**) and with respect to the temporal lag when the spatial lag is set to 0 (**Right**) for hourly O3

There are four sets of offsets and therefore there are four resulting covariance models corresponding to each of the transformed data. To assess which of the 4 offsets should be selected to transform hourly ozone concentrations for the BME simulation, dominance plots are created as shown below (figure 19s). According to the dominance plots, the intermediate offset ($a_r = 50 \text{ km}$ and $a_t = 10 \text{ day}$) was chosen because it contains a larger spatial autocorrelation than other offsets, while only sacrificing a small amount

of variance compared with the one with the short offset but it contains a larger temporal autocorrelation than the one with the short offset. Thus, the intermediate offset offers an attractive tradeoff of lowering variance while maintaining autocorrelation in the transformed data. As stated above, the offset selected will have a combination of the lowest variance and the highest autocorrelation. The intermediate offset and its corresponding covariance model are used in the subsequent BME simulation.

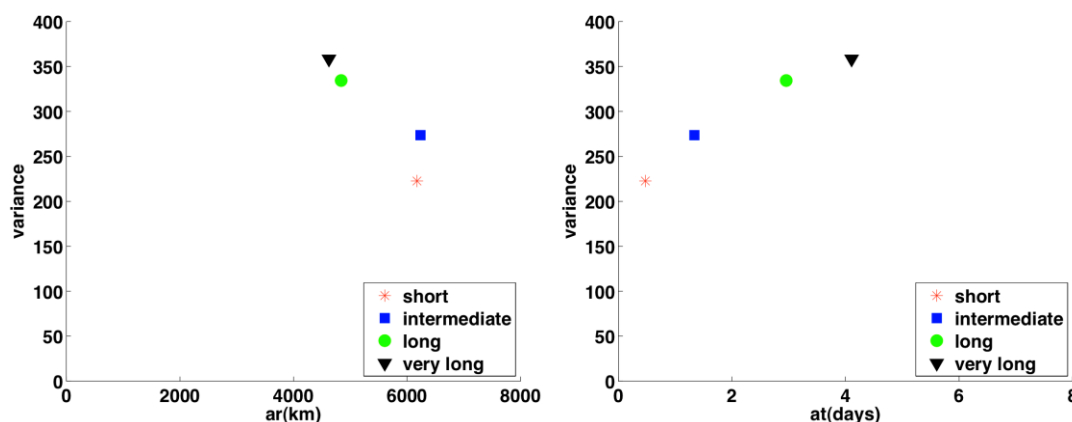


Figure B. 18s: Dominance plot showing how the variance changes with respect to the spatial range (**Left**) and with respect to the temporal range (**Right**) for covariance model corresponding to each offset range for hourly O₃

B.3 Additional results of BME ozone estimates

The main manuscript shows the maps of BME ozone estimates from two up-scaling methods USM1 and USM2 through estimation scenario RAMP on a selected day (11-July-2005). Here we provide additional results to show the BME ozone estimate through estimation scenario OBS demonstrating the differences between two up-scaling methods USM1 and USM2. Figure 20s shows for 11-July-2005 the BME estimates of DM8A and D24A ozone concentrations obtained from methods USM1 and USM2 through estimation scenario OBS. The differences in daily ozone concentrations between these two methods are below 5ppb for most part of the continental U.S. For the DM8A, the averaged difference across the continental U.S is 4.4ppb with the largest difference at 37.7 ppb. For the D24A, the averaged difference is 3.5 ppb with the largest difference at 30.0 ppb.

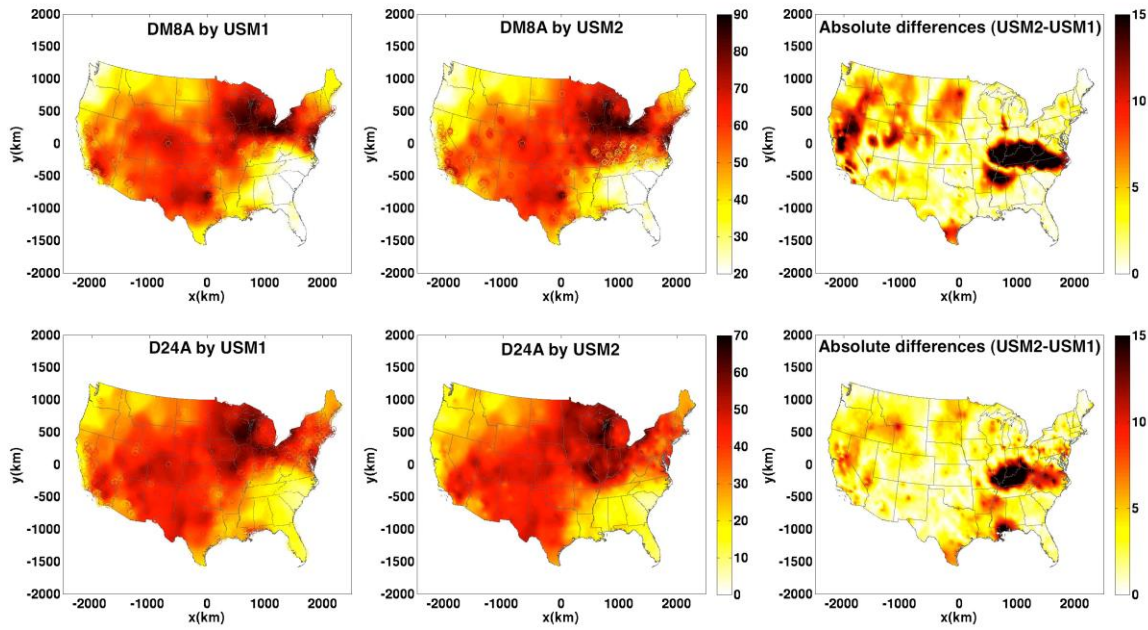


Figure B. 19s: Maps of scenario OBS BME mean estimates of the DM8A (**Top**) and the D24A (**Bottom**) ozone concentrations (ppb) on day Jul-11-2005 obtained from upscaling method USM1 (**Left**) and USM2 (**Middle**), and the absolute differences between these two methods (**Right**).

The computational costs and efforts of generating the maps using USM2 is substantially higher than the ones using USM1. The following table summarizes the CPU run time for generating a map covering the continental U.S. on one day.

Table B. 4s. Computational costs for generating BME ozone maps USM1 vs USM2[#]

Ozone metrics/upscaling method	Averaged CPU run time for one space/time location (seconds)	Averaged total CPU time for BME map on one day (seconds)	Averaged total CPU time (hours)
DM8A/USM1	2.9	16576*2.9=48070.4	13.4
DM8A/USM2	23.1	16576*23.1*32=12252979.2	3403.6
D24A/USM1	2.9	16576*2.9=48070.4	13.4
D24A/USM2	23.1	16576*23.1*24=9189734.4	2552.7

[#] The maps are created having the same domain as the CTM modeling with a total number of 16576 grid cells, using a Linux-based computing system, each compute node with 12-core, 2.93 GHz Intel processors, 12M L3 cache and 96GB memory for a total of 1776 processing cores at 2:1 ratio IB interconnect.

B.4 Additional results for the validation analysis

B.4.1 Additional error statistics for the validation analysis

Besides the error statistics of RMSE, R^2 and PCR^2 shown in the main manuscript, we also used other commonly used error statistics to quantify the estimation accuracy, which is a function of the validation radius r_v . They consist of the Mean Normalized Bias (MNB) and the Mean Normalized Gross Error (MNGE) between observations and re-estimated values. Table 5s show the MNB and MNGE error statistics as a function of cross-validation radius r_v . When comparing USM1 with USM2, we can see USM1 has a smaller estimation error, with lower MNB and MNGE, for both daily metrics of DM8A and D24A.

Table B. 5s. Validation statistics for BME data integration scenarios OBS and RAMP

Validation radius r_v (km)	USM1				USM2			
	0	36	72	108	0	36	72	108
MNB_{OBS} (%)	7.595	10.809	11.286	11.576	8.924	14.884	15.540	16.134
MNB_{RAMP} (%)	8.202	10.785	11.072	11.171	9.375	12.878	12.802	12.768
$MNGE_{OBS}$ (%)	19.278	22.257	22.923	23.553	21.197	25.800	26.605	27.598
$MNGE_{RAMP}$ (%)	19.329	21.786	22.373	22.858	21.844	25.652	26.375	26.745

B.4.2 The additional validation analyses show the influence of the grid cell resolution on the accuracy of the BME estimates

To demonstrate the influence of the CTM grid cell resolution on the accuracy of the BME estimates, we also calculated the error statistics, including R^2 and RMSE, between the observed and re-estimated daily ozone concentrations on each monitoring site (888 sites in total) where the validation analysis was conducted with two sets of soft data, one from CTM with $36 \times 36 \text{ km}^2$ and the other from CTM with $12 \times 12 \text{ km}^2$ grid cell resolution. We map the changes in these error statistics on each monitoring site, as shown in Figure B.20s&B.21s. For the daily metric D24A, we can see that the majority of the sites (539 out of 888 sites) have a R^2 increase when using finer grid cell resolution soft data, while 676 out of 888 sites have a RMSE decrease when using soft data with $12 \times 12 \text{ km}^2$ grid cell resolution compared with the one with $36 \times 36 \text{ km}^2$ grid cell resolution. For the daily metric DM8A (Figure B.21s), we also see a significant improvement at a majority of the sites when using soft data with finer grid cell resolution.

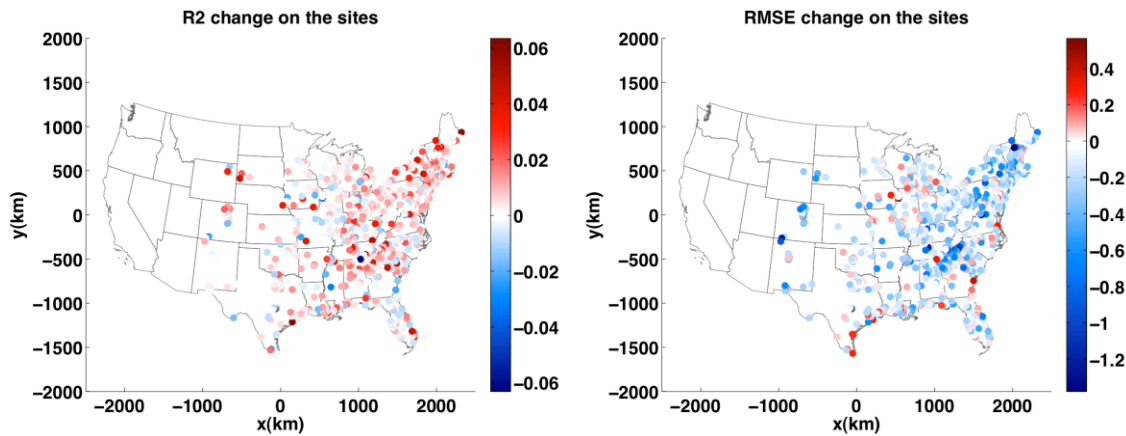


Figure B. 20s: Maps of R^2 (Left) and RMSE (Right) changes for validation analysis conducted with two sets of soft data, one from CTM with $36 \times 36 \text{ km}^2$ and the other from CTM with $12 \times 12 \text{ km}^2$ grid cell resolution for D24A at validation radius $R_v=0 \text{ km}$. Red color in the maps means R^2 and RMSE increase when change the soft data from $36 \times 36 \text{ km}^2$ to $12 \times 12 \text{ km}^2$, while blue color means R^2 and RMSE decrease.

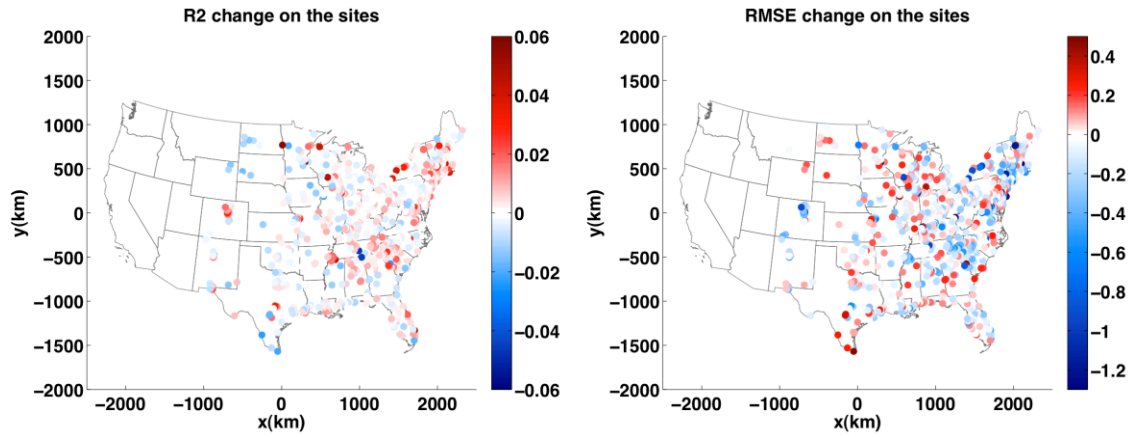


Figure B. 21s: Maps of R^2 (**Left**) and RMSE (**Right**) changes for validation analysis conducted with two sets of soft data, one from CTM with $36 \times 36 \text{ km}^2$ and the other from CTM with $12 \times 12 \text{ km}^2$ grid cell resolution for DM8A at validation radius $R_v=0 \text{ km}$. Red color in the maps means R^2 and RMSE increase when change the soft data from $36 \times 36 \text{ km}^2$ to $12 \times 12 \text{ km}^2$, while blue color

APPENDIX C-WHIMS PROJECT REPORT-CODE DOCUMENTATION AND QUALITY ASSURANCE FOR ESTIMATION OF DAILY O₃ FROM 1993 TO 2010 USING OBSERVATIONS AND CTM

This document describes the code used to calculate daily BME estimates of O₃ for the WHIMS study from 1993-2010. A description of the data set is given along with a summary of methods used to calculate estimates. The organization of all code is described along with instructions on its execution. Results of estimation and QAQC are displayed.

C.1 Introduction

The Bayesian Maximum Entropy (BME) estimation method is used to perform an interpolation of observed daily ozone concentrations and obtain at each participant location an estimate of the daily ozone concentration for years 1993 to 2010. *Data:* The hourly concentrations were downloaded from the Air Quality System (AQS) maintained by the U.S. Environmental Protection Agency (EPA) and Chemical Transport Model (CTM) hourly data obtained from CMAQ and CAMx. *Methods:* The AQS raw data were processed by the BME method as the followings: A database of observed ozone daily concentrations was obtained by getting the daily averages of hourly concentrations on each monitoring site. A database of CTM modeled ozone daily concentrations was also obtained by getting the daily averages of hourly concentrations on each grid cell. Then Localized bias-corrected CTM data were processed. A transformation of these data was used, which consisted in removing from the data an offset obtained using an exponential kernel smoothing (Lee et al., 2013) of the data. The exponential kernel smoothing was set so that the offset captured the spatial variability of the data over intermediate spatial distances and intermediate time scales. A 3-term exponential/exponential space/time covariance model was used to characterize the space/time autocorrelation in the offset removed data. The BME method was then used to estimate daily O₃ at un-sampled locations using the offset removed daily observations treated as hard data. Since the observations are treated as hard data and localized bias-corrected daily CTM data treated as soft data, the BME method reduces to the space/time Kriging with measurement error within the framework of linear geostatistics.

C.2 Materials

C.2.1 O₃ daily data

Observed data

The daily O₃ concentrations for each monitoring site/day during the study period (1993-2010) were constructed based on raw monitoring data from ozone monitoring stations measuring hourly O₃ concentrations using the procedure described here.

We downloaded hourly O₃ monitoring data (raw data) sampled during the study period (1993-2010) from the Air Quality Systems (AQS) database maintained by the U.S. Environmental Protection Agency (EPA), which is a repository of the monitoring data collected across various monitoring networks. Various monitoring methods were used to collect hourly O₃ data, including reference methods and equivalent methods. Hourly O₃ data collected at AQS database, either used EPA reference methods or valid equivalent methods, were considered to maintain the same level of data quality. O₃ hourly concentrations were reported in different units for different sites, which were marked as different unit codes in the database, including 007(parts per million), 008(parts per billion), 087(parts per ten million) and 040 (parts per 100 million). We converted all the units into a uniformed unit as parts per billion.

The longitudes and latitudes of the monitoring sites were not included in the original raw data file. The constructed hourly O₃ concentrations were joined with the geographic coordinates of the sites, which were stored in a separate site information file, by matching the unique combination of state code, county code and site ID between the data file and the site information file.

Finally we computed the daily average of hourly O₃ concentrations at each monitoring site to construct a daily O₃ concentration database from year 1993 to year 2010.

Modeled Data

Daily O₃ concentrations were also constructed from modeled hourly O₃ data from both the Community Multiscale Air Quality (CMAQ) and Comprehensive Air Quality Model with Extensions (CAMx) models. CMAQ and CAMx are Chemical Transport Models (CTMs). These air quality modeling

systems unites three kinds of models: meteorological models, emissions models, and a chemistry-transport model. They have been designed to approach air quality as a whole by including state-of-the-science capabilities for modeling multiple air quality issues. Modeled O₃ concentrations are available at a 36km resolution every hour for the years 2001, 2002, 2005 and 2007. Modeled data with finer horizontal grid cell resolution, such as 12km or 4km, are also available for part of the United States for shorter period of time. Data are projected using a Lambert Conic Conformal (LCC) projection.

C.3 Location of the study participants

In order to protect the confidentiality of the location of WHIMS participants, Dr. Whitsel provided to the Serre lab a large set of locations (N=17,461) that included within it the WHIMS participant locations (n=7479). The Serre lab was not given knowledge of which of the 17,461 locations were actual WHIMS participants locations. This in effect “hides” the participants amongst the large set of locations, which provides an added level of data protection. The locations were saved in a file named `partdata.mat` containing three columns: “partid”, “partx”, and “party”. In order to protect the confidentiality of WHIMS participants’ data, the “partid” field contains a randomized id that is generated solely for the purpose of this study and does not correspond to the actual WHIMS participant id. The “partx”, and “party” fields provide the spatial coordinates of each location record.

In order to protect the confidentiality of the location data, we present here examples that are based on 500 simulated (fake) case locations randomly located across the contiguous US. An example of these simulated (fake) case locations is shown as below.

partid	partx	party
1	-1132770.34	-659383.01
2	-1789928.05	257671.13
3	625256.73	-669169.08
4	626906.89	-847002.75
5	1603181.17	127434.18
6	-503791.48	-724386.12
7	15752.19	582908.44
8	1206674.01	-154597.59
9	-1837605.24	62486.18
10	-1765398.46	527184.83

C.4 Methods

C.4.1 Estimation of Daily O3 Concentration

BME estimation methodology

BME is a sophisticated geo-statistical method developed by Christakos. BME can perform spatial-temporal interpolation by incorporating information from many different sources and it is implemented using the *BMElib* suite of functions in MATLAB™. The implementation and performance of BME have been detailed in other works, and can be summarized as performing the following steps: 1) gathering the general knowledge base (G-KB) and site-specific knowledge base (S-KB) characterizing the Space/Time Random Field (S/TRF) $X(\mathbf{p})$ representing a process at space/time coordinate $\mathbf{p}=(s,t)$, where s is the spatial coordinate and t is time, 2) using the *Maximum Entropy* principle of information theory to process the G-KB in the form of a prior Probability Distribution Function (PDF) f_G (for this study G-KB consists of the mean trend and covariance model $G = \{m_X(\mathbf{p}), c_X(\mathbf{p}, \mathbf{p}')\}$), 3) integrating S-KB using an epistemic *Bayesian* conditionalization rule on data f_S with and without measurement error to create a posterior PDF f_K and 4) creating space/time estimates based on the analysis calculated at \mathbf{p}_k .

Offset functions

The offset is considered a deterministic function of space and time that can be mathematically calculated for any space/time point \mathbf{p} without error. The offset is used to transform the daily O3 data into residual offset-removed data. Since the BME analysis will be performed on these transformed data in later stage, ideally an optimal offset is chosen to ensure that the transformed data have low variance for the accuracy of the estimation and have high autocorrelation to ensure that neighboring data locations are informative to the estimation location. In this study we selected several sets of offset parameters to capture variability of O3 at varying spatial and temporal scales, and we pick one that meets the above criteria most closely. The offset is calculated by using a space/time composite kernel smoothing function from the data within the local neighborhood (Lee et al., 2013).

Covariance modeling

The covariance model for the homogeneous/stationary S/TRF $X(\mathbf{p})$ is developed from the experimental covariance of the transformed data $\mathbf{x}_h = \mathbf{z}_h - o_z(\mathbf{p}_h)$. Several covariance models were evaluated and a 3-structured exponential model was selected because it provided the best overall performance with respect to least square fit and model interpretability. The intermediate offset was chosen because it offers the best tradeoff of lowering variance while maintaining autocorrelation in the transformed data, and this intermediate offset and its corresponding covariance model is used in the subsequent BME simulation.

BME estimation

The BME estimation method is used to perform an interpolation of observed daily O₃ concentrations and obtain at each participant location an estimate of the daily O₃ concentration for each day from 1993 to 2010. The BME method was then used to estimate daily O₃ at unsampled locations using the offset removed daily observations treated as hard data. The BME (simple kriging) mean estimates are in good agreement with the observed data, and the BME (simple kriging) variance show that the estimation is least accurate in areas where monitoring stations are sparse.

Estimation accuracy

To assess the estimation accuracy of the BME model simulation, a 10-fold estimation was conducted. We divided the ozone monitoring sites into 10 equal partitions. Each fold consists of a test set, which is 10% of the monitoring sites and also a training set, which is the remaining 90% of the monitoring sites. For each fold, we performed BME estimation (without recalculating the offset or the covariance model) by using the training set to predict the daily O₃ concentrations at the locations in the test set.

Quality Assurance

To confirm that the estimation of daily O₃ concentrations were performed correctly, quality assurance plots were created showing the concentrations at randomly selected estimation locations along with the concentrations from the closest 5 monitoring stations. The estimates from the randomly selected locations match well with the concentrations from the 5 closest neighboring sites, implying that the estimation was performed correctly.

C.5 Numerical implementation

C.5.1 Data and analysis folders

All data and analysis are housed in the folder “D:\AirCTMneuro\O₃est”. All steps of this analysis, including data preparation, offset analysis, covariance model, and estimation at the participant’s locations, QA/QC, 10-fold estimation and results have been documented in the folders as below.

Folder		The contents in the subfolder
00_CodeDocAndQualityAssurance		CodeDocandQA_dailyO3_Ver2.0.docx
01_prepdata	01a_matfiles	Hourly O3 and daily O3 data files in *.mat format
	01b_mfiles	MATLAB functions and/or shell scripts to perform formatting and aggregating hourly O3 data
02_projection	02a_matfiles	Projected US maps in *.mat format
	02b_mfiles	MATLAB functions to preform converting longitude and latitude in degrees into Lambert Conformal Conic projection
03_offset	03a_matfiles	Offset calculation results in *.mat format, including offsets on the monitoring sites, offsets on the participant locations, offsets on 10-fold estimation locations and also offsets on the grid for the US.
	03b_mfiles	MATLAB functions and/or shell scripts to perform offset calculations
	03c_maps_figures	Maps and time series plots for visualizing the offsets
04_covariance	04a_matfiles	Covariance calculation results in *.mat format, including the experimental covariance and also the covariance model after removing different offsets
	04b_mfiles	MATLAB functions to preform covariance calculation and model fitting
	04c_maps_figures	Figures for visualizing the experimental covariance and model fitting
05_estimation	05a_matfiles	The estimates for daily O3 in *.mat format, stored by year
	05b_mfiles	MATLAB functions and/or shell scripts to perform BME estimation with observational O3 data
	05c_maps_figures_csv	The maps and figures for visualizing the final estimates; the *.csv files that contain the final estimates for daily O3
06_QAQC	06a_matfiles	QA/QC outputs in *.mat files
	06b_mfiles	MATLAB functions to perform QA/QC on simulated locations and randomly selected participant locations
	06c_maps_figures	The time series plots for visualizing the daily O3 estimates and also the hard data on the closest sites
07_validation	07a_matfiles	10-fold estimation results in *.mat format
	07b_mfiles	MATLAB functions to perform 10-fold estimation
	07c_maps_figures_csv	The maps and figures for visualizing 10-fold estimation and also the *.csv files that contain the error statistics

Intermediate_matfiles		All the intermediate *.mat files during the analysis are stored in this folder for debug purposes and also improving computational efficiency
10_soft_data_construction	01_predata	Prepare CTM data and paired data
	02_test_parameters	Exploratory analysis to select parameters for soft data construction
	03_cal_soft_data	Generate soft data
	04_projection	Projected US maps in *.mat format and mapping functions
	05_QAQC	MATLAB functions to perform QA/QC on soft data construction
	Intermediate files	All the intermediate *.mat files during the analysis are stored in this folder for debug purposes and also improving computational efficiency

C.5.2 Instructions to estimate O3 concentration from year 1993 to 2010

To ensure the work can be completed in a reasonable timeline, all the analysis and model simulations were executed on a Linux cluster. Some of the MATLAB codes even needed to be executed in parallel. In that case, shell scripts were prepared for submitting multiple jobs at a time. In order to run the shell script, use the following command.

```
sh (shell script name)
```

In our analysis, MATLAB jobs and/or shell scripts were executed in the following sequential order. As all the intermediate files were saved in the corresponding sub-folders, the scripts can also be re-ran without regenerate the files in the previous steps.

Folder	Name of Shell script or MATLAB functions
01_prepdata/01b_mfiles	sh_to_run_01.sh
03_offset/03b_mfiles	sh_to_run_03_on_sites.sh sh_to_run_03_on_partlocs.sh runall_TS_daily_offsets_on_sites.m
04_covariance/04b_mfiles	runall_04.m
10_soft_data_construction	sh_to_run_cal_E_var_on_grid.sh
07_validation/07b_mfiles	sh_to_run_07_create_10folds.sh sh_to_run_07_10fold_est.sh
05_estimation/05b_mfiles	sh_to_run_05_cal_residual_hard.sh sh_to_run_05_on_partlocs.sh sh_to_run_05_write_csv.sh
06_QAQC/06b_mfiles	sh_to_run_06_QA.sh sh_to_run_06_post_QA.sh

C.6 Results

The results are saved under the sub-folder “/ 05_estimation/05c_maps_figures_csv/csv_files”, Each record in the data file “dailyO3_hardonly_Case_YYYY.csv” (or “dailyO3_hardsoft_Case_YYYY.csv”) has the data fields described below, where YYYY is the 4 digit year ranging from 1993-2010.

Field Name	Description
id	participants’ identification number
O3m_XXXXMMDD	BME mean estimate of daily O3 concentration on YYYY (4 digit year) MM (2 digit month) DD (2 digit day). This BME estimate is obtained using only AQS observations treated as hard data, in which case BME is also referred to as space/time Simple Kriging (SK). The date-time ranges from YYYY0101 to YYYY1231. Each column corresponds to a particular day.
O3sd_XXXXMMDD	Corresponding BME standard deviation of daily O3 concentration on YYYY (4 digit year) MM (2 digit month) DD (2 digit day). Since BME only uses observed values treated as hard data, then the BME variance is the same as the kriging variance of space/time SK. Each column corresponds to a particular day. The date-time range from YYYY0101 to YYYY1231.

C.7 QA/QC

To confirm that the estimation of daily O3 concentrations were performed correctly, quality assurance plots were created. Below are plots showing the concentrations at randomly selected estimation locations along with the concentrations of the closest 5 monitoring stations. Intuitively, BME estimations should be close to the values of its surrounding stations. In each figure, the predicted concentrations match well with the observational ozone data from the 5 surrounding monitors, implying that the estimation was performed correctly.

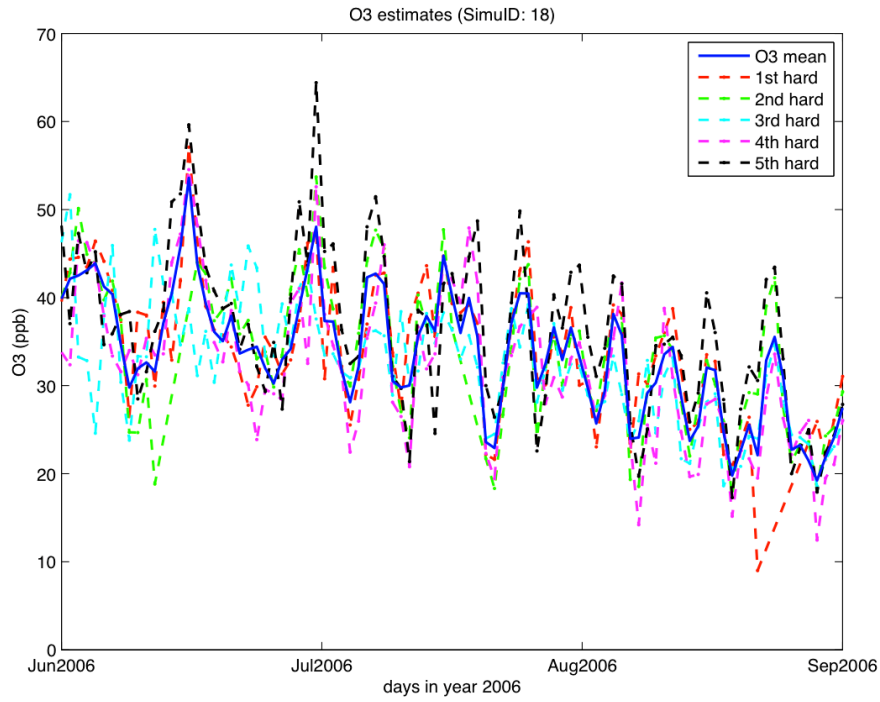


Figure C. 1: The time series plot showing the BME mean estimates on simulated ID18 and the hard data in the neighborhood

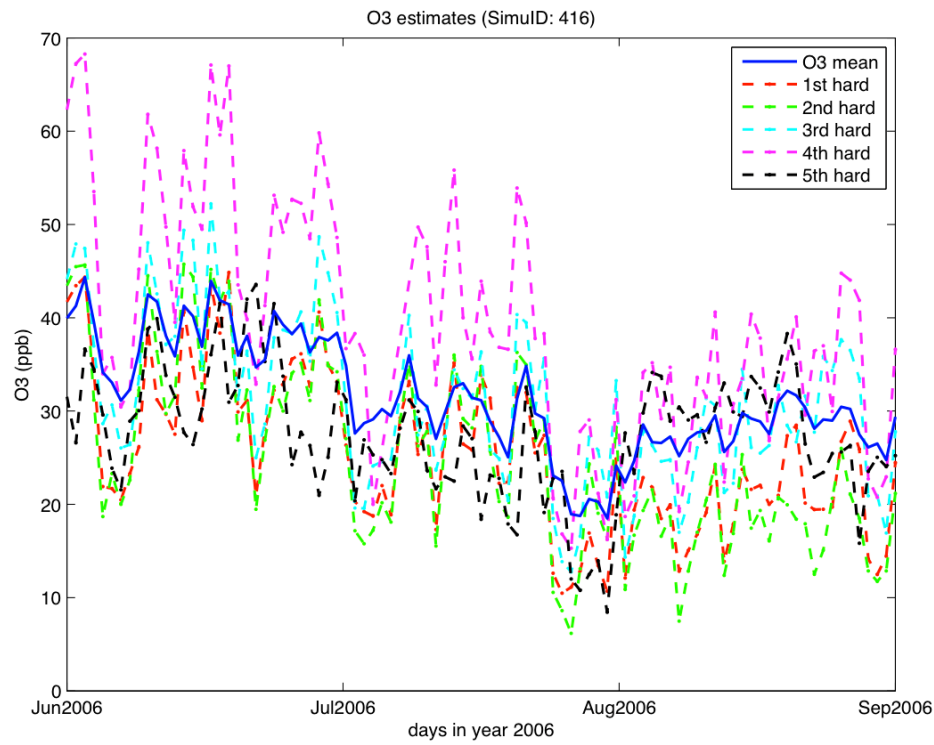


Figure C. 2: The time series plot showing the BME mean estimates on simulated ID416 and the hard data in the neighborhood

C.8 Date and version number

We obtained the study participants' location data from Dr. Eric Whitsel in a file named *will_marc_10_21_2013.sas7bdat* on October 25, 2013. Yadong Xu completed the estimation of daily O₃ concentrations for these locations by using AQS ozone observational data only in March 2014. Results of estimation were copied to txt files named *dailyO3_hardonly_Case_YYYY.csv*, which were delivered to Eric Whitsel in March 2014 as version 1.0; In August 2014, Yadong Xu completed the estimation of daily O₃ concentrations for these locations by incorporating both AQS ozone observational data and CTM data. Results of estimation were copied to txt files named *dailyO3_hardsoft_Case_YYYY.csv*, which were delivered to Eric Whitsel in August 2014 as version 2.0.

REFERENCES

1. EPA, U.S., *Final Report: Integrated Science Assessment of Ozone and Related Photochemical Oxidants*. U.S. Environmental Protection Agency, Washington, DC, 2013. **EPA/600/R-10/076F**(2013).
2. Lin, S., et al., *Ambient ozone concentration and hospital admissions due to childhood respiratory diseases in New York State, 1991–2001*. Environmental Research, 2008. **108**(1): p. 42-47.
3. Tager, I.B., et al., *Chronic Exposure to Ambient Ozone and Lung Function in Young Adults*. Epidemiology, 2005. **16**(6): p. 751-759.
4. Pride, K.R., et al., *Association of short-term exposure to ground-level ozone and respiratory outpatient clinic visits in a rural location – Sublette County, Wyoming, 2008–2011*. Environmental Research, 2015. **137**: p. 1-7.
5. Moore, K., et al., *Ambient Ozone Concentrations Cause Increased Hospitalizations for Asthma in Children: An 18-Year Study in Southern California*. Environmental Health Perspectives, 2008. **116**(8): p. 1063-1070.
6. Ferdinands, J.M., et al., *Breath acidification in adolescent runners exposed to atmospheric pollution: A prospective, repeated measures observational study*. Environmental Health: A Global Access Science Source, 2008. **7**: p. 11.
7. Bell, M.L., R.D. Peng, and F. Dominici, *The Exposure–Response Curve for Ozone and Risk of Mortality and the Adequacy of Current Ozone Regulations*. Environmental Health Perspectives, 2006. **114**(4): p. 532-536.
8. Franklin, M. and J. Schwartz, *The impact of secondary particles on the association between ambient ozone and mortality*. Environmental health perspectives, 2008. **116**(4): p. 453.
9. Appel, K.W., et al., *Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance: Part I, ÅiOzone*. Atmospheric Environment, 2007. **41**(40): p. 9603-9615.
10. Appel, K.W., et al., *Performance Summary of the 2006 Community Multiscale Air Quality (CMAQ) Simulation for the AQMEII Project: North American Application*. Air Pollution Modeling and its Application XXI, 2012. **4**: p. 505-511.
11. Tesche, T.W., et al., *CMAQ/CAMx annual 2002 performance evaluation over the eastern US*. Atmospheric Environment, 2006. **40**(26): p. 4906-4919.
12. Eder, B. and S. Yu, *A performance evaluation of the 2004 release of Models-3 CMAQ*. Atmospheric Environment, 2006. **40**(26): p. 4811-4824.
13. Odman, M.T., et al., *Quantifying the sources of ozone, fine particulate matter, and regional haze in the Southeastern United States*. J Environ Manage, 2009. **90**(10): p. 3155-68.

14. Morris, R.E., et al., *Preliminary evaluation of the Community Multiscale Air Quality model for 2002 over the Southeastern United States*. J Air Waste Manag Assoc, 2005. **55**(11): p. 1694-708.
15. Cohan, D.S., Y. Hu, and A.G. Russell, *Dependence of ozone sensitivity analysis on grid resolution*. Atmospheric Environment, 2006. **40**(1): p. 126-135.
16. Arunachalam, S., et al., *A quantitative assessment of the influence of grid resolution on predictions of future-year air quality in North Carolina, USA*. Atmospheric Environment, 2006. **40**(26): p. 5010-5026.
17. Fuentes, M. and A.E. Raftery, *Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models*. Biometrics, 2005. **61**(1): p. 36-45.
18. Berrocal, V., A. Gelfand, and D. Holland, *A Spatio-Temporal Downscaler for Output From Numerical Models*. Journal of Agricultural, Biological, and Environmental Statistics, 2010. **15**(2): p. 176-197.
19. McMillan, N.J., et al., *Combining numerical model output and particulate data using Bayesian space-time modeling*. Environmetrics, 2010. **21**(1): p. 48-65.
20. Nazelle, A.d., S. Arunachalam, and M.L. Serre, *Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina*. Environmental Science & Technology, 2010. **44**(15): p. 5707-5713.
21. Yu, H.L., et al., *BME estimation of residential exposure to ambient PM10 and ozone at multiple time scales*. Environ Health Perspect, 2009. **117**(4): p. 537-44.
22. Houtekamer, P.L., et al., *Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations*. Monthly Weather Review, 2005. **133**(3): p. 604-620.
23. Elbern, H., et al., *4D-variational data assimilation with an adjoint air quality model for emission analysis*. Environmental Modelling & Software, 2000. **15**(6-7): p. 539-548.
24. Adhikary, B., et al., *A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique*. Atmospheric Environment, 2008. **42**(37): p. 8600-8615.
25. Breton, C.V., et al., *Childhood Air Pollutant Exposure and Carotid Artery Intima-Media Thickness in Young Adults*. Circulation, 2012. **126**(13): p. 1614-1620.
26. Christakos, G., *Modern Spatiotemporal Geostatistics*. 2000, Cary, NC, USA: Oxford University Press, USA.
27. Corporation, E.I., *Comprehensive Air Quality Model with Extensions (CAMx) User's Guide Version 5.30*. WWW.camx.com, 2010.
28. EPA, U.S., *Air Quality Modeling Final Rule Technical Support Document*. <http://epa.gov/airtransport/pdfs/AQModeling.pdf>, 2011.

29. Akita, Y., J.-C. Chen, and M.L. Serre, *The moving-window Bayesian maximum entropy framework: estimation of PM_{2.5} yearly average concentration across the contiguous United States*. J Expos Sci Environ Epidemiol, 2012. **22**(5): p. 496-501.
30. Lee, S.-J., et al., *Comparison of Geostatistical Interpolation and Remote Sensing Techniques for Estimating Long-Term Exposure to Ambient PM_{2.5} Concentrations across the Continental United States*. Environmental Health Perspectives, 2012. **120**(12): p. 1727-1732.
31. Hansen, S. *Monitoring and Data Analysis Issue Area Report*. in *Annual Members Meeting*. 2009. Houston, TX.
32. Jerrett, M.P., et al., *Long-Term Ozone Exposure and Mortality*. The New England Journal of Medicine, 2009. **360**(11): p. 1085-95.
33. Neidell, M. and P.L. Kinney, *Estimates of the association between ozone and asthma hospitalizations that account for behavioral responses to air quality information*. Environmental Science & Policy, 2010. **13**(2): p. 97-103.
34. Xu, Y.S., Marc;Reyes,Jeanette;Vizuet,William, *Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions : A National Application*. Environmental Science & Technology, 2016.
35. Bell, M.L. and F. Dominici, *Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities*. Am J Epidemiol, 2008. **167**(8): p. 986-97.
36. Hoek, G., et al., *Estimation of long-term average exposure to outdoor air pollution for a cohort study on mortality*. J Expo Anal Environ Epidemiol, 2001. **11**(6): p. 459-69.
37. Gilliam, R.C., et al., *Impact of inherent meteorology uncertainty on air quality model predictions*. Journal of Geophysical Research: Atmospheres, 2015. **120**(23): p. 12,259-12,280.