Incorporating Space-borne Observations to Improve Biogenic Emission Estimates in Texas

Report Type
Final – AQRP Project 14-017

Prepared For:
Dr. Elena C. McDonald-Buller
Texas Air Quality Research Program
The University of Texas at Austin

Authors
Arastoo Pour Biazar
RichardT. McNider
Andrew White
Earth System Science Center
University of Alabama in Huntsville
Phone: (256) 961-7970
Fax: (256) 961-7755
E-mail: biazar@nsstc.uah.edu

Daniel S. Cohan
Rui Zhang
Dept. of Civil & Environmental Engineering
Rice University
Phone: (713) 348-5129
Fax: (713) 348-5203
cohan@rice.edu

Report Period
7/8/2014 – 9/30/2015

Texas Air Quality Research Program
Task Order: **UTA14-000540-UAH-RP14-017-TO1**

QA Requirements: **Audits of Data Quality - 10% Required**

EXECUTIVE SUMMARY

Problem Statement

The purpose of this project is to advance our understanding of Texas Air Quality (AQ) by utilizing satellite observations and the new advances in biogenic emissions modeling to improve biogenic emission estimates used in Texas State Implementation Plan (SIP) modeling activities.

One of the challenges in understanding Texas air quality has been the uncertainties in estimating the biogenic hydrocarbon emissions. Biogenic volatile organic compounds (BVOCs) play a critical role in atmospheric chemistry, particularly in ozone and particulate matter (PM) formation. In southeast Texas, BVOCs (mostly as isoprene) are the dominant summertime source of reactive hydrocarbon. Despite significant efforts by the State of Texas in improving BVOC estimates, the errors in emission inventories remain a concern. This is partly due to the diversity of the land use/land cover (LU/LC) over southeast Texas coupled with a complex weather pattern, and partly due to the fact that isoprene is highly reactive and relating atmospheric observations of isoprene to the emissions source (vegetation) relies on many meteorological factors that control the emissions, chemistry, and atmospheric transport.

BVOC emissions depend on Photosynthetically Active Radiation (PAR) reaching the canopy and temperature. However, the treatment of temperature and PAR is not uniform across emission models and still poses a problem when evaluating the emission inventories. Recent studies (e.g., Guenther et al., 2012) show that the largest source of uncertainty in BVOC estimates is the model solar radiation estimates and that using satellite-based PAR would be preferable. Emissions from soils also remain as one of the poorly quantified sources of nitrogen oxides (NOx) in most air quality models. Soils can be the largest source of NOx in rural regions where low-NOx conditions make ozone production efficiency especially high, contributing to background ozone levels.

Summary of the Project

This project specifically addressed two priority areas; namely improving biogenic emission estimates and improving the simulation of clouds in air quality models. In particular, a new satellite-based PAR estimate, retrieved from the Geostationary Operational Environmental Satellites (GOES) visible imager, for Texas was produced, evaluated, and used for BVOC estimates. The study episodes included selected periods of summer 2006 and September 2013. The 2013 period coincides with the Deriving Information on Surface Conditions from COlumn and VERtically Resolved Observations Relevant to Air Quality (Discover-AQ) Texas campaign. Also, a new soil NOx scheme, the Berkeley-Dalhousie Soil NOx Parameterization (BDSNP), which provides more mechanistic representation of how emissions respond to nitrogen deposition, fertilizer application, and changing meteorology, was incorporated into the Community

Multiscale Air Quality (CMAQ) model. BDSNP replaced the default Yienger and Levy 1995 (YL95) algorithm and was tested during the same periods. Furthermore, BDSNP was modified to be used as a stand-alone tool for estimating soil nitrogen monoxide (NO) emissions in any air quality modeling practice without conducting the atmospheric chemistry modeling. Stand-alone BDSNP along with a User's Manual will accompany this report.

The new satellite-based PAR product went through several iterations to fine tune the retrieval algorithm. The final product was evaluated against surface pyranometer observations from the Surface Radiation Budget Network (SURFRAD), the Soil Climate Analysis Network (SCAN), as well as Texas local broadband radiation monitoring stations for August 2006 and August-September 2013. The new PAR product was also compared against another satellite-based PAR product generated by the University of Maryland (UMD) (which is now discontinued) for August 2006. SURFRAD is operated by National Oceanic and Atmospheric Administration (NOAA) and is the only available direct continuous measurement of PAR at seven sites nationwide. SCAN is operated by the US Department of Agriculture and has continuous solar radiation measurements at more than 100 stations located in 40 states. In this project, 40 sites from the SCAN network, 7 sites from the SURFRAD network, and 47 sites from local Texas network were chosen to do performance evaluation of the satellite insolation/PAR retrieval products.

The new UAH PAR product was in good agreement with the UMD product. UAH product generally exhibited a small positive bias with respect to surface observations, while the UMD product showed a negative bias. However, both satellite retrieval products substantially outperformed Weather Research and Forecasting (WRF) model simulations. Compared to surface observations, correlation coefficients for satellite products were R=0.96~0.97 versus R=0.93 for the model. Satellite products also had smaller normalized mean error (NME) of 20.7%~20.1% compared to NME=35.5% for WRF. The UMD retrieval underestimated PAR with a normalized mean bias (NMB) of -12.4% while the University of Alabama in Huntsville (UAH) retrieval overestimated PAR with NMB of 10.2%.

Cloud assimilation in the Weather Research and Forecasting (WRF) model significantly reduced the normalized mean bias with respect to surface observations. NMB was reduced from 22.2% for control simulation to 8.9% for WRF with cloud assimilation. Additional evaluation of the results over 47 Texas Commission on Environmental Quality (TCEQ) sites (Broadband Radiation Monitoring networks) also indicated that cloud assimilation was able to significantly reduce the over-prediction by WRF. This result shows that WRF with cloud assimilation significantly improved the location and timing of clouds over Texas.

By using the new satellite-based PAR in the Model of Emissions of Gases and Aerosols from Nature (MEGAN), emission estimates indicated that the highest emission regions for isoprene in Texas are East Texas (2754 tons/day), North Central Texas (2036 tons/day), and Edwards Plateau (1199 tons/day). For terpenes, the highest emission regions are East Texas (1011 tons/day), Trans-Pecos (615 tons/day), and North Central Texas (562 tons/day). The results indicate that using GOES satellite retrievals on average reduced isoprene emission estimates by 20% and terpene emission estimates by 5% during August-September 2013 compared to the control case. In some regions these differences were as high as 29%. The emission algorithm for estimating terpenes in MEGAN is more impacted by the surface temperature than by PAR. The lowest emission estimates for terpenes over Texas were observed on August 15 and September 20 under overcast conditions.

Overall, the estimated isoprene emissions (for control case) by MEGAN resulted in simulated concentrations being 2-3 times higher than the observed values. Using satellite-based PAR reduced the emissions by about 30%. However, this reduction was not enough to correct the large model bias. The model was also unable to explain the observed diurnal variation of isoprene. The model results for isoprene in this project are consistent with findings from other AQRP investigations (e.g., AECOM project testing the sensitivity of different mechanisms to changes in BVOC emission estimates, or the University of Texas at Austin (UT Austin) project testing the impact of different land use/land cover on BVOC emission estimates). This suggests that the 2011 global BVOC emission factors in the current MEGAN release (v2.10), at least for isoprene, may have high uncertainty over Texas. Alex Guenther's group has been contacted with respect to this issue.

With respect to the other component of this project addressing soil NOx emission estimates, soil emission rates estimated by BDSNP module were consistently higher than the estimates by YL95 algorithm. The spatial patterns for the two algorithms were also quite different. YL95 estimated high NO emissions around Houston, while for BDSNP the highest emissions occurred near the state boundary between Texas, Louisiana and Arkansas. This contrast may be due to the combined contributions from different soil biome types, fertilizer implementations, and the different response curve for soil temperature and moisture in the two soil NO schemes. BDSNP estimated 674 moles/s over the East Texas region. This estimate is about 2.3 times higher than the corresponding YL95. But for South Texas, YL95 estimates were about 3 times higher than BDSNP (34.8 moles/s for YL95 versus 12.9 moles/s for BDSNP). Overall, the estimates by BDSNP were about 21% higher than YL95 over Texas. The stand-alone BDSNP module along with the user's manual accompanies this final report.

A series of CMAQ simulations was also performed to evaluate the impact of emission and meteorological changes on air quality predictions. The base case used the MEGAN outputs derived by default WRF simulation, the case marked "analytical" in this

report used the MEGAN outputs derived by WRF simulation with cloud assimilations, and the case named “UAHPAR” used the MEGAN outputs derived by satellite PAR retrievals and the soil NO emission from BDSNP scheme. The preliminary results from these simulations do not show a significant difference in ozone predictions. Compared to observations from 38 evaluation sites over Texas the correlation coefficients were around 0.75-0.76. The base case overestimated ozone concentration by 2-3%, which is consistent with previous modeling studies (Song et al., 2008; Kota et al., 2015). By using the cloud assimilation in WRF, the ozone performance was marginally better, decreasing the mean bias from 2.8% to 2.4%, the root mean square error (RMSE) from 14.5 ppbV to 14.2 ppbV, and the NMB from 18.2% to 17.0%. Model performed better over the Dallas-Fort Worth metroplex (DFW) region with correlation coefficients of 0.75-0.77 and NME of 28.9%-30.6%. On average, the model underestimated mean ozone over Austin, DFW, East Texas and El Paso regions while overestimating mean ozone over Corpus Christi, HGB and San Antonio regions. The ozone predictions over Corpus Christi exhibited the highest NMB in all three simulations (105-111%).

Due to the large bias in biogenic VOC emission estimates by MEGAN, even after a 30% reduction in BVOC emissions (resulted from using satellite data) most of the modeling domain over east Texas was saturated with VOC and remained sensitive to NOx. Therefore, the CMAQ simulations used in this study did not show a substantial improvement in ozone prediction over east Texas. These simulations should be repeated using a biogenic emissions model that better represent clear sky emission estimates.

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY II

PROBLEM STATEMENT II
SUMMARY OF THE PROJECT II

TABLE OF CONTENTS VI

LIST OF FIGURES IX

LIST OF TABLES XIII

LIST OF ACRONYMS XIV

1. INTRODUCTION 1

1.1. DEFINITION OF PROBLEM 2
1.2. CURRENT STUDY 3

2. PHOTOSYNTHETICALLY ACTIVE RADIATION (PAR) 4

2.1. APPROACH 4
2.2. BIAS CORRECTION FOR INSOLATION PRODUCT 7

3. EVALUATION OF PAR PRODUCTS 12

3.1. OBSERVATIONS USED FOR PAR EVALUATION 12
3.2. EVALUATION OF PAR PRODUCT BEFORE BIAS CORRECTION 12
3.3. EVALUATION OF PAR PRODUCT AFTER BIAS CORRECTION 14

4. WRF SIMULATIONS WITH CLOUD ASSIMILATION (AUG-SEP 2013) 19

4.1. WRF MODEL PERFORMANCE EVALUATION 22
4.1.1. Performance Statistics for Domain 1 22
4.1.2. Performance Statistics for Domain 2 26
4.1.3. Performance Statistics for Domain 3 29

5. BVOC AND SOIL NOX EMISSION ESTIMATES IN MEGAN 32

5.1. INITIAL BVOC AND SOIL NOX EMISSION ESTIMATES IN MEGAN 33
5.2. MEGAN SIMULATIONS WITH SATELLITE-BASED PAR DURING AUGUST-SEPTEMBER 2013 36
5.3. DEVELOPING OFFLINE BDSNP MODULE FOR SOIL NOX EMISSION ESTIMATES 45
5.3.1. Developing a New Soil Biome Spatial Map Based on 12km CONUS 40-category 2006 NLCD-MODIS Land Use Classification (NLCD40) 49
5.3.2. Test Case for soil NO emissions using stand-alone BDSNP scheme 54
5.4. SOIL NO EMISSION ESTIMATES FOR 2013 USING BDSNP SCHEME 55

6. CMAQ SIMULATIONS USING DIFFERENT MEGAN OUTPUTS WITH SATELLITE PAR 59

7. CONCLUSIONS AND RECOMMENDATIONS 68

8. AUDITS OF DATA QUALITY 70

9. PAPERS AND PRESENTATIONS 71

REFERENCES 72

APPENDIX A: DEFINITIONS OF EVALUATION METRICS USED IN THIS STUDY 75

APPENDIX B: STASTICS FOR INSOLATION SIMULATION/RETREIVAL FOR DIFFERENT CASES AT 47 TCEQ NETWORK SITES 78

LIST OF FIGURES

Page

Figure 1. Variation of PAR conversion factor with respect to solar zenith angle, optical depth, and water vapor (adapted from Frouin and Pinker, 1995). 5

Figure 2. Conversion factor as a function of optical depth for two different zenith angles as presented in Figure 1. The insert is from Frouin and Pinker, 1995. 7

Figure 3. Scatter plots and the corresponding best linear fit showing GOES insolation retrievals versus pyranometer observations from Soil Analysis Climate Network (SCAN) for the month of September 2013. 8

Figure 4. Scatterplots showing GOES insolation retrievals against Soil Analysis Climate Network (SCAN) data for September 2013. The figure to the left shows the data before applying bias correction to GOES retrievals, and the figure to the right shows the scatter plot after applying bias correction. 10

Figure 5. A snapshot of satellite-derived insolation and PAR at 19:45 GMT, September 1, 2013. PAR is produced in units of W/m^2 as well as micro-mol/m^2/s. 11

Figure 6. Location of SCAN and SURFRAD sites used for PAR/insolation evaluation 12

Figure 7. Spatial distribution of normalized mean bias at each SCAN site during September 2013 (UAH product before bias correction). 14

Figure 8. Three domains for WRF-MCIP simulation in this study and locations of the insolation/PAR evaluation sites at SURFRAD network (red) and TCEQ broadband radiation network (blue) 16

Figure 9. Scatter plots showing four different hourly simulated/retrieved PAR (vertical axis) and observed PAR (horizontal axis) during August 2006 at 7 SURFAD sites 16

Figure 10. Performance of WRF incoming solar radiation (RGRND) from simulation case 'cntrl' (left) and 'analytical' (right) as well as UAH insolation retrievals (right) at TCEQ sites. The upper panel shows the correlation coefficient (R) and the lower panel shows the normalized mean bias (NMB). 19

Figure 11. Daily agreement index for CNTRL and ASSIM 36 km WRF simulations over August-September 2013 using a 10% cloud albedo threshold. 24

Figure 12. Average hourly AI for August-September 2013. Assimilation performs better in the latter part of the day. 24

Figure 13. Comparing model cloud to satellite observation for CNTRL (left) and ASSIM (right) simulations for August 21, 2013 at 17 UTC: Green indicates the model and GOES were clear, Red indicates model cloud over-predictions, Orange indicates model cloud under-predictions, and Grey indicates locations where

the model and GOES are cloudy. Agreement Index for a) CNTRL (AI=59.9%) b) ASSIM (AI=73.4%). 25

Figure 14. Wind speed and temperature bias results for CNTRL and ASSIM simulations. 25

Figure 15. Daily agreement index for CNTRL and ASSIM 12 km WRF simulations over August-September 2013 using a 10% cloud albedo threshold. 27

Figure 16. Average hourly AI for 12-km domain during August-September 2013..... 28

Figure 17. Comparing model cloud to satellite observation for CNTRL (left) and ASSIM (right) simulations for August 27, 2013 at 22 UTC: Green: model and GOES were clear, Red: model cloud over-predictions, Orange: model cloud under-predictions, and Grey: locations where the model and GOES are cloudy. Agreement Index for a) CNTRL (AI=59.0%) b) ASSIM (AI=73.6%). 28

Figure 18. Temperature and mixing ratio bias results for 12-km CNTRL and ASSIM simulations..... 29

Figure 19. Daily agreement index for CNTRL and ASSIM 4 km WRF simulations over August-September 2013 using a 10% cloud albedo threshold. 30

Figure 20. Average hourly AI for 4-km domain during August-September 2013..... 30

Figure 21. Comparing model cloud to satellite observation for CNTRL (left) and ASSIM (right) simulations for September 16, 2013 at 19 UTC: Green: model and GOES were clear, Red: model cloud over-predictions, Orange: model cloud under-predictions, and Grey: locations where the model and GOES are cloudy. Agreement Index for a) CNTRL (AI=44.5%) b) ASSIM (AI=56.1%). 31

Figure 22. Wind speed and temperature bias results for CNTRL and ASSIM for 4-km domain. 32

Figure 23. Spatial distribution of estimated ISOP and TERP emission rate by MEGAN using different PAR inputs data (WRF versus UAH satellite retrievals)..... 34

Figure 24. Total ISOP (left) and TERP (right) emission rate estimates by MEGAN using different PAR inputs data (WRF versus UAH satellite retrievals) for each climate region in United States. 35

Figure 25. Demonstration of spatial different of soil NO emission rate estimates using BDSNP or YL95 scheme..... 35

Figure 26. Comparison of the spatial patterns of the monthly mean isoprene (ISOP) emission rate using different PAR inputs for WRF control case (cntrl), WRF cloud assimilation case (analytical) and PAR satellite retrievals (PAR) in MEGAN over Texas domain during August (left) and September (right) 2013..... 37

Figure 27. Comparison of the spatial patterns of the monthly mean monoterpene (TERP) emission rate using different PAR inputs for WRF control case (cntrl), WRF cloud

assimilation case (analytical) and PAR satellite retrievals (PAR) in MEGAN over Texas domain during August (left) and September (right) 2013. 38

Figure 28. Geographic distribution of the 10 climate divisions in Texas by the National Weather Service (adapted from http://www.nass.usda.gov/Statistics_by_State/Texas/Charts_&_Maps/cwmap.htm) 39

Figure 29. Comparsion of average isoprene (ISOP, top) and monoterpene (TERP, bottom) emission rate (tons/day) over the 10 climate divisions of Texas during August and September 2013 by MEGAN using different PAR inputs. 40

Figure 30. Comparison of the daily variation of isoprene (ISOP, top) and monoterpene (TERP, bottom) emission rate (tons/day) over East Texas during August and September 2013 by MEGAN using different PAR inputs. 42

Figure 31. US surface weather map at 7 a.m. E.S.T on September 20, 2013 from NOAA (left); and the daily mean PAR retrieval from UAH on September 20, 2013 (right). 43

Figure 32. Time series of average daily mean surface temperature (degrees Celsius, top), PAR (W/m^2, middle) and 24hr accumulated precipitation (cm, bottom) over East Texas during August and September 2013 by MEGAN using PAR satellite retrievals from UAH. 44

Figure 33. Domain-wise sum of estimated isoprene (ISOP) and monoterpene (TERP) emission strength over Texas area using different PAR inputs in MEGAN during September 2013. 44

Figure 34. BDSNP soil NOx emission scheme given by Hunman et al. (2012) 45

Figure 35. Flow chart of the inline/off line BDSNP soil NO_x emission model. 46

Figure 36. Soil N reservoir from deposition (NDEPRES) used in off-line BDSNP model which is calculated from 2005 CMAQ simulation results. 47

Figure 37. Impact of different BDSNP processes on the Soil NO_x emission estimates. 48

Figure 38. Comparison of soil NOx emission estimates using the inline (left) and offline (right) BDSNP scheme on Jun 28, 2011. 48

Figure 39. Spatial map of five climate zones over CONUS 12km domain (bottom) based on Köppen-Geiger climate classification (top). 52

Figure 40. Comparison of the spatial pattern of 2011 National Land Cover Database (30m resolution, top) with the soil biome type developed either from GEOS-Chem (0.25 degree resolution, middle) or from MODIS NLCD40 classification in CMAQ (12km resolution, bottom). 53

Figure 41. Spatial pattern difference of soil NO base emission simulated from BDSNP module using the global GEOS-Chem soil biome (control), updated regional soil

biome based on NCLD40 (new Biome), and North American specified emission factors (NA EF) over the continental US. 54

Figure 42. The difference in spatial pattern between daily mean soil NO emission rate (g/s) from MEGAN default YL95 scheme (left) and BDSNP scheme (right) on August 1, 2013 over the Texas domain. 55

Figure 43. Comparison of the spatial patterns of the monthly mean soil NO emission rate using different meteorology inputs, WRF control (cntrl) and WRF cloud assimilation (analytical) and different soil NO emission algorithms, YL95 and BDSNP, over Texas domain during August 2013. 56

Figure 44. Comparison of the spatial patterns of the monthly mean soil NO emission rate using different meteorology inputs, WRF control (cntrl) and WRF cloud assimilation (analytical), and different soil NO emission algorithms, YL95 and BDSNP, over Texas domain during September 2013. 57

Figure 45. Comparsion of average soil NO emission rate (moles/s) over the 10 climate divisions of Texas during August and September 2013 by different algorithms and meteorological fields. 58

Figure 46. Flow chart to prepare daily emission input files for CMAQ run. 60

Figure 47. Spatial pattern of daily ground gridded SO2, NOx, CO, PM2.5 and VOC emission prepared for CMAQ runs over Texas 12km domain on Sep 1, 2013 (tons/day). (The daily ground VOC emission pattern is given in three versions with different PAR inputs in MEGAN model). 61

Figure 48. Average spatial pattern of simulated maximum daily 8-hr average ozone concentrations (MDA8 O3) for three cases during September 1-15, 2013. 63

Figure 49. Average spatial pattern of simulated daily mean isoprene concentrations for three cases during September 1-15, 2013. 63

Figure 50. The locations of selected TCEQ CAMS sites for ozone (left, black markers) and isoprene (right, red markers) for CMAQ performance evaluation 64

Figure 51. Performance of CMAQ hourly ozone from simulation case ‘cntrl’ (left) and ‘analytical’ (middle) as well as UAH insolation retrievals (right) at TCEQ sites. The upper panel shows the correlation coefficient (R) and the lower panel shows the normalize mean bias (NMB). 66

Figure 52. Time series comparison of ISOP concentration simulation for different CMAQ cases at site CAMS 26 (top), CAMS 1007 (middle) and CAMS 69 (bottom). 67

LIST OF TABLES

Page

Table 1. Case arrangements for PAR and the observational data sources used in evaluation. 15

Table 2. Summary of statistics for PAR simulation/retrievals for different cases at 7 SURFRAD network sites[1] 17

Table 3. Summary of statistics of insolation simulation/retrievals for different cases at 47 TCEQ network sites. 18

Table 4. WRF domain setup for 36-, 12-, and 4-km grid spacing used for UAH simulations. 20

Table 5. WRF configuration for 2013 satellite cloud assimilation runs. 21

Table 6. Contingency table used for evaluating model cloud simulation. 22

Table 7. Configuration of WRF-MEGAN used in the initial study. 33

Table 8. MEGAN configuration used in the 2013 study. 36

Table 9. Comparison of daily average isoprene (ISOP) and monoterpene (TERP) emission rate (tons/day) over 10 climate zones of Texas from MEGAN using different PAR inputs. 41

Table 10. Comparison of CPU time usage for inline and offline BDSNP. 49

Table 11. Mapping used to create the soil biome map based on NLCD40 MODIS land use/land cover categories. 49

Table 12. Different input combinations used in CMAQ simulations. 59

Table 13. Configuration of CMAQ simulation in this study. 62

Table 14. Statistics for the CMAQ hourly ozone simulation performance for the three cases over 38 TCEQ CAMS sites (grouped by 7 regions). 64

Table 15. Statistics for model isoprene predictions for three cases over 18 TCEQ CAMS sites. 66

LIST OF ACRONYMS

ACM2	Asymmetrical Convective Model version 2
ADP	Automated Data Processing
AECOM	Architecture, Engineering, Consulting, Operations, and Maintenance
AI	Agreement Index
AMD	Atmospheric Modeling Division
AQ	Air Quality
AQRP	Air Quality Research Program
ARW	Advanced Research WRF
ASSIM	WRF simulations with satellite cloud assimilation performed in this study
BDSNP	UC-Berkeley and Dalhousie University Soil NOx Parameterization
BEIS	Biogenic Emission Inventory System
BVOC	Biogenic Volatile Organic Compound
CAA	Clean Air Act
CAMx	Comprehensive Air Quality Model with Extensions
CB05cl	Carbon Bond-05 mechanism with chlorine
CEM	Continuous Emissions Monitoring
CF	Correction Factor
CLM	Community Land Model
CMAQ	Community Multiscale Air Quality
CMAS	Community Modeling and Analysis System
CNTRL	Control: base WRF simulations performed in this study
CONUS	Continental United States
DFW	Dallas-Fort Worth
Discover-AQ	Deriving Information on Surface Conditions from COlumn and VERtically Resolved Observations Relevant to Air Quality
DSS	Decision Support System
EDAS	Eta-based 4-D Data Assimilation System
EGU	Electric Generating Units
EOS	Earth Observing System
EPA	(US) Environmental Protection Agency
FDDA	Four-Dimensional Data Assimilation
FINN	Fire Inventory from NCAR
GCM	Global Climate Model
GHRC	Global Hydrology Resource Center
GLOBEIS	Global Biosphere Emissions and Interactions System
GOES	Geostationary Operational Environmental Satellites
GPGS	GOES Product Generation System
HGB	Houston-Galveston-Brazoria
HPMS	Highway Performance Monitoring System
IA	Index of Agreement

IR group	Infrared Measurements and Water Vapor Studies Group
LAI	Leaf Area Index
LU/LC	Land Use/Land Cover
MAGE	Mean Aggregate Gross Error
MB	Mean Bias
ME	Mean Errors
MEGAN	Model of Emissions of Gases and Aerosols from Nature
MM5	The Fifth-Generation NCAR / Penn State Mesoscale Model
MODIS	Moderate Resolution Imaging Spectroradiometer
MOZART	Model for Ozone and Related Chemical Tracers
MSFC	Marshall Space Flight Center
NAAQS	National Ambient Air Quality Standards
NAM	North American Mesoscale
NARR	North American Regional Reanalysis
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NERL	National Environmental Research Laboratory
NLCD	National Land Cover Data
NLCD40	CONUS 40-category 2006 NLCD-MODIS land use classification
NMB	Normalized Mean Bias
NME	Normalized Mean Error
NOAA	National Oceanic and Atmospheric Administration
NOx	Nitrogen Oxides
NSSTC	National Space Science and Technology Center
NWS	National Weather Service
OBS	Observation Means
OD	Optical Depth
PAR	Photosynthetically Active Radiation
PFT	Plant Functional Type
PM	Particulate Matter
$PM_{2.5}$	Particulate Matter with diameter less than 2.5 microns
R	Correlation Coefficient
RADM	Regional Acid Deposition Model
RMSE	Root Mean Square Error
RRTMG	Rapid Radiative Transfer Model for GCMs
SCAN	Soil Climate Analysis Network
SIM	Simulation Mean
SIP	State Implementation Plan
SPoRT	Short-term Prediction Research and Transition Center
SURFRAD	Surface Radiation Budget Network
TCEQ	Texas Commission on Environmental Quality
TexAQS-06	Texas Air Quality Study 2006

UAH	University of Alabama in Huntsville
UC	University of California
UMD	University of Maryland
USDA	United States Department of Agriculture
USRA	Universities Space Research Association
UT	University of Texas
WRF	Weather Research and Forecasting
YL95	Yienger and Levy 1995 soil NOx algorithm
YSU	Yonsei University

1. INTRODUCTION

This project was in response to the call by the State of Texas Air Quality Research Program (AQRP) seeking studies to support Texas Air Quality by utilizing the data from the recent DISCOVER-AQ field campaign. This work specifically addressed two priority areas; namely improving biogenic emission estimates and improving the simulation of clouds in air quality models. The project also contributes to several other priority areas as the improvements in radiation field not only impacts the biogenic emissions but also improves the overall photochemical simulation and leads to a better understanding of ozone and PM formation. The project devised a new algorithm to derive photosynthetically active radiation (PAR) from geostationary satellite observations to be used in biogenic emission estimates. The project also employed a new soil NO emission scheme in an air quality model.

As stated in the AQRP 2012 State of the Science report, biogenic volatile organic compounds (BVOCs) play a critical role in formation of ozone and secondary organic aerosols in east Texas. Previous studies have shown that isoprene is the dominant BVOC in southeast Texas. However, due to high reactivity of isoprene, indirect evaluations of emission estimates through comparing the simulated isoprene concentration with measurements have proven to be a challenging task. In this project, satellite observations that directly impact photochemical activity as well as BVOC emissions were assimilated and thereby created a simulated atmosphere that is more compatible with the measurements during DISCOVER-AQ.

The University of Alabama in Huntsville (UAH) currently generates a set of products from the Geostationary Operational Environmental Satellite (GOES) that includes surface incident short-wave radiation as well as cloud albedo and cloud top temperature. Under this activity, UAH also started producing satellite-based PAR retrievals that are needed in the estimation of biogenic hydrocarbon emissions. Satellite-derived PAR was evaluated against previous satellite-based products as well as surface observations during Texas Discover-AQ campaign. Furthermore, the new PAR retrievals were used in the Model of Emissions of Gases and Aerosols from Nature (MEGAN) to generate BVOC emissions. Also under this activity, Berkeley-Dalhousie Soil NOx Parameterization (BDSNP) was implemented into the Community Multiscale Air Quality (CMAQ) model to replace the default Yienger and Levy 1995 (YL95) algorithm. BDSNP provides a more mechanistic representation of how emissions respond to nitrogen deposition, fertilizer application, and changing meteorology. A series of sensitivity simulations were performed and evaluated against Discover-AQ observations to test the impact of satellite-derived PAR and the new soil NOx emission model on air quality simulations.

The Weather Research and Forecasting (WRF) model and CMAQ modeling system were used for air quality simulations. WRF simulations took advantage of improved cloud simulation by applying a technique developed at UAH under a previous TCEQ funded project (TCEQ Grant No. 582-12-10111). The technique uses GOES cloud observations to dynamically correct cloud fields in WRF.

1.1. Definition of Problem

One of the challenges in understanding Texas air quality has been the uncertainties in estimating the biogenic hydrocarbon emissions (Allen et al., AQRP State of the Science 2012 report). BVOCs play a critical role in atmospheric chemistry, particularly in ozone and particulate matter (PM) formation. In southeast Texas, BVOCs (mostly as isoprene) are the dominant summertime source of reactive hydrocarbon (Wiedinmyer et al., 2001). Despite significant efforts by the State of Texas in improving BVOC estimates, the errors in emissions inventories remain a concern. This is partly due to the diversity of the land use/land cover (LU/LC) over southeast Texas coupled with a complex weather pattern (Song et al., 2008), and partly due to the fact that isoprene is highly reactive and relating atmospheric observations of isoprene to the emissions source (vegetation) relies on many meteorological factors that control the emission, chemistry, and atmospheric transport.

BVOC estimates depend on LU/LC, the amount of Photosynthetically Active Radiation (PAR), and temperature. There have been many efforts in developing high resolution LU/LC data sets to better represent the diversity of vegetation over the State of Texas (Wiedinmyer et al., 2001; Byun et al., 2005). However, the treatment of temperature and PAR is not uniform across emissions models and still presents a problem when evaluating the inventories. Guenther et al., 2012, argued that the largest source of uncertainty in BVOC estimates is the model solar radiation estimates and that using satellite-based PAR would be preferable.

Warneke et al., 2010, compared several BVOC emission models and showed that they agree within a factor of two. This was partly due to the differences in estimating the impact of light and temperature on emissions. Among the models used in their study, MEGAN (Guenther et al., 2006, 2012) produced higher estimates compared to measurements. Indirect evaluations of MEGAN by using satellite observation of formaldehyde also indicated that MEGAN over-estimates isoprene emissions (Palmer et al., 2006; Miller et al., 2008). But contrary to the above findings, a model study by Muller et al., 2008, showed that MEGAN under-estimated isoprene flux over the Harvard forest site. Karl et al., 2007, also found MEGAN under-predicted isoprene fluxes when compared to the flux estimates derived from aircraft measurements.

This goes to show the difficulty of evaluating the estimated inventory. This difficulty is mostly due to the high reactivity of isoprene and the need to have a reasonable representation of the physical atmosphere when comparing modeled concentrations of isoprene (or related compounds) to the observations. The emissions model estimate of isoprene is highly dependent on providing the correct PAR and temperature. But, relating the observed isoprene concentration (or derived flux) to the emissions inventory also depends on the atmospheric conditions that are regulated by radiation and temperature.

Song et al., 2008, demonstrated this difficulty when they compare modeled and observed isoprene concentrations in southeast Texas. They noted that the vertical gradient of isoprene is highly influenced by the effectiveness of boundary layer mixing. Due to high reactivity of

isoprene, less efficient mixing in the model allows for higher surface concentration of isoprene and lower concentrations aloft. Given that the surface heating is a key factor in creating efficient mixing and increased isoprene emissions, it is crucial to correctly specify the surface incident radiation in such studies. In fact, their results signified the importance of meteorology when evaluating BVOC emissions.

Therefore, to be able to use the measurements to improve the emissions inventory, it is imperative to recreate the best model representation of the atmospheric condition during the observations. The work presented here is an attempt to achieve this objective by using geostationary satellite observations to retrieve PAR for direct use in the biogenic emissions model and by improving model simulated clouds, surface incident radiation, and temperature fields used in BVOC estimates.

Emissions from soils also remain one of the most poorly quantified sources of NOx in most air quality models. Soils can be the largest source of NOx in rural regions where low-NOx conditions make ozone production efficiency especially high, contributing to background ozone levels. A new soil NOx scheme has been developed by University of California at Berkeley (UC-Berkeley) and Dalhousie University (Hudman et al., 2012), which provides a more mechanistic representation of how emissions respond to nitrogen deposition, fertilizer application, and changing meteorology. Previous studies (Hudman et al., 2010; Hudman et al., 2012) have shown that the new scheme more than doubles soil NOx emission estimates in many regions and greatly increases their episodic and interannual variability.

1.2. Current Study

In the current study, an attempt was made to use GOES Imager visible data to produce PAR and to quantify the impact of observed PAR on biogenic emission estimates and thereby on ozone predictions. After several iterations, a parametrization was devised to produce PAR from the GOES Imager visible band. The PAR products were evaluated against surface observations over Texas and also against satellite-based PAR previously generated by the University of Maryland (UMD) for 2006. PAR products were also compared with model estimates to quantify model errors. In this study, WRF was used for meteorological modeling, MEGAN for biogenic emission estimates, and CMAQ for air quality simulations.

For biogenic emission estimates, MEGAN was used to generate emissions for different input parameters. Also, the UC-Berkeley and Dalhousie University Soil NOx Parameterization (BDSNP) was adapted for soil NO estimates. WRF-CMAQ simulations for August-September 2013 (Discover-AQ period) were performed to test model sensitivity to these inputs.

In the following chapters detailed description of these efforts will be presented.

2. PHOTOSYNTHETICALLY ACTIVE RADIATION (PAR)

The uncertainties in estimating biogenic hydrocarbon emissions have been a major concern in regulatory air quality applications (Allen et al., 2012). Models have considerable difficulties in creating clouds in the right place and at the right time compared to observed clouds. This in turn significantly impacts model PAR estimates and thereby the emission estimates of BVOCs. Therefore, a meaningful evaluation of the biogenic emissions model relies on incorporating the observed radiation field in PAR estimates. Thus, to be able to use the measurements to improve the emissions inventory, geostationary satellite observations were used to retrieve PAR for direct use in the biogenic emissions model. The following describes the technique used for retrieving PAR from satellite observation, as well as the evaluation of the data product.

2.1. APPROACH

Currently, University of Alabama in Huntsville collaborates with the Infrared Measurements and Water Vapor Studies Group (IR group) at the National Aeronautics and Space Administration (NASA)/ Marshall Space Flight Center (MSFC) to generate and archive several GOES derived products including cloud albedo, surface albedo and surface insolation for the use in meteorological and air quality models (Haines et al., 2003). Over the years, these products have been evaluated and used in many air quality studies (Pour-Biazar et al., 2007; Mackaro et al., 2011; McNider et al., 1998; Haines et al., 2003). The algorithm used for the retrieval of albedo and surface insolation is the implementation of the method developed by Gautier et al. (1980) and complemented by new improvements in subsequent studies (Diak and Gautier, 1983; personal communications with George Diak).

PAR is defined as:

$$PAR = \int_{.4}^{.7} I(\lambda) d\lambda \quad (W\,m^{-2}) = \frac{1}{hc} \int_{.4}^{.7} I(\lambda) d\lambda \quad (quanta\,m^{-2} s^{-1}) \qquad (1)$$

Where, λ is the wavelength, h is the Planck constant, and c is the speed of light.

Thus, in principal insolation can be scaled to produce PAR. Meaning that a conversion factor (*CF*) can be defined to convert insolation to PAR:

$$CF = \frac{PAR}{Insolation} \qquad (2)$$

Frouin and Pinker, 1995 and Pinker and Laszlo, 1992, documented the dependency of the conversion factor on several relevant atmospheric parameters such as water vapor, total overhead ozone, optical depth (representing aerosol/cloud impact), and zenith angle. The largest variations are caused by water vapor, optical depth, and solar zenith angle (Figure 1).

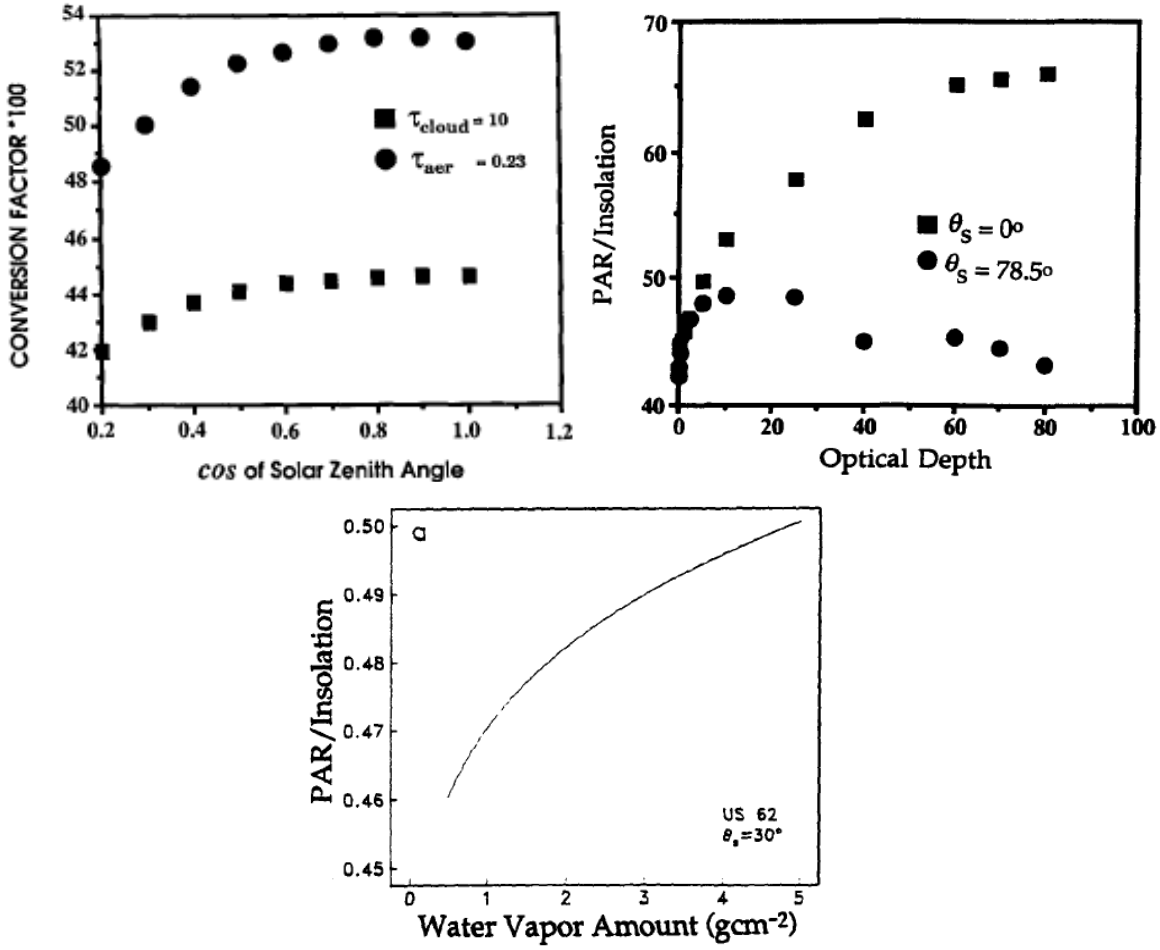


Figure 1. **Variation of PAR conversion factor with respect to solar zenith angle, optical depth, and water vapor (adapted from Frouin and Pinker, 1995).**

This variation is mostly due to the difference in the impact of direct and diffused light. Meaning that in the presence of water vapor and aerosols, a modest increase in diffused light increases the conversion factor when the sun is overhead. However, one must note that the largest increase in CF is when the insolation is drastically reduced (for optical depths greater than 10-15). This means that in the presence of opaque clouds, the sizeable reduction in surface incident radiation will offset such marginal increases in CF. Therefore, the practical variation of conversion factor hovers around 0.5. In fact many of the models used in agricultural applications use a CF of 0.5. In its default configuration, MEGAN also uses CF=0.5 when model estimates of solar radiation are used.

The approach here is to construct a simple parameterization for calculating a variable conversion factor from the current insolation product at UAH that encapsulates the impact of environmental variables on PAR. In this approach, CF is defined as a function of optical depth

and zenith angle based on the data offered in Pinker and Laszlo (1992) and Frouin and Pinker (1995) and as described above. After several iterations, a suitable functional form was devised for this purpose. The following relationship takes into account the impact of optical depth and zenith angle on conversion factor:

$$CF = \frac{PAR}{Insolation} = .42 + .28 * Cfactor * Zfactor$$
$$Where \quad Cfactor = (1 - 1.8^{-\sqrt{.1\tau}}) \qquad (3)$$
$$and \qquad Zfactor = \left(1.1 - .1\sqrt{\cos(Z)}\right)^{-\tau}$$

In Equation 3, $\tau = Optical\ Depth\ (OD)$, *Cfactor* represents the impact of cloud attenuation, and *Zfactor* represents the impact of zenith angle on *CF*. This relationship yields the rapid growth of the conversion factor for small optical depths (τ<5) due to the impact of diffused sunlight (to reflect the impact of aerosol loading and transparent clouds) and approaches a steady rate for thick clouds. The conversion factor gradually increases as the cloud opaqueness increases.

Optical depth then is estimated from satellite-derived cloud albedo. Based on Stephens 1978 and Joseph 1976 (implemented in Regional Acid Deposition Model (RADM), Comprehensive Air Quality Model with Extensions (CAMx) and CMAQ) we arrive at the following parameterization:

$$\tau = \frac{8\alpha_c}{(1-\alpha_c)^2}, \quad where \qquad \alpha_c = cloud\ albedo \qquad (4)$$

Figure 2 demonstrates how the functional form of the conversion factor performs as a function of optical depth for two different zenith angles as presented in the study of Frouin and Pinker, 1995 (see Figure 1 in their paper). The functional form of conversion factor seems to be fitting the data well and is used to convert insolation to PAR.

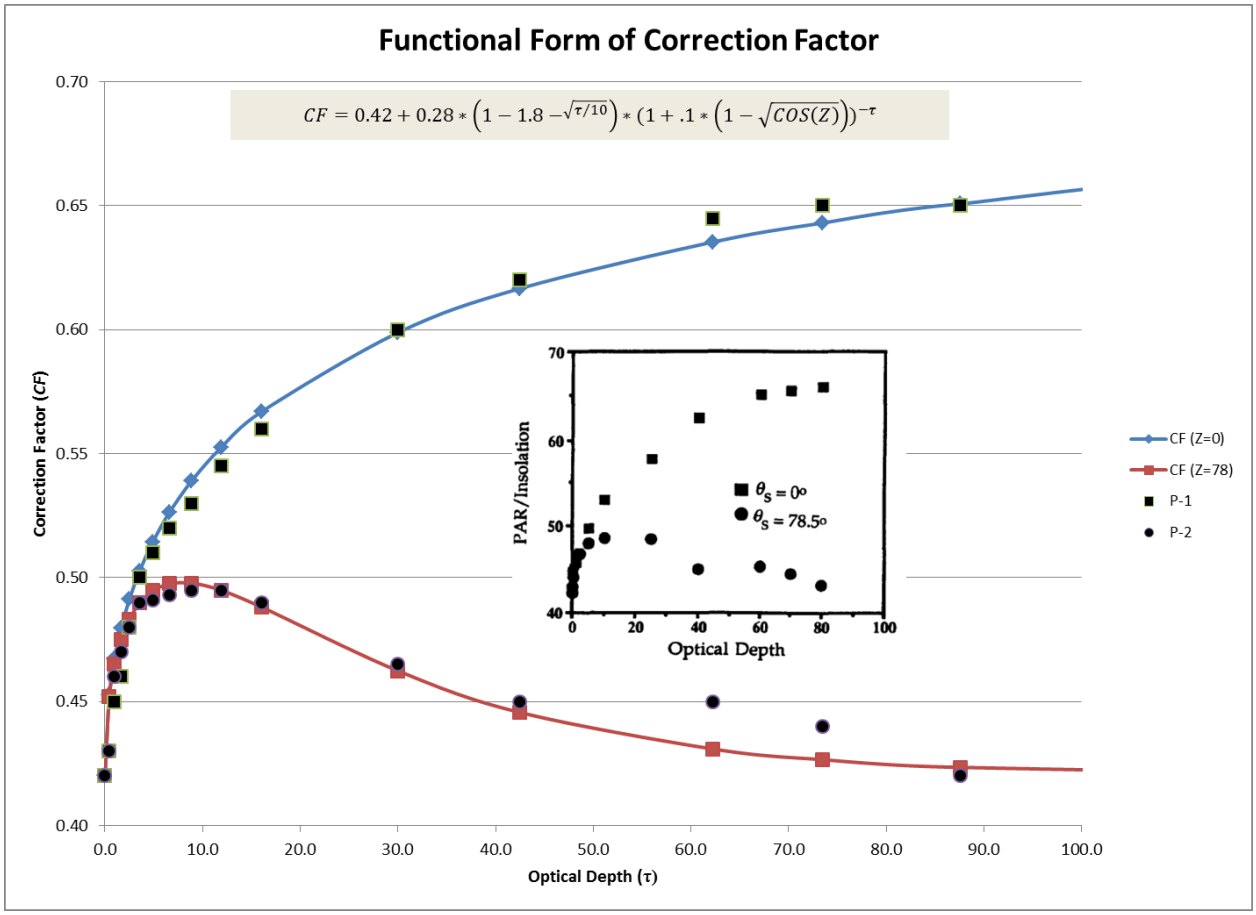


Figure 2. **Conversion factor as a function of optical depth for two different zenith angles as presented in Figure 1. The insert is from Frouin and Pinker, 1995.**

2.2. Bias Correction for Insolation Product

Since the PAR product is based on satellite insolation retrievals, any error in the insolation retrieval carries over to the PAR product. The current insolation retrieval at UAH uses a constant correction factor when accounting for the impact of precipitable water on insolation. Due to the large spatial variation of precipitable water over the continental United States (CONUS) from east to west, using a constant correction factor could be under-estimating insolation in the west and over-estimating insolation in the eastern U.S. Initial evaluation of PAR product also revealed a systematic bias for clear-sky satellite-based insolation against surface pyranometer observations.

Figure 3 shows the progression of insolation over-estimations from west to east. The figure demonstrates paired pyranometer observations from Soil Climate Analysis Network (SCAN) sites versus GOES insolation retrievals for several locations representing western, central and eastern United States for September 2013. The red dots and the corresponding best linear fit for the data indicated by the red line, show the scatter plot for a SCAN site in Virginia. The blue dots and the corresponding blue line are for the SCAN sites in Iowa, Tennessee, and Kansas. The

green dots and the corresponding green line exhibit the scatter plot for Nebraska, Wisconsin, and Colorado. As demonstrated in the figure, the bias reduces as we move from east to west (similar evaluation specific to Texas is presented in the next section).

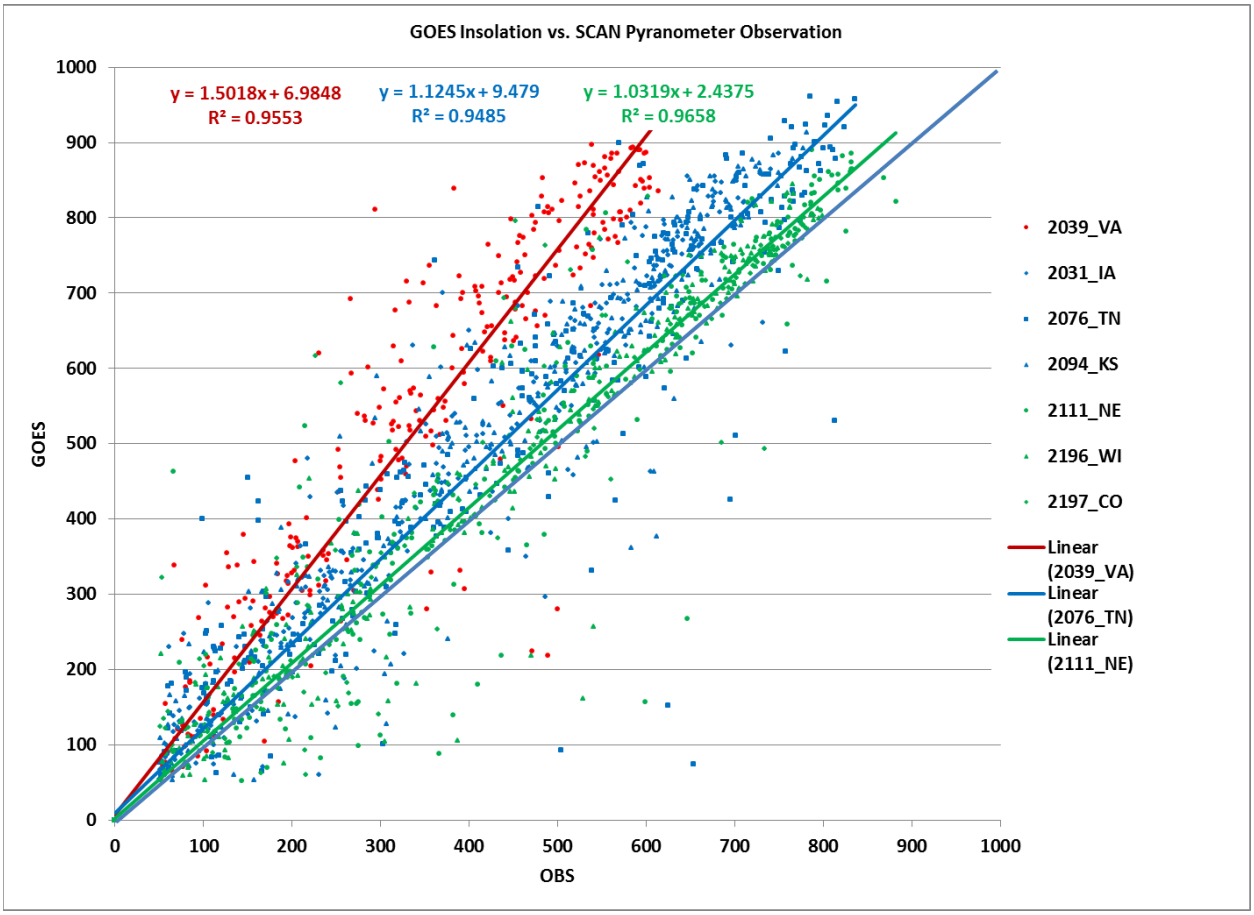


Figure 3. Scatter plots and the corresponding best linear fit showing GOES insolation retrievals versus pyranometer observations from Soil Analysis Climate Network (SCAN) for the month of September 2013.

While the large positive bias for Virginia station could be dismissed as an anomaly, the overall trend showing the progression of bias from west to east cannot be ignored. Examining the retrievals for a longer period during 2013 and the summer of 2006 also exhibited a similar pattern. The evaluation of insolation product against Surface Radiation Budget Network (SURFRAD) observations also confirmed the existence of a small bias over eastern U.S. While the use of constant precipitable water in the insolation retrieval algorithm could partly explain this bias, there are other factors in the retrieval algorithm (e.g., the construction of a composite image for surface albedo calculation) that need to be examined. Since revisiting the retrieval algorithm (something that UAH is currently addressing) was a time consuming effort, to deliver an acceptable PAR product for this project, it was decided to perform a bias correction to the insolation retrievals.

To improve the data, a simple bias correction (based on the best linear fit as compared to pyranometer) was devised and applied to the insolation data before retrieving satellite-based PAR. Figure 4 shows the GOES insolation as compared to SCAN pyranometer data before and after applying bias correction. While there is still some scatter in the data, the overall pattern shows a good correlation between pyranometer and GOES retrieval after bias correction.

As evident in the formulation for conversion factor (Equation 3), the spatial variation of PAR is very similar to the spatial pattern of insolation. The geographical areas with large reduction in insolation due to cloud cover, also exhibit smaller PAR values. Only in the areas with shallow transparent cloud cover is there a difference. Figure 5 shows a snap shot of GOES observed insolation and PAR at 19:45 GMT. PAR is produced in units of W/m^2 and micro-mol/m^2/s. As evident in the figure, the areas with insolation less than 200 W/m^2 (east/northeast part of the domain, due to cloud cover), also exhibit lower PAR values of less than 100 W/m^2 (about 50% of insolation). However, in the areas with smaller reduction in insolation (which is an indication of more transparent clouds) the corresponding PAR values are in excess of 50%.

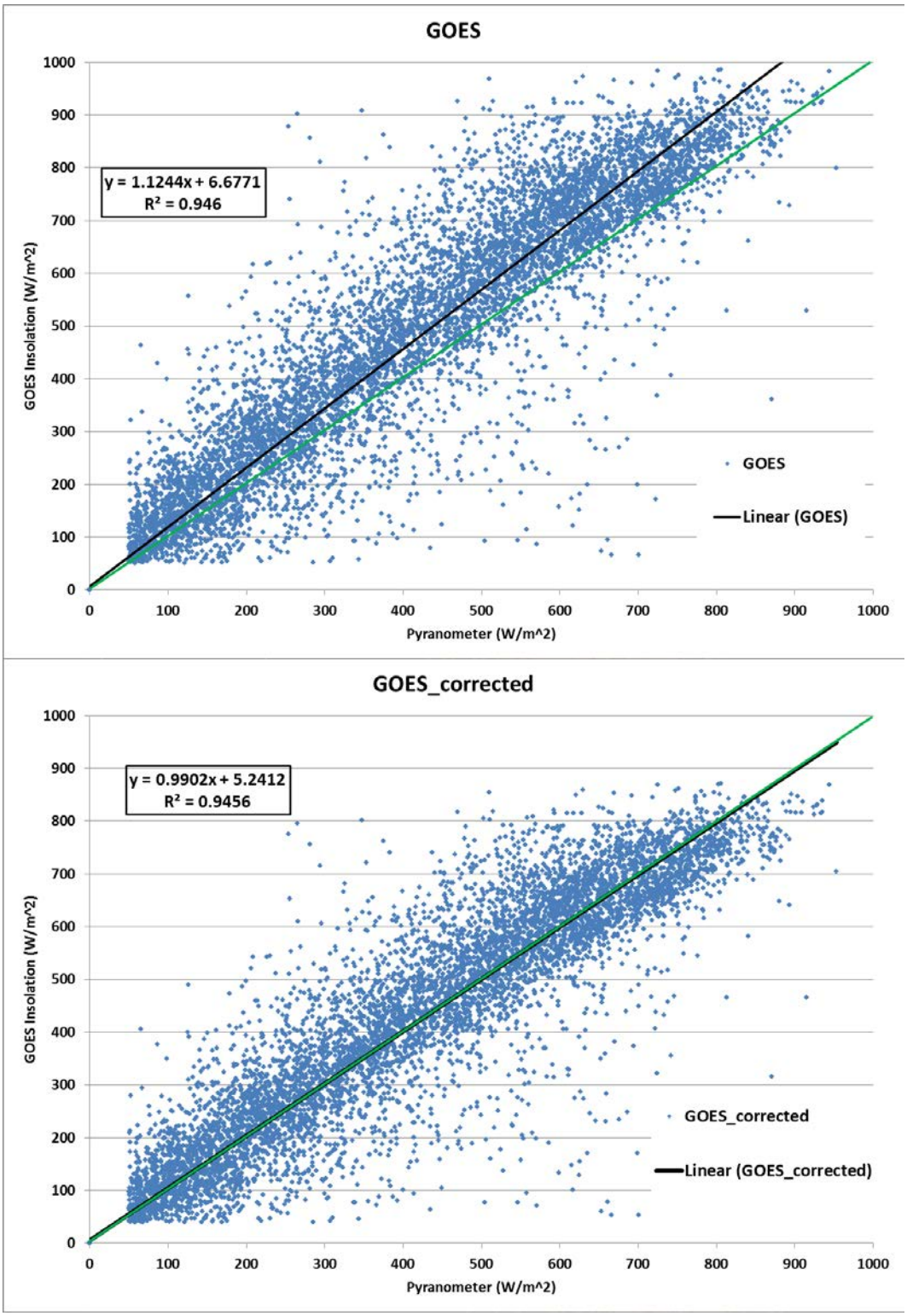


Figure 4. **Scatterplots showing GOES insolation retrievals against Soil Analysis Climate Network (SCAN) data for September 2013. The figure to the left shows the data before applying bias correction to GOES retrievals, and the figure to the right shows the scatter plot after applying bias correction.**

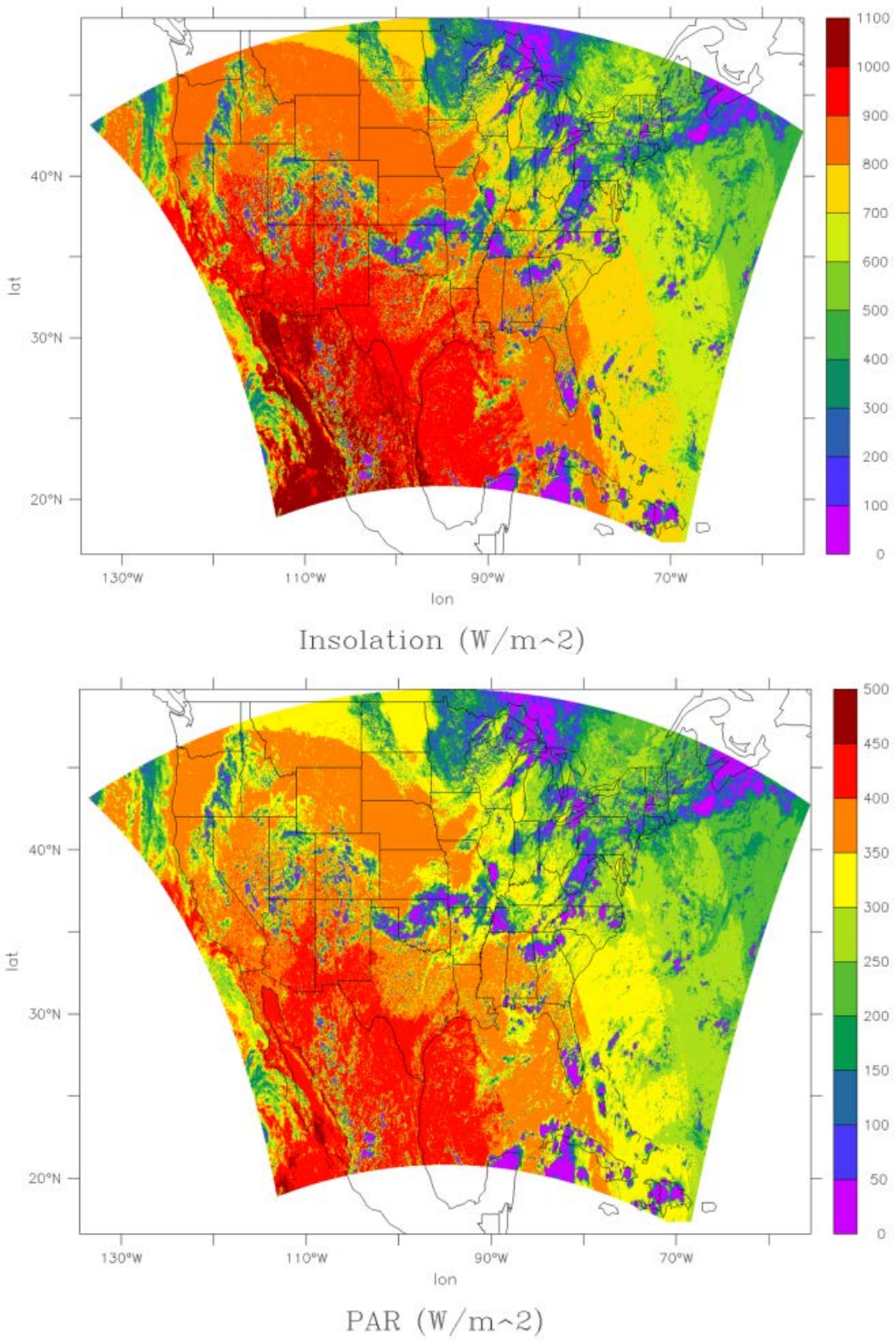


Figure 5. **A snapshot of satellite-derived insolation and PAR at 19:45 GMT, September 1, 2013. PAR is produced in units of W/m^2 as well as micro-mol/m^2/s.**

3. EVALUATION OF PAR PRODUCTS

3.1. Observations Used for PAR Evaluation

Two networks were chosen for evaluation. One is the Surface Radiation Budget Network (SURFRAD) operated by NOAA (http://www.esrl.noaa.gov/gmd/grad/surfrad/), which is the only available direct continuous measurement of PAR at seven sites nationwide. The other is the Soil Climate Analysis Network (SCAN), operated by the US Department of Agriculture (http://www.wcc.nrcs.usda.gov/scan/), which has continuous solar radiation measurements, collected by pyranometers at more than 100 stations located in 40 states. In this evaluation, 40 sites from the SCAN network and 7 sites from the SURFRAD network were chosen. The locations of these sites are shown in Figure 6. Detail evaluation of insolation/PAR retrieval products over ground observation sites in Texas is given in section 3.3.

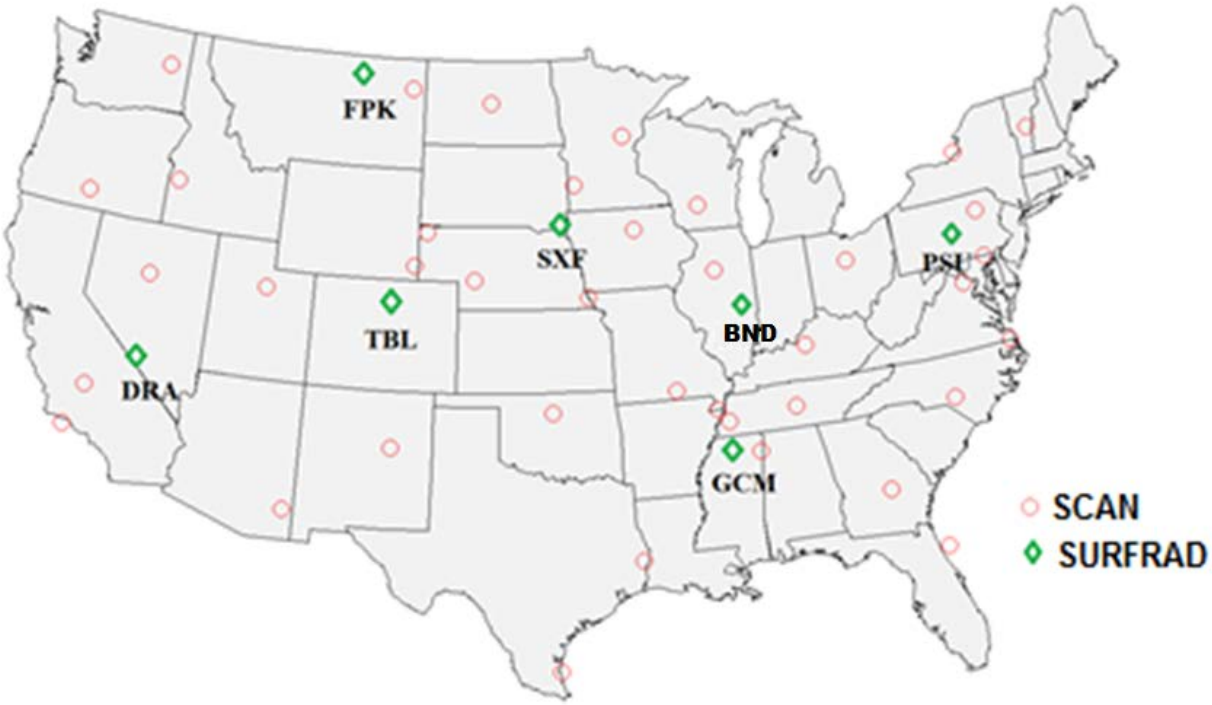


Figure 6. Location of SCAN and SURFRAD sites used for PAR/insolation evaluation

3.2. Evaluation of PAR product before Bias Correction

The month of September 2013 coinciding with the DISCOVER-AQ Houston period was chosen to evaluate the PAR retrievals by comparing with the available ground observations. The time series of PAR/insolation product at each evaluation site were pin-pointed at the nearest satellite pixel/grid (with horizontal resolution 4 km) containing the pyranometer and were

interpolated to the end of each hour (since the GOES data were instantaneous observations at 45 minutes after the hour reported on GMT time). As a reference, a 12 km CONUS WRF simulation during September 2013 was also conducted to show the typical weather model radiation performance without data assimilation from the GOES satellite.

Comparing the time series for observed PAR, PAR results from the WRF model simulation, and satellite retrievals at different SURFRAD sites show that the satellite-based PAR tends to systematically decrease the positive bias of the WRF model for the peak values during the daytime (not shown here). This was most clearly seen for periods such as Sep 17 at BND site, Sep 8 at DRA site, Sep 11-13 at PSU site, Sep 15-18 at SXF site, and Sep 13 at TBL site. The overestimation of PAR by the WRF model might be due to the incapability of its current cumulus physics module to resolve enough subgrid clouds, and because meteorological models often under-predict the amount of thin clouds.

A more detailed evaluation at each SURFRAD site by widely used statistic metrics (see Appendix A for the definition of those metrics) indicated that the satellite-based PAR product shows high correlation and good agreement with ground observations with the typically R value of 0.97-0.98 at each site. However, on average, the satellite retrieval overestimated the PAR value by 18% (Mean Bias (MB): 6.6%-28%) and deviates from the observation by 15.6 W/m^2. This over-estimation was the basis for a closer look at the insolation retrievals and the eventual bias correction applied to the data. The bias correction reduced the insolation by an average of 12% and substantially improved these statistics.

The satellite retrievals also outperformed WRF in simulating observations of insolation at SCAN network sites. On average the WRF results tend to overestimate the insolation during September 2013 by 36.5 W/m^2 at SCAN network sites and the satellite product (before bias correction) decreased the bias to 23.6 W/m^2.

Figure 7 provides the spatial comparison of insolation performance over SCAN network either by WRF results or satellite retrievals (UAH product before bias correction) in term of normalized mean bias (NMB). The satellite-derived insolation product achieved less bias in the western part of United States (with NMB -10%-10%) than in the central and southeastern part of the country (with typical NMB ~20%). There are two SCAN sites (site 2042 at Vermont and site 2039 at Virginia) that had much higher bias than the others (NMB ~50%) in the current evaluation. This is partly due to the insolation bias as described in the previous section.

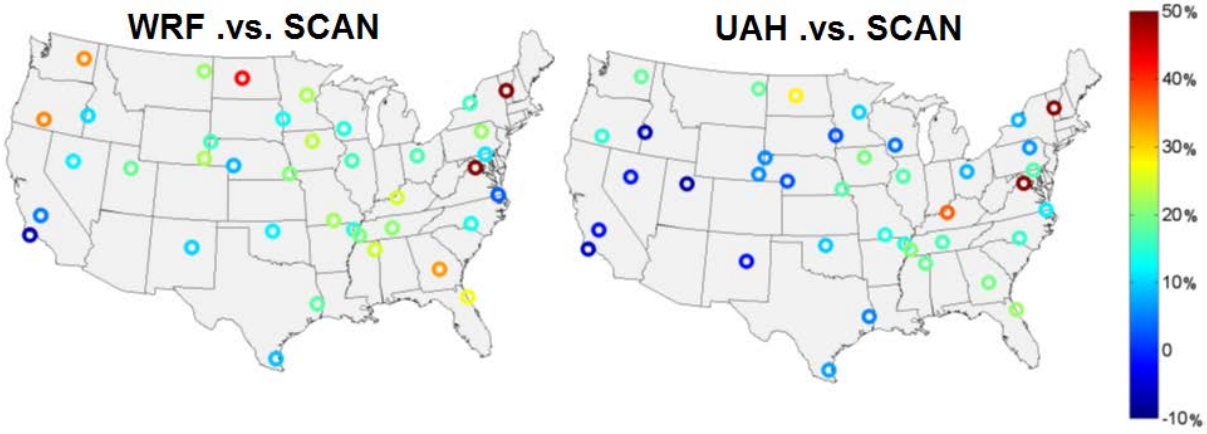


Figure 7. Spatial distribution of normalized mean bias at each SCAN site during September 2013 (UAH product before bias correction).

3.3. Evaluation of PAR product after Bias Correction

After performing bias correction as described in the section 2.2, PAR retrievals were re-evaluated for September 2013. The new retrievals reduced the bias by about 12%. Following the satisfactory results from 2013 evaluation, the project moved to the second phase of the evaluation for the summer of 2006. Summer of 2006 was chosen for evaluation for several reasons. First, another satellite-based PAR product, generated and archived by the University of Maryland (UMD), was available for 2006 for comparison. Also, the project had access to WRF simulations with and without satellite cloud assimilation that could also be used for evaluation.

Thus, satellite-based PAR was generated for the summer of 2006 and evaluated against surface observations from SCAN, SURFRAD, and TCEQ Broadband Radiation Monitoring networks. For quantifying the relative impact of satellite-based PAR, four cases for different PAR/insolation estimates were compared with surface observations:

1. WRF control case 'cntrl' with basic configurations
2. WRF cloud assimilation case 'analytical' with the cloud assimilation from GOES observations
3. PAR satellite retrievals from University of Maryland with the resolution of 0.5 degrees (http://www.atmos.umd.edu/~srb/gcip/)
4. PAR satellite retrievals from UAH with the resolution of 4 km.

PAR was computed from the WRF runs by scaling ground-reaching solar radiation (RGRND) by 50%. The details of case arrangement, as well as data sources used in performance evaluation are given in Table 1. In subsequent analyses, 'PAR_cntrl' refers to base WRF simulation, 'PAR_analytical' refers to WRF simulation with satellite cloud assimilation, 'PAR_UMD' refers to satellite-based PAR generated by UMD, and 'PAR_UAH' is the new PAR products generated in this project.

Table 1. Case arrangements for PAR and the observational data sources used in evaluation.

Case Arrangement	
1. PAR_cntrl:	Base WRF simulation to provide insolation for MEGAN
2. PAR_analytical:	Base WRF + cloud assimilation from GOES to provide insolation for MEGAN
3. PAR_UMD:	Direct use PAR retrievals from UMD, other met inputs same as case 'PAR_analytical'
4. PAR_UAH:	Direct use PAR retrievals from UAH, other met inputs same as case 'PAR_analytical'
Observational Networks	
1. NOAA SURFRAD (Surface Radiation) Network	http://www.esrl.noaa.gov/gmd/grad/surfrad
2. USDA SCAN (Soil Climate Analysis Network)	http://www.wcc.nrcs.usda.gov/scan
3. TCEQ broadband radiation monitoring Network	http://www.tceq.state.tx.us/agency/data/air_met_data.html

The four sets of PAR estimates were evaluated against observations from seven NOAA SURFRAD direct measurement sites (site locations are indicated with red dots in Figure 8) for August 2006. Since these sites represent locations from east, west, north, and southern United States, they are used to evaluate PAR over the continental U.S. Evaluation statistics are given in Table 2, and corresponding scatterplots are given in Figure 9.

Overall, both satellite retrieval products substantially outperformed the two WRF simulations, including better correlation coefficients (R=0.96~0.97 versus R=0.93) and smaller simulation errors (NME=20.7%~20.1% versus NME=32.8%~35.5%). Both satellite retrievals also achieved much lower bias than the base WRF run, though the UMD retrieval tended to underestimate PAR (NMB=-12.4%) while the UAH retrieval tended to overestimate PAR (NMB=10.2%). Despite the different sign of the bias, the performance of the two satellite products by most other metrics was similar, e.g. R=0.97 for case 'PAR_UMD' and R=0.96 for case 'PAR_UAH'; RMSE=38.5 W/m^2 for case 'PAR_UMD' and RMSE=43.0 W/m^2 for case 'PAR_UAH'; Mean Aggregate Gross Error (MAGE)=22.5 W/m^2 and MAGE=22.0 W/m^2; and NME=20.7 W/m2 and NME=21.0 W/m^2 respectively. (Table 2).

The performance of the satellite retrievals compared to the WRF runs is illustrated by the scatterplots in Figure 9, with the satellite cases (red and blue) clustering closer to the 1:1 ratio lines at all seven evaluation sites. The UAH PAR data achieved its best performance at the FPK site (Fort Peck, Montana) and its worst performance at the PSU site (Penn State, Pennsylvania) in terms of combined highest correlation value (IA=0.98) and smallest bias (NMB=3.7%), although the differences in performance across sites were not dramatic (R=0.94~0.97, RMSE=39.0W/m^2~45.4W/m^2 and NME=13.4%~28.1%). The scatter plots indicate that while the bias correction has reduced the bias in the eastern U.S. and has substantially improved PAR retrievals, perhaps it has not been enough. Small positive bias (NMB=3.7%) still exists for the areas in the east coast. As evident from Figure 9 and statistics in Table 2, the UMD product has a large negative bias at FPK (NMB=-17.7%) but a small positive bias at PSU (NMB=0.7%). This means that the UMD product also shows the same west-east pattern seen in UAH product. As evident in Figure 9, the UAH PAR product exhibits less scatter than UMD product. This could be due to the fact that the UAH product represents a smaller footprint (4-km resolution) than the UMD product (.5 degree).

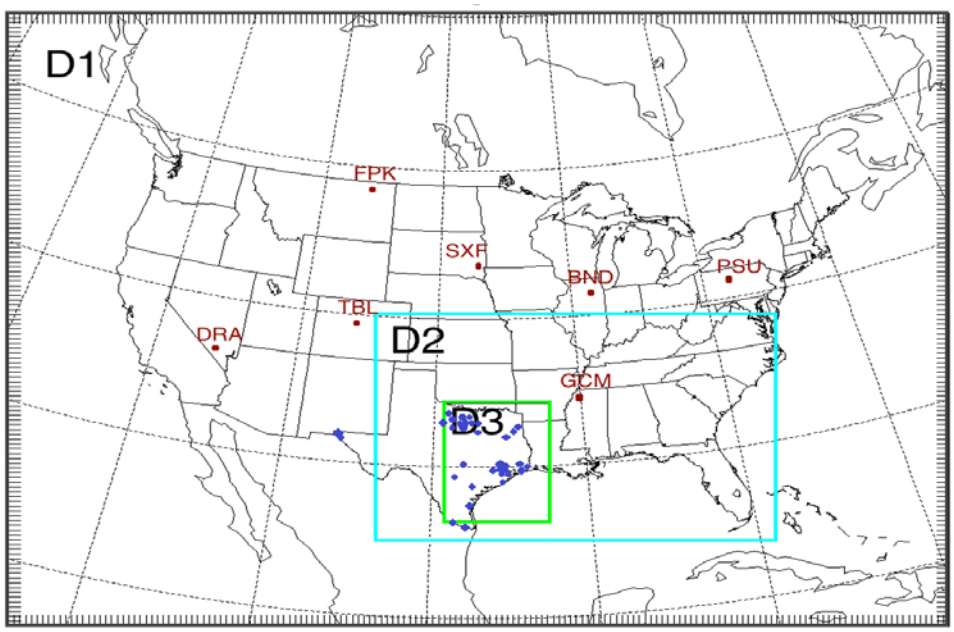


Figure 8. **Three domains for WRF-MCIP simulation in this study and locations of the insolation/PAR evaluation sites at SURFRAD network (red) and TCEQ broadband radiation network (blue).**

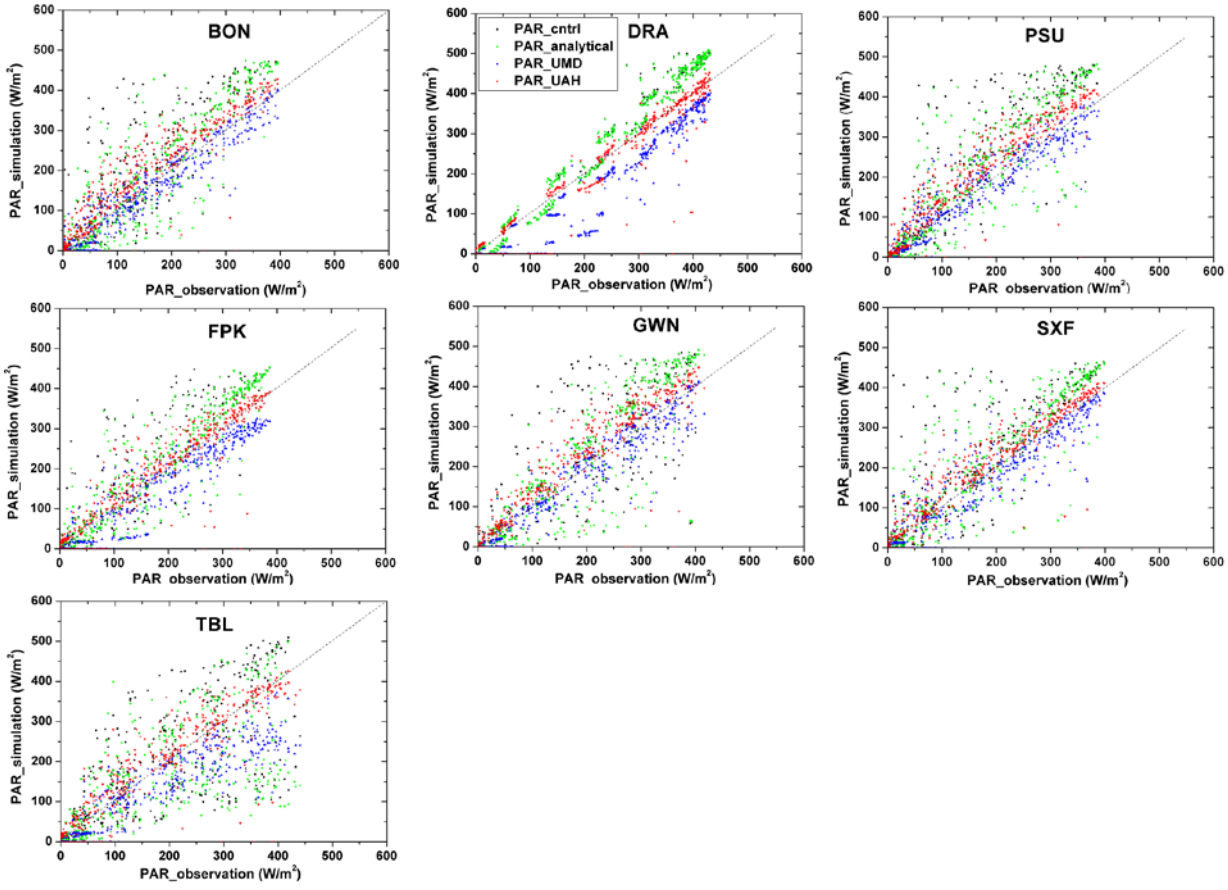


Figure 9. **Scatter plots showing four different hourly simulated/retrieved PAR (vertical axis) and observed PAR (horizontal axis) during August 2006 at 7 SURFAD sites.**

Table 2. Summary of statistics for PAR simulation/retrievals for different cases at 7 SURFRAD network sites[1]

CASE	SITE	OBS_AVE (W/m2)	SIM_AVE (W/m2)	IA	R	RMSE (W/m2)	MB (W/m2)	MAGE (W/m2)	NMB (%)	NME (%)
PAR_cntrl	BND	92.6	119.8	0.94	0.93	68.1	27.1	38.5	29.2	41.5
	DRA	140.8	162.1	0.99	0.99	41.6	21.4	26.0	15.2	18.5
	FPK	109.8	130.2	0.96	0.96	56.2	24.2	32.8	22.0	29.9
	GCM	111.0	132.7	0.94	0.92	71.2	21.8	40.5	19.6	36.5
	PSU	92.1	133.9	0.93	0.94	77.0	37.2	43.0	40.4	46.7
	SXF	101.6	126.1	0.95	0.93	65.5	24.4	33.4	24.0	32.8
	TBL	105.9	113.9	0.92	0.86	77.7	7.9	45.4	7.5	42.9
	average	107.7	131.2	0.95	0.93	65.3	23.4	37.1	22.6	35.5
PAR_analytical	BND	92.6	107.5	0.95	0.94	58.3	14.9	34.4	16.0	37.1
	DRA	140.8	162.0	0.99	0.99	41.3	21.3	25.8	15.2	18.4
	FPK	109.8	121.9	0.97	0.96	49.8	16.3	29.7	14.8	27.1
	GCM	111.0	124.8	0.96	0.94	61.9	13.9	36.4	12.5	32.8
	PSU	92.1	118.1	0.94	0.93	65.4	21.8	37.0	23.6	40.2
	SXF	101.6	119.2	0.95	0.93	60.2	17.5	32.0	17.2	31.4
	TBL	105.9	93.7	0.89	0.81	80.7	-12.4	45.0	-11.7	42.5
	average	107.7	121.0	0.95	0.93	59.6	13.3	34.3	12.5	32.8
PAR_UMD	BND	92.6	87.2	0.99	0.98	25.8	-5.4	15.3	-5.9	16.5
	DRA	140.8	108.2	0.97	0.97	51.7	-32.2	32.5	-22.9	23.1
	FPK	109.8	87.1	0.97	0.97	39.0	-19.4	24.8	-17.7	22.6
	GCM	111.0	103.1	0.99	0.98	30.0	-7.9	16.5	-7.1	14.9
	PSU	92.1	95.6	0.98	0.97	28.9	-0.1	15.9	0.7	17.2
	SXF	101.6	94.3	0.98	0.96	36.5	-7.3	20.4	-7.1	20.1
	TBL	105.9	78.5	0.94	0.95	57.2	-27.4	32.1	-25.9	30.3
	average	107.7	93.4	0.97	0.97	38.5	-14.2	22.5	-12.4	20.7
PAR_UAH	BND	92.6	110.6	0.97	0.96	44.1	19.4	25.2	21.3	27.6
	DRA	140.8	130.4	0.98	0.96	44.0	-10.0	18.8	-7.1	13.4
	FPK	109.8	112.8	0.98	0.96	39.9	4.2	19.4	3.7	17.2
	GCM	111.0	127.3	0.98	0.97	44.2	16.6	23.7	15.0	21.4
	PSU	92.1	117.1	0.97	0.97	44.2	21.1	25.7	23.0	28.1
	SXF	101.6	114.9	0.98	0.96	39.1	11.2	19.5	10.8	18.8
	TBL	105.9	111.4	0.97	0.94	45.4	5.1	22.0	4.8	20.7
	average	107.7	117.8	0.97	0.96	43.0	9.7	22.0	10.2	21.0

Note: [1]The 7 SURFRAD sites are: BND (Bondville, IL), DRA (Desert Rock, NV), FPK (Fort Peck, MT); GCM (Goodwin Creek, MS), PSU (Penn. State Univ., PA), SXF (Sioux Falls, SD), and TBL (Table Mountain, CO).

IA-index of agreement, R-correlation coefficient, RMSE-root mean square error, MB-mean bias, MAGE-mean aggregate gross error, NMB-normalized mean bias, and NME-normalized mean error.

The insolation outputs from the two WRF runs (based on the WRF variable 'RGRND' for solar radiation reaching surface) and the UAH satellite retrieval were compared for August 2006 with 47 available broadband radiation monitoring stations over Texas (the locations are indicated in Figure 8 with blue dots). The solar radiation observation data are provided by TCEQ. Table 3 summarizes the overall model performance for insolation for the three cases.

Table 3. Summary of statistics of insolation simulation/retrievals for different cases at 47 TCEQ network sites.

	OBS_AVE (W/m2)	SIM_AVE (W/m2)	IA	R	RMSE (W/m2)	MB (W/m2)	MAGE (W/m2)	NMB (%)	NME (%)
WRF cntrl	248.6	299.8	0.95	0.91	142.3	53.9	74.7	22.2	30.7
WRF analytical	248.6	266.8	0.95	0.91	143.9	20.3	74.9	8.9	30.7
UAH satellite	248.6	263.6	0.96	0.96	123.2	17.3	71.8	7.5	29.5

Note: IA-index of agreement, R-correlation coefficient, RMSE-root mean square error, MB-mean bias, MAGE-mean aggregate gross error, NMB-normalized mean bias, and NME-normalized mean error.

Using satellite data substantially reduces the overprediction bias of the WRF control run, reducing NMB from 22.2% to 8.9% for the cloud-assimilated WRF run and 7.5% for the UAH retrieval. While cloud assimilation was able to reduce the overprediction bias in WRF, the UAH retrieval strongly outperformed both WRF runs in terms of error and correlation (Table 3). This indicates that WRF with cloud assimilation has been able to significantly improve the location and timing of clouds over Texas.

Figure 10 provides the spatial maps of the correlation coefficient (R, upper panel) and normalized mean bias (NMB, lower panel) for the individual 47 TCEQ sites among the three different cases. In terms of correlation coefficient R, the WRF 'cntrl' case performs worst (R~0.85) near the coastal sites especially around the greater Houston area. It performs better at inland sites near the Dallas region(R~0.92). The cloud assimilation WRF run slightly improves model correlation at most of the sites. UAH satellite retrieval insolation products significantly improve the model correlation with the R values at inland sites approaching 1 and at the coastal sites mostly around 0.95. In terms of NMB, all three products performed better around Dallas and other inland sites than near Houston, and the satellite-based cases far outperformed the base WRF simulation. Cloud assimilation in WRF reduces the normalized mean bias significantly, but slightly improves the correlation. This could be due to an inherent problem in WRF that generally produces more opaque clouds. Appendix B provides the detail statistics for the insolation/simulation retrievals for the three different cases at each of the 47 TCEQ network sites for reference.

A much smaller NMB for the satellite product and significantly higher correlation for sites in eastern Texas indicate the robustness of satellite retrievals. The large difference in performance between base WRF simulation and satellite product with respect to radiation field in a region with high vegetation cover could have significant impact on BVOC emissions.

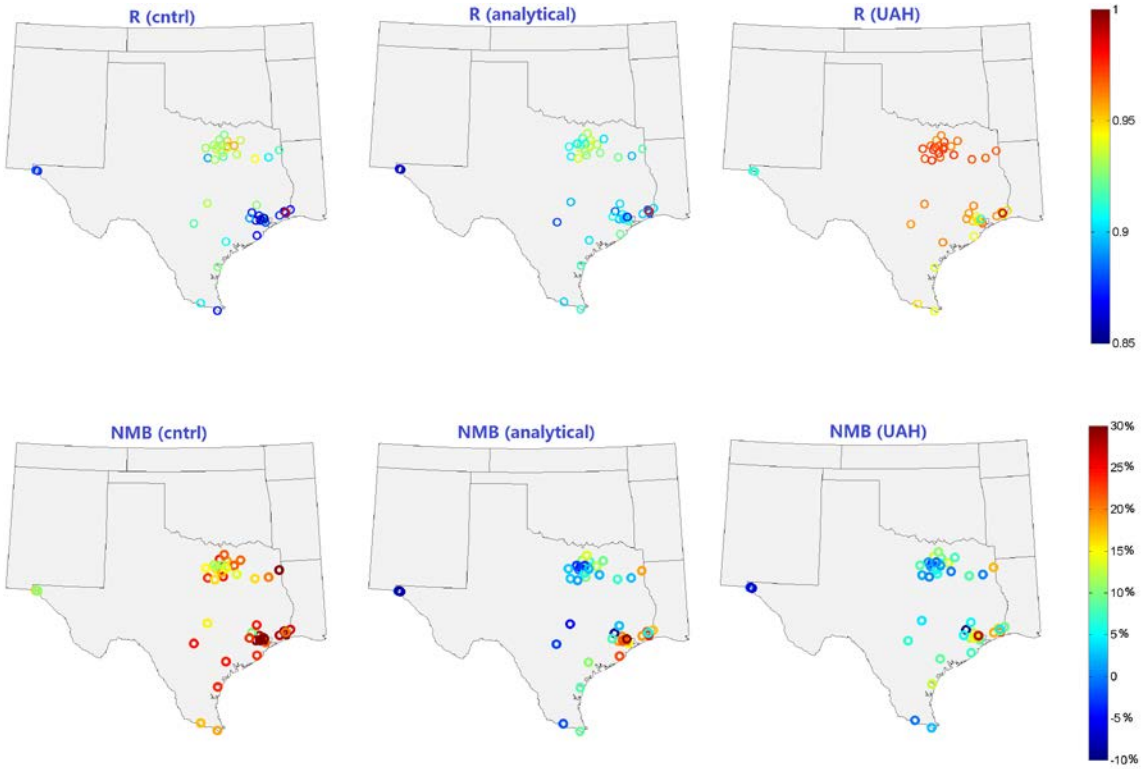


Figure 10. Performance of WRF incoming solar radiation (RGRND) from simulation case 'cntrl' (left) and 'analytical' (right) as well as UAH insolation retrievals (right) at TCEQ sites. The upper panel shows the correlation coefficient (R) and the lower panel shows the normalized mean bias (NMB).

It should be noted that during this evaluation, Spatial Allocator was used for mapping to ensure that the 4km resolution UAH satellite products mapped properly to the model grid. Spatial Allocator (www.epa.gov/AMD/Tools/spatialAllocator.html) is a community tool and has an option for mapping UAH products.

4. WRF SIMULATIONS WITH CLOUD ASSIMILATION (AUG-SEP 2013)

In addition to base WRF simulations used in the preliminary evaluation of PAR and MEGAN BVOC estimates for 2013, additional simulations using the TCEQ State Implementation Plan (SIP) domain and model configurations were also performed. This was to test the impact of improved cloud prediction on air quality simulations.

Clouds play a critical role in the production and destruction of pollutants and yet the models have difficulty in creating clouds in the right place and time compared to observed clouds. This

is especially the case when synoptic-scale forcing is weak (e.g. Stensrud and Fritsch 1994). Weak synoptic-scale forcing is often associated with air pollution events. Errors in model cloud not only impact radiative fluxes and subsequently surface temperature and boundary layer evolution, they also alter the boundary layer photochemistry, aerosol formation and recycling, heterogeneous chemistry, and wet deposition.

Under a previous TCEQ funded project (Pour-Biazar et al., 2011), UAH has developed a technique to assimilate GOES cloud observations in the WRF model that dynamically adjusts cloud fields (Allen et al., AQRP State of the Science 2012 Report). This technique has proven to improve cloud simulation in WRF. In this project, this technique was used in WRF to test the sensitivity of biogenic emission estimates and air quality simulations to improved cloud simulations.

Meteorological Model: WRF (Skamarock et al. 2008) is the meteorological model most widely used in CMAQ. The model is a joint undertaking of NOAA and NCAR (the National Center for Atmospheric Research) and is used to simulate local and synoptic scale meteorological conditions prevalent during the period of interest. The WRF model is a community model. WRF has many capabilities pertinent to the needs of this project. These include: (1) a multiple-nest capability; (2) non-hydrostatic dynamics that allow the model to be used at a scale of approximately 4 km; (3) multi-tasking capability on shared and distributed-memory machines; (4) four-dimensional data assimilation (FDDA) capability, and (5) multiple physics options (http://www.wrf-model.org/index.php). Required data inputs include topography and land cover/land use (LU/LC) data, gridded atmospheric fields of sea-level pressure, wind, temperature, relative humidity and geopotential height at defined pressure levels, and observation data including soundings and surface reports. LU/LC is used as input directly to WRF to provide surface boundary conditions such as albedo, soil moisture availability, surface roughness and canopy height.

The August and September 2013 WRF simulations were performed for three domains. The extent of these domains is shown in Figure 8. Table 4 summarizes the domain configuration used for 2013 WRF simulations.

Table 4. WRF domain setup for 36-, 12-, and 4-km grid spacing used for UAH simulations.

Domain Name	Range (km)		Number of Grid Points		Cell Size (km)	
	Easting	Northing	Easting	Northing	Easting	Northing
CONUS	(-2916,2916)	(-2304,2304)	163	129	36	36
SouthUS	(-1188,900)	(-1800,-144)	175	139	12	12
Texas	(-396,468)	(-1620,-468)	217	289	4	4

Domain 1 is a 36 km grid covering the Continental United States (CONUS) region, domain 2 is a nested 12 km grid covering part of the Southern United States (SouthUS) and domain 3 is a nested 4 km grid covering Texas. In the following model performance for all three domains is

evaluated with respect to clouds and surface measurements. The WRF configuration used for UAH simulations (both 2006 and 2013) is given in Table 5.

Table 5. WRF configuration for 2013 satellite cloud assimilation runs. WRF 2006 configuration shown in brackets where different from 2013 configuration.

	Domain 1	Domain 2	Domain 3
Simulation Period	August – September 2013 [August 2006]		
Horizontal Resolution	36 km	12 km	4 km
Time Step	90 s	30 s	10 s
Number of Vertical Levels	43 [42]		
Top Pressure of the Model	50 hPa		
Shortwave Radiation	The Rapid Radiative Transfer Model for GCMs (RRTMG) [Dudhia]		
Longwave Radiation	The Rapid Radiative Transfer Model for GCMs (RRTMG) [RRTM]		
Surface Layer	Monin-Obukhov		
Land Surface Layer	Unified Noah (4-soil layer)		
PBL	Yonsei University (YSU) scheme		
Microphysics	Thompson [LIN]		
Cumulus Physics	Kain-Fritsch (with Ma and Tan 2009 trigger function)		None
Meteorological Input Data	NAM Analysis [Eta Model Analysis]		
Analysis Nudging	Yes		Winds Only
U, V Nudging Coefficient	3×10^{-4}		3×10^{-4}
T Nudging Coefficient	3×10^{-4}		0
Q Nudging Coefficient	1×10^{-5}		0
Nudging within PBL	Yes for U and V, No for q and T		Yes for U and V

The results of two different WRF simulations will be presented: the control (CNTRL) simulation and the GOES satellite assimilation (ASSIM) simulation. Note that 'ASSIM' and 'Analytical' have been used interchangeably in this report. The control simulation only nudges in the North American Mesoscale analysis database (NAM) analysis temperature, wind, and mixing ratio data throughout the forecast time period, while the Assimilation simulation uses an analytical technique for assimilating in GOES satellite observations through the nudging field. Note that for domain 3, only winds were nudged into the model as indicated in Table 5. The main reason for this is that the nudging of analysis winds into WRF has been shown to reduce the positive wind bias that the WRF model produces (Pour-Biazar et al., 2014). This also leads to smaller errors in the wind speed and direction.

4.1. WRF Model Performance Evaluation

To evaluate the model performance of the control and assimilation simulations, the cloud agreement index (AI) and model statistics with respect to surface observations were calculated. AI is defined over the model domain as the total number of grid cells that are in agreement with satellite observation (either cloudy or clear) divided by the total number of grid cells. Table 6 shows the contingency table describing how AI is calculated.

Table 6. Contingency table used for evaluating model cloud simulation.

AI = (A+D) / (A+B+C+D)		WRF		TOTAL
		Cloudy	Clear	
GOES	Cloudy	A	B	A+B
	Clear	C	D	C+D
TOTAL		A+C	B+D	A+B+C+D

The (AI) calculates how well the model does at producing clouds in the correct place and at the correct time when compared to GOES satellite observations. Thus, it will be used as the metric to rate the model cloud performance. Therefore a value of AI closer to 1 is desirable. The AI was calculated for each hour in the range 15:00-22:00 UTC in the August-September 2013 time frame. The time range was chosen to ensure maximum daylight coverage across the domain so that GOES imager observations are available. The hourly AIs were then averaged to produce the daily AI. To calculate model statistics with respect to surface observations, METSTAT (http://www.camx.com/download/support-software.aspx) was used to determine the model bias and root mean square error (RMSE) for wind speed, temperature, and mixing ratio.

4.1.1. Performance Statistics for Domain 1

The daily AI for domain 1 in Figure 11 shows that the ASSIM simulation has a greater AI than the CNTRL simulation for all days in the simulation time period. The average daily percentage increase in the AI from the CNTRL to the ASSIM simulation was found to be 12.71%. The individual hourly results similarly showed that the AI was greater for the ASSIM simulation than it was for the CNTRL simulation. The maximum hourly percentage increase was found to be 22.54%, while the minimum increase was 0.92%. These results show that this GOES assimilation technique, overall, does improve cloud placement in space and time relative to GOES satellite observations. The large variation in hourly AIs within a day indicates that the technique at some periods is more effective than other periods. Figure 12 shows the average hourly AI for the simulation period. The assimilation technique performs better in the latter

part of the day. For this domain (36-km domain) the hourly pattern for assimilation is not drastically different from the control simulation.

Figure 13 shows a spatial plot of the agreement index for August 21, 2013 at 17 UTC. In this figure green and grey areas indicate that both model and GOES observations agree on being clear (green) or cloudy (grey). Red areas indicate model over-prediction of clouds (as compared to satellite observation) and orange indicate model under-prediction. The figure shows that the CNTRL simulation has trouble creating clouds in locations that GOES observes them as indicated by the large coverage of orange shading. Also, the CNTRL simulation tends to produce more clouds over the ocean than what is observed by GOES, as indicated by the red shading in the figure. By assimilating GOES observations into WRF, the result is less over-prediction and under-prediction of clouds with respect to observations, as can be seen with the reduction of orange and red shading when the two images are compared. For this scene, the agreement index (AI) is increased by 13.5%.

Evaluating the model performance with respect to surface observations, assimilation simulation reduces wind speed bias, but increases the temperature cold bias. The wind and temperature statistics are shown in Figure 14. The wind speed bias is less in the assimilation simulation but the error actually increases indicating that the reduction in the bias is more due to an increase in the wind speed across the domain. The negative temperature bias is increased in the ASSIM simulation, which is due to higher cloud fractions throughout the model domain than the control simulation. It seems that the assimilation simulation improved AI mostly by creating clouds over land and removing them over ocean causing a reduction in the amount of energy (radiation) reaching surface and increasing an inherent cold bias in the model. For mixing ratio, generally the control simulation tends to be dryer, but overall closer to the observations.

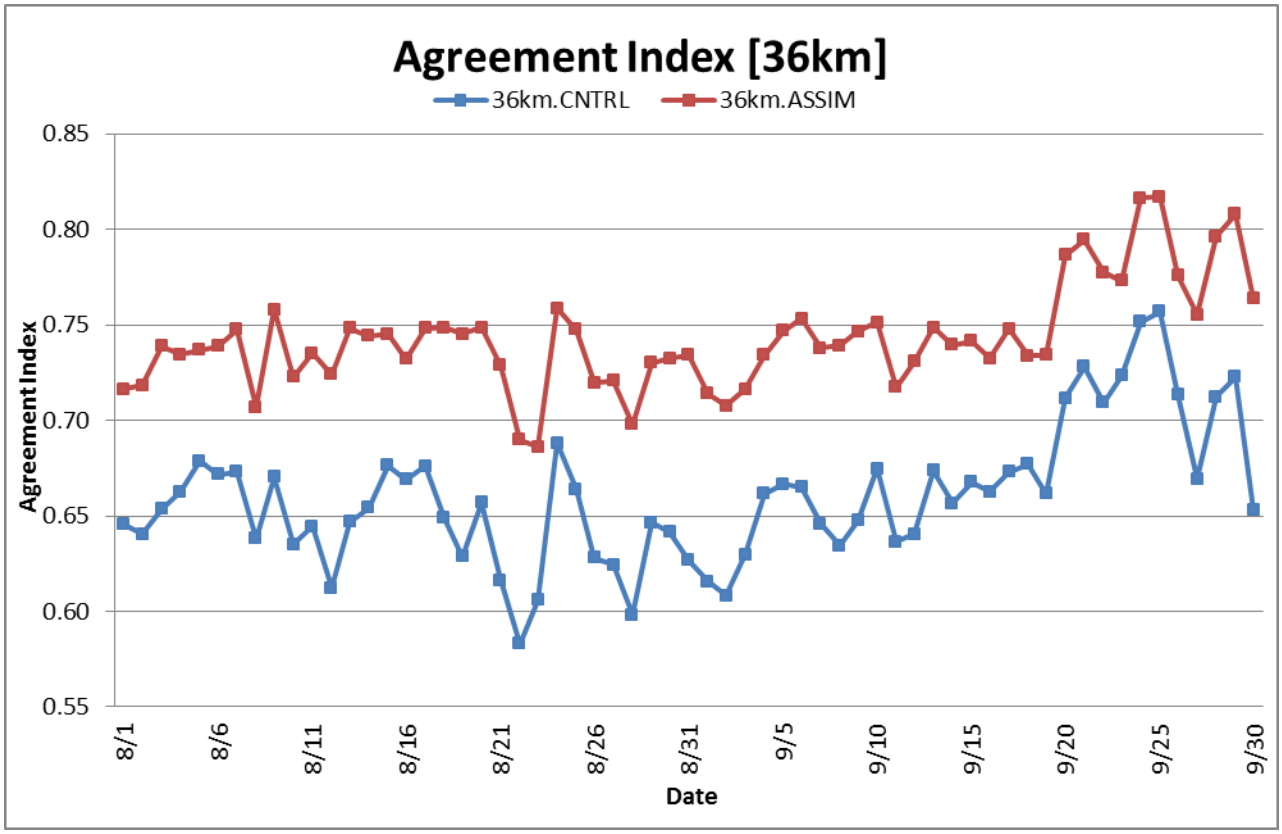


Figure 11. **Daily agreement index for CNTRL and ASSIM 36 km WRF simulations over August-September 2013 using a 10% cloud albedo threshold.**

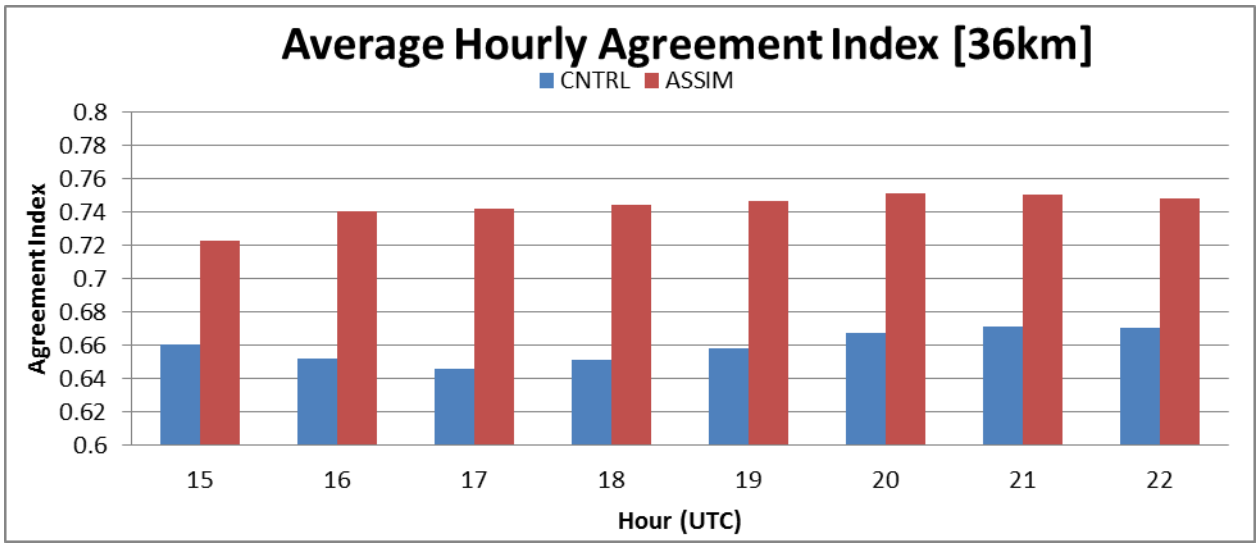


Figure 12. **Average hourly AI for August-September 2013. Assimilation performs better in the latter part of the day.**

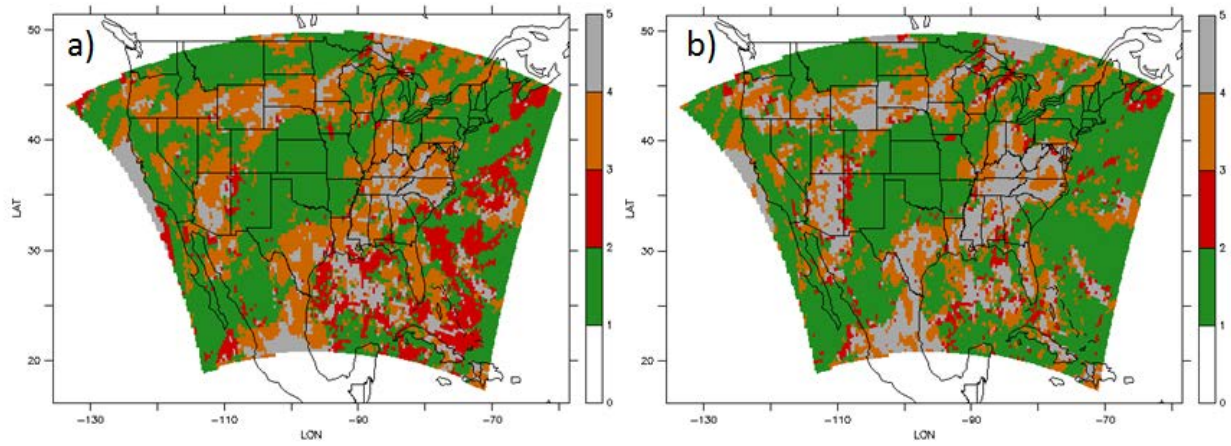


Figure 13. **Comparing model cloud to satellite observation for CNTRL (left) and ASSIM (right) simulations for August 21, 2013 at 17 UTC: Green indicates the model and GOES were clear, Red indicates model cloud over-predictions, Orange indicates model cloud under-predictions, and Grey indicates locations where the model and GOES are cloudy. Agreement Index for a) CNTRL (AI=59.9%) b) ASSIM (AI=73.4%).**

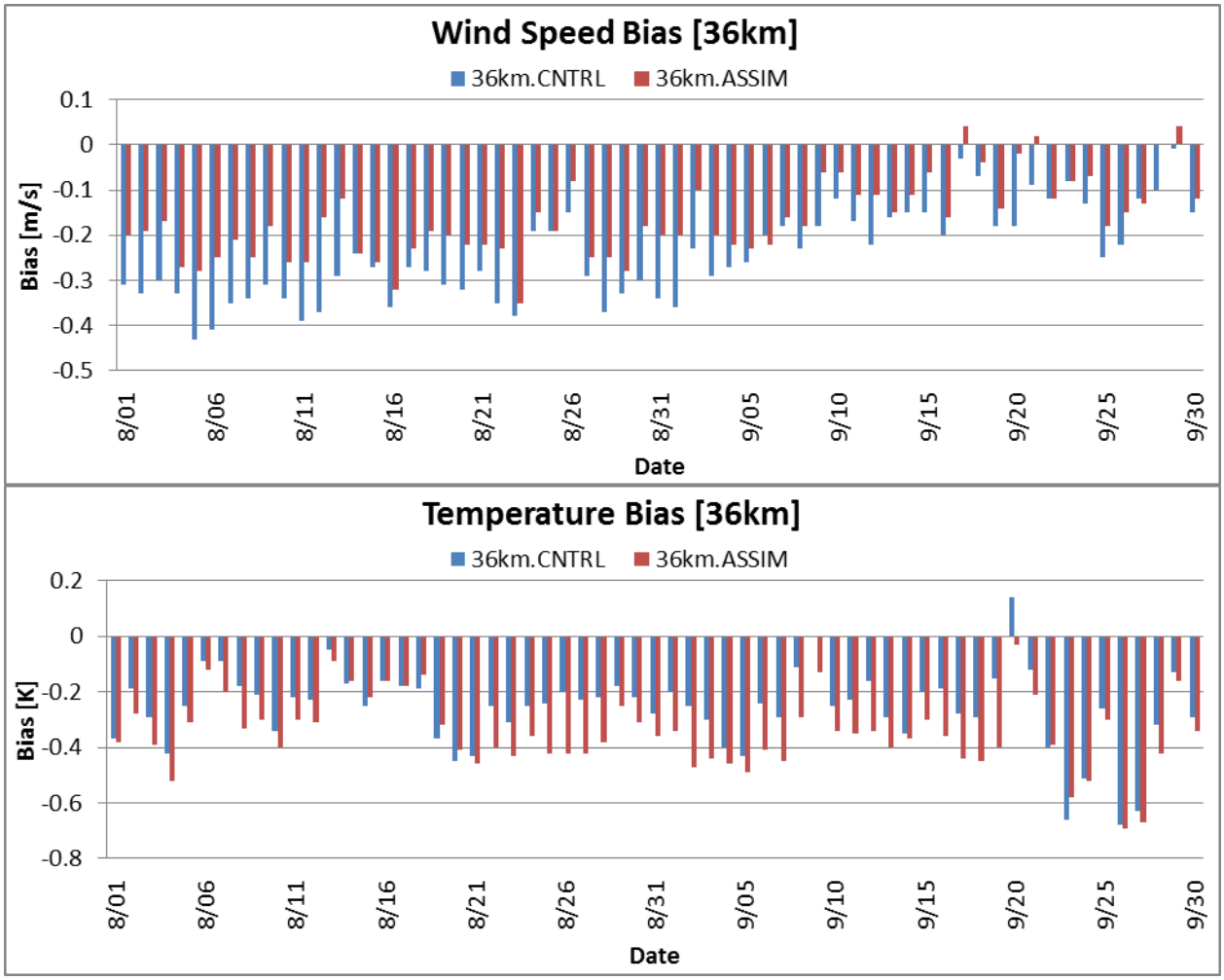


Figure 14. **Wind speed and temperature bias results for CNTRL and ASSIM simulations.**

Overall, the results of the 36 km grid indicate that the assimilation technique led to better cloud agreement. In doing so, the assimilation simulation increased the surface wind speeds across the domain and increased the surface moisture as well.

4.1.2. Performance Statistics for Domain 2

For the 12 km domain, the control and assimilation WRF simulations are compared with observations in the same manner. The daily AI results are shown in Figure 15. Similar to the results for the 36 km domain, the daily AI for the ASSIM simulation was greater than the CNTRL simulation. However, the average daily percentage increase was determined to be less at 9.65%. Also, AI exhibited greater variability in the hourly comparisons over the control simulation. The maximum hourly percentage increase was found to be 24.75%, while the minimum hourly percentage increase was found to be -4.70%. The negative sign indicates that at the particular hour, the control simulation actually performed better. However, this decrease in the AI occurred during times where the AI was already high for control simulation (AI>80%), in which corrections to disagreement areas tend to have a higher probability of disrupting areas previously in agreement with the GOES observations.

Figure 16 shows the average hourly AI for the 12-km domain during the simulation period. Similar to the 36-km domain, the assimilation technique performs better than control for all hours of the day, but the pattern is different from the 36-km domain. For the 12-km domain the hourly pattern for assimilation shows a reduction for AI as the sun moves overhead. The largest AI's are achieved in mid-morning and the latter part of the day.

Figure 17 shows a spatial plot of the AI for August 27, 2013 at 22 UTC. For this particular hour, the AI for the CNTRL was low with a large amount of under-prediction (orange) by the model centered over Texas and over-prediction (red) by the model over the Gulf of Mexico up into Mississippi. The ASSIM simulation was able to efficiently clear the over-prediction areas and increase the cloud coverage over the state of Texas where there was under-prediction. This resulted in better overall agreement between the model and GOES observations. Thus, while on the average AI for 12-km domain was less than 36-km simulation, there are particular hours when the assimilation technique substantially improve cloud placement for 12-km domain.

Evaluation of model performance for the 12-km grid with respect to surface observations shows that the ASSIM simulation performs better with respect to temperature. Both bias and RMSE are reduced for temperature. For the mixing ratio, the result is mixed. Bias in the mixing ratio is slightly increased for some days and reduced for other days. The ASSIM simulation tends to increase the mixing ratio but not as consistently as was seen in domain 1. The temperature and mixing ratio biases are shown in Figure 18. The wind speed bias is less in the ASSIM simulation throughout the simulation period but the error fluctuates. The reduction in temperature RMSE for ASSIM simulation is due to the correction of clouds within the model.

Overall, the results for the 12-km grid indicate that the assimilation technique led to improved cloud agreement. In doing so, the ASSIM simulation improved the model performance for surface temperature. As the grid size was reduced, the wind speed statistics for the ASSIM simulation improved, but there are still times where the error in the wind statistics is less for the CNTRL simulation. The mixing ratio results also started to become more variable between the two simulations with the CNTRL simulation having less error on some days. With the reduction in the spatial coverage of the domain, it is likely that analyzing more spatial patterns would reveal the exact cause of these variations and perhaps will lead to identifying the scenarios in which satellite cloud assimilation would be more effective. Future research also will try different model configurations to identify which configuration is more conducive to the adjustments introduced by this technique.

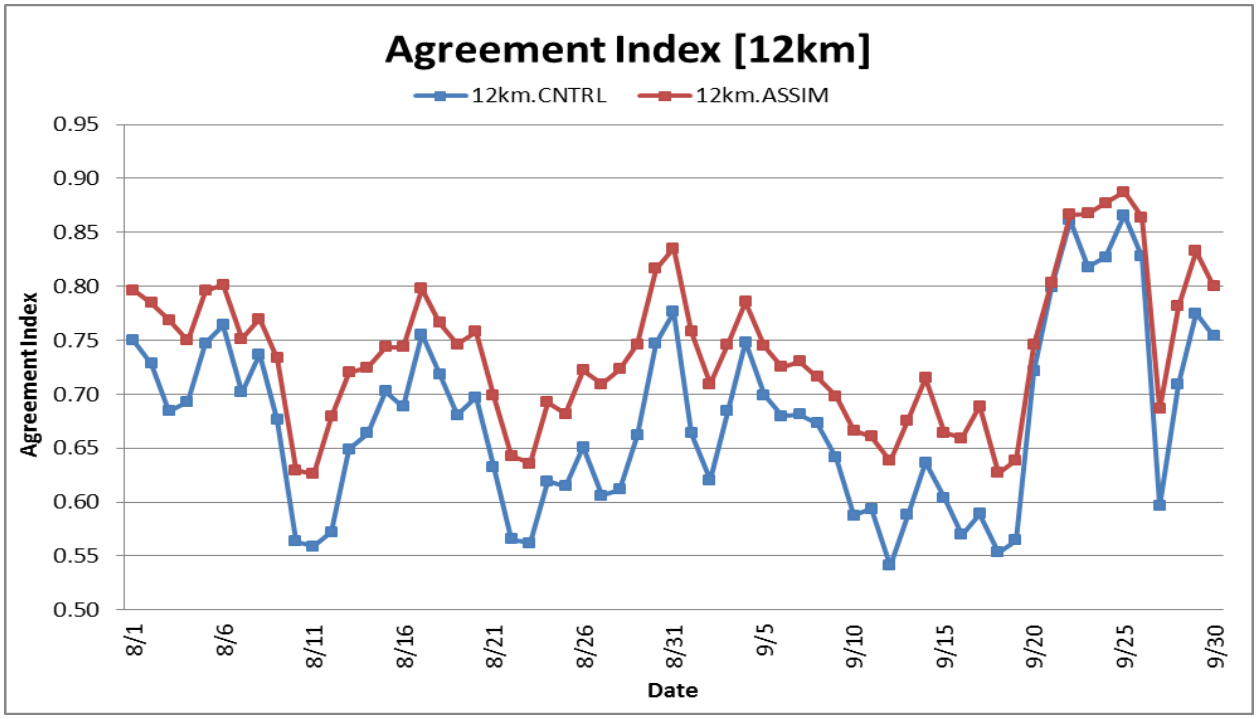


Figure 15. Daily agreement index for CNTRL and ASSIM 12 km WRF simulations over August-September 2013 using a 10% cloud albedo threshold.

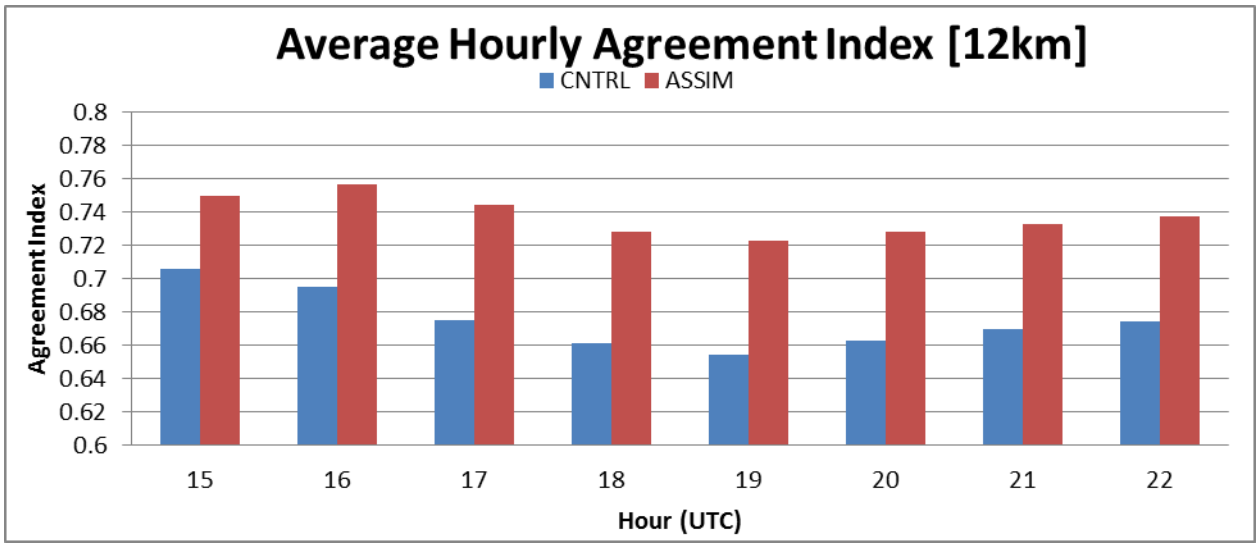


Figure 16. Average hourly AI for 12-km domain during August-September 2013.

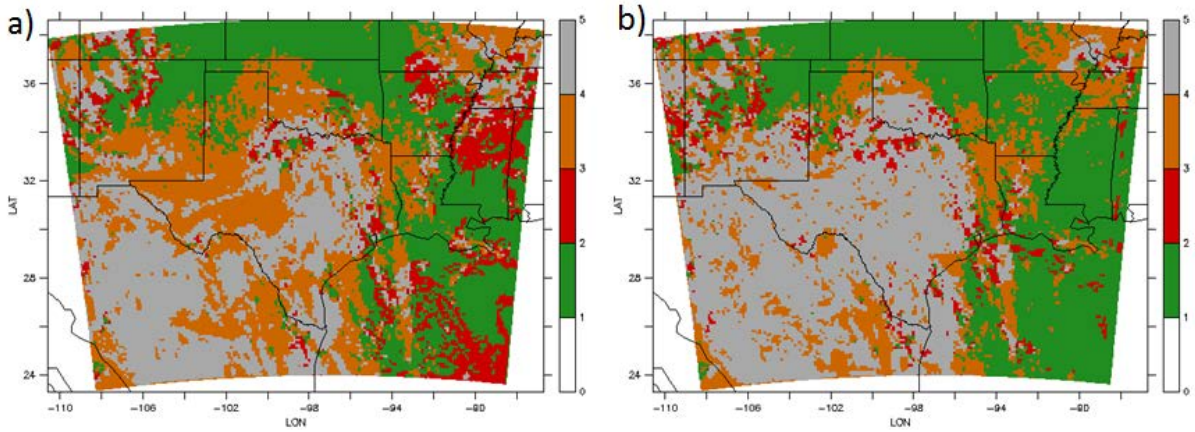


Figure 17. Comparing model cloud to satellite observation for CNTRL (left) and ASSIM (right) simulations for August 27, 2013 at 22 UTC: Green: model and GOES were clear, Red: model cloud over-predictions, Orange: model cloud under-predictions, and Grey: locations where the model and GOES are cloudy. Agreement Index for a) CNTRL (AI=59.0%) b) ASSIM (AI=73.6%).

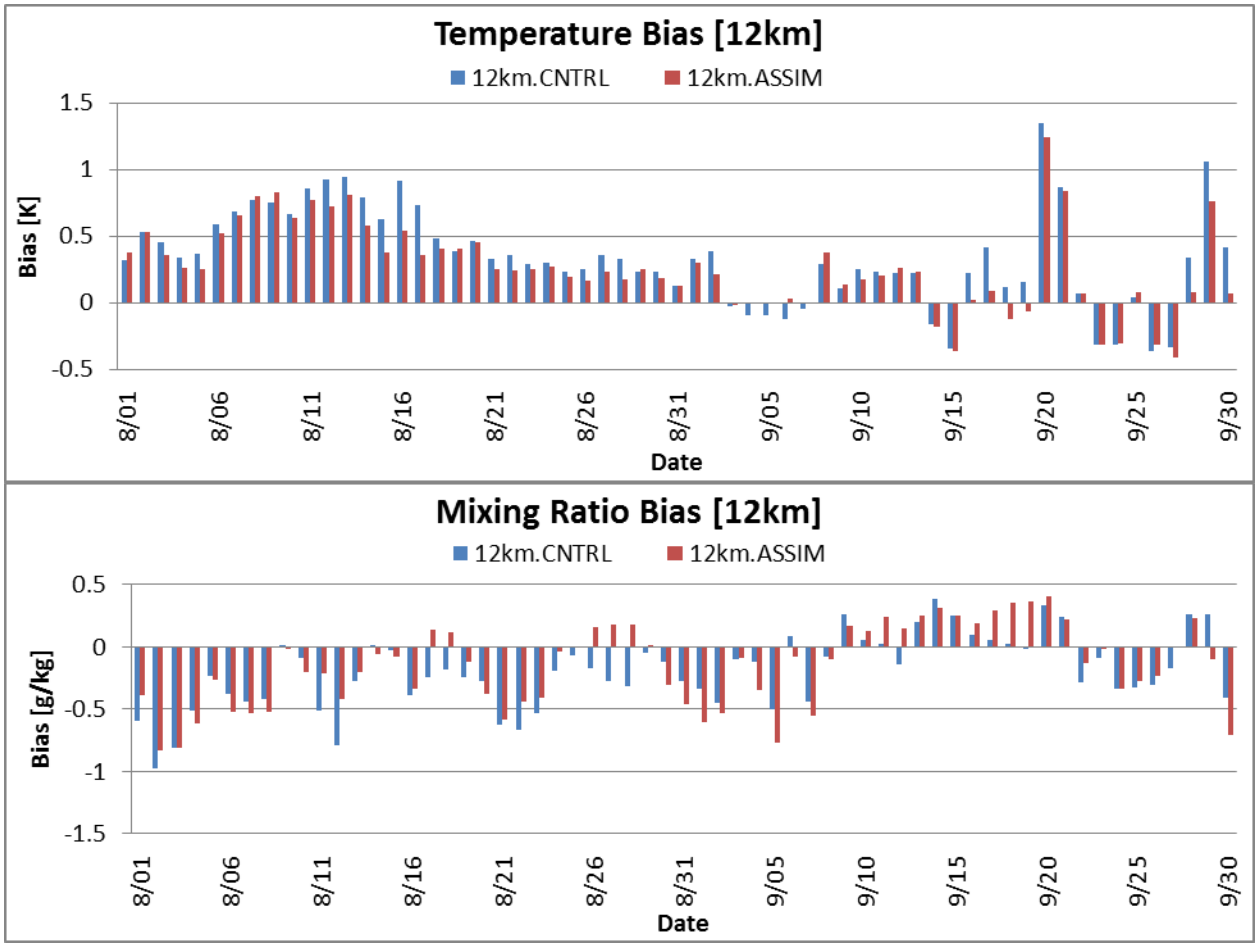


Figure 18. Temperature and mixing ratio bias results for 12-km CNTRL and ASSIM simulations.

4.1.3. Performance Statistics for Domain 3

For the 4-km domain, the improvement in the AI was less significant. Figure 19 shows the daily averaged AI for the 4-km domain during August and September, 2013. The average daily percentage change in the AI due to cloud assimilation was 6.28%. However, for domain 3 there were three days where the daily AI for the CNTRL and the ASSIM simulation were almost similar. Much like the domain 2 results, the hourly percentage change in the AI for domain 3 between the two simulations fluctuated from day to day. The maximum hourly percentage increase was found to be 26.07%, while the minimum hourly percentage increase was found to be -9.13%. The negative percentage change indicates that the control simulation performed better during those hours. It should be noted, however, that for domain 3, the degradation in the hourly AI performance was not only during times that the initial CNTRL AI was high, unlike what was found for domain 2. This can likely be attributed to one of two factors. The first is that the advection of the clouds between 4-km grid cells in the model is more likely to occur within an hour which is the time scale of satellite observations used here. Thus, a cloud created in the correct place might move to a surrounding grid quicker than what observation can

explain. The second is that the correction scheme can correct a grid but at the same time change the AI in the neighboring grids.

Figure 20 shows the average hourly AI for the 4-km domain during the simulation period. While the assimilation technique performs better than control for all hours of the day, the pattern shows a larger reduction in AI as the sun moves overhead. As for the 12-km domain, the largest AI's are achieved in mid-morning and the latter part of the day.

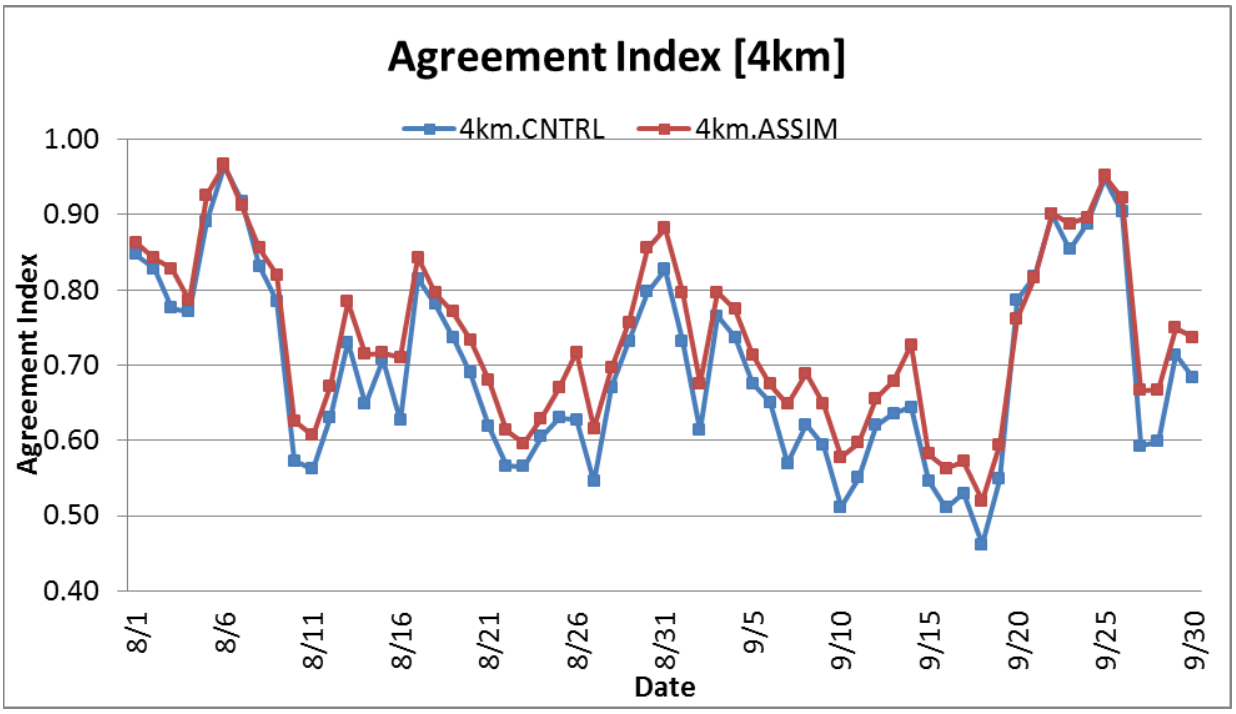


Figure 19. Daily agreement index for CNTRL and ASSIM 4 km WRF simulations over August-September 2013 using a 10% cloud albedo threshold.

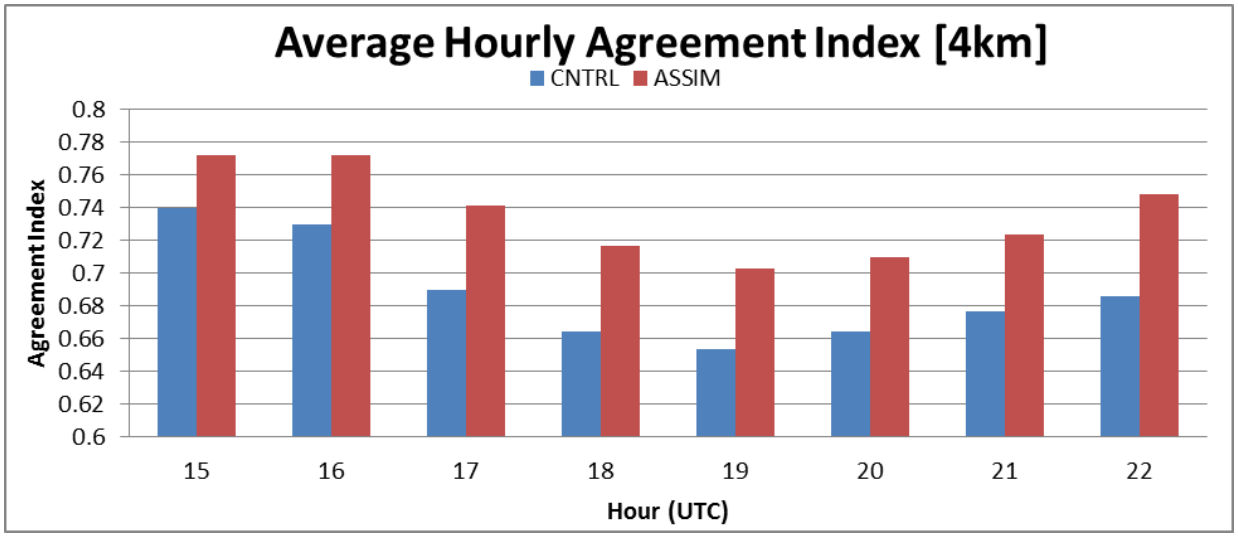


Figure 20. Average hourly AI for 4-km domain during August-September 2013.

The cloud correction scheme also was capable of improving the AI for hours in which the CNTRL simulation had high AI (AI>90%). The capability to correct cloud placement when the original simulation performs both good and bad likely indicates that the assimilation technique becomes more dependent on the atmospheric conditions as the grid resolution is increased and the spatial extent is decreased. As an example, a spatial plot of the AI for September 16, 2013 at 19 UTC is shown in Figure 21. For this hour, it can be seen that the CNTRL simulation had under-predicted clouds across a large portion of Texas but the ASSIM simulation was able to correct a rather large portion of this area, improving the AI by greater than 10%, but unable to correct all the error. These results indicate that the assimilation technique on average improves the cloud placement in space and time, but not as significantly as was seen for domains 1 and 2.

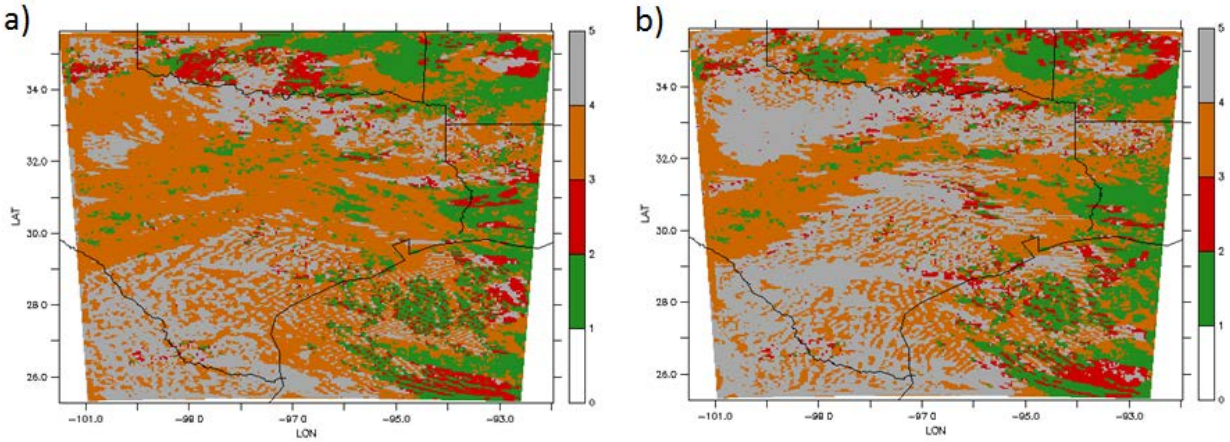


Figure 21. Comparing model cloud to satellite observation for CNTRL (left) and ASSIM (right) simulations for September 16, 2013 at 19 UTC: Green: model and GOES were clear, Red: model cloud over-predictions, Orange: model cloud under-predictions, and Grey: locations where the model and GOES are cloudy. Agreement Index for a) CNTRL (AI=44.5%) b) ASSIM (AI=56.1%).

Evaluating the model performance of the 4-km grid with respect to surface observations, the ASSIM simulation still performs better with respect to temperature but the results for wind speed and mixing ratio (not shown here) are highly variable. The wind speed bias is less in the ASSIM simulation throughout the simulation period but the error fluctuates as was seen in domain 2. Cloud assimilation reduced the RMSE of the temperature, which is likely due to the correction of clouds within the model. For mixing ratio, the ASSIM simulation tends to reduce the model bias and error for some days but increase it for others. Figure 22 shows the wind and temperature bias results.

Overall, the results of the 4 km grid indicate that the assimilation technique still led to better cloud agreement. In doing so, the ASSIM simulation once again improved the model performance for surface temperature. As the grid size was reduced, we also saw some further improvement in the wind speed statistics of the ASSIM simulation. However, there are still times where the error in the wind statistics is less for the CNTRL simulation. With the further

reduction in the spatial coverage of the domain, it is likely that analyzing more spatial patterns would allow distinguishing the physical processes responsible for the differences in the surface statistics. It is also apparent that there are distinct periods of time where one simulation out performs the other. This indicates that more analysis is needed to determine the cause.

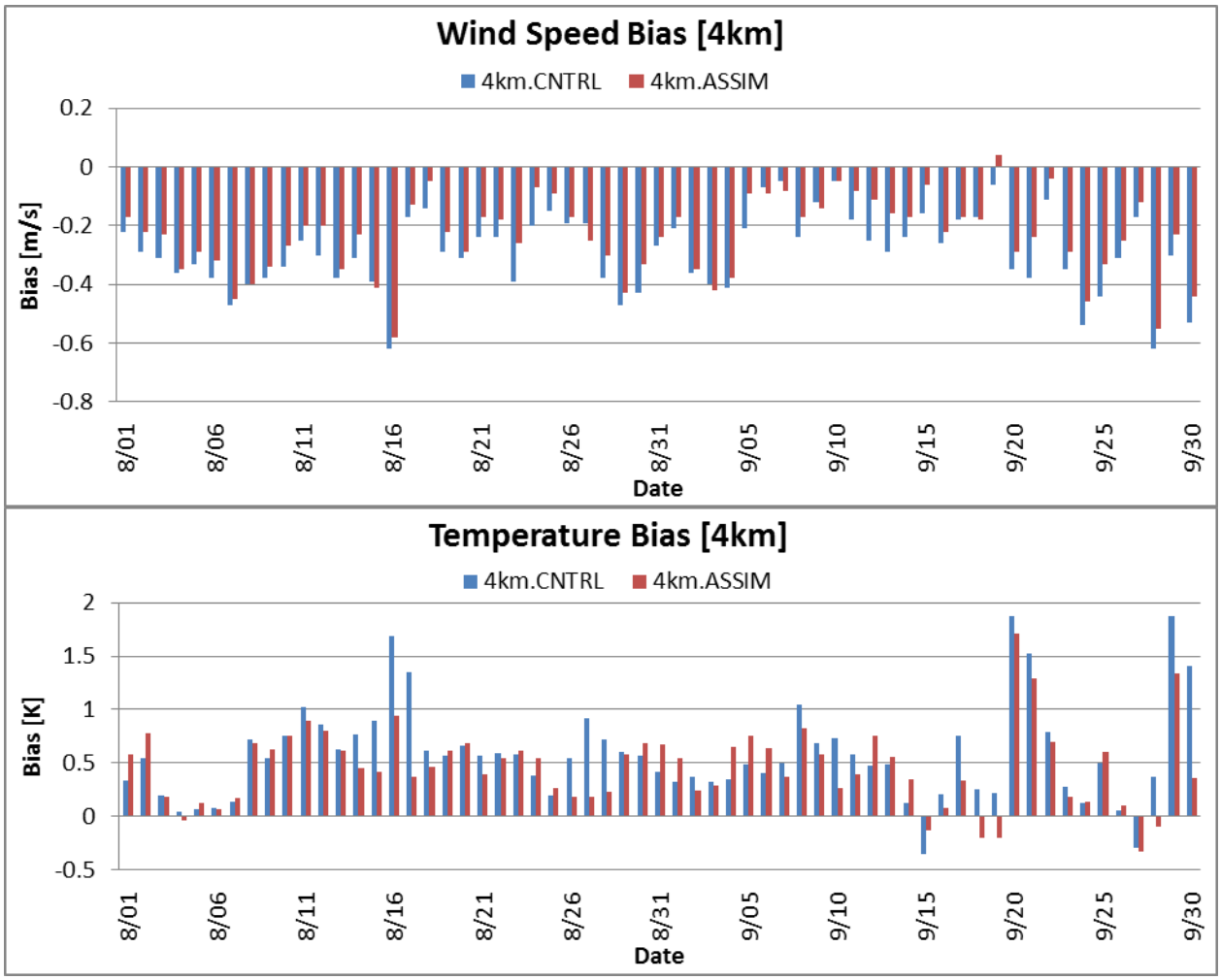


Figure 22. Wind speed and temperature bias results for CNTRL and ASSIM for 4-km domain.

5. BVOC AND SOIL NOX EMISSION ESTIMATES IN MEGAN

TCEQ has historically simulated biogenic emissions with the Global Biosphere Emissions and Interactions System (GLOBEIS), a model based on the BEIS (Biogenic Emission Inventory System) family of models developed by Alex Guenther and others in the 1980s and 1990s. TCEQ is now considering MEGAN (Model of Emissions of Gases and Aerosols from Nature, Guenther et al., 2006, 2012). This model applies more sophisticated algorithms to represent the responsiveness of biogenic emissions to changing conditions and which continues to be updated in ongoing model development efforts by Guenther and others.

In this project MEGAN was applied with different WRF simulations from the episodes of interest in 2006 and 2013 and with 8-day MODIS-based leaf area index (LAI) data to replicate the baseline approach used by TCEQ. MEGAN was also applied with photosynthetically active radiation (PAR) data from two satellite-based retrievals: the product developed by Dr. Rachel Pinker of University of Maryland, and the new PAR data created by University of Alabama-Huntsville.

Also in this project, the new Berkeley-Dalhousie soil NO parameterization scheme (BDSNP) was implemented into CMAQ. A stand-alone version of BDSNP was also constructed. With the offline capability, the influence of wet and dry N depositions to nitrogen reservoir in BDSNP module were estimated by using archived CMAQ or CAMx simulation results instead of by online calculation along with the time consuming atmospheric chemistry module. We will also explore the influence of using temperature data from the satellite retrievals to influence the soil NOx emission rates. All of the MEGAN estimates of biogenic VOCs and soil NOx will be converted into a CAMx ready format and substituted for the biogenic emissions component of the base CAMx runs.

5.1. Initial BVOC and Soil NOx Emission Estimates in MEGAN

The WRF-MEGAN modeling framework was implemented to quantify the sensitivity of BVOC emission estimates to different PAR inputs (WRF versus satellite-based). The September 2013 period on the 12km CONUS domain was used for the initial testing. Details of the WRF and MEGAN configurations are provided in Table 7.

Table 7. Configuration of WRF-MEGAN used in the initial study.

WRF			
Version:	ARW[1] V3.6.1	**Shortwave radiation:**	RRTMG[2] scheme
Horizontal resolution:	D1 (CONUS, 36km); D2 (SW US, 12km)	**Surface layer physic:**	Pleim-Xiu surface model
	D3 (E Texas, 4km)	**PBL scheme:**	ACM2[3]
Vertical resolution:	42 layer (first layer height ~ 37 m)	**Microphysics:**	Morrison double-moment scheme
Boundary Condition:	NARR[4] 32km	**Cumulus Parameterization:**	Kain-Fritsch scheme
Initial condition:	NCEP-ADP[5]	**Analysis nudging:**	NCEP-ADP[5]
Longwave radiation:	RRTMG[2] scheme		Temp., wind, moisture above boundary layer
MEGAN			
Version:	V2.10	**Emission factor:**	Global emission factor (ver. 2011)
Horizontal resolution:	Same as WRF	**Leaf area index:**	30 sec, MODIS[6] 8 day average
Plant functional type:	16 CLM[7] PFT[8] types, 30 sec	**Gas-phase mechanism:**	CB6[9]

Note: 1, ARW: Advanced Research WRF; 2. RRTMG: Rapid Radiative Transfer Model for GCMs; 3. ACM2: Asymmetrical Convective Model version 2; 4. NARR: North American Regional Reanalysis; 5. NCEP-ADP: National Centers for Environmental Prediction-Automated Data Processing; 6. MODIS: Moderate Resolution Imaging Spectroradiometer; 7. CLM: Community Land Model; 8. PFT: Plant Functional Type; 9. CB6: Carbon Bond Mechanism version 6.

By default, the MEGAN model scales the insolation data from WRF uniformly by half (CF=0.5) to represent the PAR value. MEGAN simulations with the satellite retrievals directly used satellite-based PAR estimates, which had already been computed by condition-specific conversion factors. Data from the UAH retrievals were aggregated from nine 4km pixels to

represent one PAR value on the 12km grid used in the MEGAN simulations. MEGAN BVOC emission outputs for isoprene (ISOP) and monoterpenes (TERP) with the two types of radiation input were compared. Figure 23 shows the spatial pattern of emissions and the percent difference between the two estimates. The base cases with WRF inputs are shown on the left, and the percent changes caused by satellite-based PAR are shown on the right.

In terms of magnitude, the estimated ISOP emission rate is much larger than TERP, with hotspots appearing at Southeast states with the typical value of 30 mol/s/gridcell, while the corresponding typical value for TERP is only 5 mol/s/gridcell during the evaluation period. However, due to the different plant functional types and different temperature response curve between ISOP and TERP, the geographic distribution of TERP emission is wider than ISOP; in other words, there is less spatial heterogeneity for TERP compared with ISOP. Isoprene emission is more sensitive to PAR inputs with the highest increase region in the Northeast (> 30%) and decrease in the Northwest (> 20%). The relative change for monoterpene emission is modest (-10% to 5%).

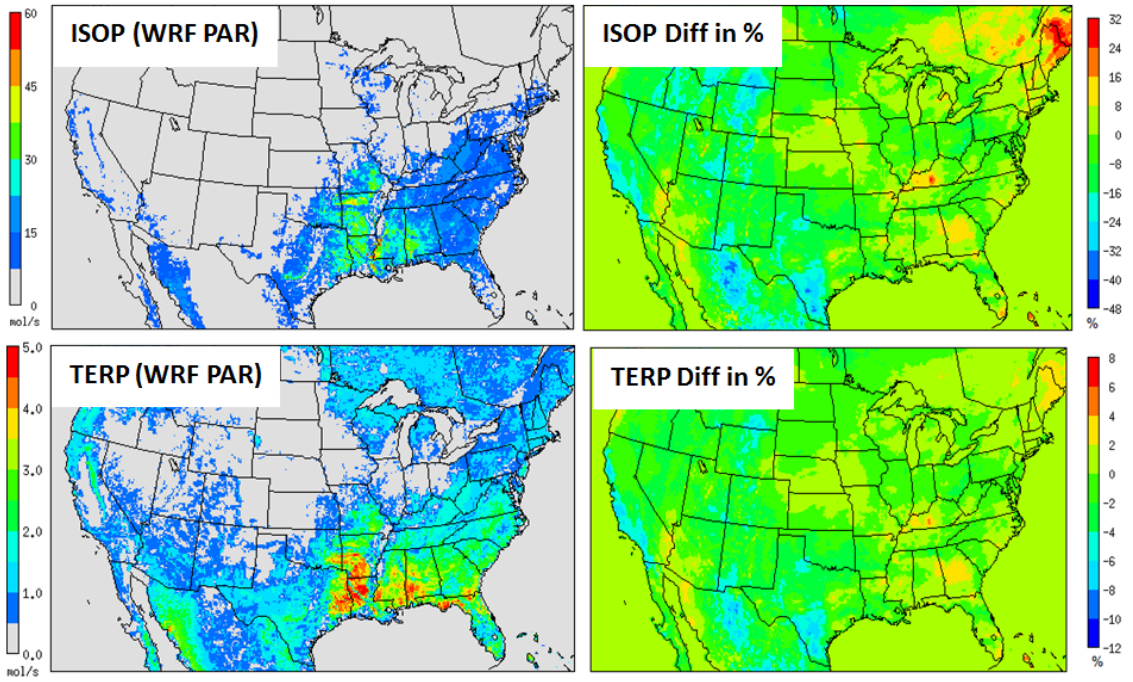


Figure 23. Spatial distribution of estimated ISOP and TERP emission rate by MEGAN using different PAR inputs data (WRF versus UAH satellite retrievals)

The domain-wide sum of ISOP and TERP emission rates based on different climate regions using different PAR inputs is presented in Figure 24. The figure shows that the South and Southeast region is the largest contributor to BVOC emission. Emission rate estimates using satellite PAR data is projected to increase in the Northeast by 4%, Southeast by 1% but decrease in the Northwest by 7%, West in 7%, and South region in 8% for both isoprene and monoterpene.

Also, in the initial testing, the Berkeley Dalhousie soil NOx parameterization scheme (BDSNP) was implemented in MEGAN as an alternate option to the old Yienger-Levy 1995 parameterization (YL95). The implementation in CMAQ better represented the soil NOx response to nitrogen deposition, fertilizer application and changing meteorology. Figure 25 demonstrates the daily mean difference of soil NO emission estimates by using the BDSNP scheme at one test case in July 2011 at 12 km CONUS domain. It can be seen that for agricultural areas in United States, the soil NO emission is projected to increase by more than 5 mg/s/grid with the highest increase appearing at Kansas and Northern Texas.

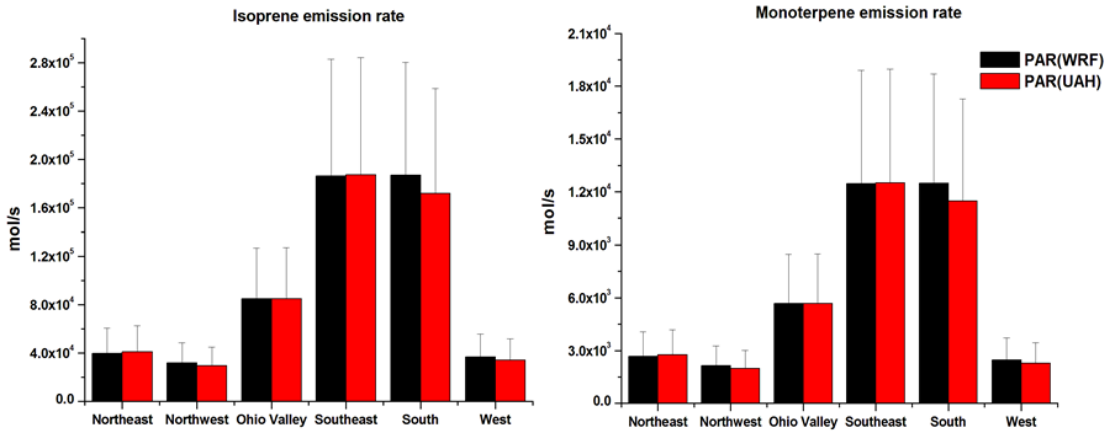


Figure 24. Total ISOP (left) and TERP (right) emission rate estimates by MEGAN using different PAR inputs data (WRF versus UAH satellite retrievals) for each climate region in United States.

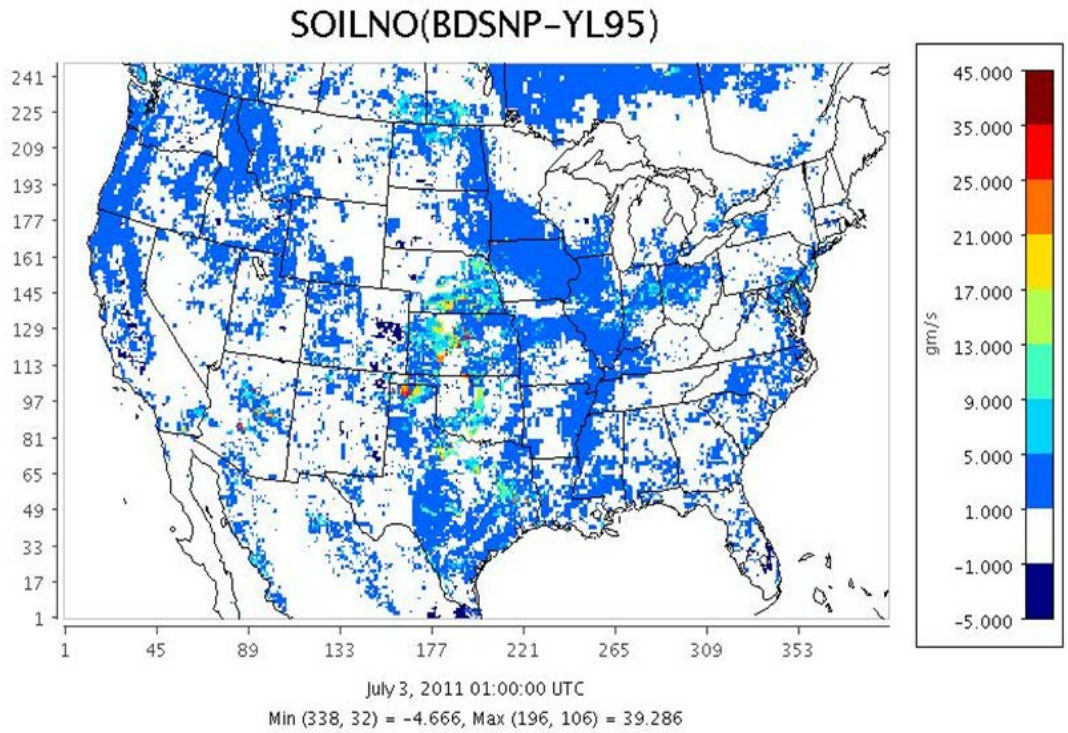


Figure 25. Demonstration of spatial different of soil NO emission rate estimates using BDSNP or YL95 scheme.

Since these initial results were satisfactory, the project proceeded to perform a more detailed study for August-September 2013.

5.2. MEGAN Simulations with Satellite-Based PAR during August-September 2013

Three sets of MEGAN runs were carried out to quantify the impact of different PAR inputs on biogenic emission estimates over the TCEQ SIP domains during September 2013. These inputs were: 1) satellite-based PAR retrievals from GOES imager, 2) control WRF simulation (cntrl) with basic WRF configuration, and 3) WRF with cloud assimilation (analytical) that assimilated clouds from GOES observations. The details of the WRF-MEGAN model configurations, simulation case arrangement, and the simulation time period selection are given in Table 8. The modeling domains are shown in Figure 8.

Table 8. MEGAN configuration used in the 2013 study.

MEGAN			
Version:	V2.10	**Emission factor:**	Global emission factor (ver. 2011)
Horizontal resolution:	Same as WRF	**Leaf area index:**	30 sec, MODIS[1] 8 day average
Plant functional type:	16 CLM[2] PFT[3] types, 30 sec	**Gas-phase mechanism:**	CB-05[4]
Simulation Case Arrangement			
1. PAR_cntrl:	Base WRF simulation to provide insolation for MEGAN		
2. PAR_analytical:	Base WRF + cloud assimilation from GOES[5] to provide insolation for MEGAN		
4. PAR_UAH:	Direct use PAR retrievals from UAH, other met inputs same as case 'PAR_analytical'		
Simulation Time Period			
	Sep 1-30, 2013		

Note: 1. MODIS: Moderate Resolution Imaging Spectroradiometer; 2. CLM: Community Land Model; 3. PFT: Plant Functional Type; 4. CB-05: Carbon Bond Mechanism version 5; 5. GOES: Geostationary Operational Environmental Satellites.

The raw UAH 4km CONUS PAR retrieval products were mapped to the three TCEQ SIP simulation domains (36km for CONUS, 12km for Texas and 4km for east Texas) using the revised utility codes based on the UNC Spatial Allocator. The regridded PAR products directly replaced the calculated PAR. In default mode, MEGAN assumes PAR to be half of the solar radiation reaching the surface (RGRND).

The biogenic VOC emissions from MEGAN were lumped according to CB05 chemical mechanism and were archived in the NetCDF format. (NOTE: The total disk storage of the two months MEGAN runs is 2.9 GB for D1, 5.4 GB for D2 and 16.8 GB for D3. It is ready to share with the CAMx Fortran binary input format using the CMAQ2CAMx interface program provided by Ramboll-Environ, http://www.camx.com/getmedia/a9e648b7-2b2d-487d-9243-2f363a6feea4/cmaq2camx-4sep13.tgz.aspx).

Since the focus of this study is to provide useful information for TCEQ SIP modeling, only MEGAN results from domain 2 (the Texas domain, D2) is presented here. Figures 26 and 27 show the comparison of the spatial patterns of the monthly mean isoprene (ISOP) and monoterpene (TERP) emission rates for the three different PAR inputs during August and September 2013.

For the ISOP simulations during August 2013, the general emission pattern for the three PAR inputs case is quite similar. Hot spots over the Texas territory mainly concentrated over the Edwards Plateau and the eastern Texas boundary adjacent with the Louisiana and Arkansas, where the broadleaf evergreen tree or shrub is the dominate plant functional type.

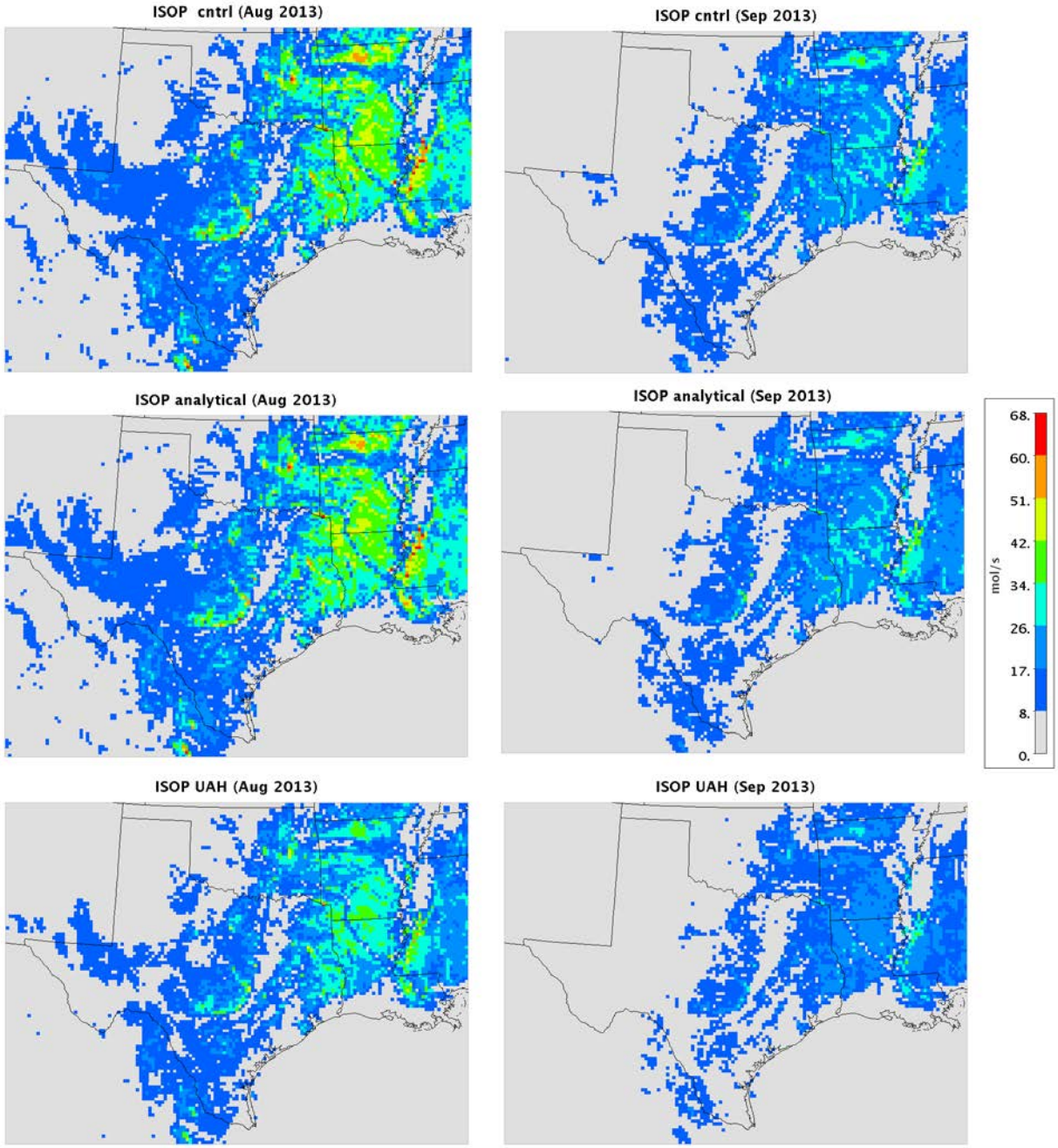


Figure 26. Comparison of the spatial patterns of the monthly mean isoprene (ISOP) emission rate using different PAR inputs for WRF control case (cntrl), WRF cloud assimilation case (analytical) and PAR satellite retrievals (PAR) in MEGAN over Texas domain during August (left) and September (right) 2013.

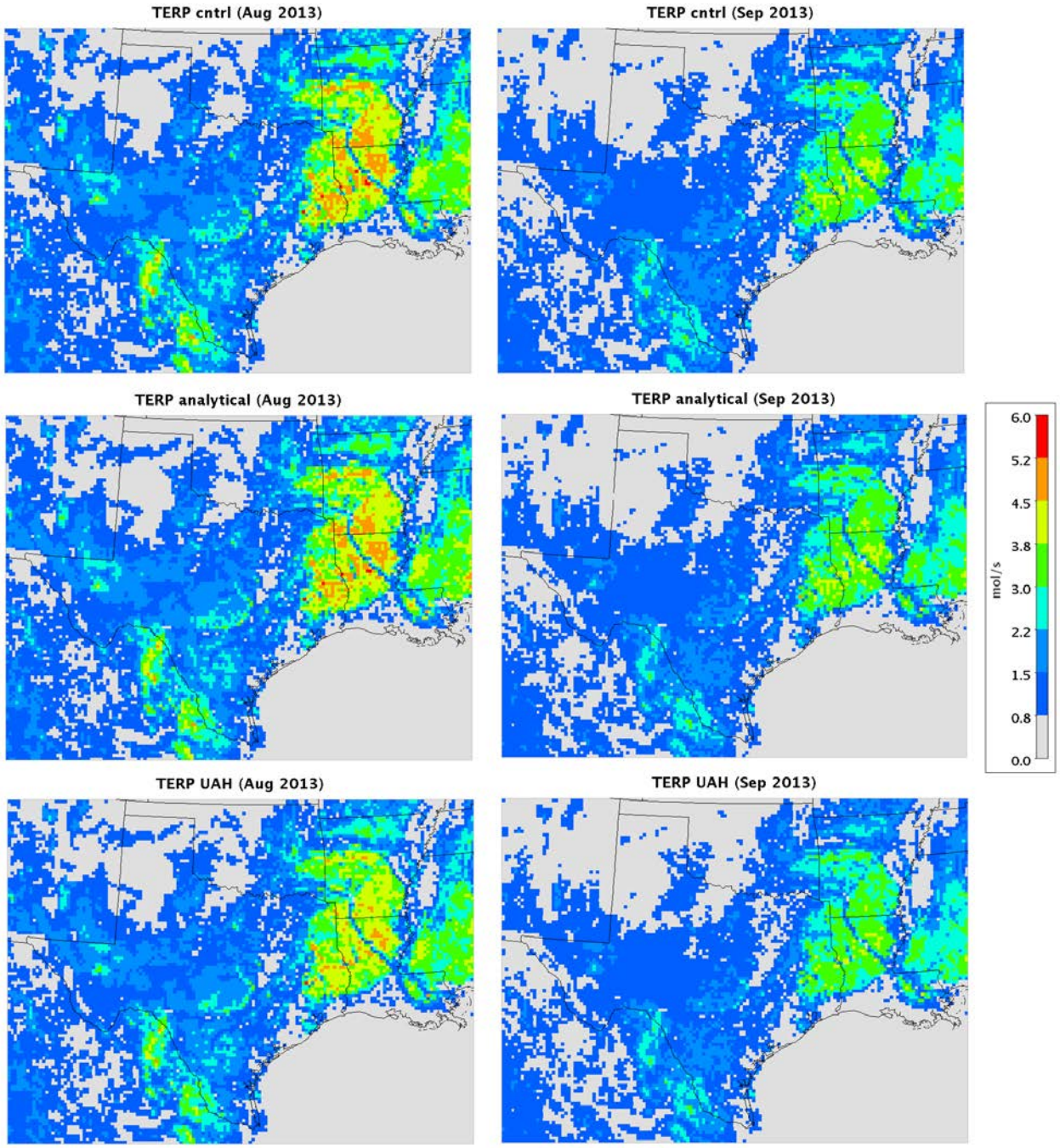


Figure 27. Comparison of the spatial patterns of the monthly mean monoterpene (TERP) emission rate using different PAR inputs for WRF control case (cntrl), WRF cloud assimilation case (analytical) and PAR satellite retrievals (PAR) in MEGAN over Texas domain during August (left) and September (right) 2013.

In terms of the magnitude, the 'UAH' case is the lowest with the maximum value 54 moles/s/grid, followed by the 'analytical' and 'cntrl' case. For September, the base ISOP emission is lower than that of in August 2013 due to the lower mean surface temperature and

smaller leaf area index value input from MODIS. The pattern for TERP during the two months in 2013 shows additional hot spots near the south Texas boundary adjacent to Mexico. The overall magnitude of mean TERP emission rate is much smaller than for ISOP, with TERP ranging 0-6 moles/s/grid and ISOP 0-68 moles/s/grid.

In order to characterize BVOC emission pattern from different MEGAN simulations over the heterogeneous plant functional type over Texas, the average monthly emission rates over the 10 climate divisions in Texas were calculated separately. The climate classification is based on historical climate analyses (1895-2013) for the monitored drought, temperature, precipitation and heating/cooling degree day values over the continental US (http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-divisions.php). The geographic locations of these 10 climate divisions in Texas are illustrated in Figure 28. An area mask file consistent with the TCEQ domain configurations were generated based on the climate division boundary polygon shapefiles provided by NCAR (http://www.ncl.ucar.edu/Applications/Data/cdf/climdiv_polygons.nc).

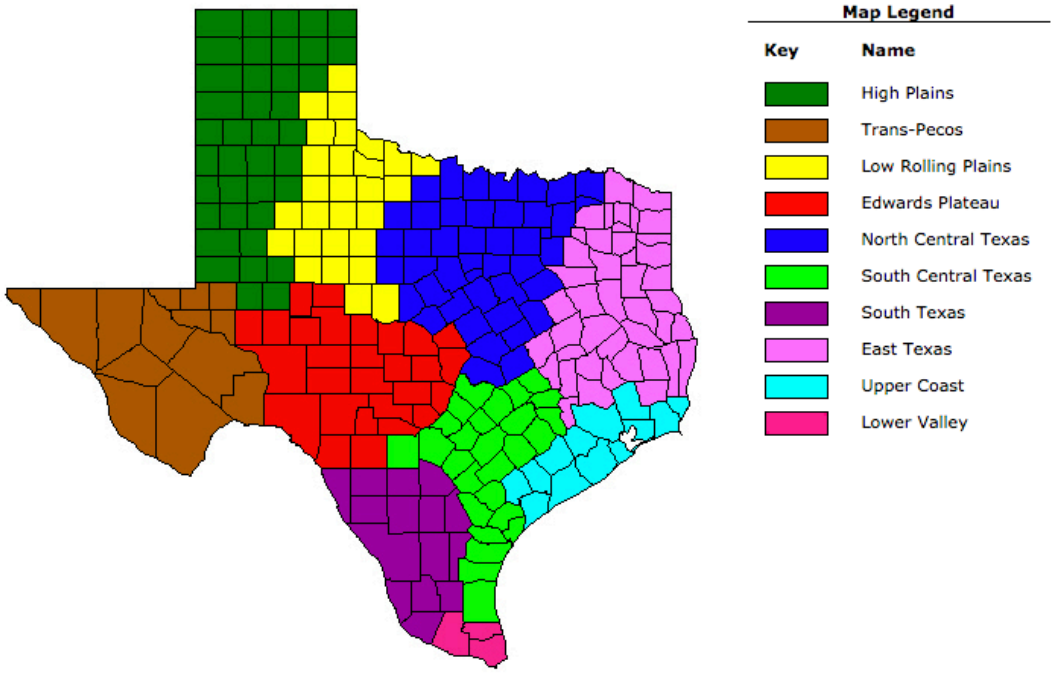


Figure 28. Geographic distribution of the 10 climate divisions in Texas by the National Weather Service (adapted from http://www.nass.usda.gov/Statistics_by_State/Texas/Charts_&_Maps/cwmap.htm)

The ISOP and TERP results are given as the histogram comparison plots in Figure 29 and detailed statistics in Table 9.

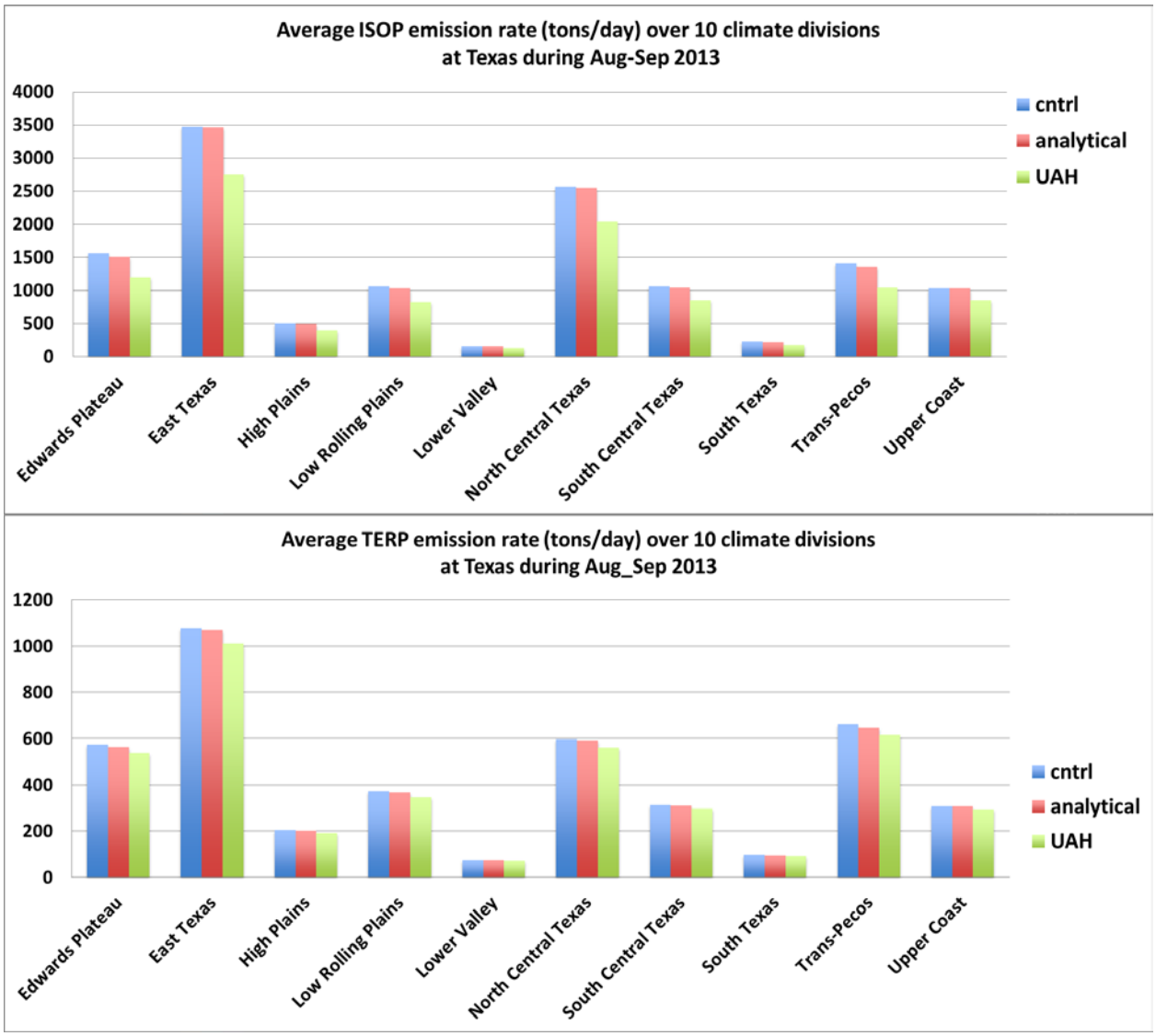


Figure 29. Comparsion of average isoprene (ISOP, top) and monoterpene (TERP, bottom) emission rate (tons/day) over the 10 climate divisions of Texas during August and September 2013 by MEGAN using different PAR inputs.

For ISOP, the top 3 highest emission regions in Texas are East Texas (2754 tons/day for case 'UAH'), North Central Texas (2036 tons/day for case 'UAH'), and Edwards Plateau (1199 tons/day for case 'UAH') separately. For TERP, the top 3 highest emission regions are East Texas (1011 tons/day for case 'UAH'), Trans-Pecos (615 tons/day for case 'UAH'), and North Central Texas (562 tons/day for case 'UAH').

Table 9. Comparison of daily average isoprene (ISOP) and monoterpene (TERP) emission rate (tons/day) over 10 climate zones of Texas from MEGAN using different PAR inputs.

Aug 2013										
			ISOP					TERP		
climate zone	cntrl	analytical	PAR	relative diff. (analytical-cntrl)	relative diff. (PAR-cntrl)	cntrl	analytical	PAR	relative diff. (analytical-cntrl)	relative diff. (PAR-cntrl)
	(tons/day)	(tons/day)	(tons/day)	%	%	(tons/day)	(tons/day)	(tons/day)	%	%
Edwards Plateau	1794.6	1736.4	1374.0	-3.2	-24.2	645.8	640.2	600.9	-0.9	-7.0
East Texas	3876.0	3837.2	3062.0	-1.0	-21.2	1207.4	1196.2	1128.8	-0.9	-6.5
High Plains	600.8	596.9	451.6	-0.6	-25.0	239.2	238.7	222.4	-0.2	-7.0
Low Rolling Plains	1217.1	1199.6	926.4	-1.4	-24.2	423.4	420.6	393.1	-0.7	-7.2
Lower Valley	173.3	171.3	137.8	-1.1	-20.7	80.9	80.3	76.4	-0.7	-5.6
North Central Texas	2887.6	2829.5	2248.1	-2.0	-22.6	673.9	662.7	629.0	-1.7	-6.7
South Central Texas	1203.5	1176.4	955.4	-2.3	-21.1	355.5	351.6	335.3	-1.1	-5.7
South Texas	257.8	248.1	197.7	-3.8	-24.2	110.1	107.3	102.2	-2.6	-7.1
Trans-Pecos	1655.1	1614.4	1189.9	-2.5	-28.8	769.4	758.2	707.6	-1.5	-8.0
Upper Coast	1135.7	1118.2	922.5	-1.5	-19.1	343.7	340.3	326.9	-1.0	-4.9
Sep 2013										
			ISOP					TERP		
climate zone	cntrl	analytical	PAR	relative diff. (analytical-cntrl)	relative diff. (PAR-cntrl)	cntrl	analytical	PAR	relative diff. (analytical-cntrl)	relative diff. (PAR-cntrl)
	(tons/day)	(tons/day)	(tons/day)	%	%	(tons/day)	(tons/day)	(tons/day)	%	%
East Texas	3060.0	3084.5	2445.6	0.8	-19.9	943.5	943.5	892.9	0.0	-5.4
High Plains	411.3	396.6	325.6	-3.6	-21.6	169.6	166.2	161.2	-2.0	-5.0
Low Rolling Plains	905.8	869.0	702.6	-4.1	-23.4	320.1	313.4	302.1	-2.1	-5.6
Lower Valley	137.6	139.0	111.9	1.1	-18.5	68.0	68.5	65.1	0.8	-4.1
North Central Texas	2247.3	2257.1	1823.8	0.4	-18.8	521.2	519.5	495.3	-0.3	-5.0
South Central Texas	908.2	896.0	744.2	-1.3	-18.3	271.3	268.4	259.5	-1.0	-4.3
South Texas	192.4	183.4	149.8	-4.7	-23.2	86.5	84.8	82.0	-1.9	-5.2
Trans-Pecos	1162.8	1089.4	893.5	-6.3	-24.7	553.2	536.9	522.8	-2.9	-5.5
Upper Coast	918.0	942.5	773.6	2.7	-15.3	272.9	274.6	263.4	0.6	-3.5

The results indicate that the case 'UAH' using GOES satellite retrievals on average predict 21% less ISOP than the base WRF case ('cntrl') during August 2013 and 19% less during September 2013. The cloud assimilation WRF case ('analytical') predicts slightly less ISOP than case 'cntrl' with the mean value around -2% during August 2013 and -3% during September 2013. The impact of more realistic insolation data from satellite on TERP is modest and on average is about 5% reduction. The TERP emission algorithm in MEGAN is more directly connected with the surface temperature instead of PAR. At least for the evaluated two months in 2013, the most sensitive climate region for ISOP emission estimation in Texas due to different PAR inputs is Trans-Pecos, with the relative difference compared to base case -28.8% during August 2013 and -24.7% during September 2013.

Figure 30 plots the time series comparison of the daily mean ISOP and TERP emission rate at the highest BVOC emission climate division (East Texas) in Texas during the two months simulation period. It can be seen that both ISOP and TERP emission experienced the decreasing trend with the highest emission rate appearing at the first week of August (~ 5500 tons/day for ISOP and ~ 1300 tons/day for TERP for case 'UAH') and the lowest emission rate appearing at the third week of September (~ 200 tons/day for ISOP and ~600 tons/day for TERP for case 'UAH'). The two lowest emission days during the simulation period, namely August 15 and September 20, correlate well with the regional weather pattern of low surface temperature, cloudy sky and major rain events. For Example, the weather chart and the corresponding satellite-based PAR for September 20 presented in Figure 31 indicate the presence of clouds and rain in east Texas causing the sharp drop in PAR. Daily mean PAR for this day is only 13 W/m^2 compared to typical 120 W/m^2 for other days. Satellite-based PAR of less than 18 W/m^2 over Arkansas and East and north central Texas corresponds well with the low-pressure trough and large rain belt (in green area) from ground observations.

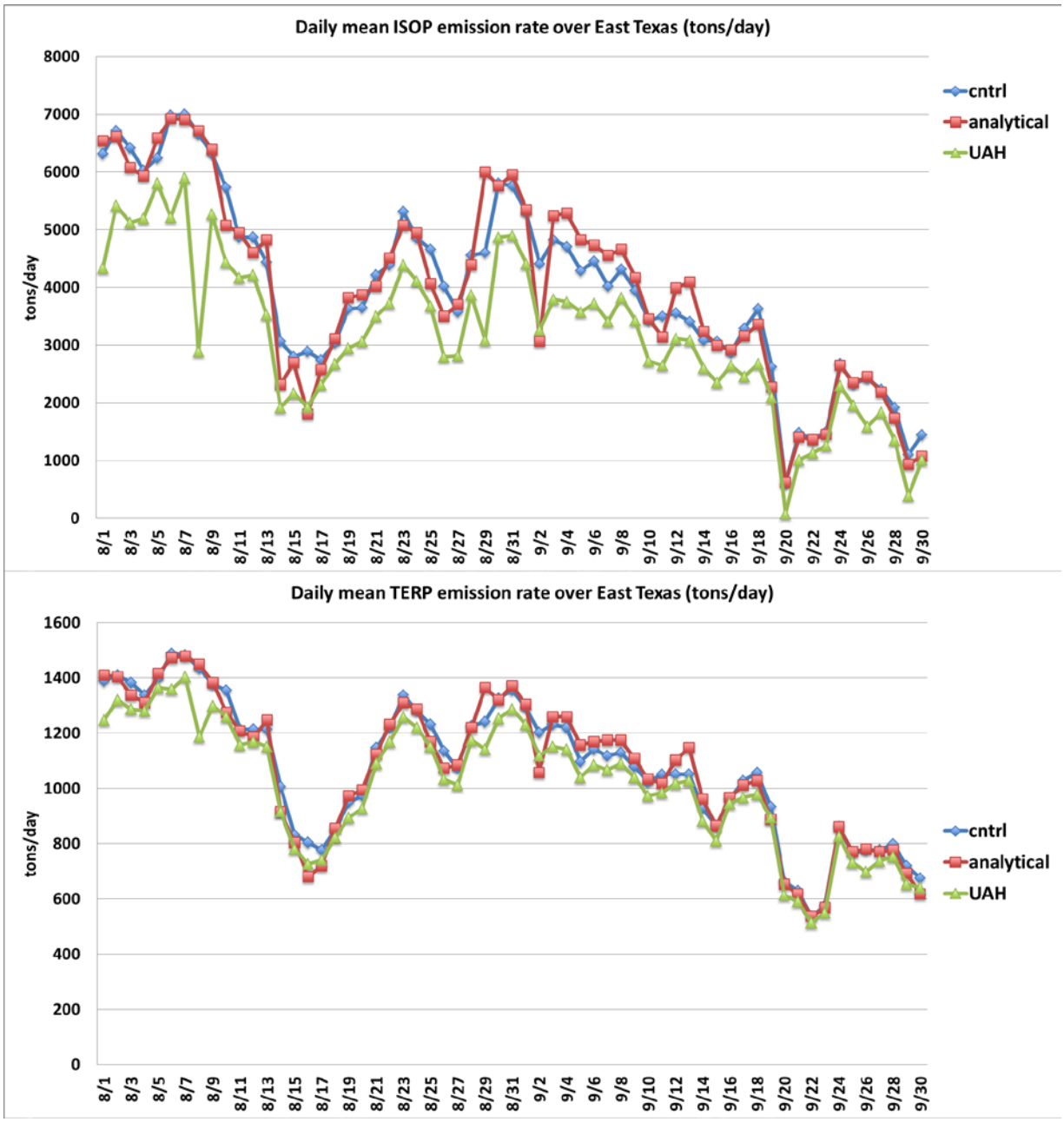


Figure 30. **Comparison of the daily variation of isoprene (ISOP, top) and monoterpene (TERP, bottom) emission rate (tons/day) over East Texas during August and September 2013 by MEGAN using different PAR inputs.**

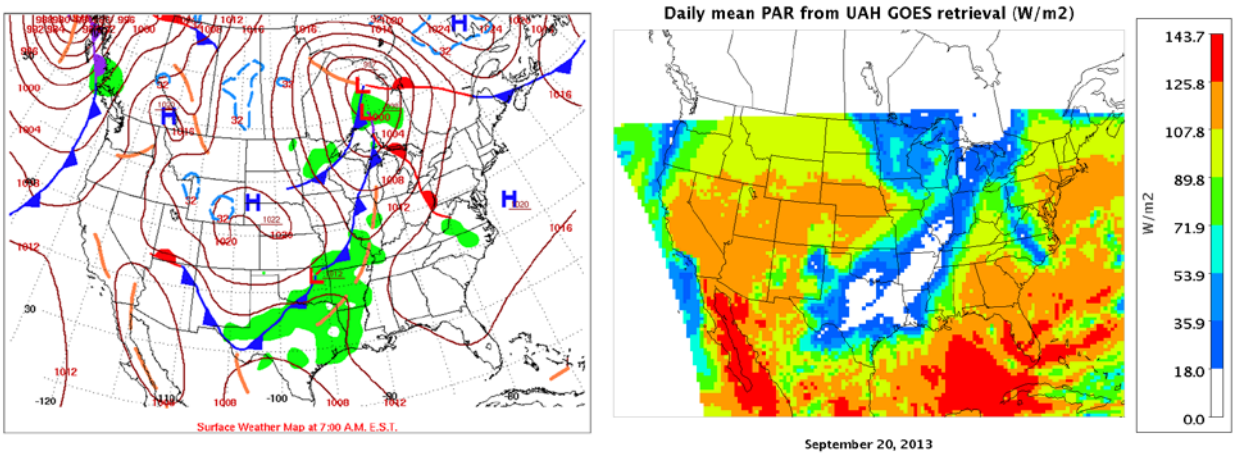


Figure 31. US surface weather map at 7 a.m. E.S.T on September 20, 2013 from NOAA (left); and the daily mean PAR retrieval from UAH on September 20, 2013 (right).

Overall, the good correlation between emission rates based on satellite-derived PAR and PAR calculated from the model field for these two months, indicate that the model, to a large degree, has been able to explain the synoptic scale pattern responsible for the reduction in emissions. One exception is August 8, 2013, when for both 'cntrl' and 'analytical' case, the predicted ISOP emission rate is around 7000 tons/day while the corresponding 'UAH' value is less than half of that. The large contrast is mainly due to a sudden drop of the PAR retrieval value from 140 W/m^2 to 80 W/m^2. Figure 32 shows the variation of temperature, PAR and the 24-hr accumulated precipitation for the entire simulation period. As shown in the figure, for August 8, observed PAR shows a reduction indicating the presence of clouds, but the model does not reflect that in temperature and precipitation fields.

Figure 33 provides the domain-wide sum of daily isoprene and monoterpene emission strength over the Texas for the three PAR inputs. It can be seen that the PAR retrieval case predicted nearly one fourth lower ISOP emission compared with the WRF simulation runs (5463 moles/s for case 'cntrl', 5377 moles/s for case 'analytical' and 3698 moles/s for case 'PAR'). No significant change for TERP emission for the three cases were observed, since the monoterpene emission algorithm in MEGAN is not directly linked with PAR/insolation but more response with the surface temperature.

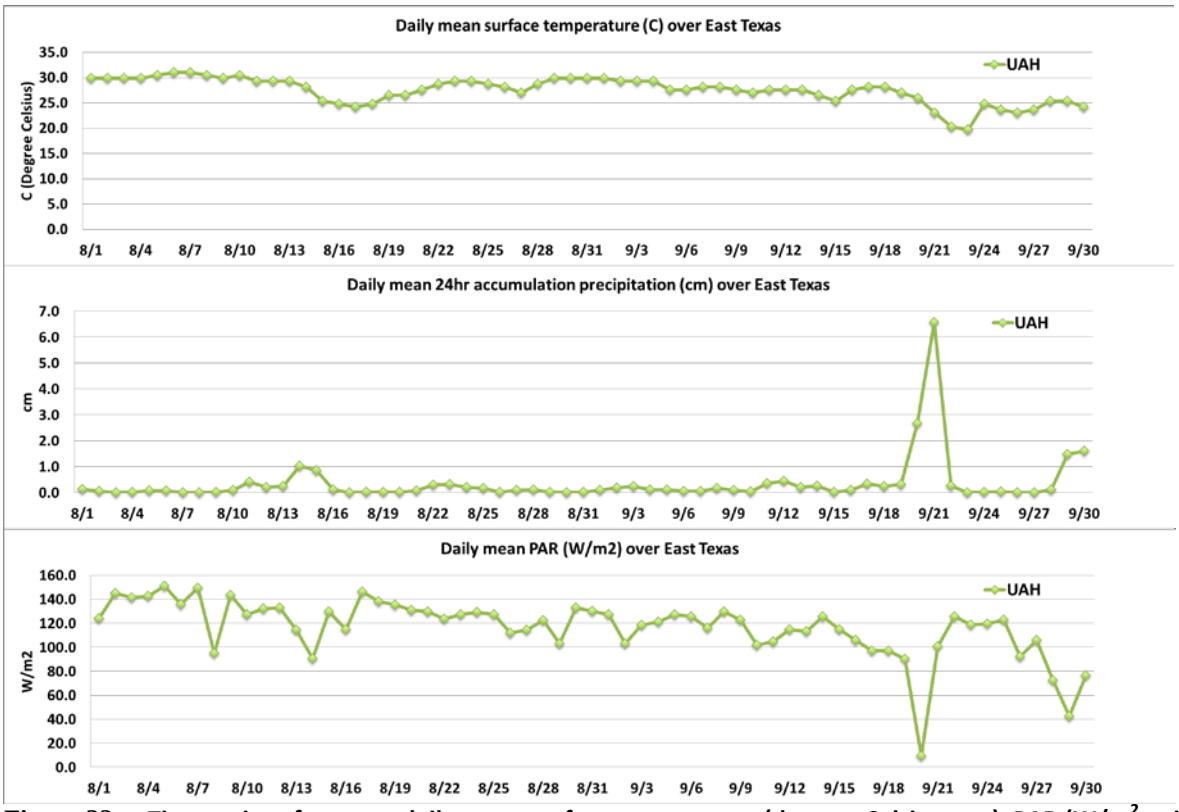


Figure 32. Time series of average daily mean surface temperature (degrees Celsius, top), PAR (W/m^2, middle) and 24hr accumulated precipitation (cm, bottom) over East Texas during August and September 2013 by MEGAN using PAR satellite retrievals from UAH.

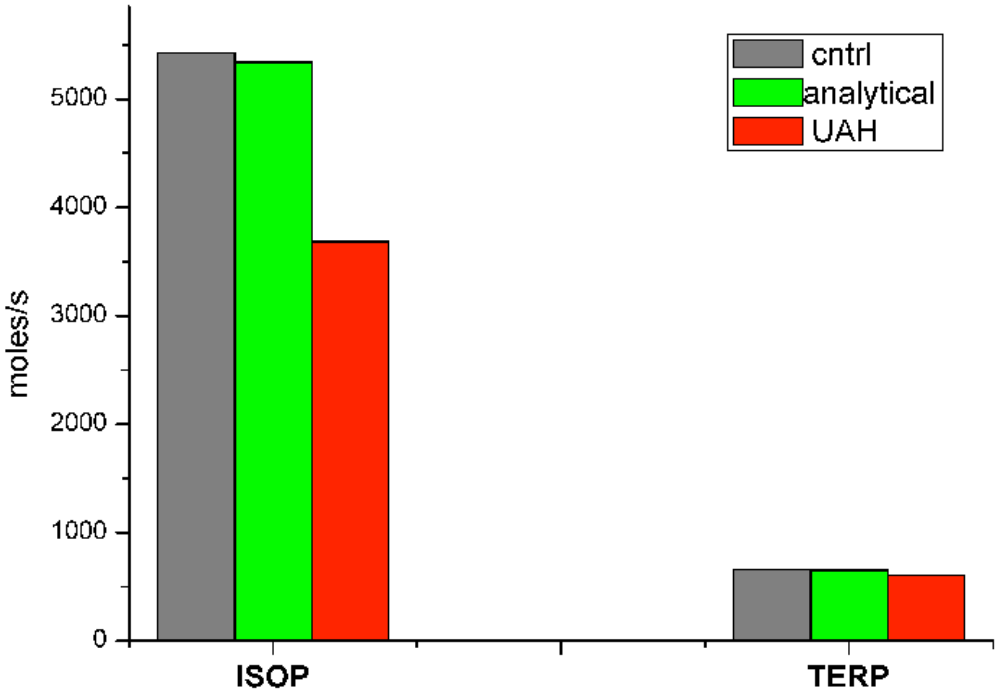


Figure 33. Domain-wise sum of estimated isoprene (ISOP) and monoterpene (TERP) emission strength over Texas area using different PAR inputs in MEGAN during September 2013.

5.3. Developing Offline BDSNP Module for Soil NOx Emission Estimates

In addition to implementing the Berkeley-Dalhousie Soil NOx Parameterization (BDSNP) (Hudman et al., 2012) in CMAQ, a stand-alone version of BDSNP was also developed. Compared to the soil NO emission module used in most air quality models (Yienger and Levy 1995), the BDSNP scheme generates higher estimates of emissions and provides a more physically realistic representation of the dependence of emissions on soil temperature and moisture. BDSNP has previously been implemented into the GEOS-Chem global model (Hudman et al., 2010, 2012), and into CMAQ. However, the inline BDSNP module in CMAQ is computationally expensive and is not cross-platform applicable, so in this project an offline version of the BDSNP module was also developed.

In the BDSNP scheme, soil NO emission estimates at each location are determined based on a biome-specific base emission factor and an available soil nitrogen pool originating from fertilizer application and nitrogen deposition from the atmosphere. Emission rates are modulated based on response functions to soil temperature and soil moisture, a soil pulsing factor when precipitation follows a dry period, and a canopy reduction factor (see Figure 34).

NO Emissions =

– A(Biome, Soil Nitrogen) x
f(T) x g(θ)
Pulse(Dry Period) x
Canopy Reduction

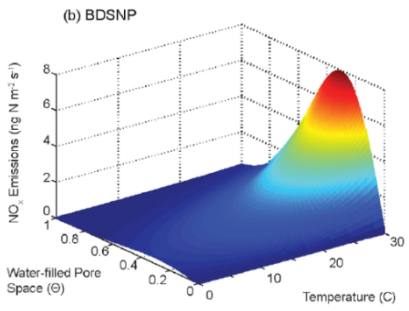


Figure 34. BDSNP soil NOx emission scheme given by Hunman et al. (2012).

From the software engineering point view, Figure 35 provides the flow chart of the BDSNP scheme implementation with the option to run inline or offline with the air quality model. Static input files such as arid/non-arid climate zone, soil biome type (must be consistent with the type of the soil emission factors given by Steinkamp and Lawrence (2011)) and global fertilizer pool from Potter et al. (2010) are needed to determine the soil base emission value at each modeling grid.

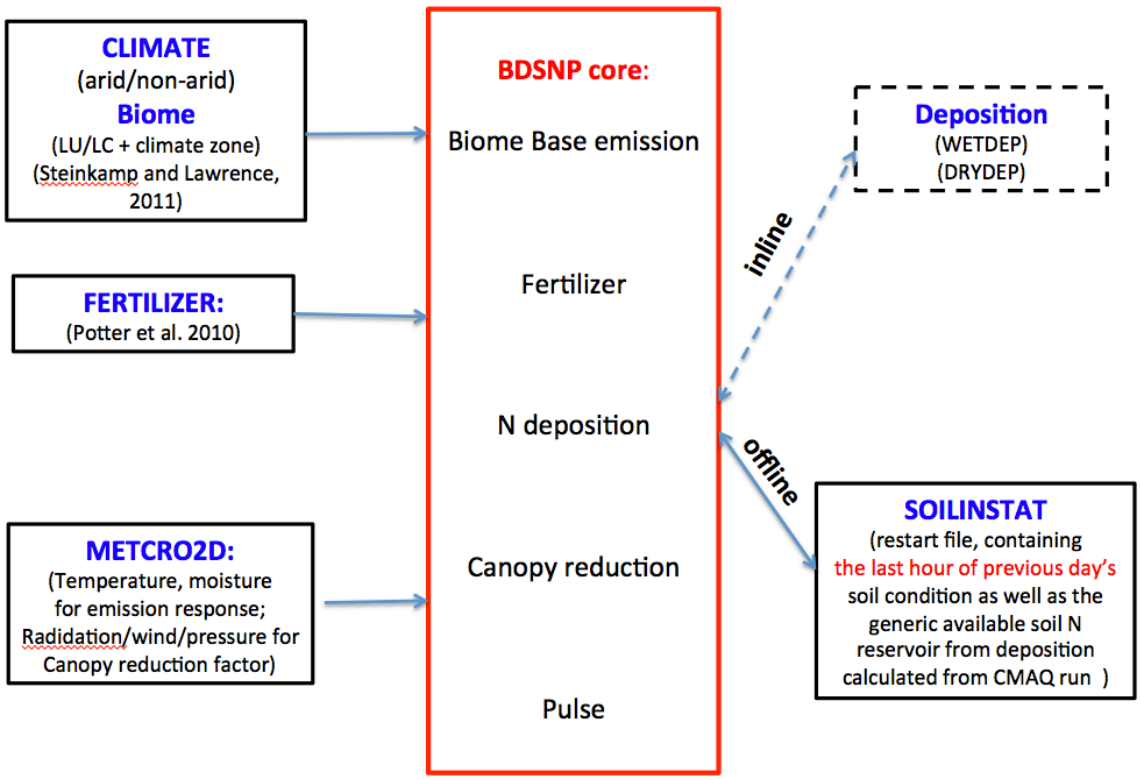


Figure 35. Flow chart of the inline/off line BDSNP soil NO_x emission model.

Soil moisture and temperature, as well as some meteorological variables such as radiation, wind and air pressure are needed (from weather models) to drive the BDSNP to represent the occurrence of pulsing and the process of canopy reduction. The major difference between the inline and offline options in BDSNP model is the approach to deal with the available nitrogen pool from the soil. For the inline BDSNP model (the dashed line), the dry or wet deposition process considered in the air quality model will continuously update the available nitrogen from the atmosphere to the ground; while for the offline BDSNP model (the sold line), only the generic daily nitrogen pool from the deposition process are used. Since the soil nitrogen pool needs a long time to build up and reach the quasi-steady state in the model, a new series of SOILINSTATE files are needed in the offline BDSNP model to provide the generic daily variation of N deposition. A full year CMAQ simulation over the 12km US continental modeling domain in 2005 was performed to record the daily total soil N reservoir for further usage.

As a test, a one week simulation (Jun 26-Jul 3, 2011) for soil NO emission estimates with both inline and offline options of BDSNP was performed. Figure 36 demonstrates the available soil N reservoir from deposition at Jun 28, 2011 (this is needed as one of the input files for the off-line BDSNP model). The "generic" available soil N reservoir from deposition process in this test case is the CMAQ deposition simulation results for the same time period. The N deposition at Jun 28 is concentrated in the central United States and California.

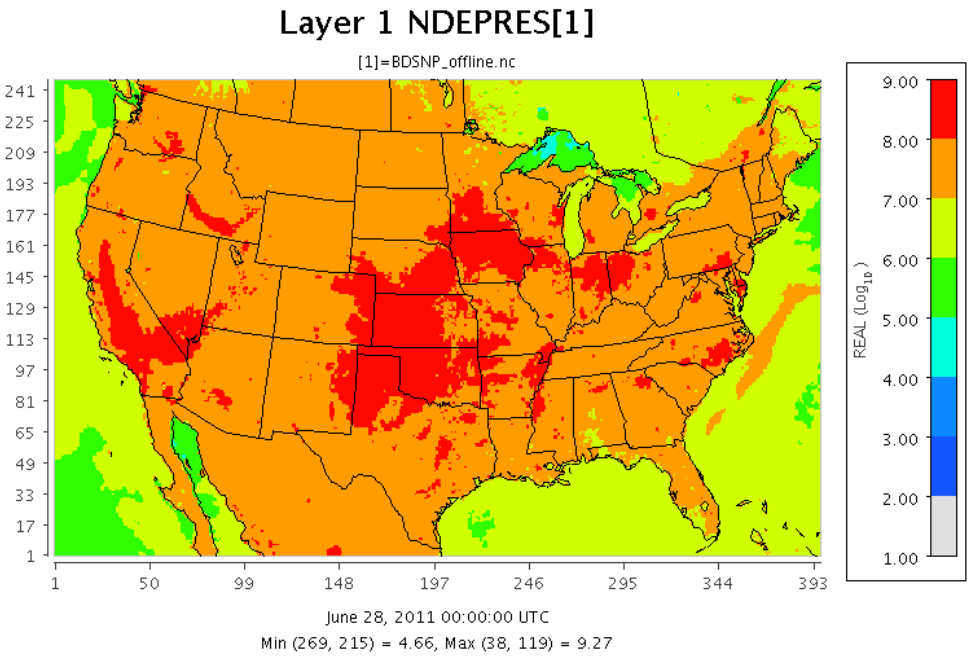


Figure 36. Soil N reservoir from deposition (NDEPRES) used in off-line BDSNP model which is calculated from 2005 CMAQ simulation results.

Figure 37 provides the breakdown of each process in the off-line BDSNP model on the final soil NO emission estimates for Jun 28, 2011. These are the biome base emission (A_DIAG), fertilizer application amount (AFERT_DIAFG), soil nitrogen pool from deposition (NDEPRES), pulse factor (PFACTOR) and the canopy reduction factor (CRFAVG). For each modeling grid, a higher biome base emission factor plus higher fertilizer application and higher available N from deposition will result in higher potential to produce soil NO. Pulsing occurs following a precipitation event, with the strength of the pulsing depending on the length of the antecedent dry period. The scaling factor CRFAVG represents the loss of NO to the plant canopy.

Figure 38 provides the comparison of the soil NOx emission estimates with the inline and offline options. It is obvious that with the nearly identical inputs files, the two options yield a quite similar result in terms of general spatial pattern and peak values.

For periods that the pre-existing CMAQ N deposition fields are not available, full year 2005 CMAQ deposition results can be used as a surrogate. The assumption is that the N deposition pattern in 2005 is comparable with the corresponding simulation date.

In terms of computation time, the offline BDSNP module involves a far smaller burden than the inline option, which requires running the full CMAQ. Table 10 gives the CPU time estimates for the one-week test case run. The CPU time decreases from 282 min (using multiple processor parallel run) for inline BDSNP module to 3 min (single processor) for offline BDSNP module.

This work is further improved generating a new soil biome spatial map based on the finer resolution land use/ land cover definition (NLCD40).

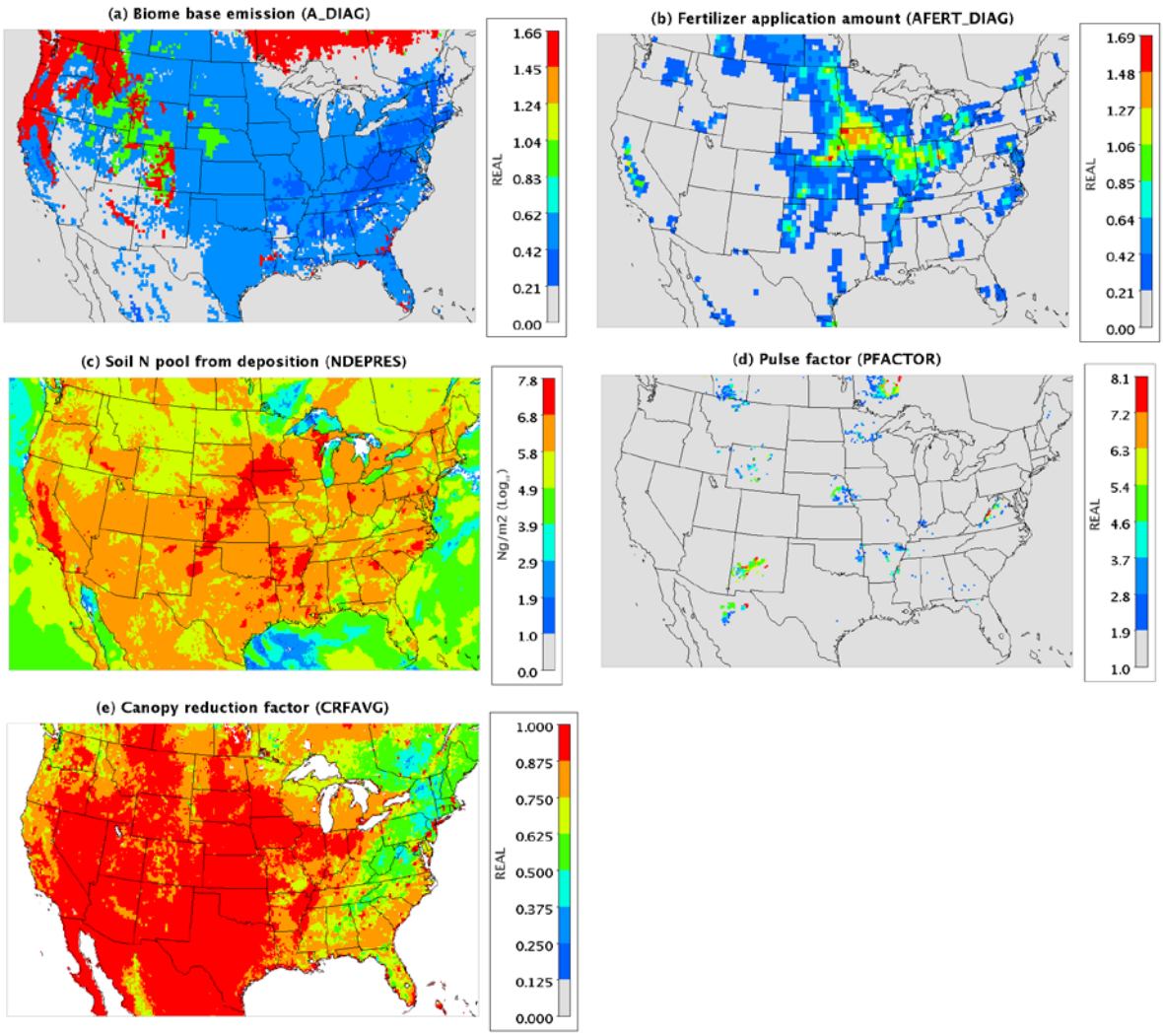


Figure 37. Impact of different BDSNP processes on the Soil NO_x emission estimates.

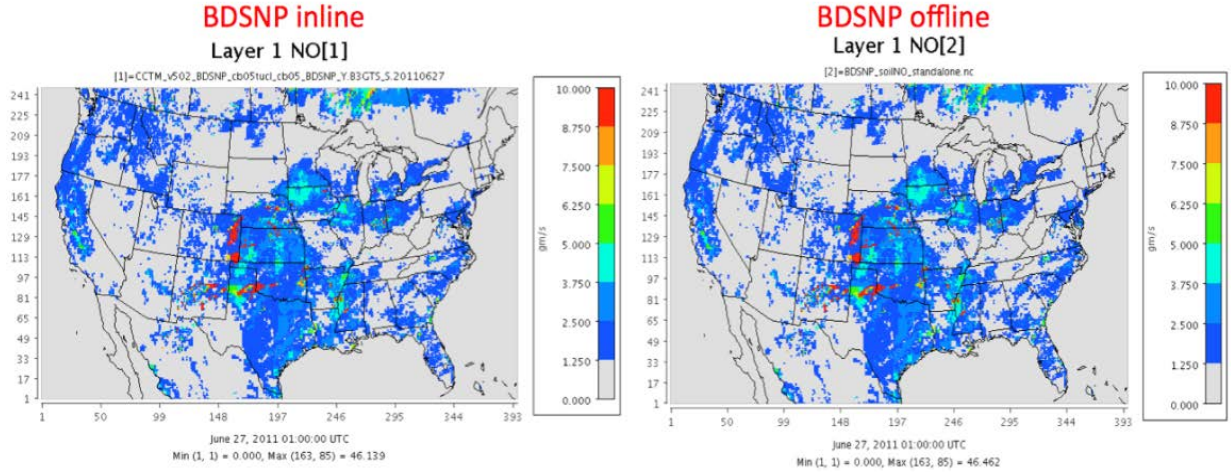


Figure 38. Comparison of soil NOx emission estimates using the inline (left) and offline (right) BDSNP scheme on Jun 28, 2011.

Table 10. Comparison of CPU time usage for inline and offline BDSNP.

	Grid resolution	Total grid	CPU used	CPU speed	Memory	CPU time
inline BDSNP module	12kmX12km CONUS	396X246	144	2.83GHz	48G	283 min
offline BDSNP module	12kmX12km CONUS	396X246	1	2.83GHz	8G	3 min

5.3.1. Developing a New Soil Biome Spatial Map Based on 12km CONUS 40-category 2006 NLCD-MODIS Land Use Classification (NLCD40)

The current offline Berkeley-Dalhousie Soil NOx Parameterization (BDSNP) module developed by Rice University uses the soil biome map directly re-gridded from global atmospheric chemistry model GEOS-Chem, which is too coarse for regional model implementation. The biome map is related to the land use/ land cover (LU/LC) classification and Köppen-Geiger climate zone definition (Kottek et al., 2006) and will determine the base soil NOx emission strength. In this project, a new soil biome spatial map based on 12km CONUS 40-category 2006 NLCD-MODIS land use classification (NLCD40) and climate zone definition was developed. This new biome map replaced the GEOS-Chem biome map to better represent the up-to-date LU/LC change with finer details.

The first step in this process was the construction of a mapping table to transfer the 40 categories of NLCD40 at each modeling grid to the 24 biome types for which soil NO emission factors are available from Steinkamp and Lawrence (2011). Table 11 details how the NLCD40 has been mapped into the 24 biome type for soil NO emission estimates. For the categories that have identical names in both categories, , such as 'evergreen needle leaf forest', 'deciduous need leaf forest', 'mixed forest', 'savannas' and 'grassland', the mapping is direct. Further separation will be done for the biome categories with different emission factors at different climate zone. For the categories in NLCD40 with more detail definitions than corresponding biome category, such as water and urban lands, they will be consolidated into one category in biome by addition. For example, 'soil and ice' in soil biome category is equal to the addition of 'permanent snow and ice' and 'perennial ice-snow' in NLCD40 MODIS category; 'urban and build-up lands' in soil biome category is equal to the addition of 'developed open space', 'developed low intensity', 'developed medium intensity' and 'developed high intensity'. For the categories appearing only in NLCD40, the mapping algorithm is according to the CMAQ mapping scheme, which is documented in each the CSQY_DATA_* under the MECHS/ directory in the CMAQ source code release. One such example is to map 'lichens' and 'moss' in NLCD40 to the category 'grassland' in soil biome.

Table 11. Mapping used to create the soil biome map based on NLCD40 MODIS land use/land cover categories.

ID	NLCD40 MODIS CATEGORY (40)	ID	SOIL BIOME CATEGORY (24)
1	Evergreen Needleleaf Forest	19	Evergr. Needle. Forest
2	Evergreen Broadleaf Forest	16 and 21	Evergr. Broadl. Forest
3	Deciduous Needleleaf Forest	18	Dec. Needle. Forest

ID	NLCD40 MODIS CATEGORY (40)	ID	SOIL BIOME CATEGORY (24)
4	Deciduous Broadleaf Forest	17 and 20	Dec. Broadl. Forest
5	Mixed Forests	15	Mixed Forest
6	Closed Shrublands	7	Closed shurb
7	Open Shrublands	8 and 9	Open shrubland
8	Woody Savannas	14	Woody savannah
9	Savannas	11 and 12	Savannah
10	Grasslands	10 and 13	Grassland
11	Permanent Wetlands	2	Permanent Wetland
12	Croplands	22	Cropland
13	Urban and Built Up	23	Urban and build-up lands
14	Cropland-Natural Vegetation Mosaic	24	Cropland/nat. veg. mosaic
15	Permanent Snow and Ice	3	Snow and ice
16	Barren or Sparsely Vegetated	6	Barren
17	IGBP Water	1	Water
18	Unclassified	1	Water
19	Fill value	1	Water
20	Open Water	1	Water
21	Perennial Ice-Snow	3	Snow and ice
22	Developed Open Space	23	Urban and build-up lands
23	Developed Low Intensity	23	Urban and build-up lands
24	Developed Medium Intensity	23	Urban and build-up lands
25	Developed High Intensity	23	Urban and build-up lands
26	Barren Land (Rock-Sand-Clay)	24	Cropland/nat. veg. mosaic
27	Unconsolidated Shore	24	Cropland/nat. veg. mosaic
28	Deciduous Forest	16 and 21	Evergr. Broadl. Forest
29	Evergreen Forest	19	Evergr. Needle. Forest
30	Mixed Forest	15	Mixed Forest
31	Dwarf Scrub	8 and 9	Open shrubland
32	Shrub-Scrub	8 and 9	Open shrubland
33	Grassland-Herbaceous	10 and 13	Grassland
34	Sedge-Herbaceous	14	Woody savannah
35	Lichens	10 and 13	Grassland
36	Moss	10 and 13	Grassland
37	Pasture-Hay	24	Cropland/nat. veg. mosaic
38	Cultivated Crops	22	Cropland
39	Woody Wetlands	2	Permanent Wetland
40	Emergent Herbaceous Wetlands	2	Permanent Wetland

For the second step, a model resolution compatible Köppen-Geiger climate zone classification is needed. This is to allocate the different emission factor for the same biome type

(e.g. ‘grassland’) at different locations. For the 12km CONUS domain, Spatial Allocator was used to generate the 5 category climate zone map as shown in Figure 39. These categories (A: equatorial, B: arid, C: warm temperature, D: snow, and E: polar) are based on the county level climate zone documentation text file as the surrogate (http://koeppen-geiger.vu-wien.ac.at/data/KoeppenGeiger.UScounty.txt). The 12km climate zone product (bottom figure) matches well with original file (top figure) in which most of southern US is classified as the ‘warm temperature’ region. Only a slight portion of Rocky mountain summit is classified as ‘polar’ climate and the south corner of Florida State is classified as ‘equatorial’ climate.

Figure 40 provides the comparison between the new soil biome spatial map based on the finer resolution LU/LC definition used in current CMAQ simulations and the old soil biome spatial map based on coarse GOES-Chem LU/LC setting. As the reference, the independent 30m resolution 2011 NLCD spatial map (http://www.mrlc.gov/nlcd2011.php) based on the Landsat satellite is also provided here. In order to be comparable with the NLCD 2011 classification system, similar color legend was used to visualize the biome soil type. It can be seen that the new CMAQ 12km soil biome has much more detailed texture and closer geolocation correspondence than the old soil biome derived from GOES-Chem. For example, the new biome map has more identified ‘cropland’ in the central US states (e.g. Oklahoma, Kansas) while the old biome map are all ‘grassland’ at the same places. Also, the new biome map identifies a lot of ‘wetland’ near the southern coastal line area, which matches with the NLCD2011 classification for ’90 woody wetlands’ and ’95 emergent herbaceous wetlands’.

The new soil biome map was used to run the soil BDSNP model and evaluate its impact on soil NO emission estimates.

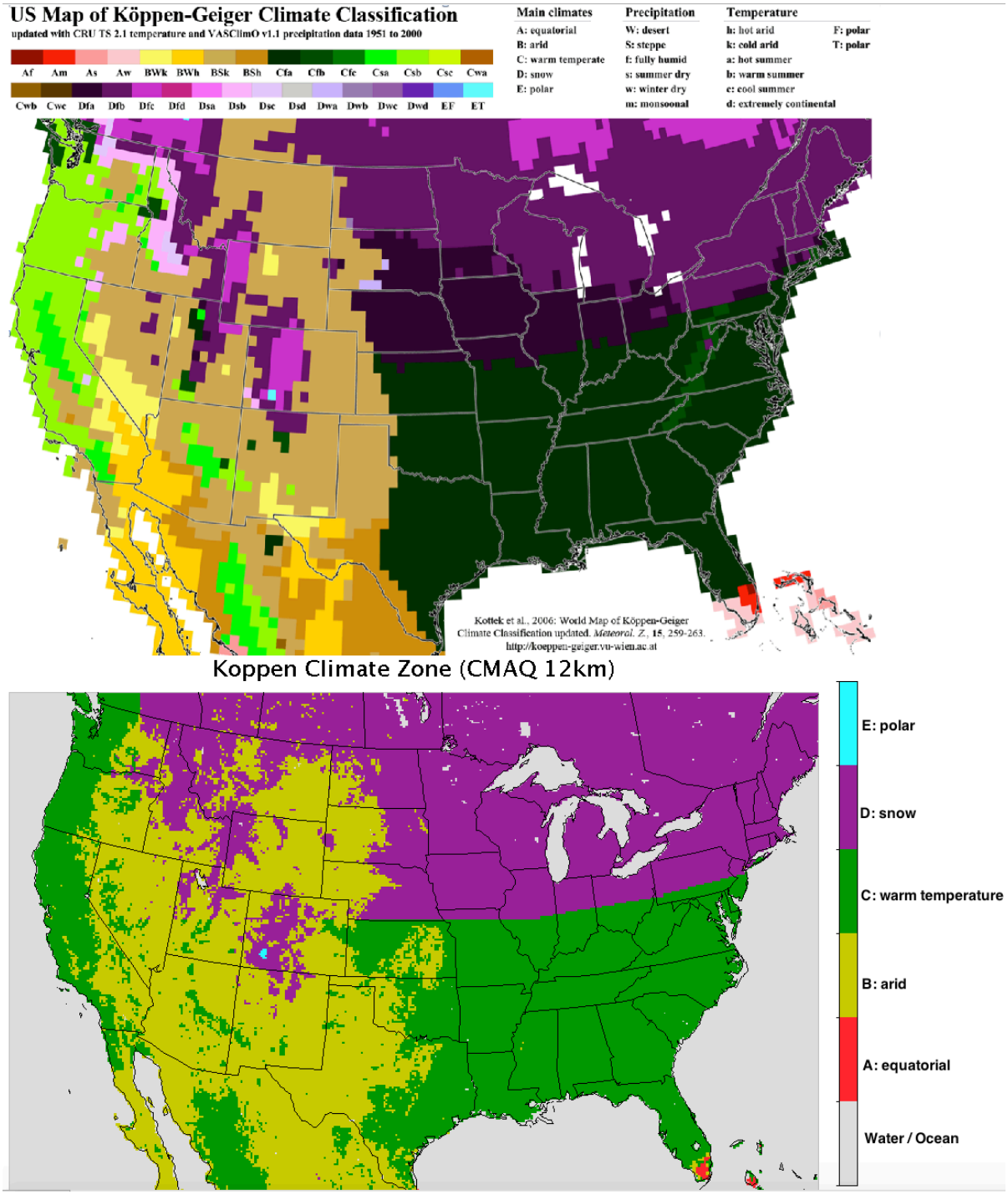


Figure 39. Spatial map of five climate zones over CONUS 12km domain (bottom) based on Köppen-Geiger climate classification (top).

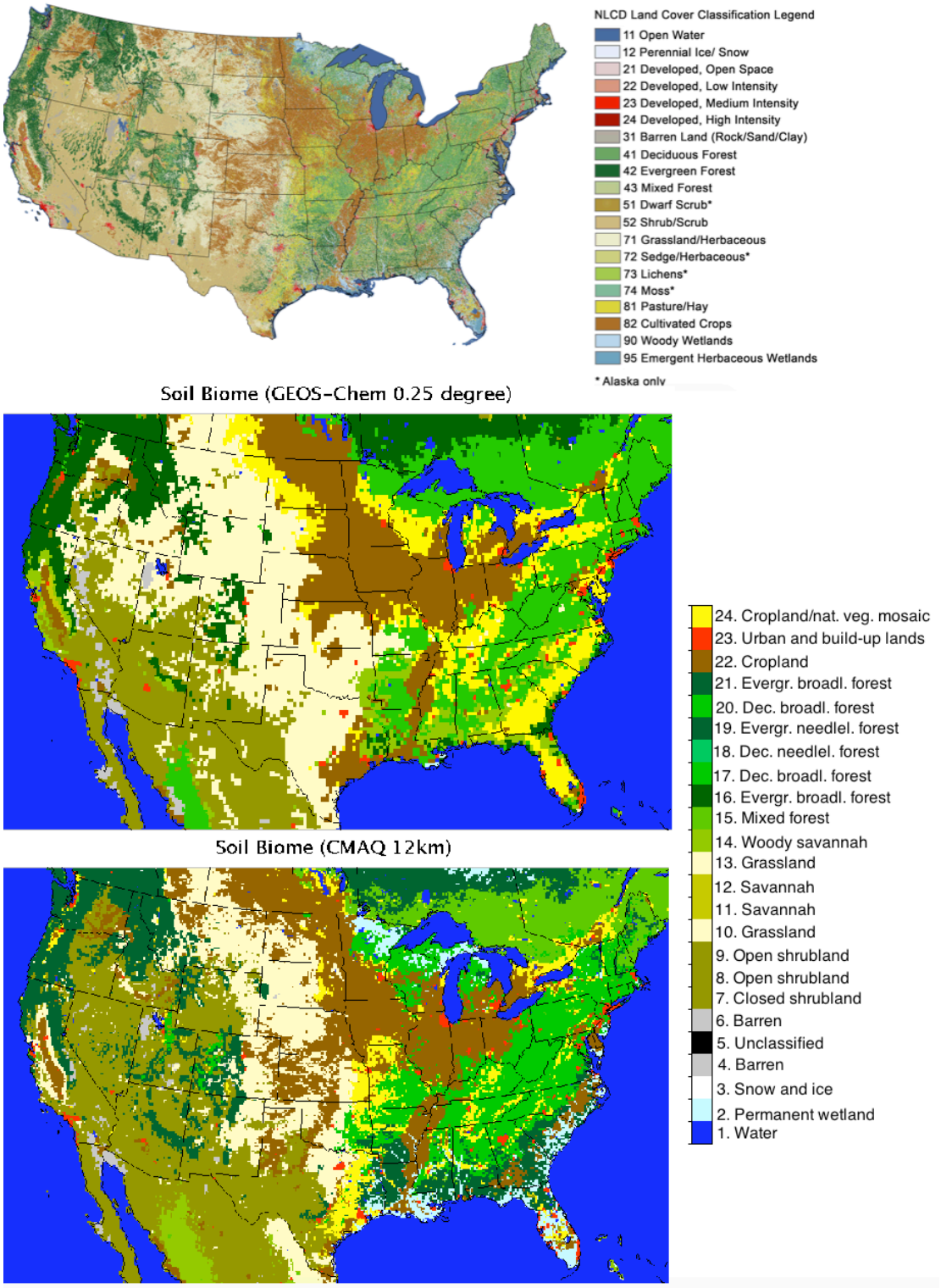


Figure 40. Comparison of the spatial pattern of 2011 National Land Cover Database (30m resolution, top) with the soil biome type developed either from GEOS-Chem (0.25 degree resolution, middle) or from MODIS NLCD40 classification in CMAQ (12km resolution, bottom).

5.3.2. Test Case for soil NO emissions using stand-alone BDSNP scheme

Functionality of stand-alone soil NO emission module with BDSNP scheme and the new soil biome map using the 12km resolution NLCD40 land use classification and Köppen-Geiger climate classification map was tested. With the high efficiency of the stand-alone version, more sensitivity tests can be carried out by switching the key input parameters for soil NO emission in the BDSNP module (e.g. different soil biome, different base emission factors, and different fertilizer pools). Figure 41 provides the difference in spatial pattern between soil NO base emission simulated by this stand-alone model using global GEOS-Chem soil biome (control), updated regional soil biome based on NCLD40 (new Biome), and North American specified emission factors (NA EF) over continental US.

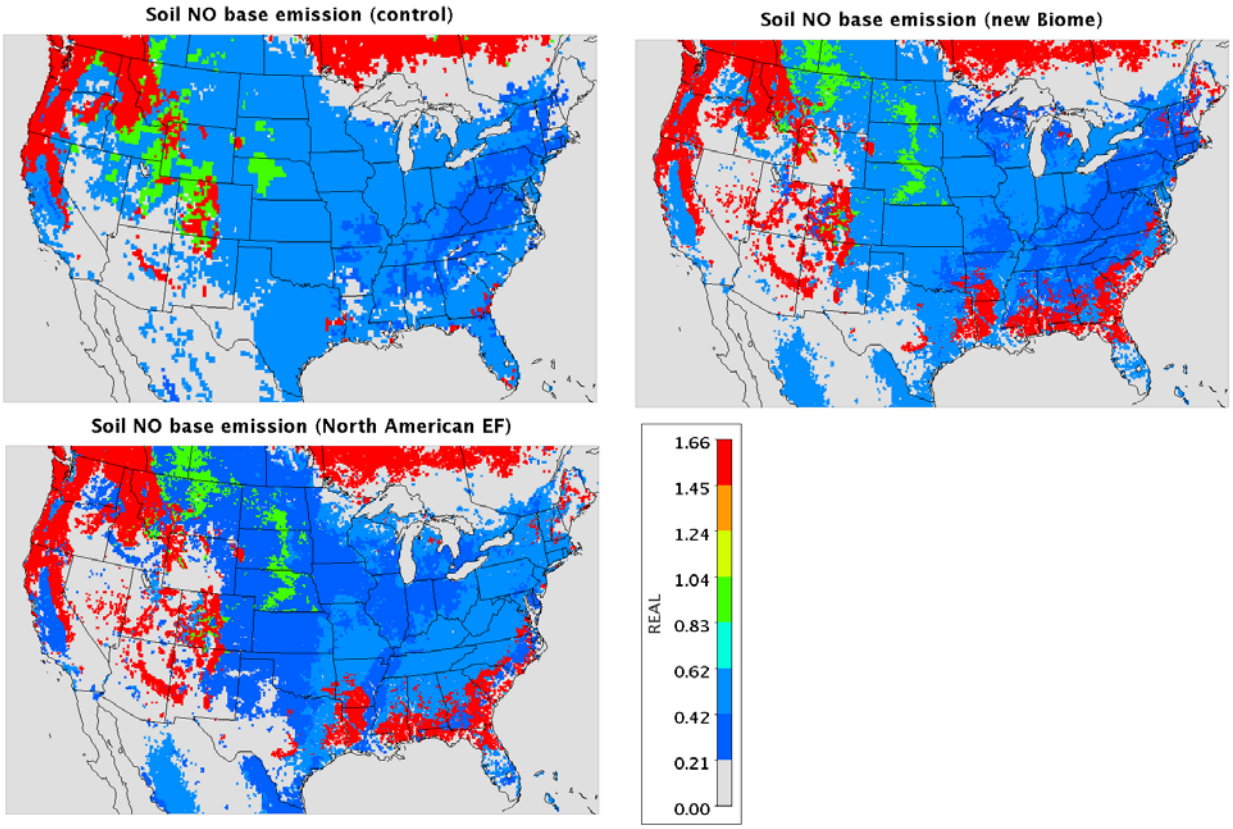


Figure 41. Spatial pattern difference of soil NO base emission simulated from BDSNP module using the global GEOS-Chem soil biome (control), updated regional soil biome based on NCLD40 (new Biome), and North American specified emission factors (NA EF) over the continental US.

In comparison to the 'control' case, the soil NO base emission pattern from case 'new Biome' has much detail texture due to the usage of higher resolution biome map and better representation of geographic locations for cropland over Midwest and evergreen board leaf forest along the South Eastern coastal areas. The original implementation of soil NO BDSNP module used the global average biome type specific emission factors, which is 2-3 times higher than the local US measured values for the category such as cold savannah. Using the local emission factor intend to produce more realistic results in this project.

The soil NO emission rate in the default MEGAN model is based on the Yienger and Levy 1995 (YL95) scheme. All the BDSNP input files including biome type map, fertilizer pool map, arid/non arid map, nitrogen deposition from dry and wet process were re- gridded to be consistent with the TCEQ modeling domains. The daily magnitudes of nitrogen deposition pool used are from the 2005 CMAQ simulation results. Figure 42 demonstrates the difference in spatial pattern for daily mean NO emission rate between YL95 and BDSNP on August 1, 2013 over the TCEQ Texas domain (D2).

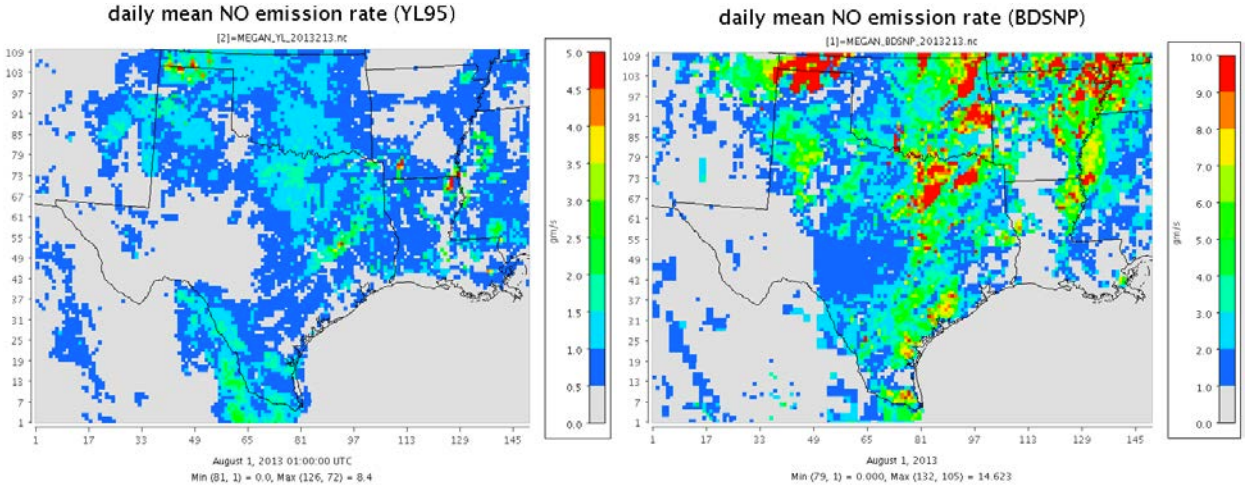


Figure 42. The difference in spatial pattern between daily mean soil NO emission rate (g/s) from MEGAN default YL95 scheme (left) and BDSNP scheme (right) on August 1, 2013 over the Texas domain.

Note the different color scales in the figures. The magnitude of soil NO emission predicted from BDSNP for this day is generally 2-3 times higher than that from YL95, with the maximum value 14.6 gm/s versus 8.4 gm/s. The spatial pattern for the two cases is also quite different due to the combined contributions from different soil biome types, fertilizer implementations, and the different response curve for soil temperature and moisture in the two soil NO schemes. The two-month soil NO emissions simulated with BDSNP scheme by using the two set of WRF runs (case 'cntrl' and case 'analytical', presented in the following section) will be archived separately along with MEGAN results and hand over to TCEQ for further test. The documentation and user's manual for the stand-alone soil NO BDSNP module also will accompany the final report.

5.4. Soil NO emission estimates for 2013 using BDSNP scheme

The stand-alone BDSNP module was used to estimate soil NO emissions for August and September 2013 over the TCEQ SIP domain. Two sets of WRF outputs ('cntrl' and 'analytical') were used to drive the BDSNP module.

Figure 43 show the spatial patterns for the monthly mean soil NO emission rates using different WRF runs and different soil NO algorithms for the month of August 2013. Figure 44 shows the similar patterns for September 2013.

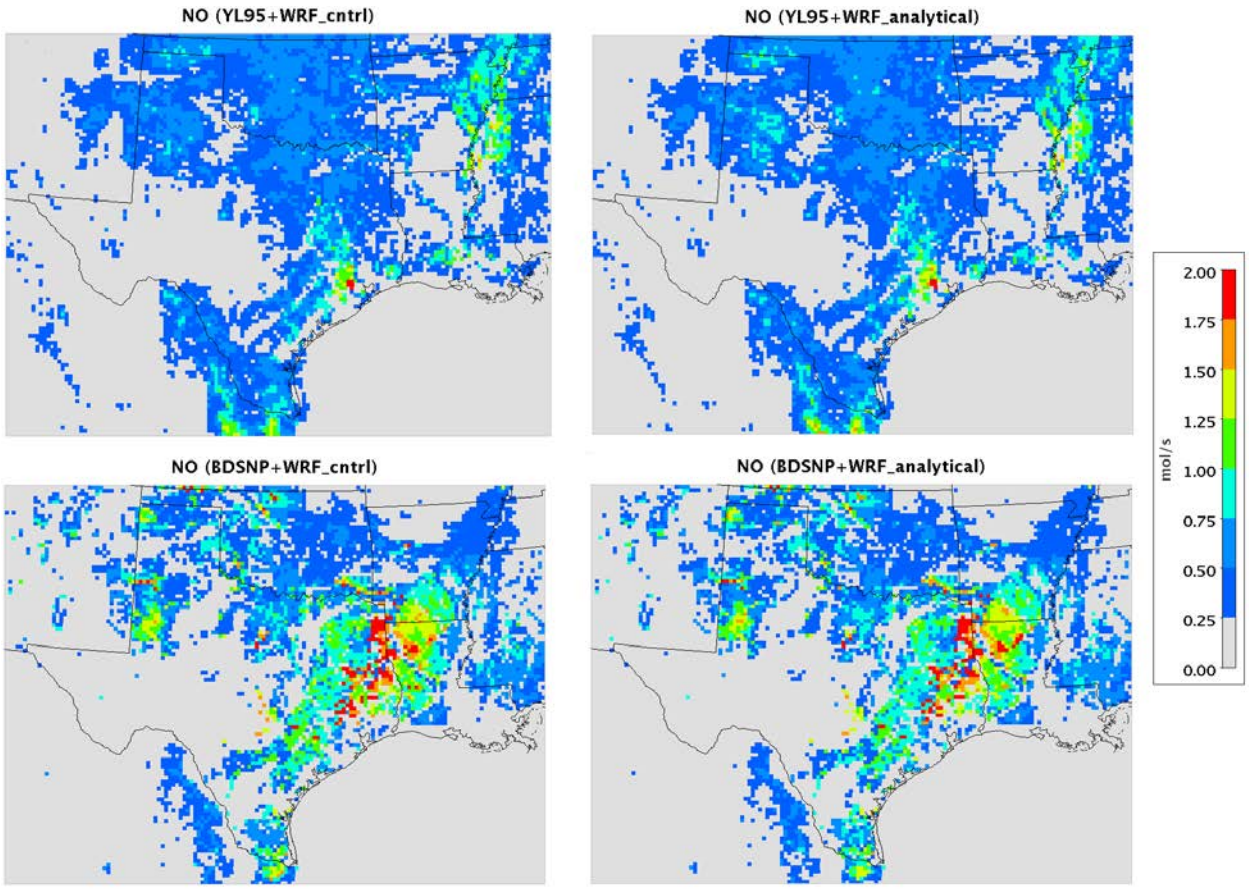


Figure 43. Comparison of the spatial patterns of the monthly mean soil NO emission rate using different meteorology inputs, WRF control (cntrl) and WRF cloud assimilation (analytical) and different soil NO emission algorithms, YL95 and BDSNP, over Texas domain during August 2013.

As expected, the soil emission rates estimated by BDSNP module are consistently higher than the results estimated by YL95 algorithm. However, the spatial patterns for the two algorithms are quite different. For instance, the hotspot emission regions are predicted to be near the suburbs of Houston for YL95 case, while for BDSNP case, the hotspots appear to be near the state boundary between Texas and Louisiana and Arkansas. Again, those contrast may due to the combined contributions from different soil biome types, fertilizer implementations, and the different response curve for soil temperature and moisture in the two soil NO schemes. Note that contrary to the test case, in these plots the same color scale has been used.

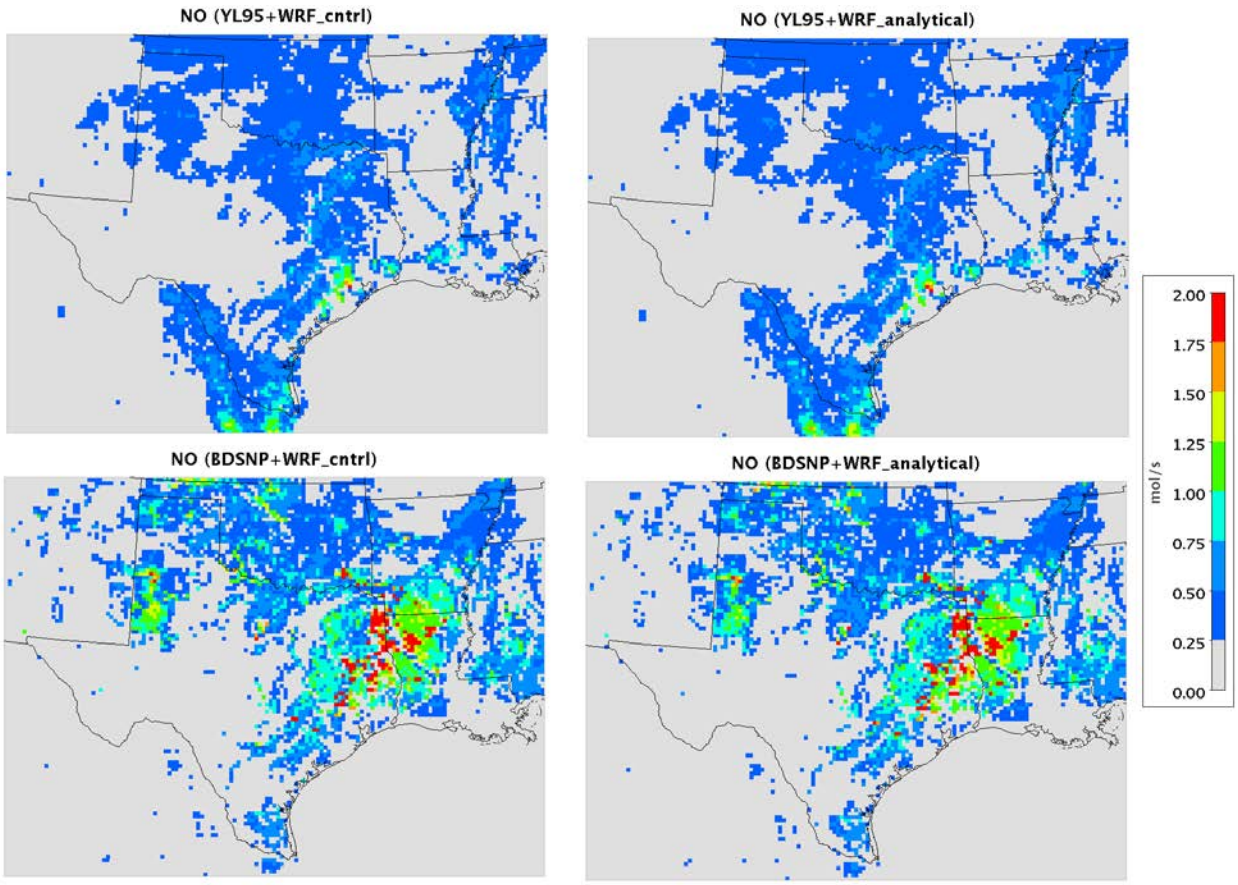
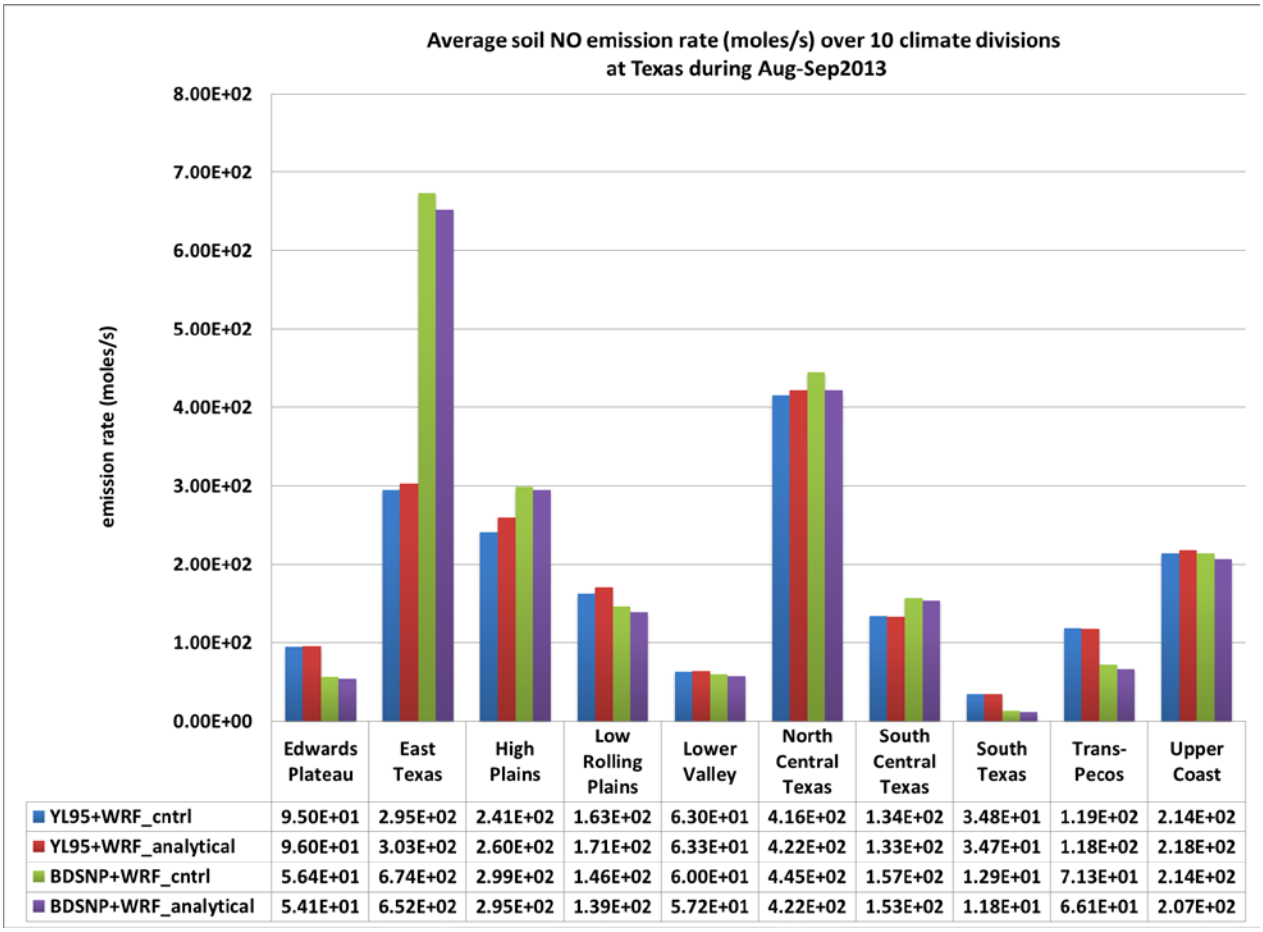


Figure 44. Comparison of the spatial patterns of the monthly mean soil NO emission rate using different meteorology inputs, WRF control (cntrl) and WRF cloud assimilation (analytical), and different soil NO emission algorithms, YL95 and BDSNP, over Texas domain during September 2013.

There is no obvious difference due to the use of different meteorology fields (with or without cloud assimilation). For August 2013 using YL95 algorithm, the maximum emission rate decreased from 2.51 moles/s to 2.21 moles/s by adding cloud assimilation; while for the BDSNP case, the trend is opposite, the maximum emission rate increased from 2.95 moles/s to 3.31 moles/s. A similar pattern is seen for September 2013 monthly mean values. It should be noted that for YL95 case, the emission rate for September is generally lower than the value in August, while for BDSNP case the values are comparable for the two months. The mean air temperature in September 2013 is lower than that of August 2013, which directly causes the low scaling factor due to soil temperature in YL95 algorithm. However, in BDSNP module, the response of NO emission to soil temperature is more complicated (and in this case modest).

Similar to characterization of BVOC emission patterns from MEGAN over ten climate divisions in Texas, average soil NO emission rates over different climate divisions using YL95 or BDSNP algorithms are shown as the histogram in Figure 45. For both YL95 and BDSNP cases, the top three emission regions are East Texas (1st for BDSNP case, 2nd for YL95 case), North Central Texas (2nd for BDSNP case and 1st for YL95 case) and High Plains (3rd place for both case).



	Edwards Plateau	East Texas	High Plains	Low Rolling Plains	Lower Valley	North Central Texas	South Central Texas	South Texas	Trans-Pecos	Upper Coast
YL95+WRF_cntrl	9.50E+01	2.95E+02	2.41E+02	1.63E+02	6.30E+01	4.16E+02	1.34E+02	3.48E+01	1.19E+02	2.14E+02
YL95+WRF_analytical	9.60E+01	3.03E+02	2.60E+02	1.71E+02	6.33E+01	4.22E+02	1.33E+02	3.47E+01	1.18E+02	2.18E+02
BDSNP+WRF_cntrl	5.64E+01	6.74E+02	2.99E+02	1.46E+02	6.00E+01	4.45E+02	1.57E+02	1.29E+01	7.13E+01	2.14E+02
BDSNP+WRF_analytical	5.41E+01	6.52E+02	2.95E+02	1.39E+02	5.72E+01	4.22E+02	1.53E+02	1.18E+01	6.61E+01	2.07E+02

Figure 45. Comparsion of average soil NO emission rate (moles/s) over the 10 climate divisions of Texas during August and September 2013 by different algorithms and meteorological fields.

For BDSNP case, the emissions over the East Texas region are dominant (674 moles/s, 2.3 times the corresponding YL95 case). The lowest emission region remains to be South Texas for both soil NO algorithms. However, the absolute value for YL95 case is almost three times higher than BDSNP case (34.8 moles/s for YL95 case versus 12.9 moles/s for BDSNP case, using the control WRF meteorology fields). Summing over the 10 climate divisions for the two-month period, the soil emission rate over Texas predicted by BDSNP module is 21% higher using the base WRF inputs and 14% higher using the WRF with cloud assimilation. Still, no significant difference (less than 5%) can be found when using the different meteorology inputs with the same soil NO algorithm.

A user's manual has been prepared for the stand-alone BDSNP module, which documents in detail the model structures and how to install and operate it for specific applications. Further possible improvements, such as replacing the default global fertilizer map from Potter et al. (2010) with more dynamic and up-to-date fertilizer fields from EPIC outputs by considering the different farming management scenarios, are also provided in the user's manual. The manual as

well as the tested benchmark case and the 4km resolution soil biome map will be provided to TCEQ as part of final delivery.

6. CMAQ SIMULATIONS USING DIFFERENT MEGAN OUTPUTS WITH SATELLITE PAR

To test and evaluate the impact of meteorological and emission changes on air quality simulations during August-September 2013, a series of CMAQ simulations were performed. CMAQ simulations using different biogenic emissions estimated by MEGAN with different PAR inputs (either from control WRF run, or WRF run with cloud assimilation, or UAH's PAR retrieval product from GOES satellite imager) and soil NO emission either from YL95 or BDSNP schemes were performed. The initial conditions and boundary conditions for domain 1 were extracted from NCAR's global Model for Ozone and Related Chemical Tracers (MOZART) model outputs at 3 hourly intervals. The anthropogenic emissions were provided by TCEQ for 2011 base year in CAMx FORTRAN binary format.

Due to the multiple combinations of inputs for CMAQ and time limitation, only three cases were finally chosen to assess the impact of different biogenic emissions on ozone during 2013 DISCOVER-AQ Houston campaign period. These configurations are listed in Table 12.

Table 12. Different input combinations used in CMAQ simulations.

Case #	MET input	ICON/BCON input	EMIS input		
			Anthropogenic	Biogenic	
				BVOC	Soil NO
C1	WRF ('cntrl')	MOZART	TCEQ(2011)	MEGAN('cntrl')	YL95
C2	WRF('analytical')	MOZART	TCEQ(2011)	MEGAN('UAH')	YL95
C3	WRF('analytical')	MOZART	TCEQ(2011)	MEGAN('UAH')	BDSNP

To show the quantitative impact of BVOC on ozone formation, only results from CMAQ simulations over D1 (CONUS, 36km resolution) and D2 (Texas, 12km resolution) domains for the first half of September 2013 (Sep 1-15, 2013, when the measured ozone and isoprene concentration data are available for model performance evaluation) will be presented here.

Figure 46 provides the detailed flow chart showing the preparation of the emissions for CMAQ simulations. Since the gridded ground-level emission files in seven source categories (exclude the biogenic part) provided by TCEQ are in CAMx binary format, the utility 'CAMx2CAMQ' was used to convert the emissions into the CMAQ I/O API format. The converter utility 'CAMx2CMAQ' was developed by Ramboll-Environ and can be obtained from the following site (https://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/pm/5821110365FY1420-20130830-environ-camx2cmaq_source_code.tgz). The emissions then were combined with the different MEGAN outputs.

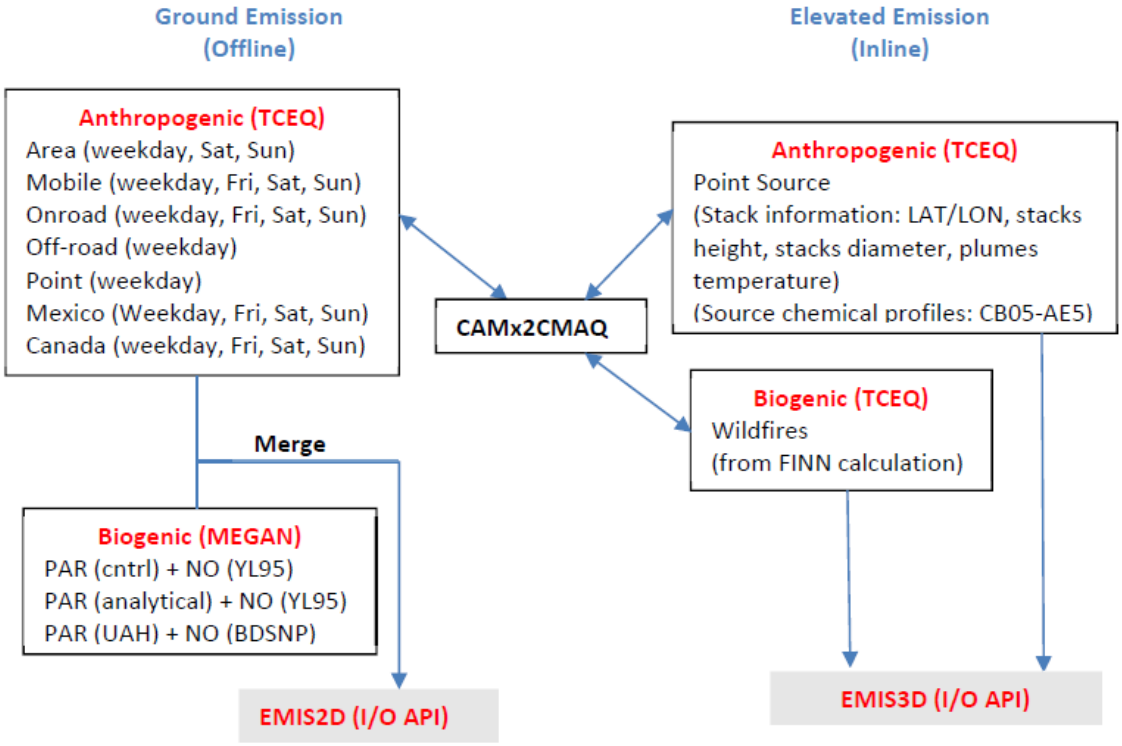


Figure 46. Flow chart to prepare daily emission input files for CMAQ run.

For elevated sources (points and fires) calculated inline by CMAQ, the daily emission files provided by Jim Mackay from TCEQ were used to provide the daily specific stack information and emission profiles. The wildfire emissions were estimated by the FINN (Fire Inventory from NCAR) model and used the MODIS hotspots to allocate the locations. Those files were also given in CAMx binary format and were converted to IO/API format by 'CAMx2CAMQ' utility.

The daily spatial pattern of the final emissions for key pollutants (SO2, NOx, CO, PM2.5 and VOC emissions) for September 1st, 2013, over the Texas domain (D2) in units of tons/day is shown in Figure 47. The color scales in these figures are different for each key pollutant and were scaled to best show the representative anthropogenic activity and plants distributions over Texas. The ground emissions are merged from seven emission source categories, namely 'area', 'Canada', 'low_points', 'Mexico', 'nonroad', 'offroad', and 'onroad'. (ftp://amdaftp.tceq.texas.gov/pub/TX/camx/basecase/bc12_12jun.reg3a.2012_wrf361_p2a_i2_a/input/ei/Components/). For each source category, there are different individual versions to represent the temporal variation, namely weekdays, Friday, Saturday and Sunday. A script was created to find the corresponding day-of-the-week for the simulation period and locate the corresponding files to merge.

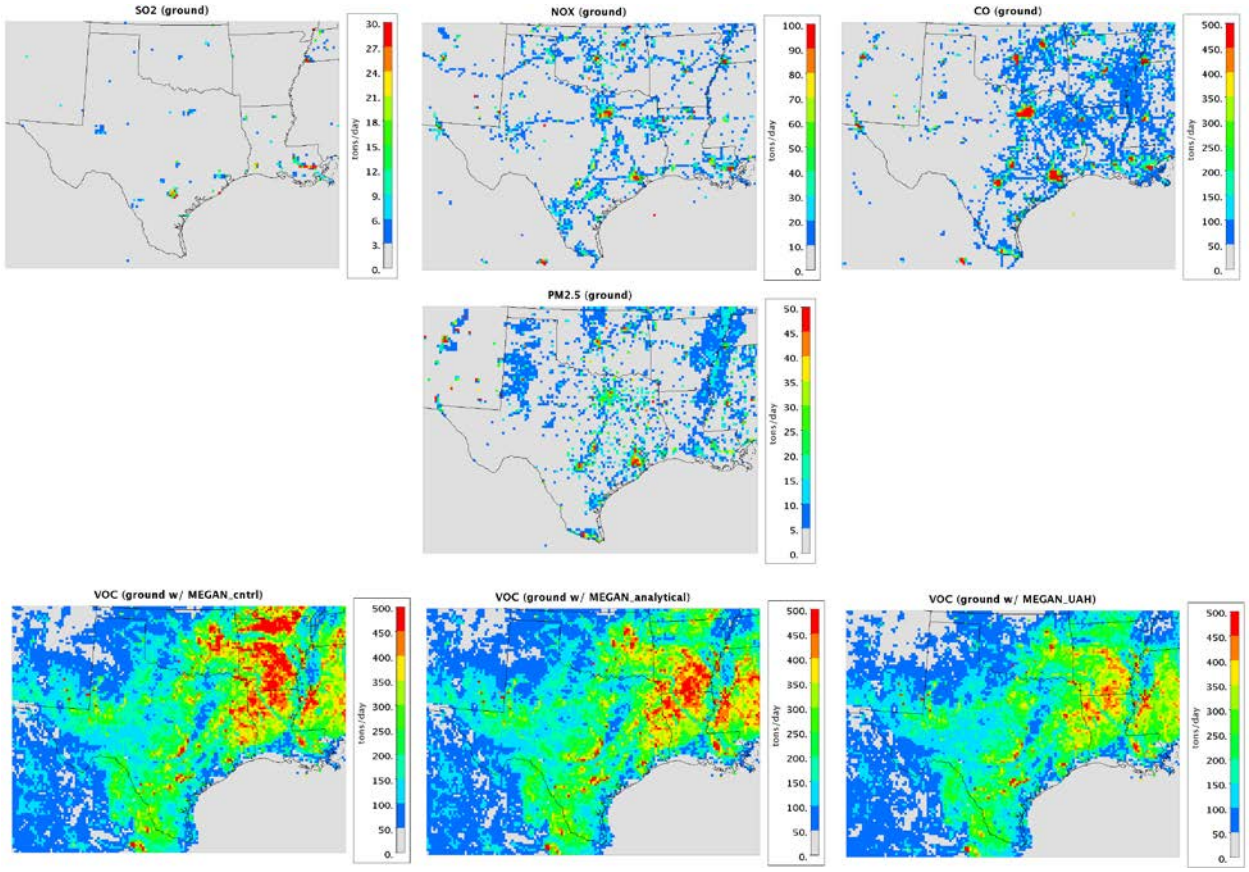


Figure 47. Spatial pattern of daily ground gridded SO2, NOx, CO, PM2.5 and VOC emission prepared for CMAQ runs over Texas 12km domain on Sep 1, 2013 (tons/day). (The daily ground VOC emission pattern is given in three versions with different PAR inputs in MEGAN model).

For NOx, CO and $PM_{2.5}$ emissions, the majority of hotspots correlate well with the Houston, Dallas-Fort Worth, San Antonio and Austin metropolitan area in Texas and the major road network. The typical daily emission rates for NOx, CO and PM2.5 over the three metropolitan areas are 100 tons/day, 500 tons/day, and 50 tons/day (per grid cell) respectively. For SO2 emission, some hotspots (> 30 tons/day/grid-cell) correlate with the coal-fired power plant locations (from source category 'point' on the ground).

For VOC emissions, the elevated values are mostly dominated by biogenic emissions over East Texas (adjacent to the border area of Louisiana and Arkansas), with the typical daily emission rate greater than 500 tons/day/grid-cell. Three versions of merged VOC emissions are presented in the figure. Each spatial pattern shows the impact of different MEGAN outputs ('cntrl', 'analytical' and 'UAHPAR'). The case 'UAHPAR' has about 30% less isoprene emission and 5% less monoterpene emission compared with the base case due to the use of PAR from satellite retrievals (see Table 9).

Table 13 summarizes the detail CMAQ configuration used in these simulations. The boundary and initial conditions were provided by the NCAR's global chemical MOZART model

outputs with the daily BCON files given in 3 hourly interval for the CONUS domain (for D2 simulation, the BCON files were extracted using the D1's concentration field). A spin-up period of 3 days, starting from Aug 29, 2013, was used to create the initial concentration fields used for September 1. The Carbon Bond-05 mechanism with chlorine (CB05cl) gas phase chemical mechanism, as well as aero5 aerosol module was selected. CMAQ version 5.0.2 was used for these simulations. The photolysis rates were calculated inline and the flags for windblown dust and inline lightning NOx were turned off due to the lack of input data. In total three sets of CMAQ runs were conducted using different biogenic emissions provide by MEGAN estimates.

Table 13. Configuration of CMAQ simulation in this study.

Input Variables	Configurations
Domain Setting	D1 (CONUS, 36km, 148(C^1)x112(R^2)X29(L^3)) D2 (Texas, 12km, 149(C)X110(R)X29(L))
Emissions	Ground emission (anthropogenic from TCEQ, biogenic from MEGAN v2.10) Elevated emission (point source and wildfire from TCEQ)
Boundary conditions	MOZART (3hr interval)
Initial conditions	MOZART[4] (for Aug 29, 2013) Spin-up days: 3day
Chemical Transport Model	CMAQv5.0.2 Advection scheme: WRF Diffusion scheme: multiscale (horizontal), ACM2[5] (vertical) Gas mechanism: cb05cl[6] Aerosol module: aero5 Cloud module: ACM2[2] Photolysis: inline Lighting NOx: off Wind blow dust: off NH3 bi-directional flux: off Inline deposition velocity: on
Simulation Period	15 days (Sep 1-15, 2013)

Note: 1. C: columns; 2. R: rows; 3. L: layers; 3. MOZART: Model for Ozone and Related Chemical Tracers; 4. ACM2: Asymmetrical Convective Model version 2; 5. Cb05cl: Carbon Bond gas phase mechanism version 5 with chlorine chemistry extension.

Figure 48 shows the average spatial pattern of simulated maximum daily 8-hr average ozone concentrations (MDA8 O3) for the three cases over CONUS domain (D1) during September 1-15, 2013. In Texas, the predicted highest MDA8 O3 areas are located near Dallas-Fort Worth with values greater than 65 ppbV. The typical MDA8 O3 value over Houston area is around 50-55 ppbV during the DISCOVER-AQ campaign period. By using the BVOC emission estimated by UAH PAR in MEGAN to drive the CAMQ, the simulated MDA8 O3 value are on average 3-7 ppbV less than that of the base case. This reduction is due to more accurate cloud representations in MEGAN (by decreasing the PAR input for isoprene emission and ground

temperature input for monoterpene emissions). This is also shown in the CMAQ daily mean isoprene concentrations shown in Figure 49 (with the typical predicted value over East Texas, decreasing from 4.8 ppbV for base case to 2.4 ppbV for UAHPAR case). The decreasing trend of ozone concentration in CMAQ over CONUS corresponds to the reduction in BVOC emission and is more pronounced over the regions where ozone formation is VOC-limited.

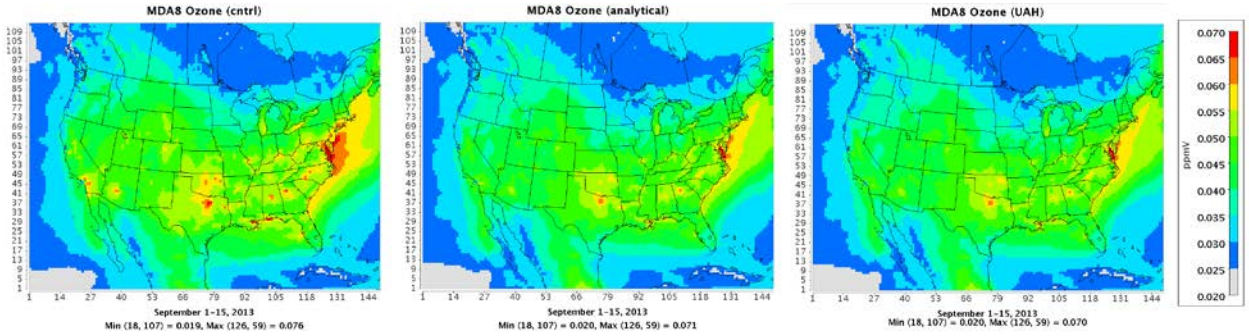


Figure 48. Average spatial pattern of simulated maximum daily 8-hr average ozone concentrations (MDA8 O3) for three cases during September 1-15, 2013.

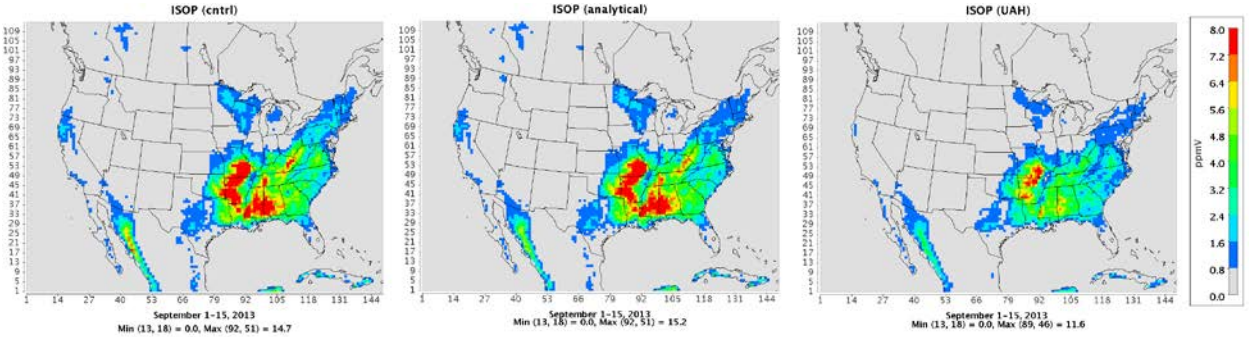


Figure 49. Average spatial pattern of simulated daily mean isoprene concentrations for three cases during September 1-15, 2013.

In order to evaluate the CMAQ performance over Texas, the predicted hourly ozone and isoprene concentrations were compared with the available ground measurements extracted from TCEQ's GeoTAM (Geographical Texas Air Monitoring) interface database webpage (https://gisweb.tceq.texas.gov/geotam3/).

Figure 50 shows the geographic location of selected observation sites for comparison. In total, 38 CAMS sites were chosen for ozone evaluation and 18 CAMS sites were chosen for isoprene comparison.

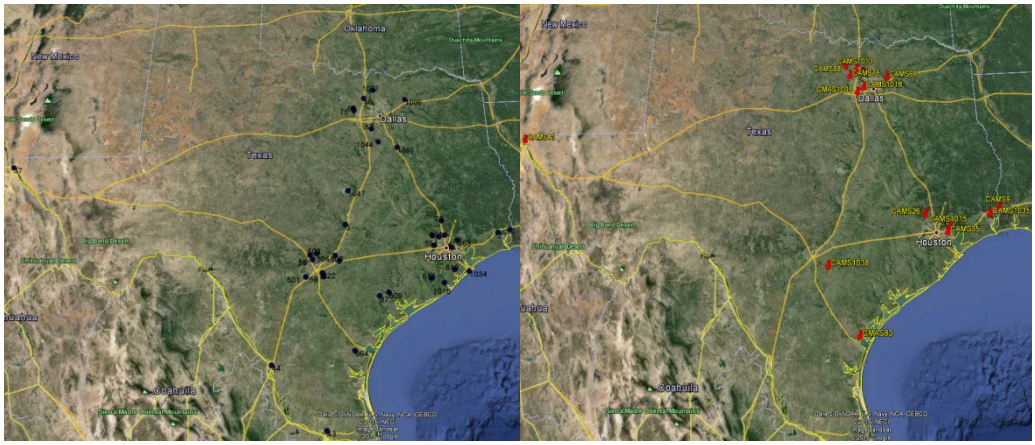


Figure 50. The locations of selected TCEQ CAMS sites for ozone (left, black markers) and isoprene (right, red markers) for CMAQ performance evaluation.

Hourly CMAQ simulation results from domain 2 were extracted for the monitoring sites and compared with the observations (at local, central, staring time, UCT+06). Routine statistical metrics such as observation mean (OBS), simulation mean (SIM), correlation coefficient (R), root mean square error (RMSE), mean bias (MB), mean error (ME), normalized mean bias (NMB) and normalized mean error (NME) were calculated for the three cases. The statistics are listed in Table 14. In order to differentiate the possible spatial heterogeneity for ozone performance, the evaluation sites were grouped into 7 regions, namely Austin, Corpus Christ, Dallas-Fort Worth (DFW), east Texas (Beaumont), El Paso, Houston-Galveston-Brazoria (HGB), and San Antonio. Note that the numbers in the parentheses denote how many sites were averaged in this region.

Table 14. Statistics for the CMAQ hourly ozone simulation performance for the three cases over 38 TCEQ CAMS sites (grouped by 7 regions).

Case	Group Sites	OBS_AVE (ppbV)	SIM_AVE (ppbV)	R	RMSE (ppbV)	MB (ppbV)	MAGE (ppbV)	NMB (%)	NME (%)
	Austin (1)	44.2	33.9	0.70	17.4	-10.2	14.4	-23.1	32.5
	Corpus Christi (1)	16.9	34.8	0.69	20.1	18.8	18.9	111.2	111.7
	DFW (6)	42.1	38.3	0.75	15.4	-3.4	12.5	-5.8	30.6
cntrl	East Texas (1)	39.8	30.4	0.79	15.4	-9.0	12.5	-22.5	31.4
	El Paso (1)	36.0	29.3	0.55	14.7	-6.6	11.7	-18.2	32.4
	HGB (7)	22.4	28.1	0.78	13.1	7.8	10.5	36.3	47.8
	San Antonio (6)	25.1	34.1	0.75	13.9	9.2	11.5	37.3	46.2
	Total	30.6	32.7	0.75	14.5	2.8	11.9	18.2	42.6
	Austin (1)	44.2	32.9	0.71	17.8	-11.3	14.7	-25.6	33.3
	Corpus Christi (1)	16.9	33.7	0.70	19.2	17.9	18.0	105.7	106.3
	DFW (6)	42.1	37.6	0.77	14.8	-4.3	11.9	-8.2	29.1
analytical	East Texas (1)	39.8	30.4	0.76	16.1	-9.1	12.9	-22.8	32.4
	El Paso (1)	36.0	28.8	0.56	14.8	-7.1	11.6	-19.8	32.3

Case	Group Sites	OBS_AVE (ppbV)	SIM_AVE (ppbV)	R	RMSE (ppbV)	MB (ppbV)	MAGE (ppbV)	NMB (%)	NME (%)
	HGB (7)	22.4	27.9	0.78	13.0	7.5	10.4	35.1	47.5
	San Antonio (6)	25.1	34.0	0.78	13.4	9.0	11.0	36.8	44.3
	Total	30.6	32.4	0.76	14.2	2.4	11.6	17.0	41.7
	Austin (1)	44.2	32.8	0.70	18.0	-11.3	14.8	-25.6	33.6
	Corpus Christi (1)	16.9	33.7	0.70	19.2	17.9	18.0	105.7	106.3
	DFW (6)	42.1	37.8	0.77	14.7	-4.1	11.9	-7.7	28.9
UAHPAR	East Texas (1)	39.8	30.4	0.76	16.1	-9.1	12.9	-22.8	32.5
	El Paso (1)	36.0	28.8	0.56	14.8	-7.1	11.6	-19.7	32.2
	HGB (7)	22.4	27.9	0.77	13.0	7.5	10.4	35.3	47.7
	San Antonio (6)	25.1	34.1	0.78	13.5	9.1	11.1	37.3	44.7
	Total	30.6	32.5	0.76	14.2	2.5	11.6	17.3	41.8

The three CMAQ cases did not show a significant difference with respect to ozone simulation. Mean R values for all three cases over the 38 sites were around 0.75-0.76. On the average, the model overestimated mean ozone concentration by 2.4-2.8 ppb.

By using the cloud assimilation (case 'analytical' and 'UAHPAR'), the ozone performance is systematically better than that of the base case by decreasing the mean bias from 2.8% to 2.4%, the RMSE from 14.5 ppbV to 14.2 ppbV, and the NMB from 18.2% to 17.0%. No significant difference was observed in model performance (for ozone) for case 'UAHPAR' compared with case 'analytical', at least for the present results. Examining the results for each sub-region, the DFW region has the best performance in terms of having the combined high R values (0.75-0.77) and lower NME values (28.9%-30.6%). The model performed better over the HGB region with respect to correlation coefficient (R=0.78) but overestimated mean ozone concentration by 7.5-7.8 ppbV. On average, the model underestimated mean ozone over Austin, DFW, East Texas and El Paso region, while overestimating mean ozone over Corpus Christi, HGB and San Antonio region. The Corpus Christi shows the highest NMB of 105-111% compared to the observations.

The poor performance for ozone simulation in Corpus Christi may be caused by the combined effect of poor marine boundary condition representation in the model, missing halogen chemistry in current CMAQ configuration and missing/incorrect emissions sources. Figure 51 shows the spatial map of statistic metrics R and NMB for all 38 TCEQ CAMS sites during the 15 days simulation period. Similar to the tabulated information, along the coastal areas model performs poorly. The cases 'analytical' and 'UAHPAR' significantly improve the model overestimation in sites near San Antonio region and northern suburbs near HGB region.

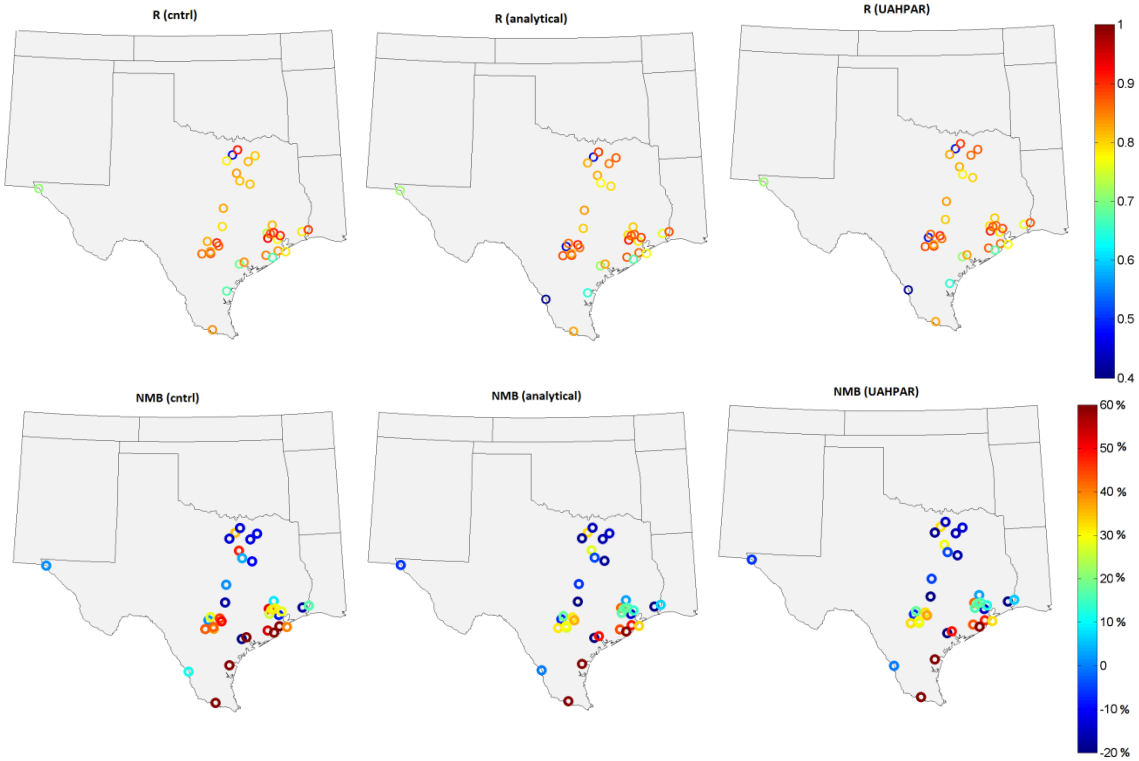


Figure 51. Performance of CMAQ hourly ozone from simulation case 'cntrl' (left) and 'analytical' (middle) as well as UAH insolation retrievals (right) at TCEQ sites. The upper panel shows the correlation coefficient (R) and the lower panel shows the normalize mean bias (NMB).

For isoprene, the model results were compared to measurements from 18 TCEQ sites and the statistics are summarized in Table 15. CMAQ performance for ISOP was particularly poor, with the typical R value from 0.36-0.40. The current model configuration significantly overestimates the isoprene concentration by a factor of 2.3-3.1 (the observation mean value during the simulation period is 0.23 ppbV while the simulation value is 0.47-0.61 ppbV).

Table 15. Statistics for model isoprene predictions for three cases over 18 TCEQ CAMS sites.

Case	OBS_AVE (ppbV)	SIM_AVE (ppbV)	IA	R	RMSE (ppbV)	MB (ppbV)	MAGE (ppbV)	NMB (%)	NME (%)
cntrl	0.23	0.59	0.37	0.36	0.69	0.39	0.49	292	326
analytical	0.23	0.61	0.37	0.37	0.72	0.42	0.51	311	342
UAHPAR	0.23	0.47	0.41	0.40	0.69	0.29	0.41	225	271

Even though the case 'UAHPAR', which introduces a more accurate PAR input, moves the model in the right direction, correcting model PAR error is not enough to correct this large model bias. Figure 52 shows the time series of predicted isoprene concentration from different CMAQ runs against observations at three individual sites. For site CAMS 26, which is located at the northern suburb of HBG region, the simulated ISOP concentration for most days are 2-3 times higher than the observed value for the all three cases (peak value 3.5 ppbV versus 1

ppbV). Case 'PARUAH' shows improvement during the second half of simulation period (i.e. the red line for Sep 10-15, 2013), but improvements are not consistent.

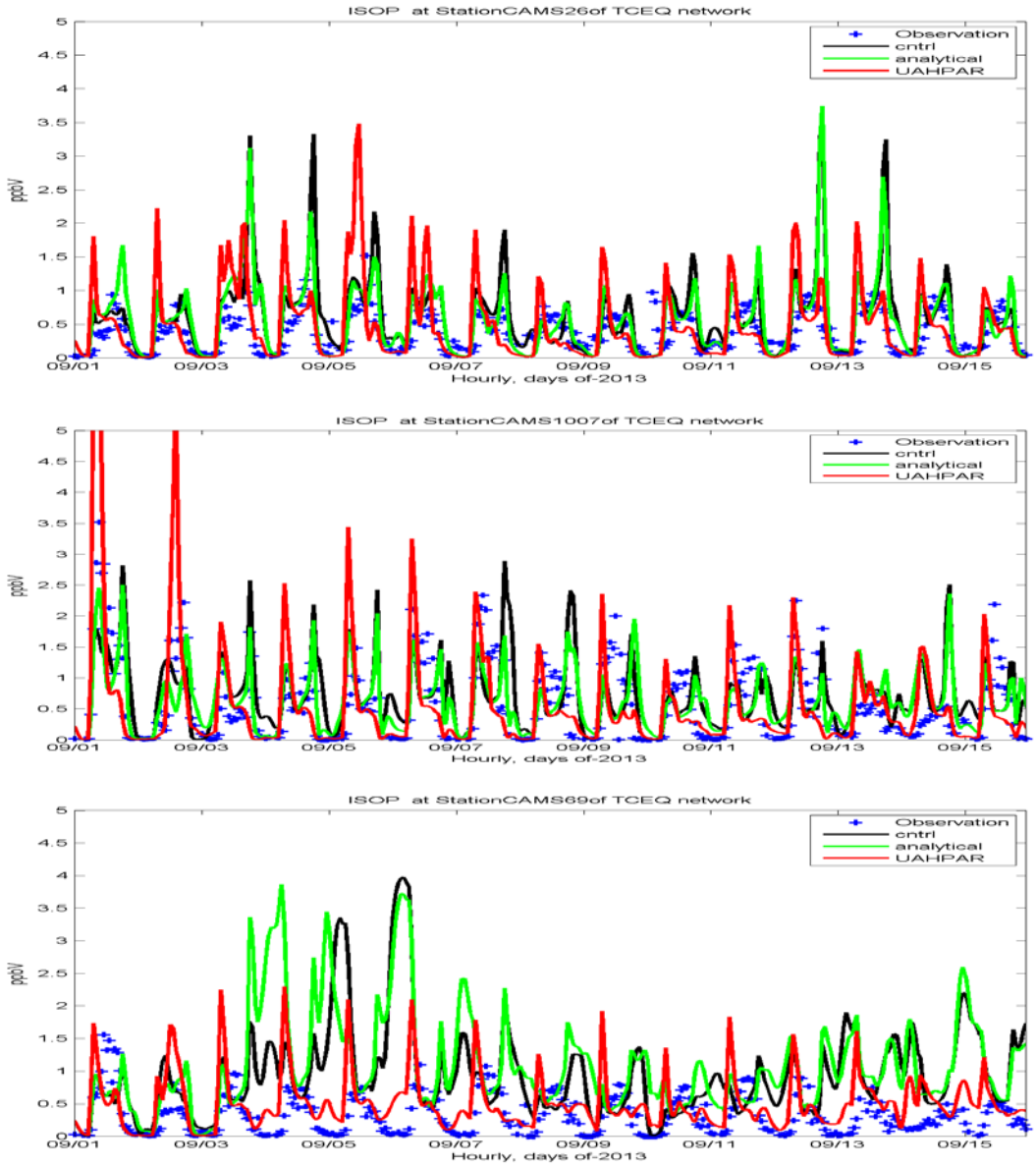


Figure 52. Time series comparison of ISOP concentration simulation for different CMAQ cases at site CAMS 26 (top), CAMS 1007 (middle) and CAMS 69 (bottom).

The model performance for case 'PARUAH' at some sites near the DFW region, as shown for CAMS1007 and CAMS69, are much better than the case 'cntrl' or 'analytical' in terms of replicating the magnitude of observation (especially during Sep 4-7, 2013 at site CAMS69). However, the model shows a diurnal pattern peaking during morning hours but suddenly dropping during afternoons on some days, which does not agree with observed pattern. The isoprene simulation results presented here are also consistent with other investigations

(personal communication) for similar AQRP projects (e.g. AECOM results testing the sensitivity of different mechanisms to changes in BVOC emission estimates, UT Austin results testing the impact of different land use on BVOC emission estimates). This suggests that the 2011 global BVOC emission factors used in the current MEGAN release (http://lar.wsu.edu/megan/docs/GlobalEmissionFactor/), at least for isoprene, may have high uncertainty over the study domain. Alex Guenther's group has been contacted to request the latest BVOC emission factors for further tests.

7. CONCLUSIONS AND RECOMMENDATIONS

This project took advantage of UAH's GOES insolation and cloud retrieval products to test the sensitivity of air quality simulations over Texas to BVOC emission estimates during the 2013 DISCOVER-AQ Houston campaign period by using the WRF-MEGAN-CMAQ modeling platform.

Compared with the ground pyranometer observations, satellite-retrieved PAR/insolation values tend to systematically correct the radiation overestimation problem in WRF, which is probably due to the model's inability to create clouds at correct location and time. This conclusion holds for all the evaluations conducted in this project against different networks (i.e. SURFRAD, SCAN and Texas local broadband radiation monitoring stations) and during different time periods (i.e. August 2006 and August-September 2013). Therefore, radiation information retrieved from satellite data is highly recommended to drive biogenic emission models. The UAH PAR products were also compared with another set of discontinued satellite retrieval products from UMD and the results show comparable statistical performance in terms of correlation and bias. However, the UAH PAR has much finer spatial texture than UMD PAR due to the higher resolution (4km).

In addition to the control WRF simulation, WRF simulations using the UAH satellite cloud assimilation technique were also performed. The results from these simulations (called 'analytical' in this report) were also used in BVOC emission estimates to quantify the impact of improved model cloud simulation. The results were independently evaluated with respect to cloud observations, surface radiation observations, and standard surface meteorological observations. The results from these simulations indicated significant improvements with respect to all the metrics used in the evaluation (ref. Chapter 4). While the use of observed satellite-based PAR is still advantageous over model data, the improvements achieved from cloud assimilation will have a broader impact on the air quality simulations as having better model cloud field impacts temperature, moisture, boundary layer development, and the transport of pollutants.

Three sets of MEGAN simulations over the TCEQ SIP modeling domains (D1 for the CONUS 36km domain, D2 for the Texas 12km domain and D3 for the East Texas 4km domain) were conducted during August and September 2013 by using different PAR inputs, namely PAR from the control WRF run (cntrl), PAR from WRF cloud assimilation run (analytical), and PAR from the GOES satellite retrieval using the new algorithm developed by UAH (UAH). Those results were

further analyzed to show the spatial heterogeneity over the Texas domain with respect to isoprene (ISOP) and monoterpene (TERP) emissions. For the 'UAH' case, the estimated ISOP emission rate by MEGAN is about 30% less than the 'cntrl' case and the estimated TERP emission rate is about 5% less than the 'cntrl'. With respect to spatial heterogeneity, the region most impacted in terms of reductions in the magnitude of concentrations is East Texas. The Trans-Pecos region of Texas (far west Texas, west of the Pecos River) was impacted most in terms of percentage. The significant differences show the importance of using observed cloud/insolation data for BVOC emission estimates used for SIP modeling activities.

A stand-alone version of the soil NO emissions model using the BDSNP scheme was developed in this project. The stand-alone model facilitates the ease of cross-platform implementation for air quality modeling (e.g., use in both CMAQ and CAMx). For the stand-alone BDSNP version, the soil NO emissions estimation process does not undergo the atmospheric chemistry calculation, therefore it saves more than 90% of CPU time compared with the standard air quality modeling. Furthermore, a new soil biome spatial map based on finer resolution land cover data and climate zone classification over the continental U.S. (CONUS) was also developed. The new map links with published estimates for biome-specific base emission rates to produce more detailed NO emission estimates. The spatial pattern of this new soil biome map matches with the latest CONUS land cover GSI database from the USGS and has much finer textural representations. The new map better represents the biome types such as grasslands, evergreen broadleaf forest, and cropland. The soil emission rates estimated by the BDSNP module are 14-21% higher than the results estimated by the YL95 algorithm but with quite different spatial patterns, which are determined by different soil biome map inputs, different soil temperature-moisture response curves, and different pulsing factor parameterizations. For future implementation, the soil biome base emission factors in the BDSNP model with the values derived from the measurements over specific North American sites is highly recommended since they better reflect the local soil characteristics.

Three sets of 15-day (Sep 1-15, 2013) CMAQ simulations were conducted to demonstrate the quantitative impact of cloud assimilation and biogenic emissions on ozone formation over Texas. The results did not show a significant difference with respect to ozone simulation. Compared to observations from 38 CAMS sites over Texas, mean correlation coefficient for all three cases were around 0.75-0.76. On the average, the model overestimated mean ozone concentration by 2.4-2.8 ppb. By using satellite data, the ozone performance was systematically better than that of the base case by decreasing the mean bias from 2.8% to 2.4%, the RMSE from 14.5 ppbV to 14.2 ppbV, and the NMB from 18.2% to 17.0%. On average, the model underestimated mean ozone over Austin, DFW, East Texas and El Paso region, while overestimating mean ozone over Corpus Christi, HGB and San Antonio region. The simulated ozone for Corpus Christi exhibited the highest NMB of 105-111% compared to the observations. The predicted highest simulated maximum daily 8-hr average ozone concentrations (MDA8 O3) were near Dallas-Fort Worth with values greater than 65 ppbV. The typical model MDA8 O3 value over Houston area were about 50-55 ppbV. By using the BVOC emission estimates using UAH PAR, the simulated MDA8 O3 values on the average were reduced by about 3-7 ppbV.

Overall, the model overestimated isoprene concentrations by a factor of 2-3. While using satellite data reduced isoprene emissions by about 30%, most of the modeling domain over east Texas remained saturated with BVOC. Therefore, the improvements to ozone prediction using satellite data were marginal over east Texas. These simulations should be repeated using a biogenic emission model that better represents clear sky emission estimates so that the impact of using satellite data can be fully realized.

8. AUDITS OF DATA QUALITY

This project has produced new data sets and has used many established data sources for validation. The observations used in the evaluation work have been rigorously audited for quality by the organizations releasing the data. All data used in the evaluation work here have gone through a secondary quality control for quality assurance by the investigators in this project before being used in the evaluation. These checks included performing statistical analysis on the data to ensure that the data are within the allowable range and to omit any anomalous datum.

The main data produced under this project is the satellite-based photosynthetically active radiation (PAR). As described in Chapters 2 and 3, the data produced for 2013 have gone through several iterations and the final product was thoroughly evaluated. The final data will be delivered along with the final report. The size of GOES retrievals including PAR for the months of August and September 2013 is about 8.54 GB.

For the biogenic VOC emissions from MEGAN, the data were evaluated and exhibited a large over-estimation. However, the data is ready to be shared with other investigators. The total disk storage of the two months MEGAN runs is 2.9 GB for CONUS domain, 5.4 GB for domain 2 and 16.8 GB for domain 3. It can be used in CAMx simulations by using the CMAQ2CAMx interface program provided by Ramboll-Environ, http://www.camx.com/getmedia/a9e648b7-2b2d-487d-9243-2f363a6feea4/cmaq2camx-4sep13.tgz.aspx).

Another component of this project was the development of the stand-alone Berkeley-Dalhousie soil NO parameterization scheme (BDSNP). This model and associated data also will be delivered alongside this final report. A user's manual that documents in detail the model structure and how to install and operate it for specific applications is also provided. The manual as well as the tested benchmark case and the 4km resolution soil biome map will be provided as part of final delivery. The data have been completely audited for quality. The statistics about the data was provided in Chapter 5. In total 828Mb data were generated for soil emission estimates for two test cases covering two-month simulations by BDSNP module (196Mb for D1, 192Mb for D2, and 540Mb for D3). The results are archived individually and compared with the soil NO estimation from the YL95 algorithm in MEGAN and will be delivered to TCEQ for further testing.

All inputs for WRF simulations were checked for quality prior to model simulations. WRF outputs were also evaluated as described in Chapter 4. The anthropogenic emissions data used in CMAQ simulations were provided by TCEQ and went through a secondary quality check before their use in CMAQ simulations. All other model inputs were also checked for their accuracy. The CMAQ air quality outputs were evaluated as described in Chapter 6.

9. PAPERS AND PRESENTATIONS

The results from this project have been presented in several scientific conferences, including CMAS, AMS, AWMA and AGU. Currently, there are two papers under preparation to be submitted to peer-reviewed journals for publication. Below, is the listing of the papers and presentations related to this project:

1. Rui Zhang, Daniel S. Cohan, Arastoo Pour Biazar, and Erin Chavez-Figueroa. Probing the impact of biogenic emission estimates on air quality modeling using satellite Photosynthetically Active Radiation (PAR), the 13th CMAS conference, Oct 27-29, 2014, Chapel Hill, NC.
2. Rui Zhang, Daniel S. Cohan, Andrew White, Arastoo Pour Biazar, and Richard Mcnider. Incorporating Geostationary Operational Environmental Satellite (GOES) Insolation and Cloud Retrievals to Improve Biogenic Emission Estimates in Texas, the 14th CMAS conference, Oct 5-7, 2015, Chapel Hill, NC.
3. Rui Zhang, Daniel S. Cohan, Andrew White, Arastoo Pour Biazar, and Richard Mcnider. Incorporating GOES satellite photosynthetically active radiation (PAR) retrievals to improve biogenic emission estimates over Southern United States and Texas. *Atmospheric Environment*, 2015 (under preparation).
4. Arastoo Pour Biazar, Andrew White, Rui Zhang, Daniel Cohan, Richard T. McNider, Bright Dornblaser, Mark Estes. Use of Satellite Observations for Improved Air Pollution Exposure Estimates. The 95th AMS Annual Meeting, Phoenix, AZ, 4-8 January 2015.
5. Arastoo Pour-Biazar, Andrew White, Rui Zhang, Richard T. McNider, Dan Cohan, Mark Estes, Bright Dornblaser. Use of Satellite Cloud Observations for Improved Biogenic Emissions, A&WMA's 108th Annual Conference & Exhibition, June 22-25, 2015, Raleigh, NC.
6. Arastoo Pour-Biazar, Andrew White, Rui Zhang, Dan Cohan, Richard T. McNider, Mark Estes, Bright Dornblaser. Estimating photosynthetically active radiation (PAR) from geostationary satellite observations. Atmospheric Environment, 2015 (under preparation).

REFERENCES

Allen, C.W. (1963). Astrophysical Quantities, 291 pp., Athlene Press, London, England, 1963.

Allen, David, Elena McDonald-Buller, and Gary McGaughey, 2012: AQRP State of the Science Report, 2012, http://aqrp.ceer.utexas.edu/docs/FY12-13/State of the Science 2012_06042012.pdf

Byun, Daewon W., Soontae Kim, Beata Czader, David Nowak, Stephen Stetson, Mark Estes, Estimation of biogenic emissions with satellite-derived land use and land cover data for air quality modeling of Houston-Galveston ozone nonattainment area, Journal of Environmental Management, Volume 75, Issue 4, June 2005, Pages 285-301, ISSN 0301-4797, http://dx.doi.org/10.1016/j.jenvman.2004.10.009.

Coulson, K.L. (1959). Characteristics of the radiation emerging from the top of a Rayleigh atmosphere, 1 and 2, Planet Space Science, Vol. 1, pp. 256-284, 1959.

Diak, G. R. and C. Gautier (1983). Improvements to a simple physical model for estimating insolation from GOES data. J. Appl. Meteor., 22, 505–508.

Frouin, R. and R. T. Pinker (1995): Estimating photosynthetically active radiation (PAR) at the earth's surface from satellite observations. Remote Sen. Environ., 51, 98–107.

Gautier, C., G.R. Diak, S. Mass, 1980: A simple physical model for estimating incident solar radiation at the surface from GOES satellite data. J. Appl. Meteor. , 19, 1005-1012.

Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer and C. Geron (2006). "Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature)." Atmospheric Chemistry and Physics 6: 3181-3210.

Guenther, A.B., X. Jiang, C.L. Heald, T. Sakulyanontvittaya, T. Duhl, L.K. Emmons, and X. Wang (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. Geoscientific Model Development, 5, 1471-1492, DOI: 10.5194/gmd-5-1471-2012.

Haines, S. L., G. J. Jedlovec, and R. J. Suggs (2003). The GOES Product Generation System. NASA Technical Memorandum, Marshall Space Flight Center.

Hudman, R. C., A. R. Russell, L. C. Valin and R. C. Cohen (2010). "Interannual variability in soil nitric oxide emissions over the United States as viewed from space." Atmospheric Chemistry and Physics 10(20): 9943-9952. doi: 10.5194/acp-10-9943-2010.

Hudman, R. C., N. E. Moore, R. V. Martin, A. R. Russell, A. K. Mebust, L. C. Valin and R. C. Cohen (2012). "A mechanistic model of global soil nitric oxide emissions: Implementation and space based constraints." Atmos. Chem. Phys. Discuss. 12: 3555-3594.

Karl, T. G., Guenther, A., Yokelson, R. J., Greenberg, J., Potosnak, M. J., Blake, D. R., and Artaxo, P.: The tropical forest and fire emissions experiment: Emission, chemistry, and transport of biogenic volatile organic compounds in the lower atmosphere over Amazonia, J. Geophys. Res., 112, D18302, doi:10.1029/2007JD008539, 2007.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World Map of the Köppen-Geiger climate classification updated. Meteorol. Z., 15, 259-263. DOI: 10.1127/0941-2948/2006/0130.

Kota, Sri Harsha, Gunnar Schade, Mark Estes, Doug Boyer, Qi Ying (2015). Evaluation of MEGAN predicted biogenic isoprene emissions at urban locations in Southeast Texas, Atmospheric Environment, Volume 110, June 2015, Pages 54-64, ISSN 1352-2310, doi:10.1016/j.atmosenv.2015.03.027.

Lacis, A.A., and J.E. Hansen (1974). A parameterization for absorption of solar radiation in the Earth's atmosphere, J. Atmospheric Science, Vol. 31, pp. 118-133, 1973.

Mackaro, S., R. McNider, and A. Pour Biazar (2011), Some physical and computational issues in land surface data assimilation of satellite skin temperatures, Pure and Appl. Geophys., 1-14, Doi: 10.1007/s00024-011-0377-0.

McNider, R. T., W. B. Norris, D. M. Casey, J. E. Pleim, S. J. Roselle, and W. M. Lapenta (1998). Assimilation of satellite data in regional air quality models XII. Edited by Gryning, Chaumerliac, and Plenum Press, pp. 25-35. 1998.

McNider, R.T., A. Song, and S.Q. Kidder, 1995: Assimilation of GOES-derived solar insolation into a Mesoscale model for studies of cloud shading effects. Int. J. Remote Sens., 16, 2207-2231.

McNider, R.T., A. Song, D.M. Casey, P.J. Wetzel, W.L. Crosson, and R.M. Rabin, 1994: Toward a dynamic-t thermodynamic assimilation of satellite surface temperature in numerical atmospheric models. Mon. Wea. Rev., 122, 2784-2803.

Miller, S. M., Matross, D. M., Andrews, A. E., Millet, D. B., Longo, M., Gottlieb, E. W., Hirsch, A. I., Gerbig, C., Lin, J. C., Daube, B. C., Hudman, R. C., Dias, P. L. S., Chow, V. Y., and Wofsy, S. C., Sources of carbon monoxide and formaldehyde in North America determined from high-resolution atmospheric data, Atmos. Chem. Phys., 8, 7673-7696, doi:10.5194/acp-8-7673-2008, 2008.

Müller, J.-F., Stavrakou, T., Wallens, S., De Smedt, I., Van Roozendael, M., Potosnak, M. J., Rinne, J., Munger, B., Goldstein, A., and Guenther, A. B.: Global isoprene emissions estimated using MEGAN, ECMWF analyses and a detailed canopy environment model, Atmos. Chem. Phys., 8, 1329-1341, doi:10.5194/acp-8-1329-2008, 2008.

Palmer, P.I., D.S. Abbot, T.-M. Fu, D.J. Jacob, K. Chance, T.P. Kurosu, A.B. Guenther, C. Wiedinmyer, J.C. Stanton, M.J. Pilling, S.N. Pressley, B. Lamb, and A.L. Sumner, 2006: Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of formaldehyde column. Journal of Geophysical Research-Atmospheres, 111, D12315, DOI: 10.1029/2005JD006689.

Paltridge, G.W. (1973). Direct measurement of water vapor absorption of solar radiation in the free atmosphere, J. Atmospheric Science, Vol. 30, pp. 15-160, 1973.

Pinker, R. T. and I. Laszlo (1992). Global distribution of photosynthetically active radiation as observed from satellites. J. Climate, 5, 56–65.

Popescu, S. C., R. Sheridan, N.-W. Ku and S. Srinivasan (2013). Urban Vegetation for Biogenic Emissions, Report to Texas Commission on Environmental Quality.

Potter, P., Navin, R., Elena M. B., and Simon D. D.: Characterizing the spatial patterns of global fertilizer application and manure production. Earth Interactions, 14, no. 2: 1-22, 2010.

Pour-Biazar, A., K. Doty, Y-H Park, R.T. McNider (2011). Cloud assimilation into the Weather and Research and Forecast (WRF) model. Submitted to Thomas C. Ho, Lamar University, Prepared for Bright Dornblaser, Texas Commission on Environmental Quality (TCEQ), 2011.

Pour-Biazar, A., R.T. McNider, A. White (2014). Cloud Assimilation into the Weather and Research and Forecast (WRF) Model: Testing Cumulus Physics Parameters. Submitted to Bright Dornblaser, Texas Commission on Environmental Quality (TCEQ), 2014.

Pour-Biazar, Arastoo, Richard T. McNider, Shawn J. Roselle, Ron Suggs, Gary Jedlovec, Soontae Kim , Daewon W. Byun, Jerry C. Lin, Thomas C. Ho, Stephanie Haines, Bright Dornblaser, Robert Cameron. (2007), Correcting photolysis rates on the basis of satellite observed clouds, J. Geophys. Res, 112, D10302, doi:10.1029/2006JD007422.

Song, Jihee, William Vizuete, Sunghye Chang, David Allen, Yosuke Kimura, Susan Kemball-Cook, Greg Yarwood, Marianthi-Anna Kioumourtzoglou, Elliot Atlas, Armin Hansel, Armin Wisthaler, Elena McDonald-Buller, Comparisons of modeled and observed isoprene concentrations in southeast Texas, Atmospheric Environment, Volume 42, Issue 8, March 2008, Pages 1922-1940, ISSN 1352-2310, http://dx.doi.org/10.1016/j.atmosenv.2007.11.016.

Steinkamp, Jörg and Mark Gary Lawrence. "Improvement and evaluation of simulated global biogenic soil NO emissions in an AC-GCM." Atmospheric Chemistry and Physics 11.12 (2011): 6063-6082.

Stockli, R., T. Rutishauser, I. Baker, M. A. Liniger and A. S. Denning (2011). "A global reanalysis of vegetation phenology." Journal of Geophysical Research-Biogeosciences 116. doi: 10.1029/2010jg001545.

Warneke, C., de Gouw, J.A., Del Negro, L., Brioude, J., McKeen, S., Stark, H., Kuster, W.C., Goldan, P.D., Trainer, M., Fehsenfeld, F.C., Wiedinmyer, C., Guenther, A.B., Hansel, A., Wisthaler, A., Atlas, E., Holloway, J.S., Ryerson, T.B., Peischl, J., Huey, L.G. and Hanks, A.T.C. (2010). Biogenic emission measurement and inventories determination of biogenic emissions in the eastern United States and Texas and comparison with biogenic emission inventories. Journal of Geophysical Research 115: doi: 10.1029/2009JD012445. issn: 0148-0227.

Wiedinmyer, Christine, Alex Guenther, M Estes, I Strange, G Yarwood, D Allen, 2001: A land use database and examples of biogenic isoprene emission estimates for the state of Texas, USA. Atmospheric Environment, 10.1016/S1352-2310(01)00429-0.

Yienger, J. J., & Levy, H. (1995). Empirical-Model Of Global Soil-Biogenic Nox Emissions. Journal of Geophysical Research-Atmospheres, 100(D6), 11447-11464. doi: 10.1029/95jd00370.

APPENDIX A: DEFINITIONS OF EVALUATION METRICS USED IN THIS STUDY

Given the simulated time series M_i and the paired observation time series O_i with available and validate N samples at same location (or the nearest location), the following definition of statistics are used in this study for model evaluation:

Group I. Agreement and Correlation

Index of Agreement (IOA):

$$\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}(\left|M_i - \overline{O}\right| + \left|O_i - \overline{O}\right|)^2} \quad \text{(A.1)}$$

IOA is monotonically increasing linear index with the range of [0 1] to indicate the extent of agreement from null ('0') to absolute the same ('1').

Correlation Coefficient (R):

$$\mathrm{R} = \frac{\sum_{i=1}^{N}(O_i - \overline{O})(M_i - \overline{M})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2 \sum_{i=1}^{N}(M_i - \overline{M})^2}} \quad \text{(A.2)}$$

R is in the range of [-1 1] with the positive unity implying perfect correlation, the negative unity implying opposite correlation, and zero implying apparently no correlation.

Group II. Difference

Mean Bias (MB):

$$MB = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) = \overline{M} - \overline{O} \quad \text{(A.3)}$$

MB is a non-symmetric metric with the same unit as observation. Positive or negative value is corresponding with overestimation and underestimation. The range for MB is $[-\overline{O},+\infty]$.

Mean Absolute Gross Error (MAGE):

$$MAGE = \frac{1}{N}\sum_{i=1}^{N}\left|M_i - O_i\right| \quad \text{(A.4)}$$

MAGE is also a non-symmetric metric with the same unit as observation. The value is positive defined. The range for MAGE is $[0,+\infty]$.

Root Mean Square Error (RMSE):

$$RMSE = [\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2]^{1/2} \quad \text{(A.5)}$$

RMSE measures the overall deviation of simulations from the trend of observation. It is also a non-symmetric metric with the unit same as the evaluated variable. The value is positive defined. The range for RMSE is $[0,+\infty]$

Group III. Relative difference

Normalized Mean Bias (NMB):

$$\mathrm{NMB} = \frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N}O_i} \times 100\% \quad \text{(A.6)}$$

NME is relative mean error normalized by mean observation. The range is $[-100\%,+\infty]$.

Normalized Mean Error (NME):

$$\mathrm{NME} = \frac{\sum_{i=1}^{N}\left|M_i - O_i\right|}{\sum_{i=1}^{N}O_i} \times 100\% \quad \text{(A.7)}$$

NME is relative mean error normalized by mean observation. The range is $[0\%,+\infty]$

APPENDIX B: STASTICS FOR INSOLATION SIMULATION/RETREIVAL FOR DIFFERENT CASES AT 47 TCEQ NETWORK SITES

1. Statistics for case 'WRF cntrl'

SITE	OBS_AVE (W/m2)	SIM_AVE (W/m2)	IA	R	RMSE (W/m2)	MB (W/m2)	MAGE (W/m2)	NMB (%)	NME (%)
C02	219.1	256.7	0.92	0.87	166.5	36.9	88.5	16.9	40.4
C04	250.4	283.4	0.96	0.93	135.9	33.0	72.8	13.2	29.1
C08	218.1	248.3	0.93	0.87	156.0	27.0	82.7	12.4	37.9
C09	228.4	250	0.93	0.87	158.3	21.6	83.0	9.4	36.3
C12	246.4	246.3	0.93	0.88	159.7	-0.1	92.6	-0.1	37.6
C13	268.5	270.8	0.97	0.94	119.6	1.8	60.1	0.7	22.4
C15	226.2	262.1	0.93	0.89	158.2	36.0	77.8	15.9	34.4
C17	279.7	271.3	0.97	0.94	122.2	-8.3	55.8	-3.0	19.9
C19	255.3	256.5	0.95	0.91	139.3	1.2	74.2	0.5	29.1
C26	268	231.1	0.93	0.87	168.9	-36.9	87.9	-13.8	32.8
C28	210.6	259.2	0.88	0.83	197.5	48.7	106.6	23.1	50.6
C31	260.4	282.8	0.97	0.95	108.1	22.5	56.4	8.6	21.7
C35	224.3	262.1	0.93	0.89	157.7	38.9	82.3	17.4	36.7
C37	260.7	246.3	0.93	0.86	173.0	-14.4	95.0	-5.5	36.4
C38	261.9	273.3	0.96	0.93	123.9	11.5	66.9	4.4	25.6
C41	262.6	246.3	0.94	0.88	160.7	-16.3	89.3	-6.2	34.0
C43	256.9	255.9	0.95	0.91	144.0	-0.9	73.2	-0.4	28.5
C45	243.5	264.1	0.94	0.88	159.4	20.6	76.7	8.5	31.5
C52	249.4	266.4	0.96	0.92	134.1	17.0	72.9	6.8	29.2
C53	221.4	248.3	0.92	0.86	165.7	26.8	91.5	12.1	41.3
C56	246.2	278.6	0.95	0.93	134.9	32.4	70.2	13.1	28.5
C58	242.3	262.4	0.95	0.92	135.6	16.5	78.3	6.8	32.3
C61	256.9	270.8	0.96	0.93	122.4	13.9	65.5	5.4	25.5
C63	264.5	282.8	0.98	0.96	101.0	18.4	53.2	6.9	20.1
C64	224.3	265.3	0.92	0.88	169.7	41.0	87.4	18.3	39.0
C69	250.2	284.6	0.97	0.96	109.6	33.6	57.3	13.4	22.9
C70	272.4	271.3	0.96	0.93	128.2	-1.0	59.8	-0.4	22.0
C71	270.6	270.6	0.97	0.94	121.4	0.0	61.6	0.0	22.8
C73	252.8	257.7	0.94	0.90	147.1	4.9	75.7	1.9	29.9
C75	276.7	278.4	0.96	0.93	128.3	4.0	60.1	1.4	21.7
C76	261.5	283.6	0.96	0.93	128.8	22.1	65.4	8.5	25.0
C77	265.5	266.4	0.97	0.93	123.6	0.9	61.6	0.3	23.2
C78	243.8	258.3	0.96	0.93	120.4	13.6	66.1	5.6	27.1
C80	255.4	262.5	0.93	0.87	166.7	7.2	81.8	2.8	32.0
C82	263.2	286.3	0.97	0.94	117.5	22.3	60.1	8.5	22.9
C85	231.8	275.2	0.94	0.92	144.8	41.8	78.2	18.0	33.7
C87	243.9	265.3	0.95	0.91	141.8	21.5	78.3	8.8	32.1
C401	269.6	273.6	0.97	0.94	120.5	3.9	59.7	1.5	22.1
C403	213.4	248.3	0.92	0.86	167.2	34.9	89.6	16.3	42.0
C416	216.7	248.3	0.92	0.87	159.0	31.6	88.3	14.6	40.7
C559	239.1	253	0.94	0.89	146.4	13.8	76.3	5.8	31.9
C643	216.3	250	0.92	0.88	159.8	33.7	85.1	15.6	39.3
C1006	248.7	267.7	0.96	0.94	120.6	19.0	64.9	7.6	26.1
C1015	205.2	262.1	0.90	0.85	183.0	56.9	94.6	27.7	46.1
C1016	236.5	254.2	0.93	0.87	163.6	17.7	83.8	7.5	35.4
C1032	251.9	279	0.96	0.93	134.0	27.1	67.3	10.7	26.7
C1035	354.3	250	0.99	0.98	84.1	15.6	56.0	4.4	15.8

Note: IA-index of agreement, R-correlation coefficient, RMSE-root mean square error, MB-mean bias, MAGE-mean aggregate gross error, NMB-normalized mean bias, and NME-normalized mean error.

2. Statistics for case 'WRF analytical'

SITE	OBS_AVE (W/m2)	SIM_AVE (W/m2)	IA	R	RMSE (W/m2)	MB (W/m2)	MAGE (W/m2)	NMB (%)	NME (%)
C02	219.1	267.9	0.93	0.89	164.4	48.5	84.8	22.1	38.7
C04	250.4	271	0.95	0.92	136.7	20.6	73.4	8.2	29.3
C08	218.1	263.1	0.94	0.91	145.1	43.1	78.5	19.8	36.0
C09	228.4	268.3	0.94	0.90	154.0	39.9	80.1	17.5	35.1
C12	246.4	240	0.92	0.86	165.6	-6.4	98.2	-2.6	39.8
C13	268.5	270.6	0.97	0.94	116.1	0.6	58.2	0.2	21.7
C15	226.2	268.9	0.94	0.90	152.5	42.7	77.0	18.9	34.1
C17	279.7	265.8	0.96	0.92	138.1	-13.9	62.7	-5.0	22.4
C19	255.3	262.9	0.95	0.90	147.1	7.6	74.8	3.0	29.3
C26	268	247	0.94	0.88	159.6	-21.0	80.1	-7.9	29.9
C28	210.6	265.3	0.91	0.88	179.5	54.7	93.6	26.0	44.4
C31	260.4	279.4	0.97	0.94	122.8	19.1	62.5	7.3	24.0
C35	224.3	268.9	0.94	0.91	148.6	45.7	78.1	20.4	34.8
C37	260.7	240	0.92	0.85	175.9	-20.7	99.5	-7.9	38.2
C38	261.9	249	0.95	0.90	144.1	-12.8	76.4	-4.9	29.2
C41	262.6	240	0.93	0.87	166.3	-22.6	94.4	-8.6	35.9
C43	256.9	252.3	0.95	0.90	144.2	-4.6	71.7	-1.8	27.9
C45	243.5	280.1	0.95	0.92	143.5	36.5	66.9	15.0	27.5
C52	249.4	275.3	0.96	0.93	130.6	25.8	70.2	10.4	28.1
C53	221.4	263.1	0.94	0.90	151.8	41.6	84.7	18.8	38.2
C56	246.2	273.3	0.95	0.92	143.3	27.1	74.7	11.0	30.3
C58	242.3	240.1	0.94	0.88	151.4	-5.9	82.3	-2.5	34.0
C61	256.9	270.6	0.96	0.93	125.2	13.8	65.7	5.4	25.6
C63	264.5	279.4	0.97	0.95	115.4	15.0	58.6	5.7	22.2
C64	224.3	267.6	0.93	0.90	158.4	43.4	81.6	19.3	36.4
C69	250.2	282.8	0.96	0.93	130.1	31.8	66.4	12.7	26.5
C70	272.4	265.8	0.95	0.91	143.0	-6.6	64.9	-2.4	23.8
C71	270.6	277	0.96	0.93	129.1	6.4	63.4	2.4	23.4
C73	252.8	260.8	0.95	0.91	141.4	8.0	74.2	3.2	29.4
C75	276.7	267.5	0.95	0.91	147.1	-7.0	68.3	-2.5	24.7
C76	261.5	270.3	0.95	0.91	142.7	8.8	70.6	3.4	27.0
C77	265.5	270	0.97	0.94	116.4	4.5	58.4	1.7	22.0
C78	243.8	251.7	0.95	0.90	139.6	7.3	74.9	3.0	30.7
C80	255.4	279.2	0.95	0.91	145.2	23.8	72.3	9.3	28.3
C82	263.2	283	0.96	0.92	134.6	18.9	65.8	7.2	25.0
C85	231.8	276.6	0.94	0.92	143.3	43.2	77.1	18.6	33.3
C87	243.9	269.8	0.94	0.90	146.5	26.0	78.4	10.7	32.2
C401	269.6	277	0.96	0.93	125.3	7.3	61.4	2.7	22.8
C403	213.4	263.1	0.93	0.90	154.2	49.7	82.2	23.3	38.5
C416	216.7	263.1	0.93	0.90	150.9	46.4	84.4	21.4	38.9
C559	239.1	258.4	0.95	0.91	137.7	19.3	74.5	8.1	31.2
C643	216.3	268.3	0.93	0.90	162.5	52.1	87.4	24.1	40.4
C1006	248.7	272.6	0.95	0.90	147.7	23.9	76.7	9.6	30.8
C1015	205.2	268.9	0.91	0.87	179.8	63.6	94.0	31.0	45.8
C1016	236.5	289.9	0.94	0.92	152.5	53.4	75.9	22.6	32.1
C1032	251.9	287.5	0.96	0.93	131.1	35.6	67.1	14.1	26.6
C1035	354.3	268.3	0.99	0.98	84.4	18.7	54.0	5.3	15.3

Note: IA-index of agreement, R-correlation coefficient, RMSE-root mean square error, MB-mean bias, MAGE-mean aggregate gross error, NMB-normalized mean bias, and NME-normalized mean error.

3. Statistics for case 'UAH statellite'

SITE	OBS_AVE (W/m2)	SIM_AVE (W/m2)	IA	R	RMSE (W/m2)	MB (W/m2)	MAGE (W/m2)	NMB (%)	NME (%)
C02	219.1	289.7	0.95	0.96	133.3	69.2	80.4	31.6	36.7
C04	250.4	309.3	0.95	0.94	144.8	58.9	79.6	23.5	31.8
C08	218.1	289.9	0.95	0.95	139.5	66.5	83.3	30.5	38.2
C09	228.4	289.9	0.96	0.95	135.4	61.0	77.6	26.6	33.9
C12	246.4	291	0.94	0.91	157.7	43.9	92.8	17.8	37.6
C13	268.5	310.2	0.98	0.97	99.0	42.5	55.8	15.8	20.8
C15	226.2	287.2	0.95	0.94	138.2	61.2	78.3	27.0	34.6
C17	279.7	309.7	0.98	0.98	88.2	30.0	47.9	10.7	17.1
C19	255.3	307.3	0.97	0.97	113.7	52.0	67.7	20.4	26.5
C26	268	295.6	0.98	0.96	102.7	27.7	51.4	10.3	19.2
C28	210.6	285.9	0.94	0.94	151.9	75.4	87.0	35.8	41.4
C31	260.4	308.7	0.98	0.97	102.2	48.3	62.1	18.6	23.9
C35	224.3	290.8	0.96	0.96	131.5	67.7	80.8	30.2	36.0
C37	260.7	291.8	0.95	0.91	153.6	30.5	86.0	11.7	32.9
C38	261.9	303	0.97	0.96	109.3	41.2	61.7	15.7	23.6
C41	262.6	291.9	0.95	0.92	146.5	28.8	83.3	10.9	31.7
C43	256.9	302.5	0.96	0.95	128.3	45.7	69.1	17.8	26.9
C45	243.5	295.6	0.97	0.97	112.6	51.5	65.3	21.1	26.7
C52	249.4	309.8	0.97	0.97	121.9	60.5	72.9	24.3	29.3
C53	221.4	288.3	0.95	0.95	135.0	67.0	81.0	30.3	36.6
C56	246.2	306.5	0.96	0.96	132.4	60.4	74.6	24.5	30.3
C58	242.3	306.9	0.96	0.96	130.0	61.0	77.7	25.2	32.1
C61	256.9	302.6	0.98	0.97	103.0	45.8	62.3	17.8	24.3
C63	264.5	307.4	0.98	0.98	98.8	43.0	59.2	16.3	22.4
C64	224.3	286.9	0.96	0.96	122.9	62.7	74.4	27.9	33.2
C69	250.2	306.1	0.97	0.96	119.3	55.2	69.6	22.0	27.8
C70	272.4	303.9	0.98	0.97	97.6	31.5	50.7	11.6	18.6
C71	270.6	309	0.98	0.97	97.0	38.4	58.5	14.2	21.6
C73	252.8	311.8	0.97	0.97	120.8	59.1	71.9	23.4	28.4
C75	276.7	308.5	0.98	0.97	99.9	34.5	52.8	12.5	19.1
C76	261.5	301.1	0.98	0.97	97.8	39.8	57.8	15.2	22.1
C77	265.5	309.5	0.98	0.98	97.8	44.0	59.4	16.6	22.4
C78	243.8	304.4	0.97	0.96	122.2	59.2	73.9	24.3	30.3
C80	255.4	303.7	0.96	0.94	131.5	48.4	71.3	19.0	27.9
C82	263.2	305.3	0.98	0.97	102.3	42.8	59.1	16.3	22.5
C85	231.8	302	0.96	0.97	132.1	69.7	80.9	30.1	34.9
C87	243.9	303.5	0.96	0.96	125.1	59.8	75.4	24.5	30.9
C401	269.6	308.8	0.98	0.97	96.7	39.2	55.2	14.5	20.5
C403	213.4	288.9	0.94	0.94	151.4	75.6	90.6	35.4	42.5
C416	216.7	294.9	0.94	0.95	150.5	78.3	92.5	36.2	42.7
C559	239.1	295.5	0.96	0.96	122.9	56.3	70.9	23.5	29.6
C643	216.3	290.2	0.95	0.96	139.7	73.5	86.5	33.9	39.9
C1006	248.7	301.8	0.97	0.97	116.4	53.1	68.5	21.3	27.6
C1015	205.2	292.5	0.92	0.91	177.8	87.3	100.0	42.6	48.7
C1016	236.5	292.1	0.96	0.94	134.7	55.7	75.0	23.6	31.7
C1032	251.9	307	0.97	0.96	117.8	54.4	69.6	21.5	27.6
C1035	354.3	289.8	0.98	0.99	108.6	73.7	74.3	20.6	20.8

Note: IA-index of agreement, R-correlation coefficient, RMSE-root mean square error, MB-mean bias, MAGE-mean aggregate gross error, NMB-normalized mean bias, and NME-normalized mean error.