The Pennsylvania State University

The Graduate School

College of Earth and Mineral Sciences

EVALUATION OF ATMOSPHERIC TRANSPORT AND ITS IMPACT ON CO_2 MIXING RATIOS

A Dissertation in

Meteorology

by

Liza Ivelisse Díaz

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2017

The dissertation of Liza Ivelisse Díaz was reviewed and approved* by the following:

Kenneth J. Davis
Professor of Meteorology
Dissertation Co-Advisor, Chair of Committee

Thomas Lauvaux
Senior Research Associate & Associate Professor of Meteorology
Dissertation Co-Advisor

Natasha L. Miles
Research Associate

Chris E. Forest
Associate Professor of Climate Dynamics

Fuqing Zhang
Professor of Meteorology
Director of Center for Advances Data Assimilation and Predictability Techniques

Andrew Carleton
Professor of Geography

Johannes Verlinde
Professor of Meteorology
Associate Head, Graduate Program in Meteorology

*Signatures are on file in the Graduate School

ABSTRACT

Atmospheric inversions are used to assess biosphere-atmosphere CO_2 surface exchanges at various scales. Recently, higher resolution inversions were performed using mesoscale models to improve the spatial and temporal resolution of these inversions, but variability among inverse flux estimates remains significant. One of the main contributors to these uncertainties is the atmospheric transport model errors. Past studies have used ensembles to understand these transport model uncertainties, but have some limitations including the small number of measurements, coarse resolution of the models, small number of members, centered only in the variation of Planetary Boundary Layer (PBL) schemes and no assessment was performed to test whether the spread of the ensemble represents the true uncertainty.

For this work, I evaluate and quantify the transport model errors with a large set of simulations generated with the Weather Research and Forecasting (WRF) mesoscale model. The large ensemble of 45-members was constructed using different physics parameterizations (i.e., land surface models (LSMs), planetary boundary layer (PBL) schemes, cumulus parameterizations and microphysics parameterizations) and initial/boundary conditions. All the different models were coupled to CO_2 fluxes and lateral boundary conditions from CarbonTracker to simulate CO_2 mole fractions. I evaluate the atmospheric transport errors over a highly instrumented area, the Mid-Continental Intensive (MCI) region, for 2008 summer period. Both modeled meteorological variables (i.e., wind speed, wind direction and PBL height) and CO_2 mixing ratios are compared to observations to evaluate the performance of the different models and the ensemble.

In Chapter 2, I performed statistical analyses to evaluate the impact of both physics parameterizations and the meteorological dataset on CO_2 mixing ratios and meteorological variables. The different model configurations show varying performances across the region that impede the selection of an optimal solution or least biased simulation for all the meteorological variables except for PBL height (PBLH). In general, physical parameterizations contribute equally to the model-to-model variability in atmospheric CO_2 and meteorological variables, with the microphysics parameterization being the exception. It was also found that daily variations in CO_2 mole fractions across the region are correlated primarily with errors in the PBLH. In Chapter 3, I introduce two calibrations (or down-selection) methodologies using Simulated Annealing (SA) and Genetic Algorithm (GA) over 2008 summer. I applied the calibration process to the multi-physics/multi-analysis ensemble of 45-members to select the optimal ensemble using the flatness of the rank histogram as the main criteria. The calibrated ensemble representing the model errors is based on all three meteorological variables. Using multiple model configurations (i.e. 45 configurations of varying physics), I show that a reduced number of simulations (less than 10 members) is sufficient to characterize the transport errors, reproducing the statistics of the model-data differences while minimizing the size of the ensemble. The CO_2 error correlations of the calibrated ensembles were compared to the large ensemble to identify any impact of the calibration. Compared to the initial error structures, the calibrated ensembles revealed sampling noise across the region which indicates that additional filtering or modeling of the errors would be required to construct the error covariance matrix for regional CO_2 inversion.

Using the multi-physics and multi-analysis ensemble, I showed the importance that other physics parameterization besides the PBL schemes has on the atmospheric CO_2 mixing ratio errors. In addition, the challenges that future atmospheric inversions still need to confront including correction of systematic biases and representation of errors, to avoid the propagation of these errors into inverse fluxes.

TABLE OF CONTENTS

List of Figures vi

List of Tables x

Acknowledgements xi

Chapter 1 INTRODUCTION 1

Chapter 2 Sensitivity and Uncertainty Analysis of Physical Parameterization and Initial Conditions on Meteorological Variables and CO_2 Mole Fractions 14

2.1 INTRODUCTION 14
2.2 METHODS 18
2.2.1 Region 18
2.2.2 Atmospheric Model Setup 18
2.2.3 Ensemble Configuration 19
2.2.3 Physics Parameterization Schemes 21
a. Land Surface Models (LSMs) 21
b. Planetary Boundary Layer (PBL) Schemes 22
c. Cumulus Parameterizations 23
d. Microphysics Parameterizations 24
2.2.4 Meteorological Initial and Boundary Conditions 24
2.2.5 CO_2 Surface Fluxes 25
2.2.6 Datasets 26
2.2.7 Data Selection 26
2.2.8 Evaluation Methodology or Analyses of the Models 27
2.3. RESULTS 29
2.3.1 Impact of physics parameterizations on the PBL dynamics 31
2.3.2 Meteorological day-to-day variability 33
2.3.3 Characterization of transport errors 36
2.3.3.1 Mean Absolute Error (MAE) 36
2.3.3.2 Root mean square error (RMSE) 39
2.3.3.3 Mean Bias Error (MBE) 42
2.3.4 Best Model Performance 48
2.3.5 Sensitivity of CO_2 mixing ratios to model configurations 52
2.4. DISCUSSION 56
2.5. CONCLUSION 59

Chapter 3 Calibration of a Multi-Physics Ensemble for Greenhouse Gas Atmospheric Transport Model Uncertainty Estimation 61

3.1 INTRODUCTION 61
3.2 METHODS 66
3.2.1 Generation of the Ensemble 66
3.2.2 Dataset and Data Selection 68
3.2.3 Criteria or Verification Methods 70
3.2.3.1 Talagrand Diagram (or Rank Histogram) and Rank Histogram Score 71

3.2.3.2 Spread-skill relationship 72
3.2.3.3 Bias and Standard Deviation 73
3.2.3.5 Ensemble-based error covariances 74
3.2.4 Calibration Methods 75
3.2.4.1 *Simulated Annealing* 76
3.2.4.1 *Genetic Algorithm* 77
3.2.5 Parameterization of the Selection Algorithms 78
3.2.6 Selection of the Optimal Reduced-sized Ensembles 79
3.3 RESULTS 81
3.3.1 Evaluation of the Large Ensemble 81
3.3.1.2 Model Skill 81
3.3.1.3 Reliability and Spread of the Ensemble 82
3.3.2 Calibrated Ensemble 86
3.3.2.1 Individual Variable Calibration 87
3.3.2.2 Multiple Variable Calibration 89
3.3.2.3 Evaluation of the Multiple Variable Calibrated Ensemble 93
3.3.3 Propagation of Transport Errors into CO_2 Concentrations 98
3.3.3.1 CO_2 Error Variances 98
3.3.3.2 CO_2 Error Correlations 101
3.4 DISCUSSION 103
3.4.1 Impact of Calibration on Ensemble Statistics 103
3.4.2 Single-variable and Multiple-variable Ensembles 103
3.4.3 Resolution and Reliability 104
3.4.4 Error Correlations 105
3.5 CONCLUSION 105

Chapter 4 CONCLUSION 107

Bibliography 114

LIST OF FIGURES

Figure 1-1. Diagram of a typical atmospheric inversion system that corrects only for errors in prior fluxes. 6

Figure 1-2. Mid-Continental Intensive (MCI) field campaign study region 9

Figure 2-1. Geographical domain used by WRF-ChemCO_2 physics ensemble. The parent domain (d01) has a 30-km resolution, the inner domain (d02) has a 10-km resolution. 19

Figure 2-2. Map of the study region, including the CO_2 towers (blue triangles) and rawinsonde sites (red circles) locations. 26

Figure 2-3. Sensitivity of CO_2 mole fractions as a function of model physics parameterizations (i.e., land surface model (LSM), planetary boundary layer scheme (PBL), cumulus parameterization (CP), microphysics parameterization (MP) and Reanalyses). The root mean square difference (RMSD) of the CO_2 mole fractions simulated at each site and for each model ensemble member was computed by varying only the type of physics parameterization noted, and keeping all other model elements constant. RMSD was averaged across sites and across model ensembles. 32

Figure 2-4. Root mean square difference (RMSD) of the PBLH (a), wind speed (b) and wind direction (c) for the different physics parameterizations (i.e., land surface model (LSM), planetary boundary layer scheme (PBL), cumulus parameterization (CP), microphysics parameterization (MP) and Reanalyses). 33

Figure 2-5. Observed (black line) and simulated (colored lines) PBL (300 m AGL) wind speed (a), wind direction (b) and PBLH (c) at time 0000 UTC from day of the year (DOY) 169 to 203 of 2008 at the Lincoln, Nebraska (OAX) rawinsonde site. 35

Figure 2-6. Maximum and minimum monthly average wind speed (a), wind direction (b) and PBLH (c) mean absolute errors (MAE) for the different sites and models. The abscissa shows the different model configurations, while the ordinate shows the MAE for the meteorological variables. Each dot represents the site with the maximum and minimum MAE for each model and variable; while the x represent the regional mean of the MAE. (See Table 2-1 for model references.) 37

Figure 2-7. Regional averages of the monthly average wind speed (a), wind direction (b) and PBLH (c) mean absolute errors (MAE) for the different models (see Table 2-1 for model references). 39

Figure 2-8. Maximum and minimum monthly average wind speed (a), wind direction (b) and PBLH (c) root mean square errors (RMSE) for the different sites and models. The abscissa shows the different model configurations, while the ordinate shows the RMSE for the meteorological variables. Each dot represents the site with the maximum and minimum RMSE for each model and variable; while the x represents the regional mean of the RMSE. (See Table 2-1 for model references.) 40

Figure 2-9. Regional averages of the monthly average of wind speed (a), wind direction (b) and PBLH (c) RMSE for the different models (see Table 2-1 for model references). 42

Figure 2-10. Monthly average wind speed (a-c), wind direction (d-f) and PBLH (g-i) MBE for rawinsonde sites ABR (first row), DVN (second row) and BNA (third row). Models are sorted from the negative to the positive bias. 44

Figure 2-11. Regional average of the monthly average of wind speed (a), wind direction (b) and PBLH (c) bias for the different models (See Table 2-1 for model references.) .. 46

Figure 2-12. Ensemble mean of the mean bias error (MBE) for PBL wind speed (a), PBL wind direction (b) and PBLH (c). 47

Figure 2-13. Observed (black line) and simulated (colored lines) DDA CO_2 mixing ratio (ppm) at Centerville (RCV) (a) and Kewanee (RKW) (b). Residuals (model-data mismatch) of DDA CO_2 mixing ratios at Centerville (RCV) (c) and Kewanee (RKW) (d). 53

Figure 2-14. Taylor Diagram comparing observations versus simulations at (a) Round Lake (RRL) and (b) Centerville (RCV), using DDA CO_2 mole fractions. Black dot represents the observation reference. 54

Figure 2-15. Tower and rawinsonde site specific spatial correlation coefficients between ensemble mean MBE of (a) wind speed, (b) wind direction and (c) PBLH and ensemble mean MBE of DDA CO_2 mole fractions. The abscissa shows the different CO_2 tower sites, while the ordinate shows rawinsonde sites. 55

Figure 3-1. Geographical domain used by WRF-ChemCO_2 physics ensemble. The parent domain (d01) has a 30-km resolution, the inner domain (d02) has a 10-km resolution. 67

Figure 3-2. Map of the study region, including the CO_2 towers (blue triangles) and rawinsonde sites (red circles) locations. 70

Figure 3-3. Diagram of the process of selection of reduced-sized ensembles. 79

Figure 3-4. Box plot of the rank histogram scores of the different sub-ensembles of 10 (a), 8(b), and 5 (c) members accepted by the SA. Each figure shows the rank histograms scores for the different variables wind speed (wspd), wind direction (wdir) and PBLH. The outliers are plotted using the + symbol. 80

Figure 3-5. Taylor diagram comparing the 0000 UTC rawinsonde observations (300 m wind speed (a), 300 m wind direction (b) and PBL height (c)) to the 45 model configurations (red circles). 82

Figure 3-6. Time series of the large ensemble for wind speed (a-b), wind direction (c-d) and PBL height (e-f) at GRB (a,c,e) and TOP (b,d,f) sites. The shaded blue area

represents the spread (i.e. RMSD) of the ensemble, the solid line the ensemble mean and the red dots the observations at 0000UTC..... 83

Figure 3-7. Rank histogram of the 45-member ensemble for wind speed (a), wind direction (b) and PBLH (c) using the 14 rawinsonde sites available over the region..... 84

Figure 3-8. Spread-skill for (a) wind speed, (b) wind direction and (c) PBL height using the 14 rawinsonde sites available over the region. Each point represents the spread-skill of the different sites at different times. A one-to-one line is plotted in black and a line of best fit is plotted in red. Correlation (r) and slope (b) of the line of best fit of the spread-skill relationship. 86

Figure 3-9. Rank histograms of the calibrated ensembles found for wind speed (a, d, g), wind direction (b, e, h) and PBL height (c, f, i) for each of the ensemble size. The upper, middle and lower panels correspond to the ensemble with 10, 8, and 5 members, respectively..... 88

Figure 3-10. Rank histograms of wind speed (a, d, g), wind direction (b, e, h) and PBL height (c, f, i) using the calibrated ensembles found with simulated annealing. The upper, middle and lower panels correspond to the ensemble with 10, 8, and 5 members, respectively..... 90

Figure 3-11. Rank histograms of wind speed (a, d, g), wind direction (b, e, h) and PBL height (c, f, i) using the calibrated ensembles found with genetic algorithm. The upper, middle and lower panels correspond to the ensemble with 10, 8, and 5 members, respectively. 92

Figure 3-12. Spread-skill for (a) wind speed, (b) wind direction and (c) PBL height using the 14 rawinsonde sites available over the region using the 5-member calibrated ensemble. Each point represents the spread-skill of the different sites at different times. A one-to-one line is plotted in black and line of best fit is plotted in red. Correlation (r) and slope (b) of the line of best fit of the spread-skill relationship. 94

Figure 3-13. Time series of simulated wind speed (a-c), wind direction (d-f) and PBL height (g-i) using the 5-, 8- and 10-member calibrated ensembles at the TOP rawinsonde site. The green shaded area represents the spread (i.e., Root Mean Square Deviation) of the ensemble, the black line is the mean of the ensemble and the red dots are the observations at 0000UTC. 95

Figure 3-14. Frequency of the physics schemes used for each of the calibrated ensemble of 10 members (a-b), 8-members (c-d) and 5-members (e). 97

Figure 3-15. Bias of individual members for wind speed (a), wind direction (b), PBL height (c) using the SA and GA calibrated sub-ensemble of five members. 98

Figure 3-16. Spread (i.e., RMSD) of the DDA CO_2 concentrations of Mead (first column a,d,g,j), WBI (middle column b,e,h,k) and WLEF (last column c,f,i,l) using Simulated Annealing calibrated ensembles. Rows from top to bottom are 45, 10, 8 and 5 member ensembles. The blue area is the spread of the 45-member ensemble,

green area is the spread is the spread of the calibrated (10-, 8- and 5-member) ensemble, the black line is the mean of the ensemble and the red dots are the observations. 100

Figure 3-17. Sum of variance of the SA (a) and GA (b)..... 101

Figure 3-18. Spatial correlation of CO_2 across the 45- (a), 10-(b), 8-(c) and 5-members (d) ensembles with respect to the location of the Round Lake tower for DOY 180. This figure use calibrated ensembles of 10-, 8-, and 5-members found by SA technique. 102

LIST OF TABLES

Table 2-1. Different model configurations used in this study. 20

Table 2-2. Parameters included in the sensitivity analysis. 21

Table 2-3. Regional average of wind speed, wind direction and PBLH mean absolute error (MAE) for each of the physics schemes. 49

Table 2-4. Regional average of wind speed, wind direction and PBLH root mean square error (RMSE) for each of the physics schemes..... 50

Table 2-5. Regional average of wind speed, wind direction and PBLH mean bias error (MBE) for each of the physics schemes..... 51

Table 3-1. Physics schemes used in WRF for the sensitivity analysis. 67

Table 3-2. Rank histogram score (δ), biases and standard deviation (σ) of the 45-member ensemble for wind speed, wind direction and PBL height computed across 14 rawindsonde sites using daily 0000 UTC observations for June 18 to July 21 of 2008 in the upper Midwest of the U.S. 84

Table 3-3. Calibrated ensembles generated by SA and a GA and their rank histograms scores and bias for each variable..... 88

Table 3-4. Ensemble members, Rank histogram scores (δ), bias, and standard deviation (σ) for wind speed, wind direction and PBLH for the calibrated sub-ensembles generated with SA. 90

Table 3-5. Ensemble members, Rank histogram scores (δ), bias, and standard deviation (σ) for wind speed, wind direction and PBL height for the calibrated sub-ensembles generated with GA. 91

ACKNOWLEDGEMENTS

My doctoral research journey has been an adventure filled with challenges and learning experiences; none of this would have been possible without the help of extraordinary people who have stimulated my personal and professional growth throughout this period of my life. In this section, I will try my very best to express my gratitude to everyone who has been part of this process.

I would like to thank my Ph.D. committee, particularly my adviser Dr. Kenneth J. Davis, co-adviser Dr. Thomas Lauvaux, as well as Dr. Natasha L. Miles, Dr. Chris E. Forest, Dr. Fuqing Zhang and Dr. Andrew Carleton, for guiding me during this project. I am also grateful to Dr. Davis for his unconditional support throughout my research and for always believing in me regardless of all the initial struggles that I had in the program as an incoming student from a different field. I also owe my deepest gratitude to Dr. Lauvaux for all the patience that he had with me during this entire process. Dr. Lauvaux's technical help and guidance have made possible the completion of this research. Thanks to the help and guidance provided by Dr. Davis and Dr. Lauvaux I have been able to accomplish one of the biggest aspirations of my life. I would also like to take this opportunity to acknowledge Dr. Marc Bocquet for providing the codes of Simulated Annealing and Genetic Algorithm used in Chapter 2, also for all the helpful discussions about this project. I would also like to thanks all my colleagues from Dr. Davis' research group for all the help and contribution to this research.

I am deeply grateful to my family and friends for their invaluable support throughout all these years of graduate school at The Pennsylvania State University. I would like to thank my parents, for the unconditional support and encouragement to pursue my dreams. I would also like to show my appreciation to my siblings, brother-in-law and sister-in-law for all the support and the

good laughs that are so needed during this process. I would like to recognize Viktoria, Natsumi and Evolet my nieces and also my small light of hope during the most difficult moment of this research.

Embarking into this journey also required me to leave family and friends, and move into a different country. During this process, I met friends like Lymaris Ortiz Rivera, Dr. Miguel Santiago Cordoba and Dr. Roberto Albandoz Carrasquillo, who have been the greatest support that I could ever have during this process and the more than friends have become my family. In this regard, I should also take the opportunity to thanks additional friends who have played supporting role in this process: Rafael Díaz, Dr. Emil Hernandez, Dr. Sandra Rodriguez, Dr. Luna Rodriguez, Daniel Sarmiento, Caroline Normile, Dr. Nella Vargas, Dr. Isamar Ortiz, Dr. Lourdes Medina and many others that have been crucial of this achievement.

This work was possible thanks to the accessibility of the data provided by NOAA Earth System Research Laboratory, University of Wyoming ‘s online data archive and for the great effort done by Dr. Miles and Dr. Richardson with in-situ CO_2 measurements data base.

This research was supported by NASA Terrestrial Ecosystem and Carbon Cycle Program, NASA’s Earth Venture Program and Alfred P. Sloan Graduate Fellowship.

Chapter 1

INTRODUCTION

The increase in atmospheric carbon dioxide (CO_2) concentration has a direct influence on the radiative budget of the Earth and is the main driver of climate change since the world's industrialization (IPCC, 2013). Global atmospheric CO_2 has increased from 280 ppm at the start of the Industrial Revolution to over 400 ppm in recent years (Canadell et al., 2007; Dlugokencky and Tans, 2015). The atmospheric CO_2 increase is primarily due to fossil fuel combustion and, to a lesser extent, land use change (Houghton et al., 1999; IPCC, 2013; Le Queré et al., 2015). Not all the CO_2 emitted remains in the atmosphere, because the terrestrial biosphere absorbs about 30% of the emissions (Battle et al., 2000; Sarmiento et al., 2010; Le Quere et al., 2015). Several studies have identified the temperate latitudes of the Northern Hemisphere as a large net sink (Tans et al., 1990; Ciais et al., 1995; Gurney et al., 2002; Houghton et al., 2002; Sarmiento et al., 2010; Pan et al., 2011; Le Queré et al., 2015). However, the specific magnitudes and distributions of terrestrial sources and sinks are still uncertain. Accurate and precise quantification of terrestrial fluxes is an important step toward successful prediction of future atmospheric CO_2.

Terrestrial fluxes over continents can be estimated using two broad classes of methods: (1) "bottom-up" and (2) "top-down". The "bottom-up" methods include either modeling of biogeochemical processes or evaluation of carbon stock changes from crop and forest inventories (e.g., King et al., 2007; Houghton et al., 2002; Pan et al., 2011; Fang et al., 2014). The "top-down" method or atmospheric inverse method uses atmospheric transport models to estimate carbon fluxes by adjusting these fluxes to be optimally consistent with observed CO_2 concentrations (e.g., Enting, 1999; Bousquet et al., 2000; Chevallier et al., 2010). Bottom-up methods estimate carbon budgets

from local-to-global scales by simulating land-atmosphere exchange processes for the terrestrial biosphere (Fung et al., 1983), collecting fossil fuel consumption data (Marland and Rotty, 1984), estimating emissions from forest fires (Van der Werf et al., 2010), and other minor sources (i.e., oxidation of other trace gases). The atmospheric inversion method is an independent approach to the "bottom-up" method, and therefore has been compared at global (Peylin et al., 2013) and regional scales (Ogle et al., 2015). Both methods yield a wide range of results (Houghton et al., 2002; Sarmiento et al., 2010; King et al., 2015), and even when flux estimates overlap between the two methods, large uncertainties still remain (Pacala et al., 200l; Janssens et al., 2003). Uncertainties are caused by the limitations of each method, and the sources of uncertainties differ. The methods differ because each incorporates different types of data, captures different processes, and uses different spatial and temporal resolutions. Many investigations have documented large uncertainty and variability among inverse flux estimates (Gurney et al., 2002; Baker et al., 2006; Sarmiento et al., 2010; Peylin et al., 2013). Different sources of uncertainty in the inverse method include limited atmospheric CO_2 data (Gurney et al., 2002), uncertain prior flux estimates (Huntzinger et al., 2012), and atmospheric transport uncertainty (Stephens et al., 2007; Gerbig et al., 2008; Picket-Heaps et al., 2011).

Early atmospheric inversions of CO_2 sources and sinks performed at global scale used observations from remote locations such as mountain tops and ocean sites, to avoid the misrepresentation of large and highly variable CO_2 fluxes over continents (e.g., Keeling et al., 1989; Tans et al., 1990). While these remote locations provide information about the composition of the "background" atmosphere, the location and magnitude of sources and sinks remain under-constrained. Continental sites are required to understand the mechanisms driving carbon sources and sinks (Denning et al., 1995). As the information content of observations increases, accurate interpretation of these data requires data assimilation systems that can simulate the spatial variability of terrestrial fluxes and the complexity of atmospheric dynamics over land (Law et al.,

2003). In the last decade, continental networks consisting of tall towers or airborne measurements have been deployed throughout Northern Hemisphere (e.g., National Oceanic and Atmospheric Administration-Earth Systems Research Laboratory (NOAA-ESRL); CarboEurope) to better constrain terrestrial vegetation fluxes. The deployment of these continental atmospheric CO_2 measurement networks helps constrain inverse fluxes mainly in the Northern Hemisphere, reducing uncertainty in carbon flux estimates from the continental biosphere (Gurney et al., 2002; Baker et al., 2006; Patra et al., 2008; Butler et al., 2010).

Adding observations to atmospheric inversions can reduce the uncertainty in inverse flux estimates. Increasing the spatial resolution of atmospheric inversions may also increase atmospheric transport uncertainties if atmospheric models are not able to represent the atmospheric processes that cause observed variability in CO_2 mole fractions over land in the atmospheric boundary layer (Gurney et al., 2002;2008, Patra et al., 2008). At regional scales, domain-limited inversions require an atmospheric CO_2 model that is capable of reproducing both large interannual and day-to-day variability in atmospheric CO_2 over land (Ahmadov et al., 2007). Such a model is required to avoid additional flux biases due to an incorrect interpretation of atmospheric signals (Law et al., 2008; Patra et al., 2008). Recent inversion systems have replaced coarse resolution models (Global Circulation models (GCM)), which caused large representation errors over heterogeneous and complex terrain (Geels et al., 2007), with mesoscale models able to capture the local variability in CO_2 observations (e.g. Lauvaux et al., 2008). To improve the simulation of both atmospheric dynamics over such terrain and high spatial variability in fluxes, higher resolution transport models have been evaluated during intensive observation campaigns of greenhouse gases (GHGs) at regional scales (Sarrat et al., 2007).

In the past few years, several inversions were performed using higher resolution transport models (e.g. Lauvaux et al., 2009a; Schuh et al., 2010; Göckede et al., 2010; Lauvaux et al., 2012) to obtain better representations of atmospheric dynamics and flux spatial variability. Although

regional inversions benefit from increased atmospheric transport model and flux resolutions (Schuh et al., 2013), several limitations affect these systems including the sparsity of regional measurement networks, errors in the boundary conditions (Göckede et al., 2010), lack of detailed prior fluxes and prior error structures (Schuh et al., 2013), and uncertainty in mesoscale models caused by different parameterizations of atmospheric physics (Lauvaux and Davis, 2014). Several regional inversions have used short-term intensive field campaigns such as CarboEurope Regional Experiment Strategy (CERES; Dolman et al., 2006), Ameriflux sites in Oregon (Göckede et al., 2010) or Mid-Continental Intensive (MCI) field campaign (Miles et al., 2012; Lauvaux et al., 2012) to solve for the increasing number of unknowns in the state vector (i.e., increased flux resolution) (Law et al., 2003). Boundary conditions (i.e. the inflow of CO_2 in the simulation domain) used in regional inversions can cause large errors in the regional carbon balance because this inflow is provided by global CO_2 models and depends on their ability to simulate continental and oceanic air masses (Göckede et al., 2010; Gourdji et al., 2010). Therefore, additional data are required to reduce the uncertainty of the boundary conditions and inform the system of any potential biases (Lauvaux et al., 2012; Schuh et al., 2013). True surface fluxes can be highly variable in space and time, and biogeochemical models and fossil fuel inventories may fail to represent the large gradients across the landscape (Lokupitiya et al., 2009). Prior flux errors can only be prescribed based on modeled error covariances in order to regularize the assimilation of local tower data (Bocquet et al., 2005). For the last critical component of inversion systems, that is atmospheric transport, higher resolution data require atmospheric models able to represent multiple drivers of the atmospheric dynamics at high resolution such as the terrain-driven dynamics (Perez-Landa et al., 2007), coastal dynamics (Sarrat et al., 2007), variable land surface properties, and the planetary boundary layer dynamics (Lauvaux and Davis, 2014). Because transport errors remain difficult to estimate, regional inversions will require new approaches to characterize their magnitude and their structures, more

advanced than the simplified description of transport errors in global inversions (i.e., variances only).

Despite the progress in the top-down methodologies (e.g., additional observations and higher model resolutions), the impact of the atmospheric transport errors on inverse flux estimates remained a major uncertainty. Inversion systems assume perfect transport models (i.e., unbiased) with frameworks that are still limited to the optimization of the biogenic surface fluxes alone. However, model-data mismatches, used to optimize the fluxes, contain the contribution of both flux and transport errors (Figure 1-1). Therefore, the atmospheric inversions will systematically attribute atmospheric CO_2 model-data mismatches to the surface fluxes without considering the potential biases in transport models. In a Bayesian framework, the atmospheric inversion assumes (1) atmospheric transport model errors are unbiased and (2) the random errors are known. Inconsistent errors (i.e., random and systematic) will be propagated into the state space by the optimization process, generating biased inverse (i.e., posterior) fluxes (Tarantola, 2004). Atmospheric inverse CO_2 flux estimates will be reliable only if random atmospheric transport errors are quantified rigorously and the transport models are unbiased.

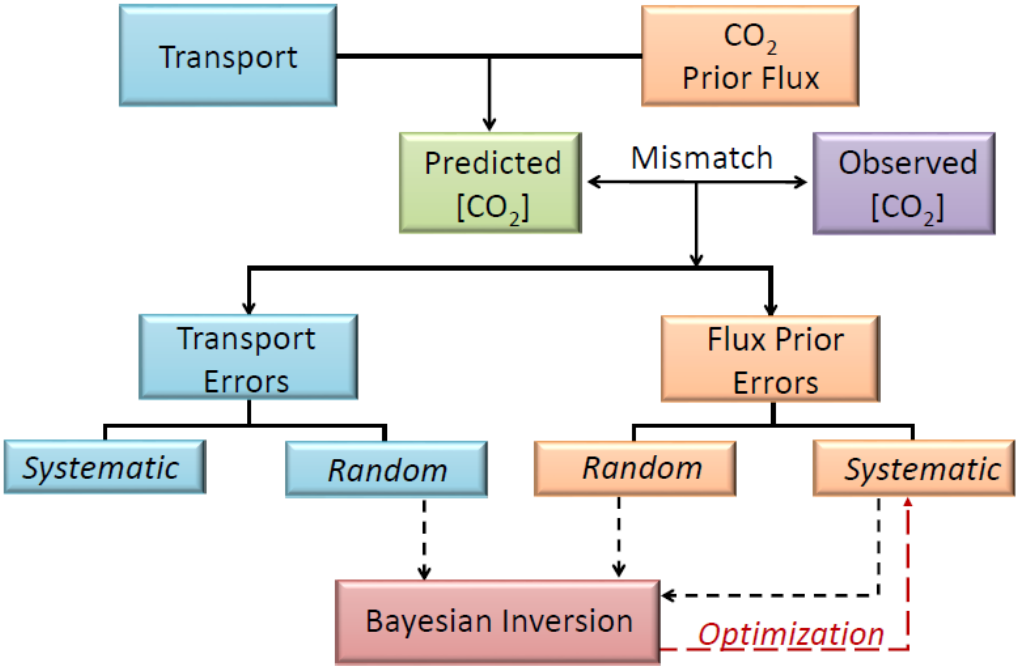

Figure 1-1. Diagram of a typical atmospheric inversion system that corrects only for errors in prior fluxes.

Few studies have quantified these transport errors thoroughly. The Atmospheric Transport Model Intercomparison Project (TransCom) has been dedicated to evaluate and to quantify the impact of the atmospheric transport errors on CO_2 fluxes using a model-intercomparison framework in a series of published research studies (Gurney et al., 2002; Baker et al., 2006; Stephen et al., 2007; Patra et al., 2008; Peylin et al., 2013). To approach the model errors, the experiment used the assumption that the spread of a model-ensemble can represent the uncertainty among inverse estimates, which is mostly due to atmospheric transport model errors. TransCom studies were not only limited to varying the transport models but also varied the number of observations and the inverse methodologies. Among other results, some of the latest Transcom studies concluded that only an atmospheric transport model capable of representing synoptic and mesoscale atmospheric dynamics will be able to extract high-resolution information from the atmospheric observations (Law et al., 2008; Patra et al., 2008). More recently, a different study clarified the

contribution of model errors by using identical surface fluxes with models of different resolutions (Díaz Isaac et al., 2014). It concluded that transport alone has a significant impact on interpretations of the atmospheric CO_2 signals that can lead to large differences in the inverse fluxes (Díaz Isaac et al., 2014). Going further into the processes responsible for these model differences, this study showed that not only the resolution of the atmospheric transport model impacts both CO_2 mixing ratios and fluxes; errors in sub-grid parameterizations such as the vertical mixing can have a significant impact on the simulated CO_2 mole fractions (Stephens et al., 2007).

Several studies that considered the misrepresentation of vertical mixing as a critical error in atmospheric transport models, assumed these errors were caused by the incorrect representation of the PBL. Therefore, efforts have been made to evaluate the mixing layer (ML) depth (Denning et al., 1995; Stephen et al., 2007; Gerbig et al., 2008; Williams et al., 2011; Kretschmer et al., 2012) to explore the vertical mixing errors. The misrepresentation of vertical mixing by TransCom's atmospheric models shown by Stephen et al. (2007) led Gerbig et al. (2008) to evaluate the uncertainty in ML depth when using a global model and found errors of several ppm in the atmospheric CO_2 mixing ratios. Sarrat et al. (2007) used an inter-comparison of five mesoscale models and identified large discrepancies in the ML depth impacted the atmospheric CO_2 mixing ratios. Most of these studies have attributed the causes of the differences between simulated and observed ML depths to the parameterizations of the PBL (i.e., PBL schemes and land surface models (LSMs)).

At the regional scale, inter-comparisons of physical parameterizations have quantified the impact of physics schemes on the vertical distribution of CO_2 mixing ratios (Kretschmer et al., 2012; Yver et al., 2013; Lauvaux and Davis, 2014; Feng et al., 2016). These studies have found systematic errors in atmospheric CO_2 that can lead to biased flux estimates when used in a CO_2 flux inversion. Nevertheless, these studies were limited in several ways. Some of them performed their evaluations with pseudo-data experiments (e.g. Lauvaux and Davis, 2014), or used a limited

number of observations (e.g. Kretschmer et al., 2012), and more importantly their primary focus was on the impact of using different PBL physics schemes (i.e., sensitivity to the physics schemes). However, the vertical mixing is not solely affected by the PBL parameterizations. Therefore, the extent of previous transport evaluation studies on vertical mixing as a function of different PBL parameterizations within atmospheric transport models is incomplete.

The ability of atmospheric transport models to simulate CO_2 mole fractions depends on different components of the models, generally described as the boundary conditions, the initial conditions, and the model physics parameterizations. Each of these components contributes to the modeling performances with varying levels of accuracy in their representation of atmospheric variability. Flux CO_2 errors are not considered here because the optimization procedure in CO_2 inversions treats them separately (i.e., prior fluxes and model-data mismatches are independent from each other). To represent model errors and propagate the uncertainties through the optimization algorithm, different approaches have been used in the carbon cycle community, as follow: (1) multi-model ensembles that encompass models from different research institutions around the world (e.g., TransCom experiment; Gurney et al., 2002; Baker et al., 2006; Peylin et al., 2013), (2) multi-physics ensembles involving different physics configurations generated by variations of different combinations of schemes (e.g., Kretschmel et al., 2012; Lauvaux and Davis, 2014; Feng et al., 2016), and perturbed ensembles consisting of multiple model simulations using different analysis fields and initial conditions (i.e. perturbations are added to the reference state) (e.g., Miller et al., 2015). While several studies have used ensembles as a technique to explore the transport errors, some of these ensembles do not vary only the transport but also the amount of observations, or the prior fluxes, and have been mostly focused on the PBL scheme parameterizations. More importantly, robust evaluations of model ensembles were missing with no guarantee that the ensemble of simulations statistically represent the transport errors.

In the present work, I will to quantify the uncertainty of atmospheric transport model in a more accurate and robust framework able to translate transport errors into uncertainty estimates on the atmospheric CO_2 mixing ratios and CO_2 surface fluxes. To explore the atmospheric transport errors, I will select a region that offers a dense network of CO_2 measurement to allow me to constrain the CO_2 fluxes and therefore avoid the lack of observation experienced in previous studies. I selected the Mid-Continental Intensive (MCI) domain, the first measurement campaign of the North American Carbon Program (NACP) that supported the deployment of a high-density measurement network (Ogle et al., 2006) over several years (2007-2009). The Midwest agricultural belt in the north central United States (U.S.) includes Iowa and surrounding states, which is also known as the "Corn-Belt" (Figure 2-2). The MCI campaign is unique for the unprecedented amount of atmospheric CO_2 mixing ratio observations and therefore an ideal experiment to constrain the carbon budget over this region with an unprecedented level of confidence.

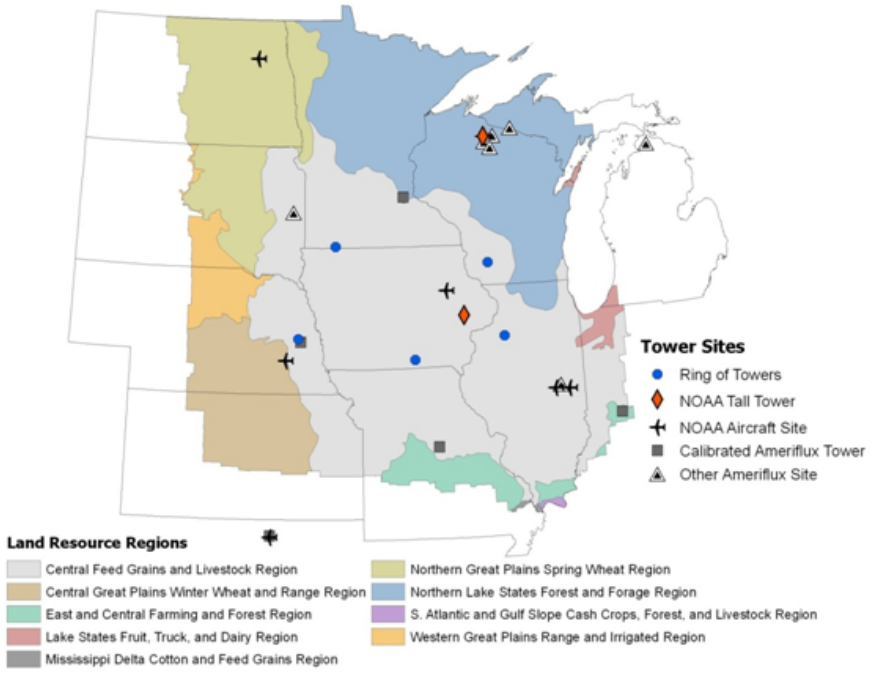

Figure 1-2. Mid-Continental Intensive (MCI) field campaign study region

Several atmospheric inversion systems implemented over the MCI showed how the large number of atmospheric CO_2 measurements was sufficient to constrain the regional carbon budget of the Corn Belt, as demonstrated by the convergence between inverse flux estimates using different prior fluxes, transport models, and prescribed errors (e.g. Lauvaux et al., 2012; Schuch et al., 2013). However, the spatial distribution of the inverse fluxes was not entirely constrained by the data, possibly due to differences caused by the definition of errors in both fluxes and transport. Similarly, Diaz-Isaac et al., (2014) showed that using the same fluxes but two different transport models (WRF versus TM5) led to large CO_2 model-data differences among the two models over the MCI region. Diverging interpretation of atmospheric signals by the models translates into different inverse fluxes across the region, as shown in Schuh et al. (2013). Based on these results, a small has been made in the characterization of transport errors in atmospheric inversions. Error propagation is non-trivial and often requires a large ensemble of simulations that are computationally expensive and often under-estimate model errors. For these reasons, transport models remain an important source of flux biases and therefore a major limitation in current regional inverse studies.

In recent years, the development of ensemble methods has improved the representation of transport uncertainty using the statistics of large ensembles to characterize the statistical spread of atmospheric forecasts (e.g. Evensen, 1994a, 1994b). However, single-physics ensemble-based statistics are highly susceptible to model error, leading to under-dispersive ensembles. Large ensembles (>50 members) remain computationally expensive and ill-adapted to the assimilation over longer time scales such as multi-year inversions of long-lived species such as CO_2 and CH_4. Smaller-size ensembles would be ideal, but most initial-condition-only perturbation methods produce unreliable and overconfident representations of the atmospheric state (Buizza et al. 2005).

To represent these errors over long time scales, the ensemble has to remain small and also has to guarantee the mass conservation of CO_2 and the atmospheric flow over time.

The research presented in this dissertation will quantify the transport errors with an ensemble of simulations created with the WRF mesoscale model using different physics parameterizations (LSMs, PBL schemes, cumulus parameterization and microphysics parameterizations) and different reanalysis data sets, avoiding the use of random perturbations that would generate incremental changes in the simulations. This multi-physics and multi-analysis ensemble of 45-members will allow me to explore the performances of these configurations by evaluating different meteorological variables affecting tower CO_2 measurements such as wind speed, wind direction, and PBL height. Additionally, these atmospheric transport errors will be reflected in the atmospheric CO_2 mixing ratios due to the offline coupling of surface CO_2 fluxes into WRF (using WRF-Chem passive tracer mode; Lauvaux et al., 2012). Throughout this dissertation, I will plan to focus on the following questions: How do the different physical parameterizations schemes affect CO_2 mixing ratios and inverse surface fluxes? Will some parameterizations be more accurate than others? Can I provide reliable random and systematic error estimates in transport models? Will the multi-physics and multi-analysis ensemble be able to represent the transport errors? Can I represent our atmospheric transport errors with a small ensemble of simulations and avoid large computational expenses? I will address these different questions by exploring the performance of the atmospheric transport model ensemble over the MCI region, for the summer of 2008.

In this dissertation, I will evaluate the multi-physics and multi-analysis ensemble in two different ways: (1) by evaluating in detail the performance of each member of the ensemble and (2) by evaluating the performance of the whole ensemble. Chapter 2 introduces how I built the 45-member multi-physics and multi-analysis ensemble used to quantify the atmospheric transport errors. This chapter shows the performance of the different model-configurations across the region,

the physics schemes that contribute to large systematic errors across the region and a sensitivity experiment that will propagate the impact of the different physics parameterizations and reanalyses into CO_2 mixing ratios. Chapter 3 shows how well the current multi-physics and multi-analysis ensemble represents the atmospheric transport errors. Additionally, in this Chapter 3, I test two techniques that will select specific members of the ensemble to better represent transport errors while considering an ensemble smaller than the original ensemble (i.e., 45-members). In this last chapter, I determine the minimal requirements for the quantification of atmospheric transport errors using an ensemble of model simulations.

In summary, this dissertation presents a unique effort to quantify the atmospheric transport errors and their potential impact on both CO_2 fluxes and mixing ratios. Both Chapters 2 and 3 will present the atmospheric transport errors from a different perspective compared to past studies; that is, using a large ensemble of model simulations and a robust calibration technique to represent accurately the model errors. Current studies use transport uncertainties based on a small number of measurements or on very limited ensembles focused on PBL schemes (e.g., Lauvaux et al., 2009b; Lauvaux and Davis, 2014; Feng et al., 2016). In this work, I use a significant number of meteorological observations able to characterize errors in horizontal and vertical transport of CO_2 and a greater number of simulations (i.e. ensemble members) that will use varying physics parameterizations to explore in detail the model uncertainties through physics configurations. My analysis will explore different meteorological observations including wind speed, wind direction, and PBL height. Past analyses mostly focused on the evaluation of CO_2 atmospheric mixing ratios with no emphasis on transport errors alone (e.g., Stephens et al., 2007; Lauvaux et al., 2009b). Past ensembles used to explore the atmospheric transport errors were small and not calibrated (e.g., Lauvaux and Davis, 2014; Feng et al., 2016). By calibration, I mean the evaluation of the ensemble statistics to test how the ensemble is representing the transport errors. In this work, I generate a large ensemble and evaluate if this ensemble represents the transport errors. A down-selection

process (or calibration) of the ensemble is presented to test if a small ensemble can represent atmospheric transport errors. Chapter 4 is a summary of the different findings and suggestion of future work that could improve and expand our knowledge of the atmospheric transport errors. In general, the proposed research will expand our ability to assess, understand and reduce transport errors in future atmospheric inversions.

Chapter 2

Sensitivity and Uncertainty Analysis of Physical Parameterization and Initial Conditions on Meteorological Variables and CO_2 Mole Fractions

2.1 INTRODUCTION

The increase in atmospheric carbon dioxide (CO_2) concentrations is a primary factor that influences the radiation budget and leads to major changes of the Earth's climate (IPCC, 2013). Atmospheric mixing ratios have increased primarily due to fossil fuel combustion and land use change. Not all CO_2 emitted remains in the atmosphere because the terrestrial biosphere absorbs about 30% of the released emissions (Le Queré et al., 2015). Terrestrial ecosystems in the temperate northern latitudes are identified as a substantial sink (Tans et al., 1990; Ciais et al., 1995; Gurney et al., 2002; Sarmiento et al., 2010; Pan et al., 2011; Le Queré et al., 2015). However, the specific magnitudes and distributions of terrestrial sources and sinks are still uncertain. Accurate and precise quantification of these fluxes is an important step towards a successful prediction of future atmospheric CO_2 and climate change mitigation.

One method used to estimate the terrestrial fluxes is the "top-down" method or atmospheric inverse method. The atmospheric inversion uses atmospheric transport models to estimate carbon fluxes (i.e., prior fluxes) by adjusting these fluxes to be optimally consistent with observed CO_2 concentrations (e.g., Enting, 1993; Bousquet et al., 2000; Chevallier et al., 2010). Uncertainties in the inverse method can be caused by the following: sparse atmospheric data (Gurney et al., 2002), uncertain prior flux estimates, limited spatial resolution in biospheric and atmospheric models, and transport model errors (Stephen et al., 2007; Gerbig et al., 2008; Picket-Heaps et al., 2011). Despite progress in top down methodologies, these sources of uncertainty have

hindered inverse estimates of sources and sinks from terrestrial ecosystems, which remain inferred from the residual term of the global carbon budget (Le Quéré et al., 2015).

Current atmospheric inversion systems are limited to the optimization of surface fluxes. However, the model-data mismatches used to optimize the fluxes contain the contributions of both flux and transport errors. Therefore, the atmospheric inversions may attribute atmospheric CO_2 model-data mismatches to surface fluxes. In a Bayesian framework, the atmospheric inversion assumes (1) atmospheric transport model errors are unbiased and (2) the random errors are known. Inconsistent errors (i.e., random and systematic) will be propagated into the state space by the optimization process, generating biased inverse (i.e. posterior) fluxes (Tarantola, 2005). The atmospheric inverse system will be reliable only if both the atmospheric transport random errors are quantified rigorously and the transport model is unbiased.

To date, only a few studies have focused on atmospheric transport errors. The Atmospheric Tracer Transport Model Intercomparison Project (TransCom) has been dedicated to quantifying atmospheric transport errors and their impact on CO_2 fluxes through model inter-comparisons (Gurney et al., 2002; Baker et al., 2006; Stephen et al., 2007; Patra et al., 2008; Peylin et al., 2013). As inter-comparison exercises, TransCom studies were not always limited to varying atmospheric transport, but at times also varied the number of observations, the inverse methodologies, and the prior fluxes that were used. Some of these studies have concluded that only an atmospheric transport model capable of representing synoptic and mesoscale atmospheric dynamics will be able to extract high-resolution information from the atmospheric observations (Law et al., 2008; Patra et al., 2008). Following these recommendations, the spatial resolution of transport models used to simulate atmospheric CO_2 mole fractions has increased to capture local-scale variability in continental observations (e.g. Ahmadov et al., 2009). Díaz Isaac et al. (2014) showed significant differences in the atmospheric CO_2 model-data mismatches when comparing a lower-resolution global transport model to a high-resolution regional transport model, but using identical surface

fluxes, suggesting that changes in the transport model resolution could lead to large differences in inverse surface flux estimates.

A critical problem in atmospheric transport resides in the representation of vertical mixing, which significantly impacts the interpretation of near-surface CO_2 mole fractions and the resulting inverse CO_2 flux estimates (Denning et al., 1995; Stephens et al., 2007). As a result, several studies have been dedicated to the evaluation of mixed layer (ML) depth (Yi et al., 2004; Gerbig et al., 2008; Kretschmer et al., 2012). An overestimation of the ML depth by an atmospheric model, for example, will cause an overestimation of the CO_2 surface flux magnitude. The misrepresentation of vertical mixing by TransCom's atmospheric models shown by Stephens et al. (2007) led Gerbig et al., (2008) to evaluate uncertainty in ML depth using a global model and find errors on the order of several ppm in ML CO_2 mixing ratios. Sarrat et al., (2007) used an inter-comparison of five mesoscale models and identified discrepancies in the ML depth that was potentially impacting the atmospheric CO_2 mixing ratios. These studies have attributed the differences between simulated and observed mixed ML height to flaws in planetary boundary layer (PBL) schemes and land surface models (LSMs). The accurate representation of the ML depth, however, is a necessary but most likely insufficient step for simulating CO_2 mixing ratios in the lower troposphere. Mixing between the ML and the rest of the atmosphere is also an important factor in the relationship between surface fluxes of CO_2 and ML CO_2 mixing ratios. It is likely that parameterizations other than PBL and LSM will influence ML CO_2 mixing ratios.

Inter-comparison of physical parameterization schemes using the Weather Research and Forecasting (WRF; Skamarock et al., 2005) mesoscale model has been explored to understand the impact of physics parameterizations on the vertical distribution of CO_2 mixing ratios (Kretschmer et al., 2012; Yver et al., 2013; Lauvaux and Davis, 2014; Feng et al., 2016). These studies have found systematic errors of several ppm in atmospheric CO_2 that can lead to biased surface flux estimates. These studies performed pseudo-data experiments or used a small number of

observations, and focused mostly on the impact of different PBL physics schemes. There is agreement among the studies that misrepresentation of vertical mixing causes biases in ML CO_2 mixing ratios, and that these biases directly affect inverse flux estimates. Vertical mixing, however, is not solely affected by the PBL parameterization. Therefore, investigations of vertical mixing of CO_2 remain incomplete. Additional parameterizations that impact the transport of air masses both horizontally and vertically should be evaluated.

In this work, we study uncertainty in an atmospheric transport model using a multi-physics approach not limited to the evaluation of the PBL schemes. This evaluation will include different LSMs, cumulus parameterizations (CP), microphysics parameterizations (MP), and initial and boundary conditions used by the WRF model. We aim ultimately to quantify the uncertainty of the atmospheric transport model and propagate these errors into the CO_2 mixing ratios and the final inverse fluxes. In this paper, we plan to focus on errors from physical parameterization schemes in a state-of-the-art mesoscale model, i.e. the WRF model. We will focus on the following questions: How do different physical parameterization schemes affect ML CO_2 mixing ratios? Are some physics parameterizations more effective/accurate than others at simulating atmospheric conditions important to interpreting CO_2 mixing ratio observations in the PBL? What are the nature and magnitude of random and systematic errors in the WRF model, and how does this depend on model configuration? We will address these questions by exploring atmospheric transport model performance over a highly instrumented area, the Mid-Continental Intensive (MCI) region (Ogle et al., 2006). Evaluating the atmospheric transport during summer, the most biologically active time of the year, is a first step toward a more rigorous and complete atmospheric inversion that quantifies random transport errors more accurately, and minimizes transport biases. Thus this work will expand our ability to assess, understand, and reduce transport errors in future atmospheric inversions.

2.2 METHODS

2.2.1 Region

The region selected for our study is the Midwest region of the United States (Figure 2-1). The Midwest of U.S. was chosen because the first multi-year (2007-2009) campaign with a high-density CO_2 measurement network was deployed in this region (Ogle et al., 2006, Miles et al., 2012). This field campaign, part of the North American Carbon Program (NACP), was called the Mid-Continental Intensive (MCI) and encompassed the agricultural belt in the north-central U.S. The MCI campaign is unique for its density of well-calibrated (Richardson et al., 2012) atmospheric CO_2 mole fraction measurements intended to constrain the region's carbon budget. We describe the operational rawinsonde and GHG tower networks over the region in Section 2.2.4. These networks provided a significant amount of observational constraint on both transport and GHG mole fractions, which allow us to evaluate and quantify the atmospheric transport errors in this study.

2.2.2 Atmospheric Model Setup

The atmospheric transport model used in this study to generate our 45-member physics ensemble is the Weather Research and Forecasting (WRF) model version 3.5.1 (Skamarock et al., 2005) and a modified chemistry module for CO_2 (called WRF-ChemCO_2, Lauvaux et al., 2012). The atmospheric column in each simulation is described with 59 vertical levels, with 40 of them within the first 2-km of the atmosphere. Two nested domains were used. The coarse domain (d01) uses a horizontal grid spacing of 30-km and the nested or inner domain (d02) uses 10-km grid spacing (Figure 2.1). The coarse domain covers most of the United States and parts of Canada and the nested domain is centered over Iowa and covers the Midwest region of United States. The

nesting method employed is the "one-way" nesting in which the outer domain constrains the inner domain through nudging of the boundary conditions that drive the meteorology once the outer domain simulation has finished (Soriano et. al., 2002). No feedback from the inner domain to the coarse domain was allowed. For our sensitivity study, only the inner domain (d02) has been analyzed as it covers the area of interest.

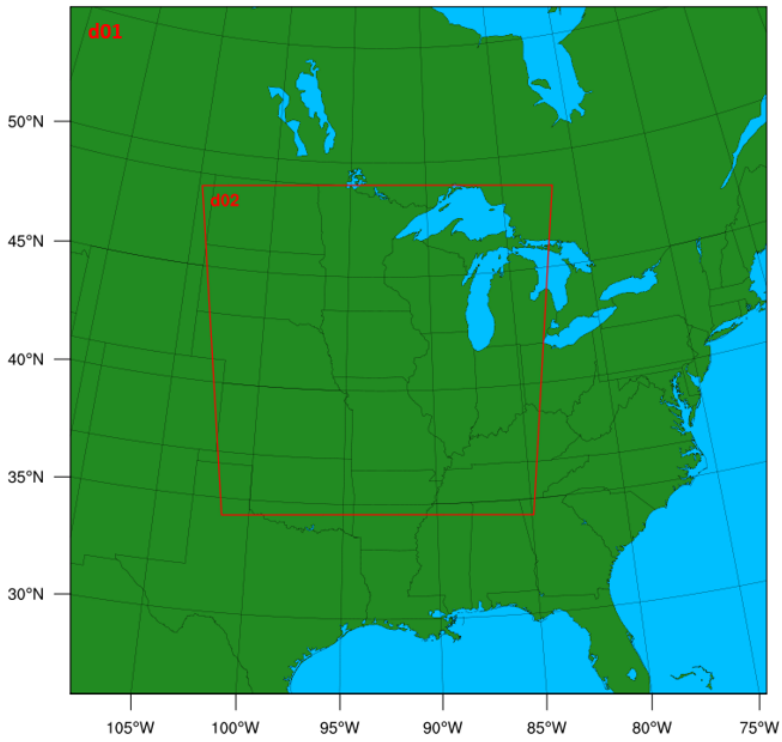

Figure 2-1. Geographical domain used by WRF-ChemCO_2 physics ensemble. The parent domain (d01) has a 30-km resolution, the inner domain (d02) has a 10-km resolution.

2.2.3 Ensemble Configuration

Similar to any domain-limited atmospheric model, transport errors arise from initial and boundary conditions and the different physics parameterizations. Therefore, we have built an ensemble of 45-members using different physical parameterization schemes and large-scale initial

and boundary conditions from reanalysis products (see Table 2-1). WRF offers multiple options for the LSM, PBL, cumulus, and microphysics schemes. The members in our multi-physics ensemble all use the same radiation schemes (both long wave and shortwave) but the land surface, surface layer, boundary layer, cumulus, and microphysics schemes are variable. In addition, we have initialized the model boundary and initial conditions with different datasets. Table 2-2 shows the different options used in this study.

Table 2-1. Different model configurations used in this study.

Num.	Cases	Reanalysis	LSM Scheme	PBL Scheme	Cumulus Scheme	Microphysics Schemes
1	CASE1 NEST CP	NARR	Noah	YSU	Kain-Fritsch	WSM 5-class
2	CASE2 NEST CP	NARR	Noah	MYJ	Kain-Fritsch	WSM 5-class
3	CASE5 NEST CP	NARR	Noah	MYNN	Kain-Fritsch	WSM 5-class
4	CASE4 NEST CP	GFS	RUC	YSU	Kain-Fritsch	WSM 5-class
5	CASE3 NEST CP	GFS	RUC	MYJ	Kain-Fritsch	WSM 5-class
6	CASE6 NEST CP	GFS	RUC	MYNN	Kain-Fritsch	WSM 5-class
7	CASE1 SLAB NEST CP	NARR	Thermal Dif.	YSU	Kain-Fritsch	WSM 5-class
8	CASE2 SLAB NEST CP	NARR	Thermal Dif.	MYJ	Kain-Fritsch	WSM 5-class
9	CASE3 SLAB NEST CP	NARR	Thermal Dif.	MYNN	Kain-Fritsch	WSM 5-class
10	CASE7 NEST CP	NARR	Noah	YSU	Grell-3D	WSM 5-class
11	CASE8 NEST CP	NARR	Noah	MYJ	Grell-3D	WSM 5-class
12	CASE11 NEST CP	NARR	Noah	MYNN	Grell-3D	WSM 5-class
13	CASE10 NEST CP	GFS	RUC	YSU	Grell-3D	WSM 5-class
14	CASE9 NEST CP	GFS	RUC	MYJ	Grell-3D	WSM 5-class
15	CASE12 NEST CP	GFS	RUC	MYNN	Grell-3D	WSM 5-class
16	CASE4 SLAB NEST CP	NARR	Thermal Dif.	YSU	Grell-3D	WSM 5-class
17	CASE5 SLAB NEST CP	NARR	Thermal Dif.	MYJ	Grell-3D	WSM 5-class
18	CASE6 SLAB NEST CP	NARR	Thermal Dif.	MYNN	Grell-3D	WSM 5-class
19	CASE13 NEST CP	NARR	Noah	YSU	Kain-Fritsch	Thompson
20	CASE14 NEST CP	NARR	Noah	MYJ	Kain-Fritsch	Thompson
21	CASE17 NEST CP	NARR	Noah	MYNN	Kain-Fritsch	Thompson
22	CASE16 NEST CP	GFS	RUC	YSU	Kain-Fritsch	Thompson
23	CASE15 NEST CP	GFS	RUC	MYJ	Kain-Fritsch	Thompson
24	CASE18 NEST CP	GFS	RUC	MYNN	Kain-Fritsch	Thompson
25	CASE7 SLAB NEST CP	NARR	Thermal Dif.	YSU	Kain-Fritsch	Thompson
26	CASE8 SLAB NEST CP	NARR	Thermal Dif.	MYJ	Kain-Fritsch	Thompson
27	CASE9 SLAB NEST CP	NARR	Thermal Dif.	MYNN	Kain-Fritsch	Thompson
28	CASE19 NEST CP	NARR	Noah	YSU	Grell-3D	Thompson
29	CASE20 NEST CP	NARR	Noah	MYJ	Grell-3D	Thompson
30	CASE23 NEST CP	NARR	Noah	MYNN	Grell-3D	Thompson
31	CASE1 NEST	NARR	Noah	YSU	No CP	WSM 5-class
32	CASE2 NEST	NARR	Noah	MYJ	No CP	WSM 5-class

33	**CASE5 NEST**	NARR	Noah	MYNN	No CP	WSM 5-class
34	**CASE4 NEST**	GFS	RUC	YSU	No CP	WSM 5-class
35	**CASE3 NEST**	GFS	RUC	MYJ	No CP	WSM 5-class
36	**CASE6 NEST**	GFS	RUC	MYNN	No CP	WSM 5-class
37	**CASE1 SLAB NEST**	NARR	Thermal Dif.	YSU	No CP	WSM 5-class
38	**CASE2 SLAB NEST**	NARR	Thermal Dif.	MYJ	No CP	WSM 5-class
39	**CASE3 SLAB NEST**	NARR	Thermal Dif.	MYNN	No CP	WSM 5-class
40	**CASE1 GFS NEST CP**	GFS	Noah	YSU	Kain-Fritsch	WSM 5-class
41	**CASE2 GFS NEST CP**	GFS	Noah	MYJ	Kain-Fritsch	WSM 5-class
42	**CASE5 GFS NEST CP**	GFS	Noah	MYNN	Kain-Fritsch	WSM 5-class
43	**CASE1 GFS SLAB NEST CP**	GFS	Thermal Dif.	YSU	Kain-Fritsch	WSM 5-class
44	**CASE2 GFS SLAB NEST CP**	GFS	Thermal Dif.	MYJ	Kain-Fritsch	WSM 5-class
45	**CASE3 GFS SLAB NEST CP**	GFS	Thermal Dif.	MYNN	Kain-Fritsch	WSM 5-class

Table 2-2. Parameters included in the sensitivity analysis.

Parameter	Options
Land Surface Model	Noah LSM Rapid Update Cycle (RUC) LSM 5-layer Thermal Diffusion
Planetary Boundary Layer (PBL) scheme – Surface Layer scheme	Yonsei University (YSU) Mellor-Yamada-Janjic (MYJ) Mellor-Yamada-Nakanishi-Niino (MYNN)
Cumulus	Kain-Fritsch (KF) Grell-3Devenyi (G3D)
Microphysics	WSM 5-class Thompson et al., (2004)
Initial & Boundary Conditions	North America Regional Reanalysis (NARR) Global Final Analysis (FNL)

2.2.3 Physics Parameterization Schemes

a. Land Surface Models (LSMs)

The land surface models (LSMs), which ingest land-surface properties, soil, and surface conditions from driver data, simulate the conditions at the land surface, including surface energy fluxes. The partitioning of these fluxes affects the structure and depth of the PBL through the turbulence parameterization, hence modifying the near-surface in situ CO_2 mole fractions. To evaluate the sensitivity of modeled mole fractions to the surface conditions, three LSM schemes

are chosen for this study: the 5-layer soil thermal diffusion model (Dudhia, 1996), the Noah land surface model (Chen and Dudhia, 2001), and the Rapid Update Cycle (RUC) (Smirnova, 2000). The three LSMs differ in several aspects, from the description of soil properties to the physical processes driving the land-surface interactions. The thermal diffusion model uses a simple thermal diffusion equation to transfer thermal energy from the ground to the atmosphere, describing the belowground profile with 5 soil layers (Dudhia, 1996). This LSM also includes snow-covered land and constant soil moisture values for a given land use type and season. The Noah LSM scheme uses time-dependent soil variables, soil temperature for four soil layers, canopy moisture, and snow cover prediction (Chen and Dudhia, 2001). The fluxes calculated within this scheme include sensible and latent heat fluxes. The RUC LSM scheme is a higher resolution soil model with six soil layers and includes the effects of vegetation, canopy water, and snow (Smirnova, 2000). This scheme also includes parameterizations for snow and frozen soil (Smirnova, 2000).

b. Planetary Boundary Layer (PBL) Schemes

The planetary boundary layer (PBL) is directly influenced by frictional drag, sensible heat flux, and evapotranspiration, all of which are responsible for generating turbulent eddies. The PBL schemes parameterize sub-grid scale turbulent vertical fluxes of heat, momentum, and moisture within the PBL and throughout the atmosphere. The three PBL schemes used in this study are the Yonsei University (YSU) (Hong et al., 2006) PBL scheme, the Mellor-Yamada-Janjic (MYJ) (Janjic, 2002) PBL scheme, and the Mellor-Yamada-Nakanishi-Niino (MYNN) PBL scheme (Nakanishi & Niino, 2004). These three PBL schemes differ in the treatment of turbulent diffusion. The YSU scheme is a non-local first order scheme that uses non-local eddy diffusivity coefficients to compute turbulent fluxes. Vertical transport is dependent on the bulk characteristics of the PBL and includes counter-gradient transport of temperature and momentum arising from large scale eddies. The YSU scheme explicitly calculates entrainment at the top of the PBL as a function of

the surface buoyancy flux. The MYJ and MYNN 2.5 PBL schemes are local closure schemes that include a prognostic equation for turbulent kinetic energy (TKE) and a level 2.5 turbulence closure approximation to determine eddy transfer coefficients. The MYJ scheme implicitly calculates the entrainment layer while the MYNN uses a more explicit calculation of entrainment at the top of the PBL (it forecasts other second order moments besides TKE) (Román-Cascón et al., 2012). The MYNN 2.5 is a variation of the MYJ PBL scheme that includes a nonlocal component of the turbulent mixing that reduces potential cold biases and increases PBL depths. The MYJ PBL scheme used in this study has been slightly modified to allow for very low turbulence regimes (e.g. nocturnal stable conditions) with a decreased minimum value for TKE.

c. Cumulus Parameterizations

The cumulus parameterization (CP) schemes are used with the aim of representing the vertical fluxes due to unresolved updraft and downdrafts and compensating motion outside the clouds. In this study we use two different cumulus parameterization schemes, Kain-Fritsch (KF) (Kain, 2004), Grell-3D (G3D) (Grell and Devenyi, 2002). The KF scheme is deep and shallow convection sub-grid scheme, which uses a simple cloud model that simulates moist updrafts and downdrafts along with detrainment and entrainment effects. The G3D cumulus scheme is based on the Grell (1993) scheme and G3D is a scheme for higher resolution domains allowing for subsidence and neighboring columns. The G3D uses a large ensemble of closure assumptions and parameters that are used in numerical models and implements some statistical techniques to determine the optimal value for feedback to the entire model (Pei et al., 2014). The cumulus parameterization is theoretically only valid for coarse grid resolutions (e.g., greater than 10 km) and should not be used when the model has a higher resolution (e.g., less than 5km) (Skamarock et al., 2005). Therefore, we are in a 'grey-zone' (e.g., 5-10km), where it is unclear if cumulus

parameterization should be used or not. For that reason, we also ran simulations that do not use a cumulus parameterization scheme in the nested domain.

d. Microphysics Parameterizations

Microphysics parameterizations (MP) schemes explicitly resolve water vapor, cloud, and precipitation processes. In this study we use two MP schemes: the WRF Single-Moment 5-class (WSM5) scheme (Hong et al., 2004) and the Thompson scheme (Thompson et al., 2004). The WSM5 scheme is a single moment parameterization that includes five species: water vapor, cloud water, cloud ice, rain, and snow, which are all treated independently. The Thompson scheme is a double moment scheme, which predicts the mixing ratio of five hydrometeors species, the number concentration of ice phase hydrometeors, and rain.

2.2.4 Meteorological Initial and Boundary Conditions

Two meteorological datasets provide the initial and lateral boundary conditions for our regional model. For initialization, WRF interpolates the coarse-resolution analysis products onto the model grid and calculates the values of the parent domain lateral boundaries. The inner grid uses the boundary conditions of the parent domain. In this study, we compare two different meteorological datasets: the North America Regional Reanalysis (NARR) (Mesinger et al., 2006), and the Global Final Analysis (FNL). The NARR dataset was developed at the Environmental Modeling Center (EMS) of the National Centers for Environmental Prediction (NCEP). NARR uses a high resolution NCEP Eta Model with a horizontal grid spacing of 32 km and includes 45 vertical levels. NARR provides both initial and boundary conditions at 3-hourly intervals. The NCEP FNL analysis data has a horizontal grid spacing of 1°×1° and is prepared operationally every

six hours. The FNL is prepared with the same model that NCEP uses in the Global Forecast System (GFS).

2.2.5 CO_2 Surface Fluxes

For this study, we used the summer 2008 posterior surface fluxes from the data assimilation system CarbonTracker[1] version 2009 (CT2009) (Peters et al., 2007). This system produces CO_2 flux estimates by integrating daily daytime averaged CO_2 mole fractions from continuous hourly observations and then minimizing the differences between the observed and modeled atmospheric CO_2 mole fractions. The Transport Model 5 (TM5) offline atmospheric tracer transport model (Krol et al., 2005) driven by the European Centre for Medium-Range Weather Forecasts (ECMWF) operational forecast model, propagates the surface fluxes to generate 3D mole fractions of CO_2 across the globe.

The CO_2 surface fluxes are represented by different sub-components, which include: fossil fuel emissions, biomass burning, terrestrial biosphere exchange, and ocean-atmosphere exchanges. The annual fossil fuel emissions used in CT2009 are from the Carbon Dioxide Information and Analysis Center (CDIAC) (Boden et al., 2009). These fossil fuel fluxes are mapped onto a 1°×1° grid and are then distributed into country totals according to the spatial patterns from the EDGAR-4 inventories (Olivier and Berdowski, 2001). Biomass burning is based on the Global Fire Emission Database version 2 (GFEDv2). The dataset consists of 1°×1° gridded monthly burned areas, fuel loads, combustion completeness, and fire emissions. Terrestrial biosphere fluxes come from the Carnegie-Ames Stanford Approach (CASA) global biogeochemical model (van der Werf et al., 2006; Giglio et al., 2006). The CASA biosphere model produces net primary production (NPP) and

[1] http://carbontracker.noaa.gov

heterotrophic respiration fluxes with a monthly time resolution at 0.5°×0.5° spatial resolution. The long-term ocean fluxes and uncertainties are derived from inversions reported in Jacobson et al., (2007). Ocean inverse flux estimates are composed of preindustrial (natural), anthropogenic flux inversions, and an additional level of biogeochemical interpretations (Gloor et al., 2003; Gruber, Sarmiento and Stocker, 1996). Similar to other CO_2 inverse systems, the fossil fuel and fire emissions are specified (i.e. remain constant) and only the oceanic and terrestrial biosphere fluxes are optimized.

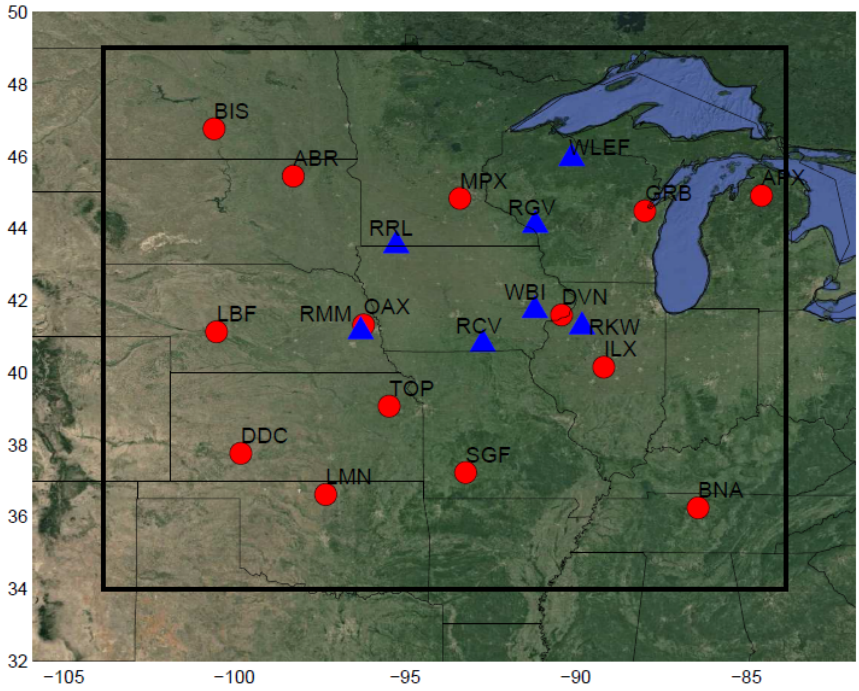

Figure 2-2. Map of the study region, including the CO_2 towers (blue triangles) and rawinsonde sites (red circles) locations.

2.2.6 Datasets

Our interest is to explore and quantify atmospheric transport errors over the Midwest U.S. using observations that we have over this region. Therefore, we will evaluate the errors over the inner domain (D2) of our models. Figure 2-2 shows the location of all the stations that provide

atmospheric CO_2 mole fractions and the meteorological observation sites that will be used. Meteorological data were obtained from the University of Wyoming's online data archive (http://weather.uwyo.edu/upperair/sounding.html) for the 14 rawinsonde stations shown in Figure 2-2. In-situ atmospheric CO_2 mole fraction data are provided by seven communication towers (Figure 2-2) (Miles et al., 2012). Five of these towers were part of an experimental network, deployed from 2007 to 2009 (Richardson et al., 2012; Miles et al., 2012). The other two towers (Park Falls-WLEF and West Branch-WBI) are part of the Earth System Research Laboratory/Global Monitoring Division (ESRL/GMD) tall tower network (Andrews et al., 2014). Each of these towers sampled air at multiple heights, ranging from 11 m AGL to 396 m AGL.

2.2.7 Data Selection

Most atmospheric inversion that use continental observations use daytime CO_2 mole fractions from continuous observations to minimize the difference between the observed and modeled atmospheric CO_2 mole fractions. Only daytime measurements are assimilated due to the difficulty in simulating strong vertical gradients in the nocturnal boundary layer, which are minimized during daytime under well-mixed boundary layer conditions (Bakwin et al., 1998). Therefore, both models and observations will be evaluated during daytime.

We analyzed CO_2 mole fractions collected from sampling levels at or above 100m AGL, which is the highest observation level across the MCI network (Miles et al., 2012). This ensures that the observed mole fractions reflect regional CO_2 fluxes and not near-surface gradients of CO_2 in the atmospheric surface layer (ASL) or local CO_2 fluxes (W. Wang et al., 2007). Both observed and simulated CO_2 mole fractions are averaged from 1800 to 2200 UTC (12:00-16:00 LST), when the daytime period when the boundary layer should be convective and the CO_2 profile well mixed

(e.g., Davis et al., 2003; Stull, 1988). This averaged mole fraction will be referred to hereafter as daily daytime average (DDA).

In this study, we will also evaluate the wind speed, wind direction, and PBL height (PBLH) from the different rawinsonde stations. Similar to the CO_2 mole fractions, we want our meteorological observations to be within the well-mixed layer. Therefore, we use the wind speed and wind direction observed approximately 300 m above ground level (AGL). CO_2 mole fraction observations were sampled at about 100m, however, the availability of meteorological observations at this height is too low to collect a sufficient number of data for our statistical evaluation. The observed PBLH was estimated using the virtual potential temperature gradient with a threshold of 0.2 K/m. We want our simulated meteorological variables to be close to the observational level, therefore we use wind speed and wind direction from level 11 (~350 m) of the model. The WRF model provides an estimate of the PBLH, but the methodology used to diagnose these values varied with the PBL scheme used in the simulation. To remain consistent, we decided to calculate the PBLH in WRF with the same potential temperature gradient method that is used for the rawinsonde data. Rawinsonde stations across this region collect data at 1200 UTC and 0000 UTC, however, our model-data evaluation will be done for daytime conditions only. Therefore, both the modeled results and data will be evaluated in the late afternoon (i.e., 0000 UTC) corresponding to well mixed conditions.

2.2.8 Evaluation Methodology or Analyses of the Models

Comparisons to measurements of wind speed, wind direction, PBLH, and DDA CO_2 mole fractions are used to inform the performance of each model configuration. Modeled data are extracted from the simulations using the nearest grid points to the locations of our observations. Each model configuration is evaluated from June 18 to July 21, 2008 for the meteorological

variables and from June 26 to July 21, 2008 for the CO_2 mole fractions. Summer in the U.S. Midwest corresponds to the peak of the growing season for both crops and most non-agricultural ecosystems (except grasslands). We focus here on the growing season because the large biogenic fluxes make this the most important time of year for understanding the relationship between fluxes and CO_2 mole fractions. Smallan et all, (2014) showed that atmospheric models simulated the day-to-day variability in CO_2 mole fraction poorly, and therefore have limited capability in constraining the regional carbon balance.

A series of statistical analyses are used to assess the performance of the different model configurations. The different metrics used include mean absolute error (MAE), root mean square error (RMSE), mean bias error (MBE), and root mean square difference (RMSD),

$$RMSD = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \mu)^2}}{N}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - o_i|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)$$

where o is the observed variable, p is the predicted variable, μ is the mean of the ensemble, N is the total number of days, and "n" is the number of members.

The MAE and RMSE represent the average daily model-observation differences. The MBE describes the model-observation difference averaged error over the entire period. These three metrics are critical to inverse flux estimates as biases can arise from day-to-day (which we will refer to as random) or longer-term (systematic) errors in the transport model. We acknowledge that the propagation of meteorological errors to mole fractions, and mole fractions errors to surface

fluxes is complex, but these metrics provide valuable insight into model performance. To these three model-data statistical metrics, we add the RMSD (model-model differences) to evaluate the impact of the physics parameterizations on both meteorological variables and CO_2 mole fractions. No observation is considered in RMSD as we compare each model configuration to the ensemble mean. Our main interest with the RMSD is to identify how much the physics parameterizations contribute to variability in CO_2 mole fraction, PBLH, wind speed, and wind direction.

We will also use Taylor diagrams and spatial correlations to evaluate the model-data mismatch in CO_2 mixing ratios. We are also interested in describing the performance of the different models using standard deviation and correlation analysis to identify particular spatial patterns among the different models. Taylor diagrams rely on three nondimensional statistics: the variance ratio (model variance normalized by the observed variance), the correlation coefficient, and the normalized centered root-mean-square (CRMS) difference (Taylor, 2001). The variance ratio or normalized standard deviation (NSD) indicates the difference in amplitude between the model and the observation. If this ratio is less than 1.0, then the model tends to underestimate the amplitude compared to the observation. The correlation coefficient measures the similarity in the temporal variations between the model and the observation, regardless of the amplitude. This correlation coefficient has a range of $-1.0 \leq R \leq 1.0$ and is insensitive to systematic errors. As R approaches 1.0, the model approaches agreement with the observation. The CRMS is normalized by the observed standard deviation and quantifies the ratio of the amplitude of the variations between the model and the observation. The CRMS is also insensitive to biases in the model-data residuals.

Temporal correlations between the modeled-observed residual in meteorological variables and CO_2 mixing ratios are used to determine the impact that meteorological errors have on the transport of CO_2. Since the meteorological and CO_2 observations are not collocated (Figure 2-2),

nearest neighbor sites are used in this evaluation. This spatial correlation will be estimated using the ensemble mean of the MBE estimated for each rawinsonde site and CO_2 tower.

2.3. RESULTS

2.3.1 Impact of physics parameterizations on the PBL dynamics

Previous studies indicate that one of the main sources of transport errors in inverse flux estimates is the misrepresentation of the PBL dynamics, which can be mainly attributed to PBL schemes (Stephens et al., 2007; Gerbig et al., 2008). The root mean square difference (RMSD) of the simulated CO_2 mixing ratio was used to explore if other physical parameterizations have a significant impact on CO_2 mole fractions compared to PBL parameterizations. The RMSD was computed for different parameterization schemes (i.e., LSM, PBL schemes, CP, and MP) and for two reanalysis products (i.e., NARR and GFS). The CO_2 mixing ratio RMSD is greatest for the LSM parameterization schemes across the different sites, followed by the PBL schemes and CP (Figure 2-3). Microphysics parameterization has the least impact on CO_2 mole fractions. Only two microphysics parameterizations are tested in this ensemble but additional tests using only two options for all the different physics produced similar results, i.e. the least impact from MP.

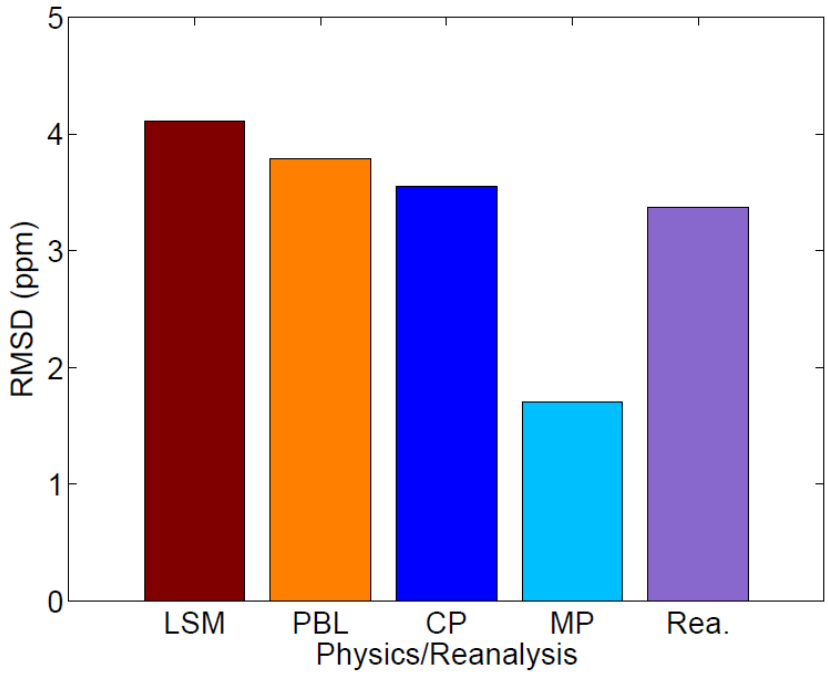

Figure 2-3. Sensitivity of CO_2 mole fractions as a function of model physics parameterizations (i.e., land surface model (LSM), planetary boundary layer scheme (PBL), cumulus parameterization (CP), microphysics parameterization (MP) and Reanalyses). The root mean square difference (RMSD) of the CO_2 mole fractions simulated at each site and for each model ensemble member was computed by varying only the type of physics parameterization noted, and keeping all other model elements constant. RMSD was averaged across sites and across model ensembles.

We also explore how much the variability in PBL winds and depth, important variables in determining CO_2 mole fractions in the PBL, are influenced by physics parameterizations. Therefore, we want to explore which parameterization impacts the PBL dynamics by comparing the RMSD of the three selected meteorological variables (i.e., PBLH, wind speed and wind direction) that define and contribute the most to the representation of the CO_2 mole fraction distribution in the PBL. Figure 2-4 shows the RMSD of PBL wind speed and direction, and PBL height over the entire simulation period. All the physics parameterizations, except for microphysics, impact all three meteorological variables. Rreanalysis has a greater impact on wind speed (Figure 2-4b) and wind direction (Figure 2-4c) than it does on PBLH (Figure 2-4a). It is worth noting that the PBLH RMSD (Figure 2-4a)

shows the same RMSD ranking (i.e. relative importance of the physics) as for CO_2 mixing ratio RMSD (Figure 2-3).

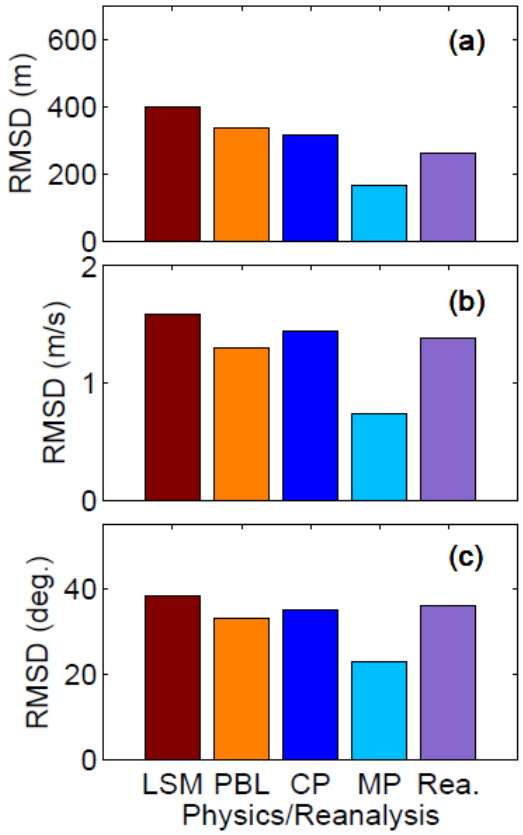

Figure 2-4. Root mean square difference (RMSD) of the PBLH (a), wind speed (b) and wind direction (c) for the different physics parameterizations (i.e., land surface model (LSM), planetary boundary layer scheme (PBL), cumulus parameterization (CP), microphysics parameterization (MP) and Reanalyses).

Based on the evaluation of the CO_2 mixing ratio, wind speed, wind direction, and PBLH RMSD, the PBL scheme impact was ranked second after the impact of the LSM. Generally, all the parameterization schemes, including the reanalysis data source, have a significant impact on each of these variables.

2.3.2 Meteorological day-to-day variability

Figure 2-5 shows a time series of the 0000 UTC observed and simulated PBL wind speed (Figure 2-5a), wind direction (Figure 2-5b), and PBLH (Figure 2-5c) from June 18 to July 21, 2008 at the Omaha, NE (OAX) rawinsonde site. Across the study region, we found that the range of residual values lie between 5.4 and 7.1 m/s for wind speed, 76 and 114 degrees in wind direction, and between 1000 and 1700 m for PBLH. Other sites have similar characteristics to Figure 2-5. At all of the sites, the model configurations followed the observed patterns in wind speed (Figure 2-5a) and wind direction (Figure 2-5b). Nevertheless, the ensemble shows less variability (or spread) for the wind speed and wind direction compared to the PBLH. The time series at each rawinsonde site shows that for certain days, the ensemble is biased (i.e. all the members either overestimate or underestimate) wind speed and wind direction (199 and 186 DOY respectively), regardless of the physics configuration. The time series of the PBLH, however, shows that simulated PBLH can vary significantly across the different physics configurations and that the ensemble is rarely biased (i.e. some members of the ensemble are able to match the observed PBLH).

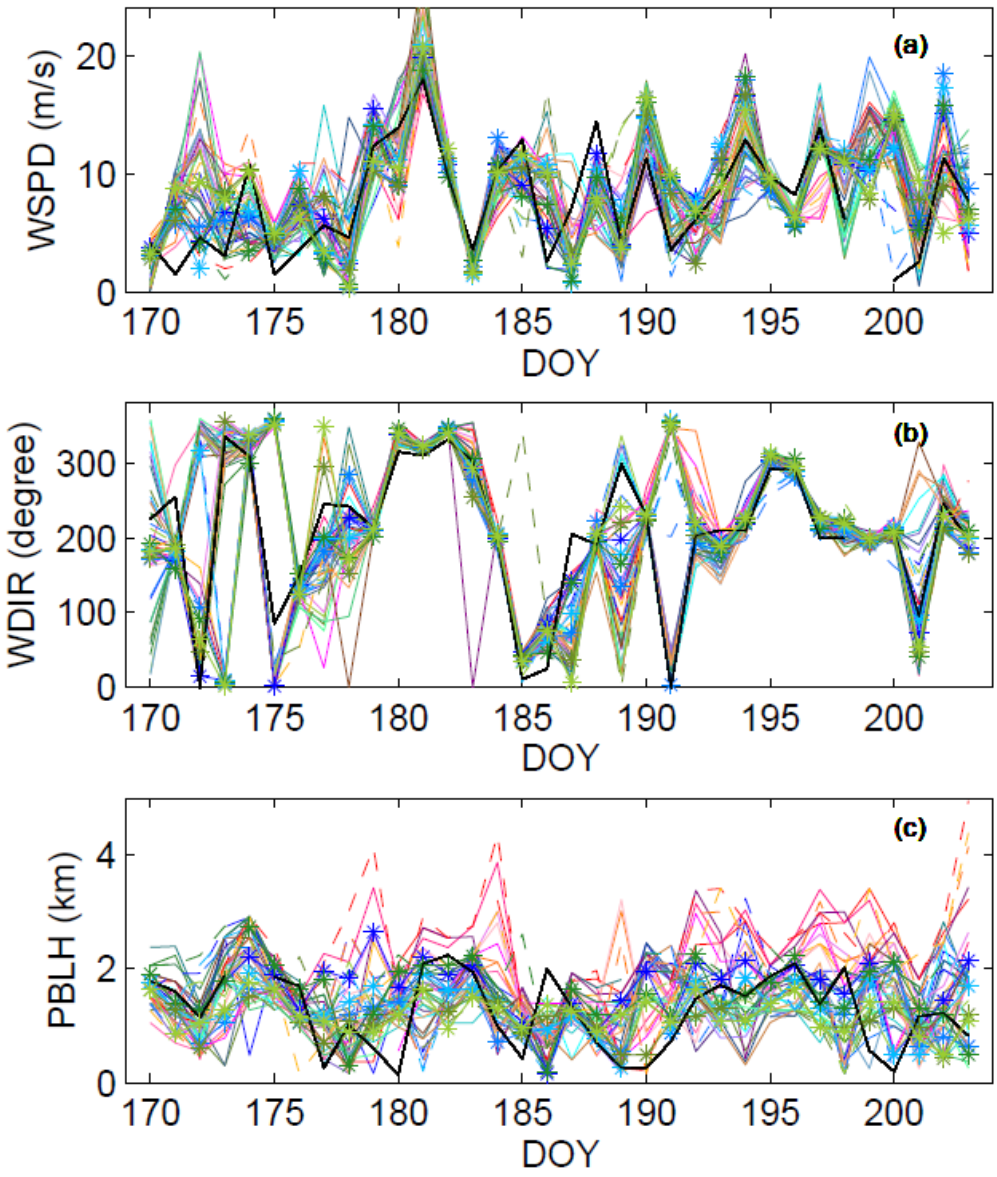

Figure 2-5. Observed (black line) and simulated (colored lines) PBL (300 m AGL) wind speed (a), wind direction (b) and PBLH (c) at time 0000 UTC from day of the year (DOY) 169 to 203 of 2008 at the Lincoln, Nebraska (OAX) rawinsonde site.

2.3.3 Characterization of transport errors

The performances of each model configuration are evaluated using the monthly average MAE, RMSE, and MBE (or bias). Each of these statistics was applied to the three meteorological variables, wind speed, wind direction, and PBLH, at a local scale and over the region.

2.3.3.1 Mean Absolute Error (MAE)

Daily model errors, represented by the MAE on Figure 2-6, show the sites with the maximum and minimum MAE for each of the model. Over the region, the MAE varies significantly from site to site for wind speed (Figure 2-6a), wind direction (Figure 2-6b), and PBLH (Figure 2-6c) regardless of the model configuration. About three (ABR, BNA, and LMN) out of the fourteen rawinsonde sites show the smallest MAEs across the different model configurations. Over the region, about two rawinsonde (DDC and TOP) sites show consistently a low MAE regardless of the model configuration. Two sites (DDC and LBF) systematically show a higher MAE for the PBLH and two additional sites (SGF and APX) show the smallest MAE for PBL. For all the three meteorological variables, the model configurations that yield the smallest MAEs varies across rawinsonde sites, impairing the selection of a best configuration over the region.

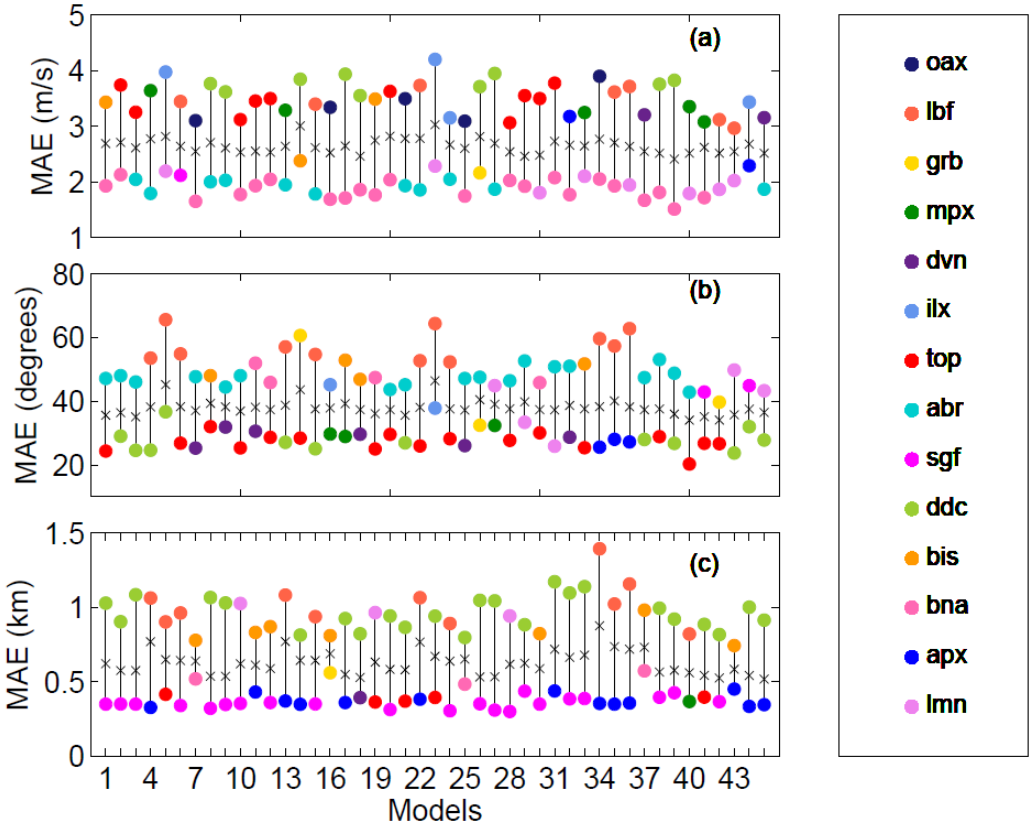

Figure 2-6. Maximum and minimum monthly average wind speed (a), wind direction (b) and PBLH (c) mean absolute errors (MAE) for the different sites and models. The abscissa shows the different model configurations, while the ordinate shows the MAE for the meteorological variables. Each dot represents the site with the maximum and minimum MAE for each model and variable; while the x represents the regional mean of the MAE. (See Table 2-1 for model references.)

The regionally—averaged MAEs reveal that some model physics parameterizations are associated with larger regional MAEs (Figure 2-7). PBL wind speed MAE averaged across the region does not vary significantly across model configurations (Figure 2-7a). The vast majority of the model configurations have a regional wind speed MAEs of 2.4 - 2.8 m/s, except for two configurations with a larger regional MAE (3 m/s). Similar to wind speed, the regional MAE for PBL wind direction is not much different (Figure 2-7b) among model configurations. Three simulations have higher regionally-averaged MAEs. Two of these configurations are the same as those with high PBL wind speed MAEs. These configurations both include MYJ as a PBL

parameterization (models 14 and 23 see Figure 2-7a,b and Table 2-1). Unlike winds, the different model configurations show significantly different regionally-averaged MAEs for PBLH (Figure 2-7c). The configurations that show a higher MAE for the PBLH use the YSU PBL scheme (models 4, 13, 22 and 34 see Figure 2-7c and Table 2-1). In addition, all model configurations with large MAEs for all three PBL meteorological variables use RUC as the LSM scheme.

Although the identification of the configurations with the lowest MAE remains unclear for wind speed and wind direction, three configurations show the lowest MAE (~ 2.4 m/s) for wind speed and two configurations show the lowest MAE (~ 34 degrees) for wind direction. For PBLH, the majority of configurations with lowest regional MAE (MAE < 600m) use the same PBL parameterization (MYNN 2.5). We were able to identify specific configurations with the lowest MAE but only for one variable at a time.

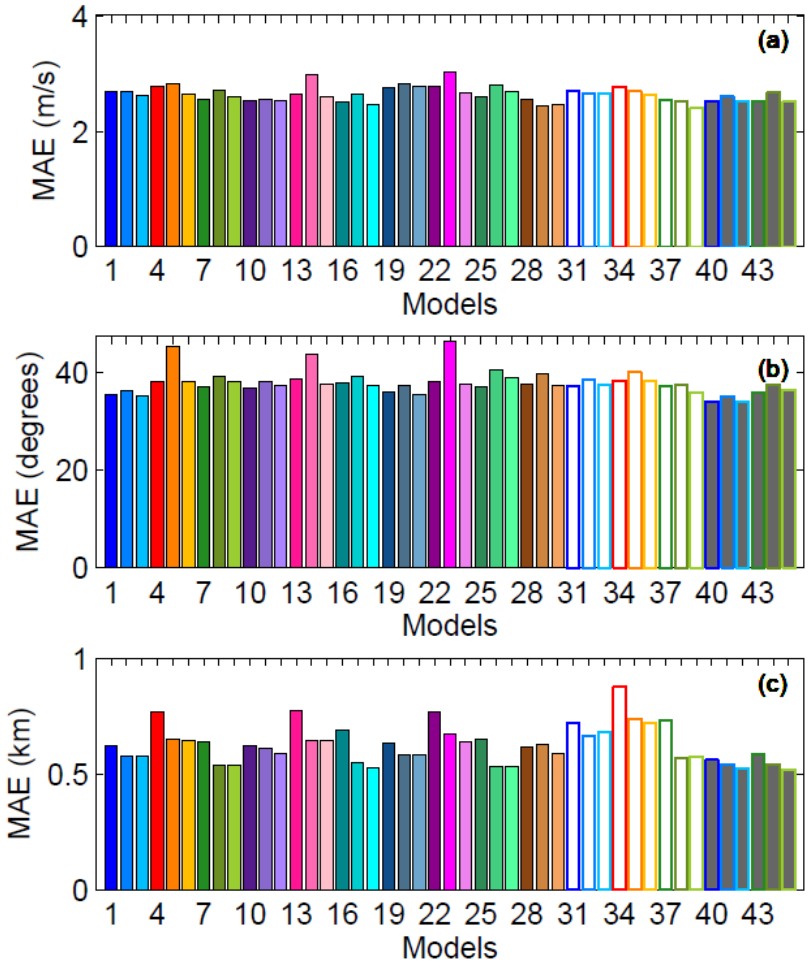

Figure 2-7. Regional averages of the monthly average wind speed (a), wind direction (b) and PBLH (c) mean absolute errors (MAE) for the different models (see Table 2-1 for model references).

2.3.3.2 Root mean square error (RMSE)

We evaluated the daily model errors by computing the RMSE (Figure 2-8) at each of the sites over a month. This analysis complements the MAE analysis but also assumes a Gaussian distribution of the errors. The RMSE results for wind speed (Figure 2-8a), wind direction (Figure 2-8b), and PBLH (Figure 2-8c) are highly variable across the region. Two sites show frequently small RMSE for both PBL wind speed (ABR and BNA) and PBLH (SGF and APX). However, there is more variability in the sites that shows a higher RMSE for the all the variables except for PBLH,

where LBF and DDC are showing the highest RMSE across the ensemble. We are able to identify some sites with specific patterns in the RMSE, however, none of these sites were having the same behavior for the three variables. After evaluating all the model configurations at each of the sites, we were unable to identify an optimal configuration that would satisfy all of the sites.

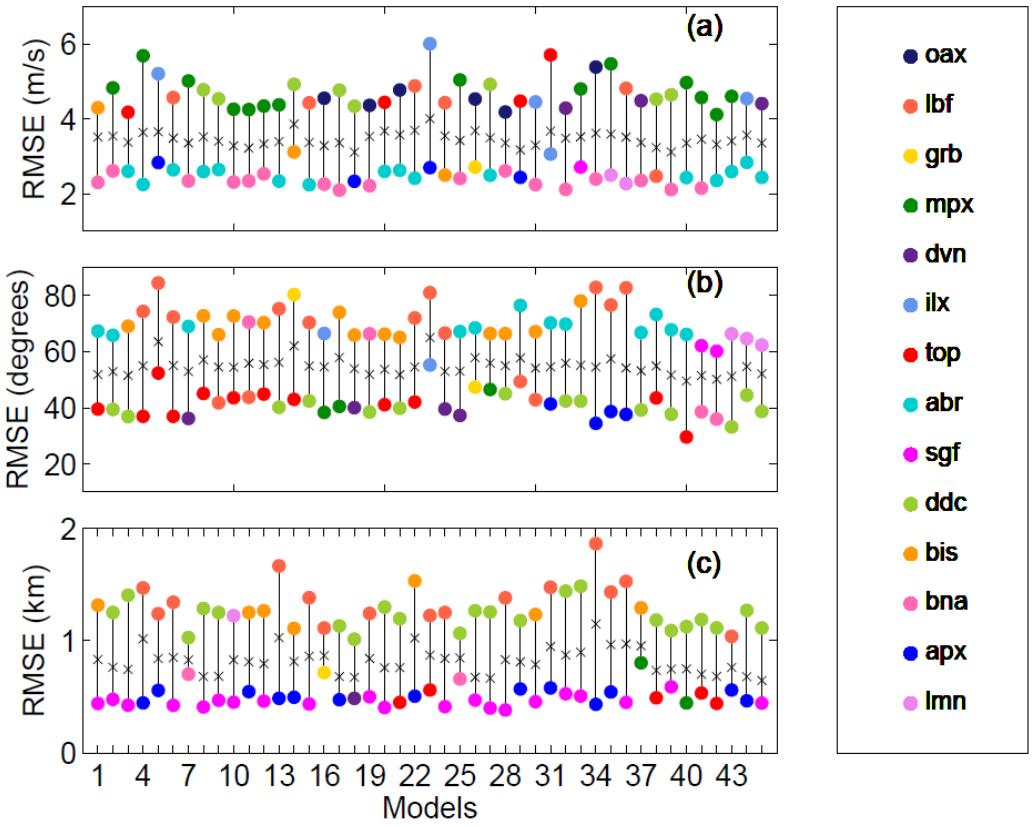

Figure 2-8. Maximum and minimum monthly average wind speed (a), wind direction (b) and PBLH (c) root mean square errors (RMSE) for the different sites and models. The abscissa shows the different model configurations, while the ordinate shows the RMSE for the meteorological variables. Each dot represents the site with the maximum and minimum RMSE for each model and variable; while the x represents the regional mean of the RMSE. (See Table 2-1 for model references.)

Because of the difficulties we encountered evaluating the RMSE for each rawinsonde site, we explore the RMSE at the regional level. Figure 2-9 shows the regional RMSE of wind speed

(Figure 2-9a), wind direction (Figure 2-9b), and PBLH (Figure 2-9c). For both PBL wind speed and wind direction, we found no significant difference in the regional RMSE. Differences among configurations are larger in the regional RMSE of the PBLH (Figure 2-9c), with most of the configurations being between 680m and 972m, except for four configurations that show higher RMSEs (1018m to 1149m). Although the regional RMSE for both wind speed and wind direction are fairly constant, the two variables have the same two model configurations with the highest RMSE. These two configurations share the same LSM scheme (RUC) and the same PBL parameterization scheme (MYJ) (models 14 and 23 see Figure 2-9a,b and Table 2-1). Similar to the MAE, the model configurations that show the highest PBLH RMSE include the same LSM scheme (RUC) and PBL parameterization scheme (YSU) (models 4, 13, 22 and 34 see Figure 2-9c and Table 2-1).

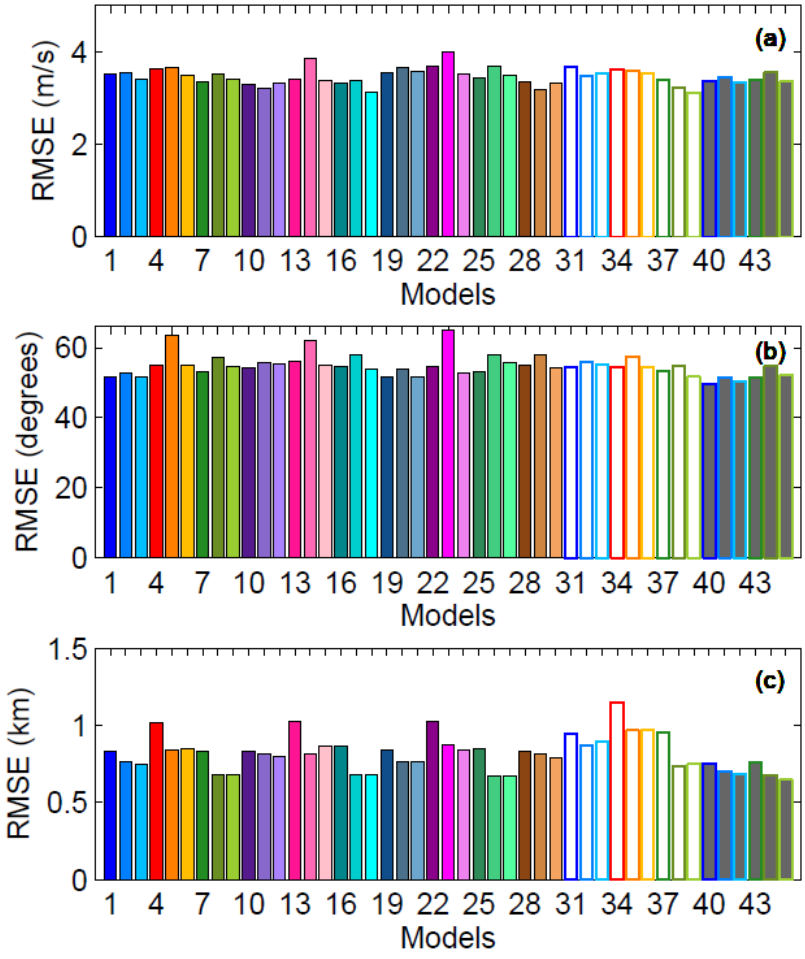

Figure 2-9. Regional averages of the monthly average of wind speed (a), wind direction (b) and PBLH (c) RMSE for the different models (see Table 2-1 for model references).

2.3.3.3 Mean Bias Error (MBE)

The average over- or under-estimation of the different model configurations is assessed by computing the MBE for each of the sites over a month (Figure 2-10). Similar to the other metrics, the MBE was computed for wind speed (Figure 2-10a-c), wind direction (Figure 2-10d-f), and PBLH (Figure 2-10g-i). In this study, a positive (resp. negative) MBE means the model configurations are systematically higher (resp. lower) than the observation. Figure 2-10 shows the MBE of three sites, however, the analysis was performed for all the sites (not shown) and they were

split up by region. The three sites shown are located in three differ rent regions of the domain: ABR in the west (Figure 2-10a, d, g), DVN which is close to the center of the domain (Figure 2-10b, e, h), and BNA in the eastern part of the domain (Figure 2-10c, f, i). Most of the model configurations show positive PBL wind speed MBE (overestimation) for the majority of the rawinsonde sites (e.g., Figure 2-10b-c), however, one site shows both positive and negative MBE for the different model configurations (e.g., Figure 2-10a). The majority of the sites shows all the models with a positive bias are located in the eastern and center areas of the domain, whereas, the sites that show both positive and negative bias for the different simulations are located in the western part of the domain. The MBE for PBL wind direction is highly variable across the different rawinsonde sites (ex. Figure 2-10d,e,f). At the majority of the sites, the simulations had both positive and negative biases, except two sites where all the models are positive or negative (not shown). The MBE of PBLH does not show any site with all the configurations being positively or negatively biased (Figure 2-10g,h,i). However, two rawinsonde sites in this region show most of the largest negative (DDC) or positive (LBF) MBE for more than half of the model configurations and these two sites are located in the western region of the domain. Again, the evaluation of the MBE does not allow us to identify a configuration with a smaller bias across all sites and all variables. However, from the PBLH MBE, we were able to identify some models that were showing a systematic bias across the different sites. All the model configurations that show the highest positive bias across all sites were models that include YSU as a PBL parameterization scheme and RUC or Thermal Diffusion as an LSM parameterization scheme (red and green bars in the highest positive bias; Figure 2-10). Conversely, model configurations with the most negative PBLH biases across the different sites include MYJ as a PBL parameterization scheme.

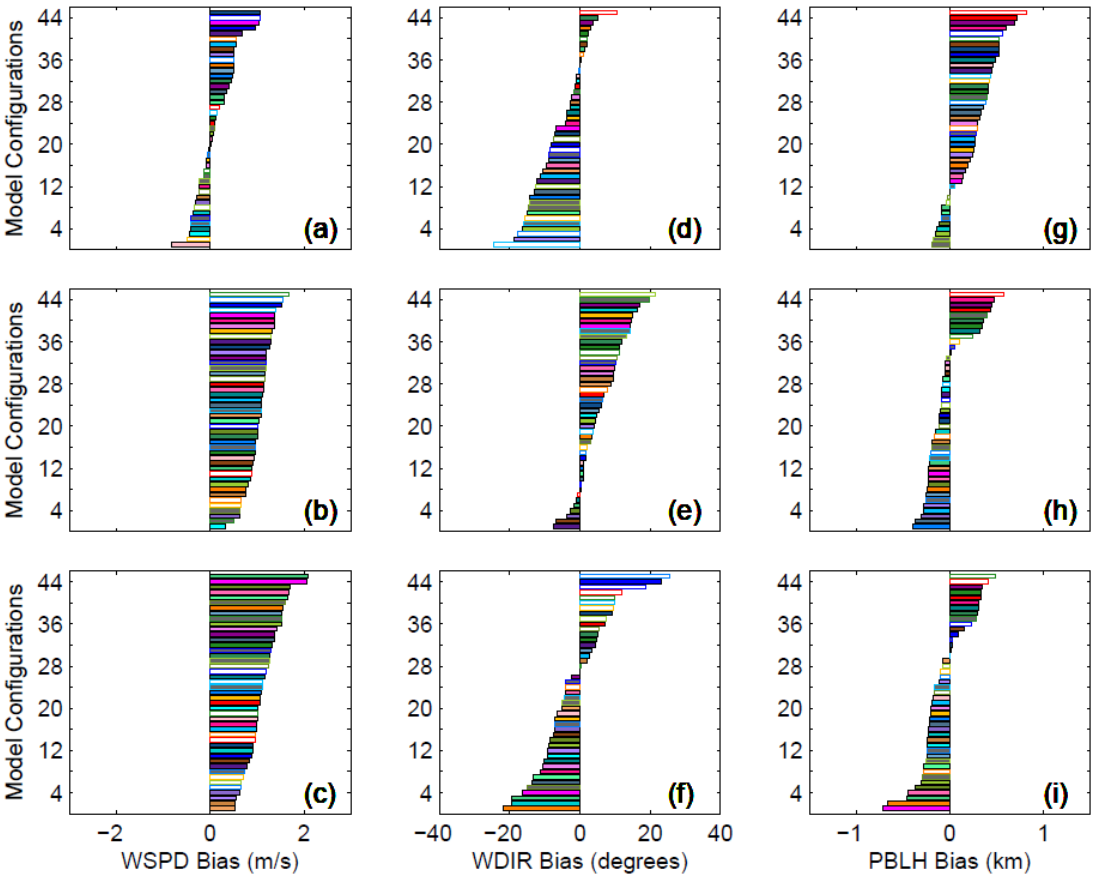

Figure 2-10. Monthly average wind speed (a-c), wind direction (d-f) and PBLH (g-i) MBE for rawinsonde sites ABR (first row), DVN (second row) and BNA (third row). Models are sorted from the negative to the positive bias.

We find remarkable variation in the regional MBE both as a function of different model configurations and across the variables of PBL wind speed (Figure 2-11a), PBL wind direction (Figure 2-11b), and PBLH (Figure 2-11c). The regional wind speed bias was less than 1.2 m/s for the entire ensemble; however, all the members show a positive bias except for one configuration (Figure 2-11a). Four model configurations show the highest regional wind speed MBE, but these biases remain between 1 and 1.2 m/s. Regional wind speed biases show that all configurations that use YSU as the PBL parameterization (e.g., models 1, 4, 7, 10 see Figure 2-11a and Table 2-1) have greater biases than the rest of the PBL schemes. The regional MBE for wind direction is highly

variable across the different model configurations. Models using YSU as PBL scheme tend to show a positive bias in the wind direction (e.g., models 1, 4, 7, 10 see Figure 2-11b and Table 2-1), whereas models that use MYJ as PBL scheme show a negative bias (e.g., models 2, 5, 8, 11 see Figure 2-11b and Table 2-1). Similar to wind direction, the regional PBLH bias is highly variable, with model configurations showing both positive and negative biases. Any model configuration that uses YSU as PBL parameterization scheme shows a positive bias, larger than the rest of the PBL schemes (e.g., models 1, 4, 7, 11 see Figure 2-11b and Table 2-1). The model configurations that do not include a cumulus parameterization (white filled bars; Figure 2-11c) also show positive PBLH biases, with one exception, regardless of the choice of LSM or PBL scheme used. The PBL wind speed analysis shows that the two model configurations with the smallest regional MBE (± 0.1 m/s) share the same LSM (SLAB) and PBL (MYNN 2.5) parameterization schemes. For PBL wind direction, two of the three model configurations with the lowest MBE (± 0.1 degrees) use the same LSM (Noah) and PBL (YSU) parameterization schemes. All 15 model configurations with the lowest MBE for PBLH (± 100 m or less) share the same PBL parameterization schemes (MYJ and MYNN 2.5). Although the configurations that provide the lowest regional MBE is not the same across all variables, we found that the MYNN 2.5 PBL scheme is the most common element of the best-performing configurations. It is possible to pick a configuration which, averaged across the month and across the region, has a very small MBE (less than 5% for both PBL wind speed and PBHL) for all three of the meteorological variables.

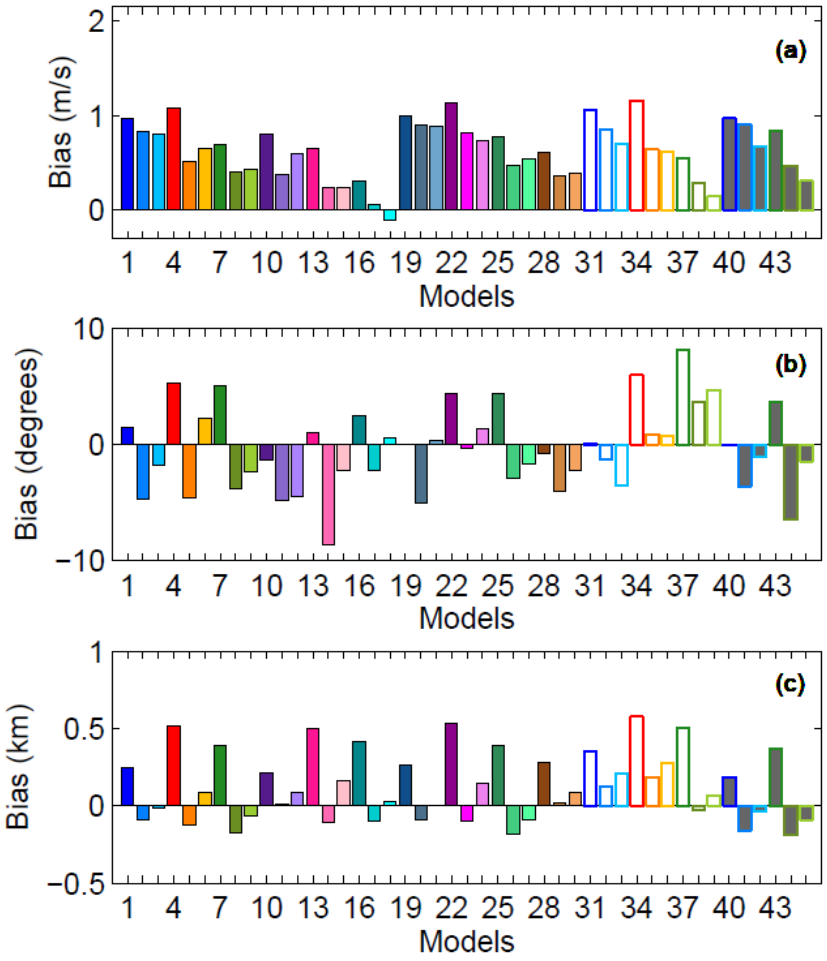

Figure 2-11. Regional average of the monthly average of wind speed (a), wind direction (b) and PBLH (c) bias for the different models (See Table 2-1 for model references.)

The MBE statistics reveals a spatial pattern in PBL wind speed and PBLH that is not present in the other metrics. The map of PBL wind speed MBE (Figure 2-12a) shows that the ensemble is positively biased in the eastern region of the domain. However, sites in the western region of the domain show that the ensemble average has either negative or near-zero PBL wind speed MBEs. The PBLH MBE map (Figure 2-12c) also shows a clear spatial pattern, with the highest values, nearly all positive, at sites located in the western part of the domain, whereas the sites in the eastern part of our domain show a smaller MBE and no distinct regional sign. PBL wind

direction does not show any spatial pattern in the ensemble mean of the MBE (Figure 2-12b). Across the region, we found that our ensemble of simulations can produce an MBE range of ±1.5 m/s in wind speed, ±20 degrees for wind direction, and ± 400 m for PBLH. Overall, the spatial patterns show that no configuration can avoid spatial biases across the region.

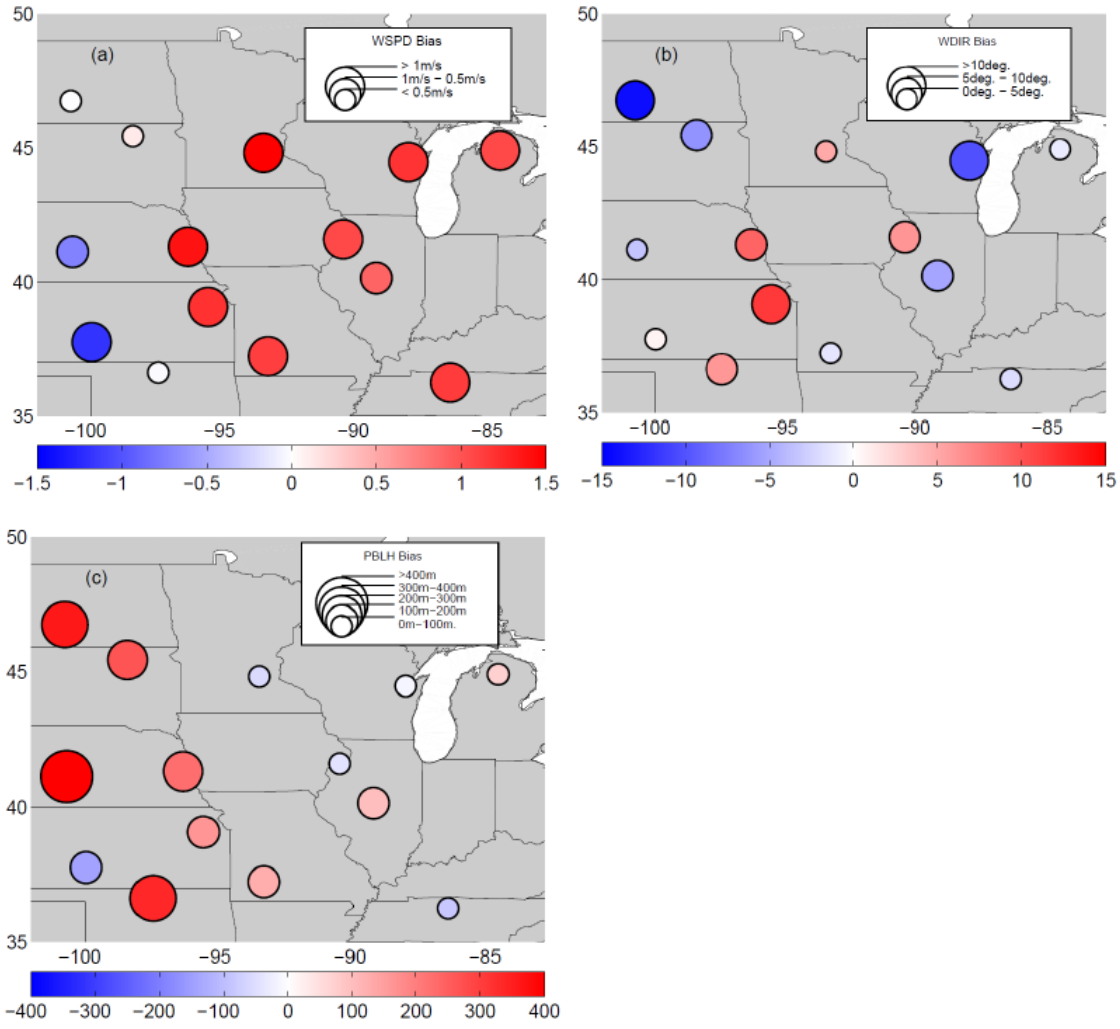

Figure 2-12. Ensemble mean of the mean bias error (MBE) for PBL wind speed (a), PBL wind direction (b) and PBLH (c).

2.3.4 Best Model Performance

Regional statistics allow us to determine the best-performing model configurations. We selected individual physical parameterization schemes to compute independent scheme-specific statistics. Table 2-3 shows the regional MAE of PBL wind speed, PBL wind direction, and PBLH for each of the physics parameterizations used. We found that changes in the physics parameterization can have a maximum impact of 0.2 m/s in the wind speed, 4 degrees in the wind direction, and 121 m in the PBLH throughout the region. Regional MAE results show that the choice of physical parameterization schemes has a small impact on PBL wind speed and wind direction but a large impact on PBLH. The regional statistics indicate that the choice of LSM, PBL, and CP parameterization scheme are the main contributors to the variability in PBLH, while the rest of the physics parameterizations show less impact. The RUC LSM scheme and MYJ PBL scheme show the highest MAE for the three meteorological variables, except for PBLH for which the YSU PBL scheme provides the highest MAE. The lowest MAE for the three variables was found with the MYNN 2.5 PBL parameterization scheme and SLAB LSM parameterization scheme, except for PBL wind direction where Noah LSM has the lowest MAE.

These results agree with section 2.3.3.1 where we found some model configurations with higher regional MAE for wind speed and wind direction using RUC as the LSM scheme and MYJ as the PBL scheme. It also agrees with the PBLH MAE results in which all the high MAE (Section 2.3.3.1) were found for configurations that use RUC as the LSM scheme and YSU as the PBL scheme. In addition, Table 2-3 shows that configurations that do not include a CP scheme have a significant error in the PBLH when compared to the other configurations. Nevertheless, for the three variables (i.e., wind speed, wind direction, and PBLH) both the choice of LSMs and PBL schemes show larger differences than the rest of the physical parameterization schemes.

Table 2-3. Regional average of wind speed, wind direction and PBLH mean absolute error (MAE) for each of the physics schemes.

Physics	MAE			
	Schemes	WSPD (m/s)	WDIR (degree)	PBLH (m)
LSM	Noah	2.67	36.23	620.57
	RUC	2.76	39.4	709.93
	SLAB	2.59	37.44	588.67
PBL	YSU	2.63	36.49	682.95
	MYJ	2.71	39.02	601.22
	MYNN 2.5	2.59	36.48	590.9
Cumulus	Kain-Fritsch	2.68	37.55	615.88
	Grell-3D	2.61	37.9	626.65
	No Cumulus	2.62	37.29	695.81
Micro	WSM-5class	2.64	37.37	613.61
	Thompson	2.7	37.91	617.68
Rean.	NARR	2.65	36.35	580.75
	GFS	2.57	35.03	544.78

Regional statistics for the meteorological variables and physics parameterizations were done for the RMSE (Table 2-4). Results show that changes in physics parameterization can affect the RMSE of PBL wind speed by 0.3 m/s, PBL wind direction by 6 degrees, and PBLH by 183 m. Similar to the regional MAE, the RMSE results show that the choice of physical parameterization schemes has greater impact on PBLH than on wind speed and wind direction. The wind speed has less sensitivity to the physical parameterization schemes than wind direction and PBLHs. Nevertheless, we conclude that the LSM and PBL schemes have a significant impact on the RMSE. For PBLH, we found that not having a cumulus parameterization can have an additional 100 m increase in the RMSE. RUC LSM scheme and MYJ PBL schemes show the highest RMSE for wind speed and wind direction. For PBLH, RUC LSM scheme, YSU PBL scheme, and the absence of a cumulus parameterization scheme produce high regional RMSE. The lowest RMSE was found

with MYNN 2.5 PBL scheme and SLAB LSM scheme, except for wind direction where Noah LSM has the lowest RMSE.

These results agree with section 2.3.3.2 where we indicate that specific model configurations show higher RMSE for wind speed and wind direction; these model configurations use RUC as the LSM scheme and MYJ as the PBL scheme. For PBLH, most of the model configurations that showed the highest RMSE were using RUC as the LSM scheme and YSU as the PBL scheme. One configuration did not use cumulus parameterization but it provides the largest RMSE out of all the model configurations. We also established that the simulations that showed the smallest regional RMSE for the three meteorological variables use the same PBL scheme (Section 2.3.3.2; MYNN 2.5), which also agrees with the results in Table 2-4.

Table 2-4. Regional average of wind speed, wind direction and PBLH root mean square error (RMSE) for each of the physics schemes.

Physics	RMSE			
	Schemes	WSPD (m/s)	WDIR (degree)	PBLH (m)
LSM	Noah	3.48	53.17	821.68
	RUC	3.61	56.58	936.72
	SLAB	3.36	54.2	753.62
PBL	YSU	3.46	52.96	901.09
	MYJ	3.53	56.61	777.65
	MYNN 2.5	3.39	53.03	774.43
Cumulus	Kain-Fritsch	3.5	54.37	804.99
	Grell-3D	3.36	55.55	818.74
	No Cumulus	3.46	54.07	916.04
Micro	WSM-5class	3.44	54.44	806.93
	Thompson	3.54	54.72	810.08
Rean.	NARR	3.45	52.9	755.89
	GFS	3.41	51.09	703.14

Table 2-5 shows the regional PBL wind speed, PBL wind direction, and PBLH MBE for each of the physics parameterizations used. In this table we can observe that all the different

parameterization schemes overestimate (i.e. create a positive MBE) in PBL wind speed over the region. These results agree with section 2.3.3.3 where all the model configurations show a positive MBE in wind speed except for one configuration (see Figure 2-12a). Most of the schemes or models show a negative MBE (underestimation) in wind direction except for two LSMs schemes (RUC and SLAB), one PBL scheme (YSU), and the simulations that do not use a cumulus parameterization. For PBLH, we found that all the choices of physical parameterization show a positive MBE (overestimation) of the PBLH, except for MYJ PBL parameterization scheme. The simulations with the highest positive PBLH bias used RUC LSM, YSU PBL scheme, and no cumulus parameterization option.

Table 2-5. Regional average of wind speed, wind direction and PBLH mean bias error (MBE) for each of the physics schemes.

Physics	MBE			
	Schemes	WSPD (m/s)	WDIR (degree)	PBLH (m)
LSM	Noah	0.81	-1.94	108.39
	RUC	0.71	0.73	219.4
	SLAB	0.38	1.4	96.72
PBL	YSU	0.84	2.74	381.44
	MYJ	0.54	-2.95	-67.49
	MYNN 2.5	0.51	-0.55	56.21
Cumulus	Kain-Fritsch	0.71	-0.19	86.3
	Grell-3D	0.35	-1.96	132.51
	No Cumulus	0.67	2.21	250.6
Micro	WSM-5class	0.68	-0.97	89.8
	Thompson	0.72	-0.4	104.15
Rean.	NARR	0.69	-0.89	49.37
	GFS	0.7	-1.2	12.98

2.3.5 Sensitivity of CO_2 mixing ratios to model configurations

In this study, we explored the sensitivity of meteorological variables to transport variations. CO_2 mixing ratios were excluded from the analysis to avoid interpreting errors from CO_2 surface fluxes and CarbonTracker global CO_2 mixing ratios. We use our different model configurations, which all share the exact same surface fluxes and identical boundary conditions, to explore the impact of transport errors on CO_2 mixing ratios. WRF transport errors will have an impact on the modeled CO_2 fields in the domain as the CO_2 is transported from the boundaries (initialized by CarbonTracker data) to the rest of the domain. However, the errors from incorrect transport from the boundaries are insignificant compared to the more localized errors associated with the transport of local CO_2 surface fluxes. Here we present the impact of model configurations on the CO_2 mixing ratios compared to the total CO_2 mixing ratio mismatches (or residuals), i.e. hourly model-data differences.

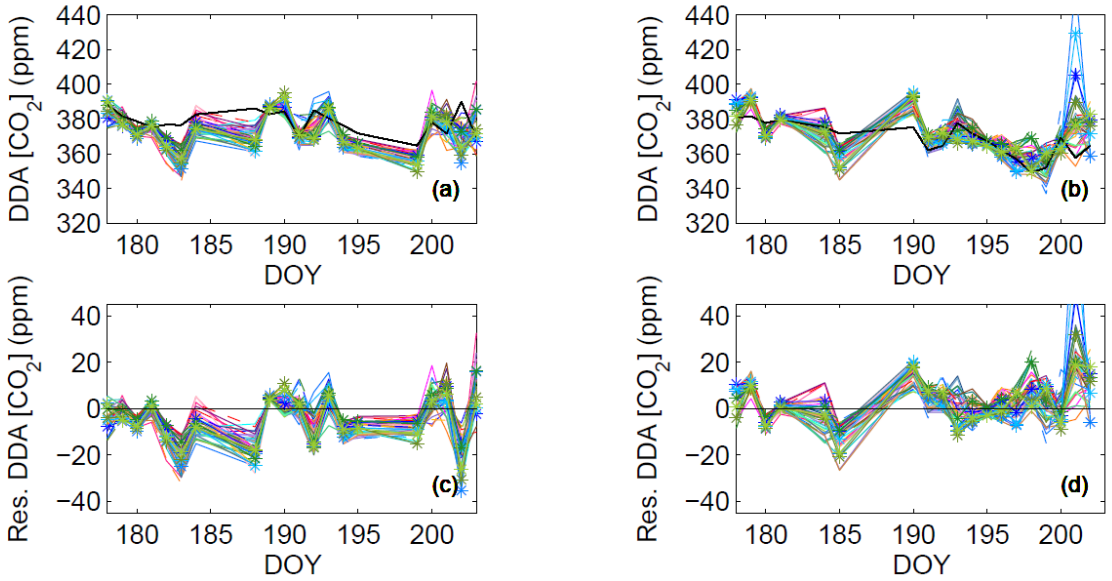

Figure 2-13. Observed (black line) and simulated (colored lines) DDA CO_2 mixing ratio (ppm) at Centerville (RCV) (a) and Kewanee (RKW) (b). Residuals (model-data mismatch) of DDA CO_2 mixing ratios at Centerville (RCV) (c) and Kewanee (RKW) (d).

Figure 2-13 shows simulated and observed atmospheric CO_2 mixing ratios for Centerville (Figure 2-13a) and Kewanee (Figure 2-13b). For this period, both sites show larger residuals compared to the ensemble spread (RMSD) for several periods (e.g. DOY 183-184 or DOY 194-199). This result suggests that transport model errors from our ensemble only represent a fraction of the total residuals. The remaining mismatches are therefore attributable to CO_2 surface fluxes or CO_2 boundary conditions. Over the region, most of the sites show that the ensemble generally underestimates the atmospheric CO_2 mixing ratios (e.g., Centerville-Figure 2-13a, c), except for one site, Kewanee (Figure 2-13b, d), where the CO_2 concentrations are overestimated for the majority of the members. We want to note here that this ensemble has not been properly calibrated and may under-estimate the total transport error. Further analyses of the ensemble will correct for the under- or over-dispersion of the ensemble.

To evaluate the performance of the different model over the month, we compute the correlation coefficient, normalized standard deviation, and the normalized root-mean-square

(RMS) difference (Taylor, 2001) for each of the sites. These results are presented as Taylor Diagrams on Figure 2-14 using the daily daytime average observed and simulated CO_2 mole fractions (Figure 2-14). All the models tend to overestimate the variability at three sites (e.g., Figure 2-14b) and the rest of the sites show that all of the models had both an overestimation and underestimation of the variability. The highest correlation ($r \geq 0.5$) for the majority of the sites is found at the Round Lake (RRL) site (Figure 2-14a), the other six sites have correlations within a range of 0.1 and 0.8 (e.g., Figure 2-14b). During this period, we did not find any model that shows a specific pattern across the region for the three statistics included in the Taylor Diagram.

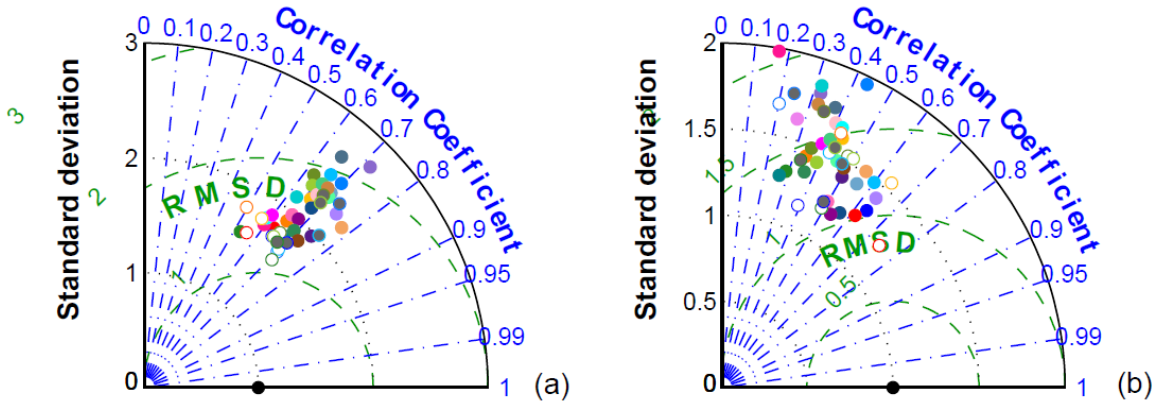

Figure 2-14. Taylor Diagram comparing observations versus simulations at (a) Round Lake (RRL) and (b) Centerville (RCV), using DDA CO_2 mole fractions. Black dot represents the observation reference.

The spatial correlation of the errors between the meteorological variables and the CO_2 mole fraction is evaluated at the different rawinsonde and CO_2 tower sites (Figure 2-15). To perform this evaluation, we decided to compare the mean ensemble MBE in CO_2 mixing ratios with mean ensemble MBE of the different meteorological variables. Both wind speed (Figure 2-15a) and wind direction (Figure 2-15a) show a low correlation coefficient, whereas PBLH (Figure 2-15a) shows a higher correlation with the CO_2 mole fraction errors among the different sites. Although the correlation varies with different variables, we did not find any change in the correlation that could

be associated with the distance among the different sites. These results suggest that the errors in the in situ CO_2 mixing ratios are directly related to the PBLH MBE, whereas correlations with wind MBE, which are strongly influenced by the magnitude and location of sources around the towers, remain low. We conclude here that first-order errors of in situ CO_2 mixing ratios are driven primarily by mean errors in the PBLH across the domain.

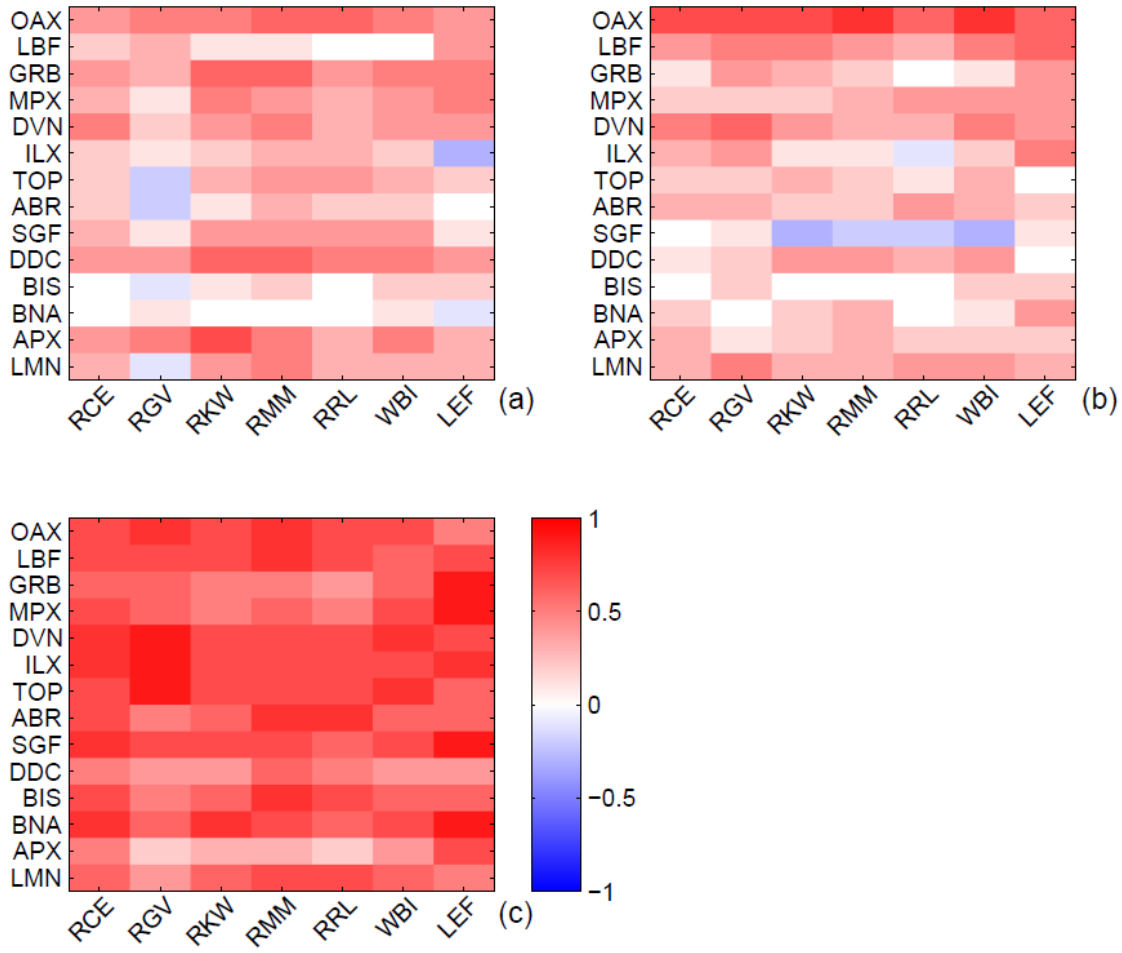

Figure 2-15. Tower and rawinsonde site specific spatial correlation coefficients between ensemble mean MBE of (a) wind speed, (b) wind direction and (c) PBLH and ensemble mean MBE of DDA CO_2 mole fractions. The abscissa shows the different CO_2 tower sites, while the ordinate shows rawinsonde sites.

2.4. DISCUSSION

The evaluation of the RMSD of CO_2 mixing ratios shows that all the physics parameterizations have a significant impact on the simulated values, except for the microphysics parameterization (Figure 2-3). Previous research has focused on the potential impact of PBL schemes errors on CO_2 mixing ratios (e.g. Kretschmer et al., 2012,2014; Lauvaux and Davis 2014). Results from our study indicate that other physics parameterization schemes including the LSM and CP generate errors of similar magnitude on simulated CO_2 mixing ratios. The influence that each physical parameterization impacts CO_2 mixing ratios (Figure 2-3) is similar to the PBLH errors (Figure 2-4c). The similarity may reflect the high correlation that exists between PBLH errors and CO_2 mixing ratio errors (see Figure 2-15c). Results from this study agree with previous studies that the misrepresentation of the PBLH plays an important role in atmospheric CO_2 errors. Figure 2-3 also shows that the reanalysis has an impact on atmospheric CO_2 mixing ratios, which indicates that even if the wind speed and wind direction errors do not show a high correlation with atmospheric CO_2 errors (see Figure 2-15a-b), these two variables can contribute to the errors in CO_2 mixing ratios due to the high impact that the reanalysis has on both of them (Figure 2-4a, b). However, we have to ponder our conclusions for three reasons: (1) we explore fewer microphysics options (only two) compared to the amount of PBL, cumulus, and LSM with three options for each, (2) we also evaluate only two reanalysis products for this evaluation, (3) this evaluation was performed over a limited period of time and location. We also note that some of the models (i.e. RUC configurations) were only run using the FNL reanalysis product, which may cause some under-estimation of the variability as this configuration contributes significantly to the errors of all the meteorological variables.

The MAE and RMSE for each of the meteorological variables do not show a specific behavior from the different model configurations at the site level. However, these statistics averaged over the region provide more insights regarding the different model configurations. For

example, the RMSE and MAE increase for the three meteorological variables when RUC is used as an LSM in the model configuration, whereas the PBL schemes that contributes to these errors varies from YSU for PBLH and MYJ for wind speed and wind direction (Zhang et. al., 2009; Yerramilli et al. 2010). The MBE, however, provides additional information for the different meteorological variables at both local and regional scale that the rest of the metrics do not provide. Over the region, we found that 10 out of the 14 rawinsonde sites show all the model configurations to be positively biased (i.e., overestimation of the wind speed). This can be shown in Figure 2-12a, where the ensemble mean bias shows all the sites in the eastern and center regions of the domain have a positive bias. Figure 2-11a shows that the wind speed is positively biased for all the models regardless of the configuration. This positive bias in the wind speed confirms what has being previously record in past studies that also had shown an overestimation of the wind speed for WRF mesoscale model (Cheng et al., 2005; Zhang et al., 2009; Yerramilli et. al., 2010). One potential reason for this positive bias can be associated with the reanalysis products, as shown in Figure 2-4b, where the meteorological driver data have a significant impact on the wind speed variability. Concerning the PBL schemes, we found that YSU at the regional scale provides the highest wind speed bias compared to the rest of the PBL schemes, this problem has been shown previously by Hu et al. (2013), especially during night time. The MBE of the PBLH also provides some other insight from the different physics schemes used. Rawinsonde sites over the western region of the domain shows certain model configurations with the highest positive bias compared to the rest. This high positive bias over the West of the domain is shown in map from Figure 2-11c where PBLH shows large positive biases in the western region of the domain and slightly negative biases in the east part of the domain. We suspect that this gradient can be associated with problems in the models in representing the gradient that exists of warmer and drier areas in the west and cooler and moister areas in the eastern portion of the domain (Molod et al., 2015). In addition, some of the models show a systematic behavior in the PBLH, where the highest positive biases were found in

configurations that use RUC as the LSM and YSU as the PBL scheme in the western region of the domain. This is unlike the eastern region of the domain, where the highest biases were dominated by configurations that use Thermal Diffusion as the LSM and YSU as the PBL scheme. To understand the potential cause of these differences in errors across the domain, we compute the MBE of the sensible heat at different eddy covariance stations available over the region. We found that Thermal Diffusion LSM tends to over-estimate the sensible heat over in the east and RUC LSM over the western region of the domain, which is reflected in the PBLH biases (not shown. However, this is not necessarily only a problem with the LSM, but also the interaction of these two LSMs with YSU PBL scheme because similar results have been shown in previous studies (e.g., Zhang et al., 2009). Finally, we note that the biases were larger for the PBLH when model configurations did not include a cumulus parameterization, creating a positive bias regardless of the model configuration. At 10 km resolution, the contribution from unresolved convection may accumulate subgrid-scale convection. In general, we were not able to find an optimal configuration but we found the MYNN 2.5 PBL scheme shows a better performance across the different variables (similar to Coniglio et al., 2013).

Over the region, we found that the different model configuration tends to underestimate the CO_2 mole fractions, except for Kewanee (RKW) site (Figure 2-13b). However, Kewanee is a site over the region that often shows a behavior that is significantly different when compared to the other sites (Díaz Isaac et al., 2014). The different models also show correlations that range between 0.1 and 0.8 for the majority of the sites, with both overestimation and underestimation of the variability, but we did not find a model that behaves optimally across the region (Figure 2-14). In our study we try to make a connection between the errors in the meteorological variables and the CO_2 mole fractions. We found a low correlation between wind speed (Figure 2-15a) or wind direction (Figure 2-15b) and the CO_2 mole fractions. However, we know errors in these two variables can impact the distribution and magnitude of the inverse CO_2 fluxes over the region. The

only variable that shows a high correlation with CO_2 mole fraction errors was the PBLH (Figure 2-15c). However, these correlations are not associated with the distance among the sites. This result is consistent with the high impact that errors in PBLH can have on CO_2 mole fraction errors (Stephens et al., 2007; Gerbig et al., 2008; Kretschmer et al., 2012).

2.5. CONCLUSION

We evaluated the atmospheric transport errors using multiple simulations with varying physical parameterization schemes and reanalysis driver data. The transport errors were quantified using observation of meteorological variables (i.e., wind speed, wind direction, and PBLH). Each of the physics parameterization schemes were tested to first determine whether the PBL schemes have the largest impact on the modeled CO_2 mole fraction errors among other parameterizations. This study shows that PBL schemes are not the only contributors of the CO_2 errors. Other physics schemes like LSMs and Cumulus parameterization have a significant impact on CO_2 errors. The meteorological variables show a similar result where all the physics parameterization schemes have a significant impact, except for microphysics. The PBLH was the only meteorological variable with ranking of the physics schemes importance identical to the CO_2 mole fractions.

The different model configurations show highly variable behaviors across rawinsonde sites for the different meteorological variables. Whereas wind speed and direction mean errors vary across sites, PBLH metrics reveal specific configurations affected by systematic errors across all the rawinsonde sites. To obtain more regional information of the different model configurations, we computed domain averages for each of the statistics. Although MAE and RMSE are fairly consistent across sites, we were able to identify specific configurations with larger errors: using RUC as the LSM and MYJ as the PBL scheme for wind speed and wind direction, and YSU PBL scheme for PBLH. For the regional MBE estimates, we observed that all the WRF configurations

show a systematic positive bias (i.e., overestimation) in wind speed across the region. The over- or under-estimation of the PBLH across the region is highly controlled by the PBL scheme, LSM, and cumulus parameterization.

The different model configurations gave us additional insights about the magnitudes of the atmospheric transport errors that can be encountered over this region. However, multiple challenges remain, such as the validity of our ensemble to represent the entire error magnitudes. We showed here that errors vary across the region. If these errors are not well represented in a regional inversion, the distribution of the inverse fluxes will be erroneous. Our ensemble has a systematic positive bias in the wind speed regardless of the configuration. These systematic errors will be propagated into the inverse fluxes, which is unavoidable unless other techniques including the assimilation of data are used. Finally, without generalizing for other areas, no optimal model configuration was found for the entire region. Therefore, random and systematic errors will remain. To understand how these errors will propagate into the fluxes, a representative ensemble of simulations that represents the atmospheric transport errors would help characterize the errors and their spatio-temporal structures in future regional inversion systems.

Chapter 3

Calibration of a Multi-Physics Ensemble for Greenhouse Gas Atmospheric Transport Model Uncertainty Estimation

3.1 INTRODUCTION

Atmospheric inversions are used to assess the exchange of CO_2 between the biosphere and the atmosphere (e.g., Gurney et al., 2002; Baker et al., 2006; Peylin et al., 2013). The atmospheric inversion or "top-down" method combines a prior distribution of surface fluxes with a transport model to simulate CO_2 concentrations and adjust the fluxes to be optimally consistent with the observations (Enting, 1993). Large differences exist among inverse flux estimates independent of the spatial scales (e.g., Gurney et al., 2002; Sarmiento et al., 2010; Peylin et al., 2013; Schuh et al., 2013). These posterior flux uncertainties arise from limited atmospheric data density (Gurney et al., 2002), uncertain prior fluxes (Corbin et al., 2008; Gourdji et al., 2010) and errors in atmospheric transport (Stephens et al., 2007; Gerbig et al., 2008; Picket-Heaps et al., 2011).

Atmospheric inversions based on Bayesian inference depend on the prior flux error covariance matrix and the observation error covariance matrix. The prior flux error covariance matrix represents the statistics of the mismatch between the true fluxes and the prior fluxes, but the limited density of flux observation limits our ability to characterize these errors (Hilton et al., 2013). The observation error covariance describes errors of both measurements and the atmospheric transport model. In the atmospheric inversions the model errors tend to dominate the measurement errors (e.g. Gerbig et al., 2003; Law et al., 2008). Additionally, the atmospheric inversion assumes that the atmospheric transport uncertainties are known, therefore, the method propagates the atmospheric transport model errors to inverse fluxes limiting their optimality. Unfortunately,

rigorous assessments of the transport uncertainties within atmospheric inversions are limited. Estimation of the atmospheric transport errors and their impact on CO_2 fluxes remains a challenge.

A limited number of studies are dedicated to quantify the uncertainty in the atmospheric transport models in detail and translate this information into the impact on the CO_2 mixing ratio and inverse fluxes. The atmospheric Tracer Transport Model Intercomparison Project (TransCom) has been dedicated to evaluate the impact of atmospheric transport models in atmospheric inversion systems (e.g., Gurney et al., 2002; Law et al., 2008; Peylin et al., 2013). These experiments, have shown the importance of the transport model resolution to avoid any misrepresentation of atmospheric signals (Law et al., 2008), especially more atmospheric observations are added. Diaz Isaac et al., (2014) showed how two models with two different resolution using the same surface fluxes can lead to large model-data differences in the atmospheric CO_2 concentrations that can yield significant errors on the inverse fluxes. However, other elements of the atmospheric transport models besides the spatial resolution has been taken into account as additional sources of error. Errors in the horizontal wind (Lin and Gerbig, 2005) and in the vertical transport (Stephen et al., 2007; Gerbig et al. 2008; Kretschmer et al., 2012) have been considered as two of the major uncertainties in the CO_2 modelling. Lin and Gerbig (2005), estimate the impact of horizontal wind error on CO_2 concentrations and conclude that uncertainties in CO_2 due to advection errors can be as large as 6ppm. Other studies have considered that critical errors of the atmospheric transport models reside in the misrepresentation of the vertical mixing (e.g., Denning et al., 1995; Stephens et al., 2007; Gerbig et al., 2008). Therefore, some studies have been dedicated to evaluate the effects that planetary boundary layer height (PBLH) has on CO_2 concentrations (Gerbig et al., 2008; Williams et al., 2011; Kretschmer et al., 2012). Some of these studies have estimated an uncertainty of approximately 3ppm on CO_2 concentrations due to PBLH errors during the summer time using different model spatial resolutions (Gerbig et al., 2008; Kretschmer et al., 2012). These studies have attributed the errors to the lack of sophisticated sub-grid parameterization, especially PBL schemes

and land surface models (LSMs). This led other studies to evaluate the impact of different parameterization specifically PBL schemes (Kretschmer et al., 2012; Lauvaux and Davis, 2014; Feng et al., 2016). These studies have found systematic errors of several ppm in atmospheric CO_2 that can generate biased fluxes. While there is an agreement that errors in the vertical mixing and advection can affect directly the inverse flux errors, these errors are significantly influenced by different components of the atmospheric models.

Atmospheric transport models have different sources of uncertainty, such as the boundary conditions, initial conditions, model physics parameterization schemes and parameter values. This has caused the ensembles to become a major tool for the quantification of atmospheric transport uncertainties. Different approaches have being taken in the carbon cycle community to represent the model uncertainty: (1) the multi-model ensembles that encompass models from different research institutions around the world (e.g. TransCom experiment; Gurney et al., 2002; Baker et al., 2006; Patra et al., 2008; Peylin et al., 2013; Houweling et al., 2010), (2) multi-physics ensembles that involve different model physics configurations generated by the variation of different parameterization schemes from the model (e.g., Kretschemer et al., 2012; Yver et al., 2013; Lauvaux and Davis 2014; Feng et al., 2016) and (3) multi-analysis that consists of running a model over the same period using different analysis fields (where perturbations can be added) (e.g., Lauvaux et al., 2009b; Miller et al., 2015). These ensembles are informative (e.g., Peylin et al., 2013; Kretshmer et al., 2012; Lauvaux and Davis 2014), but have some shortcomings. In some cases, the ensemble spread includes a mixture of transport model uncertainties and other errors such as the variation in prior fluxes or the observations used. Other studies have only varied the PBL scheme parameterizations. None of these studies have carefully assessed whether or not their ensemble spreads represent the true transport errors.

Ideally, an ensemble requires a significant number of members to avoid sampling noise and the lack of dispersion of the ensemble members (Houtekamer and Mitchell, 2001) to explore

and quantify the uncertainties. However, large ensembles are computationally expensive. The limitations in computational resources lead to restrictions including the set-up of the model (e.g., model resolution, nesting options, duration of the simulation) and the number of ensemble members. It will be desirable to generate an ensemble that is capable of representing the transport errors, that does not include any redundant information and whose members are essential to maintaining the spread of the ensemble.

Various post-processing techniques can be used to calibrate or "down-select" to a subset of ensemble members that represent the errors approximately equal to a large ensemble of 50 or more members (e.g., Alhamed et al., 2002; Garaud and Mallet, 2011; Lee et al., 2012a; 2016). Some of these techniques are principal component analysis (e.g., Lee et al., 2012a), K-means cluster analysis (e.g., Lee et al., 2012b) and hierarchical cluster analysis (e.g., Alhamed et al., 2002; Yussouf et al., 2004; Johnson et al., 2011; Lee et al., 2012b;2016). Riccio et al. (2012), applied the concept of "uncorrelation" to reduce that amount of members without using any measurement data. Solazzo and Galmarini (2014) reduced the number of members by finding a subset of members that maximize a statistical performance skill such as the correlation coefficient, the root-mean-square error or the fractional bias. Other techniques applied less commonly to the calibration of the ensembles include simulated annealing and genetic algorithms (e.g. Garaud and Mallet, 2011). All these techniques are capable of eliminating those members that are redundant, and generating an ensemble with a smaller number of members that still represents, perhaps even more faithfully uncertainty of the atmospheric transport model.

In the last two decades, the development of ensemble methods has improved the representation of transport uncertainty using the statistics of large ensembles to characterize the statistical spread of atmospheric forecasts (e.g. Evensen, 1994a, 1994b). However, single-physics ensemble-based statistics are highly susceptible to model error, leading to under-dispersive ensembles (e.g. Lee et al., 2012). But large ensembles (>50 members) remain computationally

expensive and ill-adapted to assimilation over longer time scales such as multi-year inversions of long-lived species (e.g. CO_2). Smaller-size ensembles would be ideal, but most initial-condition-only perturbation methods produce unreliable and overconfident representation of the atmospheric state (Buizza et al. 2005). Therefore, we present here an optimal selection of model simulations based on different physics configurations (i.e. unperturbed) while minimizing persistent biases that might cause the ensemble to be biased. These conditions are ideal to propagate transport errors into CO_2 mixing ratios (i.e. mass conservation and continuity in air flow) in future regional inversion systems.

In this study we start with a large multi-physics/multi-analysis ensemble of 45-members presented in (Díaz-Isaac et al., in prep) that will go through a calibration process similar to the one explained in Garaud and Mallet (2011). There are two features of the ensemble that can be explored: reliability and resolution. The reliability is the probability that a simulation has of matching on average the frequency of an observed event. The resolution is the ability of the system to predict a specific event. The main goal is to generate an ensemble that will represent the uncertainty of the transport model with respect to meteorological variables of most importance in simulating atmospheric CO_2, wind speed, wind direction and PBLH. We will focus therefore on the criterion that will measure the reliability of the ensemble, i.e. the probability of the ensemble in representing the frequency of events (i.e. the spatio-temporal variability of the atmospheric state). To calibrate the ensemble, we will use two different techniques, simulated annealing and a genetic algorithm. We will evaluate which physical parameterizations play important roles in balancing the ensembles, and evaluate how well a pure physics ensemble can represent transport uncertainty.

3.2 METHODS

3.2.1 Generation of the Ensemble

We generate an ensemble using the Weather Research and Forecasting (WRF) model version 3.5.1 (Skamarock et al., 2005), including the chemistry module modified in this study for CO_2 (WRF-ChemCO_2). The ensemble consists of 45-members that were generated by varying the different physics parameterization and meteorological data. The land surface models, surface layers, planetary boundary layer schemes, cumulus schemes, microphysics schemes, and meteorological data (i.e., initial and boundary conditions) are alternated in the ensemble (see Table 3-1). All the simulations keep the same radiation schemes, both long and shortwave.

The different simulations were run using the one-way nesting method, with two nested domains (Figure 3-1). The coarse domain (d01) uses a horizontal grid spacing of 30km and covers most of the United States and part of Canada. The inner domain (d02) uses a 10km grid spacing, is centered in Iowa and covers the Midwest region of the United States. The vertical resolution of the model is described with 59 vertical levels, with 40 of them within the first 2km of the atmosphere. This work focuses in the simulation with the higher resolution, therefore only the 10-km domain will be analyzed.

The CO_2 fluxes for summer 2008 were obtained from NOAA Global Monitoring Division's CarbonTracker version 2009 (CT2009) data assimilation system (Peters et al., 2007; http://carbontracker.noaa.gov). The different fluxes that CT2009 propagates into the models are fossil fuel burning, terrestrial biosphere exchange, and the exchange with oceans. The CO_2 lateral boundary conditions were obtained from CT2009 mole fractions. Only the atmospheric transport fields should vary between each model configuration or ensemble member.

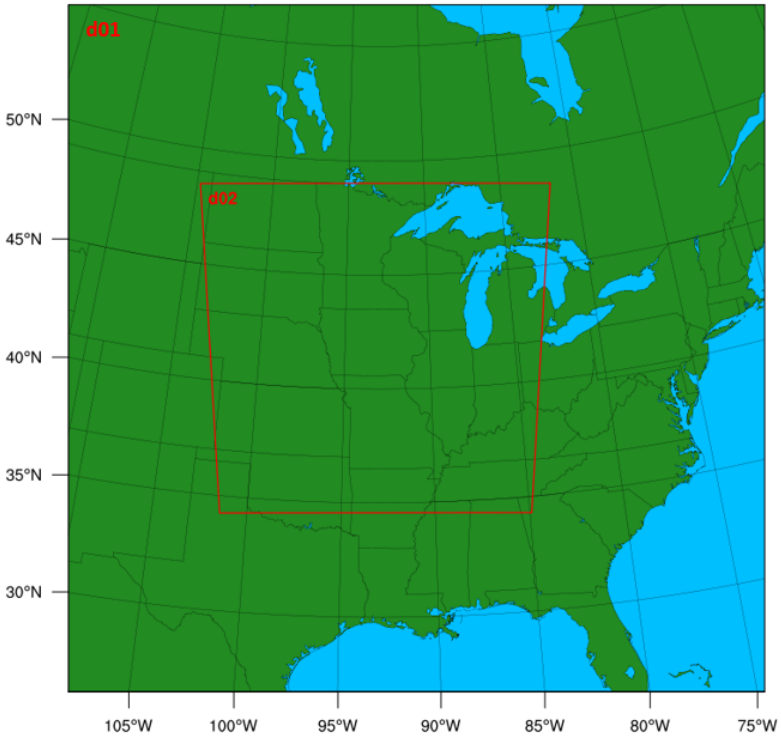

Figure 3-1. Geographical domain used by WRF-ChemCO_2 physics ensemble. The parent domain (d01) has a 30-km resolution, the inner domain (d02) has a 10-km resolution.

Table 3-1. Physics schemes used in WRF for the sensitivity analysis.

Parameter	Options
Land Surface Model	Noah LSM Rapid Update Cycle (RUC) LSM 5-layer Thermal Diffusion
Planetary Boundary Layer (PBL) scheme – Surface Layer scheme	Yonsei University (YSU) Mellor-Yamada-Janjic (MYJ) Mellor-Yamada-Nakanishi-Niino (MYNN)
Cumulus	Kain-Fritsch (KF) Grell-3Devenyi (G3D)
Microphysics	WSM 5-class Thompson et al., (2004)
Initial & Boundary Conditions	North America Regional Reanalysis (NARR) Global Final Analysis (FNL)

3.2.2 Dataset and Data Selection

Our interest is to calibrate our ensemble over the Midwest U.S. using the available observations available over this region. Therefore, the calibration of the ensemble will be done using the inner domain for each simulation. To perform the calibration we used balloon soundings collected over the Midwest region (Figure 3-2). Meteorological data were obtained from the University of Wyoming's online data archive (http://weather.uwyo.edu/upperair/sounding.html) for 14 rawinsonde stations over the U.S. Midwest region (Figure 3-2). To evaluate how the new calibrated ensemble impacts CO_2 concentrations we will use in-situ atmospheric CO_2 mole fraction data provided by seven communication towers (Figure 3-2) (Miles et al., 2012). Five of these towers were part of an experimental network, deployed from 2007 to 2009 and managed by a group at The Pennsylvania State University (Richardson et al., 2012; Miles et al., 2012). The other two towers (Park Falls-WLEF and West Branch-WBI) are part of the Earth System Research Laboratory/Global Monitoring Division (ESRL/GMD) tall tower network (Andrews et al., 2014), managed by NOAA. Each of these towers sampled air at multiple heights, ranging from 11 m AGL to 396 m AGL.

The ensemble will be calibrated for three different meteorological variables: wind speed, wind direction and planetary boundary layer height (PBLH). We will calibrate the ensemble with the late afternoon data (i.e., 0000 UTC) from the different rawinsondes. In this study, we use only daytime data, because we want to calibrate and evaluate the ensemble under the same well mixed conditions that CO_2 mole fraction are used to perform atmospheric inversions. For each rawinsonde site we will use wind speed and wind direction observations from approximately 300 m above ground level (AGL). We choose this observational level because we want the observations to lie within the well mixed layer, a similar strategy being employed for interpreting tall-tower CO_2 mole fractions. Many of CO_2 mole fractions that we will use were collected at 100 m AGL, but the data

from the radiosondes were very limited at that height. Simulated meteorological fields should be evaluated at a similar height as the observations. Therefore, simulated wind speed and wind direction are evaluated at level 11 (~350m).

The PBLH was estimated using the virtual potential temperature gradient ($\nabla\theta v$). The method identifies the PBLH as the first point above the atmospheric surface layer where (1) $\nabla\theta v$ is greater than or equal to 0.2 K/m, and (2) the difference between the surface and the threshold level virtual potential temperature is greater than or equal to 3 K($\theta_{vs} - \theta_v \geq 3K$).

WRF derives an estimated PBLH for each simulation, however the technique used to estimate the PBLH varies according to the PBL scheme used to run the simulation. For example, the YSU PBL schemes estimates PBLH using the Bulk Richardson number, MYJ PBL scheme uses the TKE to estimate the PBLH and MYNN PBL scheme uses QKE to estimate the PBLH. To avoid any errors from the technique used to estimate the PBLH, we decided to estimate the PBLH from the model using the same method used for the observations. Simulated PBLH will be analyzed at the same time as the observations, 0000 UTC, i.e., late afternoon in the study region.

We analyzed CO_2 mole fractions collected from the sampling levels at or above 100m AGL, which is the highest observation level across the MCI network (Miles et al., 2012). This ensures that the observed mole fractions reflect regional CO_2 fluxes and not near-surface gradients of CO_2 in the atmospheric surface layer (ASL) or local CO_2 fluxes (W. Wang et al., 2007). Both observed and simulated CO_2 mole fractions are averaged from 1800 to 2200 UTC (12:00-16:00 LST), when the daytime period of the boundary layer should be convective and the CO_2 profile well mixed (e.g., Davis et al., 2003; Stull, 1988). This averaged mole fraction will be referred to hereafter as daily daytime average (DDA).

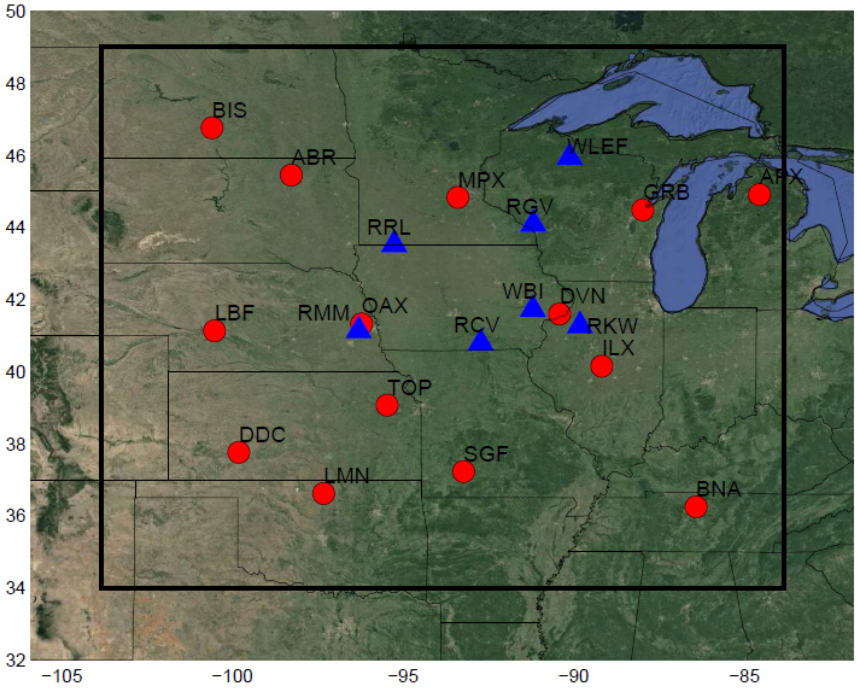

Figure 3-2. Map of the study region, including the CO_2 towers (blue triangles) and rawinsonde sites (red circles) locations.

3.2.3 Criteria or Verification Methods

The main reason to generate an ensemble is to estimate the atmospheric transport model uncertainties. In this research we want to test the performance of the ensemble and try to represent the uncertainties using an ensemble with a smaller number of members. A series of statistics analysis are used as criteria to measure the quality of the uncertainty estimation of the ensemble for the period of June 18 to July 21 of 2008. The different tools used for this evaluation include the rank histograms, rank histograms scores, Taylor diagrams, spread-skill relationship, ensemble bias and CO_2 error covariance matrices. These statistical analyses will be used to identify the characteristics of the different ensembles including the large ensemble and the calibrated ensembles.

3.2.3.1 Talagrand Diagram (or Rank Histogram) and Rank Histogram Score

The main goal is to identify a sub-ensemble that is representative of model transport uncertainties. To calibrate the ensemble, we need criteria to select the sub-ensemble. The rank histogram and the rank histogram scores are tools used to measure the spread and reliability of the ensemble.

The rank histogram (Anderson 1996; Hamill and Colucci 1997; Talagrand et al., 1999) is computed by sorting the corresponding modeled variable of the ensemble in increasing order and then a rank among the sorted predicted variable from lowest to highest is given to the observation. The ensemble members are sorted to define "bins" of the modelled variable, if the ensemble contains N members, then there will be N+1 bins. If the rank is zero then the observed variable value is lower than all the modelled variable values, and if it is N then the observation is greater than all of the modelled values. If the ensemble is reliable, the rank histogram should be flat, this happens when the probability of occurrence of the observation within each bin is equal. A rank histogram that deviates from the flat shape implies a biased, over-dispersive or under-dispersive ensemble. A "U-shaped" rank histogram indicates that the ensemble is under-dispersive, normally in this type of ensemble the observations tend to fall outside of the envelope of the ensemble and indicates a lack of variability. A "central-dome" histogram indicates that the ensemble is over-dispersive; this kind of ensemble has an excess of variability. If the rank histogram is overpopulated at either of the ends of the diagram, then this indicates that the ensemble is biased.

The rank histogram score is used to measure the deviation from flatness of a rank histogram:

$$\delta = \frac{N+1}{NM}\sum_{j=0}^{N}(r_j - \bar{r})^2 \quad (1)$$

and should ideally be close to 1 (Talagrand et al., 1999; Candille and Talagrand, 2005). In Eq.(1), N is the number of models, M is the number of observations, r_j the number of observations of rank j, and $\bar{r} = M/(N+1)$ is the expectation of r_j. In theory, the optimal ensemble has a score of one (1) when enough members are available. A score lower than one would indicate overconfidence in the results but would not affect the selection process. Nevertheless, a flat rank histogram does not necessarily mean that the ensemble is reliable or has enough spread. For example, a flat histogram can still be generated from ensembles with different conditional biases (Hamill, 2001). The flat rank histogram can also be produced when covariances between samples are incorrectly represented. Additional verification analysis has to be introduced to certify that the calibrated ensemble has enough spread and is reliable. Therefore, we present here additional metrics to evaluate the ensemble.

3.2.3.2 Spread-skill relationship

To verify that the ensemble captures the spatial and temporal structures of the atmospheric transport errors at fine scales (i.e. resolution), we computed the relationship between the spread of the ensemble and the skill of the ensemble over the entire data set (i.e. spread-skill relationship). The linear fit between the two parameters measures the correlation between the ensemble spread and the ensemble mean error or skill (Whitaker and Lough, 1998). The ensemble spread is calculated by computing the standard deviation of the ensemble and the mean error by computing the absolute difference between the ensemble mean and the observations. Ideally, as the ensemble skill improves (get smaller), the ensemble spread becomes smaller, and vice versa. Compared to

the rank histograms, spread-skill diagrams represent the ability of the ensemble to represent the errors at every time step rather than statistically represent the spatio-temporal variability of the atmosphere (i.e., englobe the observations to calibrate the ensemble). Here, the spread-skill relationship will be used as an additional verification method to assess the error representation of the ensembles.

3.2.3.3 Bias and Standard Deviation

Atmospheric inverse flux estimates are highly sensitive to biases and deviations from the mean. The bias, or the mean of the residuals (i.e. model-data mismatches), was used to assist the selection of the calibrated sub-ensemble. We identify a sub-ensemble that has minimal bias,

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(p_i) \tag{2}$$

where p_i is the difference between the modeled wind speed, direction or PBLH, and the observed value, M is the number of measurements and i sums over each of the rawinsonde measurements.

Standard deviation, used here for the ensemble evaluation, will indicate the deviation of the ensemble from the observations. We used the following definition:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i)^2} \tag{3}$$

to describe the deviation to the measurements. The equation describes the Root Mean Square Errors (RMSE) averaged across the members of any given ensemble.

3.2.3.4 Taylor Diagram

We use Taylor diagrams to describe the performance of each of the models of the large ensemble (Taylor, 2001). The Taylor diagram relies on three nondimensional statistics: the ratio of the variance (model variance normalized by the observed variance), the correlation coefficient, and the normalized center root-mean square (CRMS) difference (Taylor, 2001). The ratio of the variance or normalized standard deviation indicates the difference in amplitude between the model and the observation. The correlation coefficient measures the similarity in the temporal variation between the model and the observation. The CRMS is normalized by the observed standard deviation and quantifies the ratio of the amplitude of the variations between the model and the observations. The Taylor diagram will be used to have a general idea of the behavior of the different models with respect to the observations. An ideal Taylor diagram will be one that shows a different performance, therefore different statistical measures. This will help the calibration system to have numerous alternatives to generate the different calibrated sub-ensembles.

3.2.3.5 Ensemble-based error covariances

Transport model errors in atmospheric inversions are described in the observation error covariance matrix, assumed independent of the prior flux error covariances. Based on the calibration method, the reduced-size ensembles generated using two different optimization techniques show an optimal representation of the spatio-temporal variability in the three meteorological variables. Therefore, we examined both the ensemble spread (i.e., variances) and the structures in space and time (i.e., covariances) in the different ensembles. We evaluate the impact of the calibration on variances and covariances of CO_2. Because the limited number of members is likely to introduce sampling noise in the diagnosed error covariances, we compare the spatial extent of error structures between the full ensemble and the reduced-size ensembles. The

raw covariances were directly derived from the different ensembles to estimate the increase in sampling noise as a function of the ensemble size.

3.2.4 Calibration Methods

In this study, we want to test the ability to reduce the ensemble from 45-members to an ensemble with a smaller number of members that is still capable of representing the transport errors and does not include members with redundant information. We use the Garaud and Mallet (2011) technique to define the size of the calibrated sub-ensemble that each optimization technique will generate. Garaud and Mallet (2011) determine the size of the sub-ensemble by dividing the total number of observations by the maximum frequency in the rank histogram. We are going to generate sub-ensembles of three different size (number of members), to evaluate the impact that an ensemble size has for the representation of the atmospheric transport uncertainties. Each of the ensembles will be calibrated for the period of June 18 to July 21 of 2008.

Two optimization methods are used to select a sub-ensemble that minimizes the rank histogram score (δ), which is the criterion that each algorithm will use to test the reliability of the ensemble. These two optimization methods are simulated annealing (SA) and a genetic algorithm (GA). Each method will select a sub-ensemble that best represents the model uncertainties of the three meteorological variables selected; wind speed, wind direction and PBLH. We hypothesize that all three meteorological variables are important for CO_2 inverse flux estimates.

In this study, SA and GA techniques will generate a sub-ensemble S of size N. For the first test, we will use these algorithms to choose the combination of members that optimize the score of the reduced ensemble $J(S)$ (i.e., rank histogram score (δ as defined in Eq.(1)) for each variable. With this evaluation, we determine if each technique yields similar calibrated ensembles and if these calibrated ensembles are similar among the different meteorological variables. In the second

test, we calibrate the ensemble for all three variables simultaneously, where we use the sum of the score squared: $[J(S)]^2$ (i.e. δ^2):

$$[J(S)]^2 = [J_{wspd}(S)]^2 + [J_{wdir}(S)]^2 + [J_{pblh}(S)]^2, \qquad (4)$$

to control acceptance of the sub-ensembles. In Eq. (4), $J_{wspd(S)}$, $J_{wdir(S)}$ and $J_{pblh(S)}$ are the scores of the sub-ensemble for wind speed, wind direction and PBLH respectively. We will explore the ensemble characteristics identified by rank histogram scores for the individual meteorological variables, and for the sum of the rank histogram score for all three variables.

3.2.4.1 *Simulated Annealing*

Simulated annealing (SA) is a general probabilistic local search algorithm, described by Kirkpatrick et al. (1983) and Cerny et al. (1985) as an optimization method inspired from the process of annealing in metal work. Based on the Monte-Carlo iteration solving method, SA finds the global minimum using a cost function that gives to the algorithm the ability to jump or pass multiple local minima. In this case the optimal solution is a sub-ensemble with a rank histogram score close to 1.

The SA starts with a randomly selected sub-ensemble. The current state (i.e, initial random sub-ensemble) has a lot of neighbors states (i.e., other randomly generated sub-ensembles) in which a unit (i.e., model) is changed, removed or replaced. Let S be the current sub-ensemble and S' be the neighbor sub-ensemble. S' is a new sub-ensemble (i.e., neighbor) that is randomly built from the current sub-ensemble with one model added, removed or *replaced.* To minimize the score *J*, only two transitions to the neighbors are possible. First transition, if the score of the neighbor sub-ensembles *J(S')* is lower than the current sub-ensemble *J(S)*, then *S'* becomes the current sub-ensemble and a new neighbor sub-ensemble is generated. Second transition, if the score of the

neighbor sub-ensemble *J(S')* is greater than the current sub-ensemble J(S), moving to the neighbor S' only occurs through an acceptance probability. This acceptance probability is equal to $\exp\left(-\frac{J(S')-J(S)}{T}\right)$ and it only allows the movement to the neighbor S' if $u < \exp\left(-\frac{J(S')-J(S)}{T}\right)$. For the acceptance probability, *u* is a random number uniformly drawn from [0,1] and T is called temperature and it decreases after each iteration following a prescribed schedule. The acceptance probability is high at the beginning and the probability of switching to neighbor less at the end of the algorithm. The possibility to select a less optimal state *S'*, i.e., with higher *J(S')* is meant to escape local minima where the algorithm could remain trapped.

At the end of the process, we collect all the accepted sub-ensembles *S* and their respective scores *J(S)*. Then we choose the sub-ensemble that has the smallest score for the three variables and a lower bias compare to the sub-ensembles.

3.2.4.1 *Genetic Algorithm*

A genetic algorithm (GA) is a stochastic optimization method that mimics the process of biological evolution, with the selection, crossover and mutation of a population (Fraser and Burnell, 1970; Crosby, 1973; Holland, 1975). Let S_i be an individual; that is, a sub-ensemble, and let $P = \{S_1, \dots, S_i, \dots, S_{N_{pop}}\}$ be a population of N_{pop} individuals. As a first step in the GA a random population is generated (denoted P^0). Then this population will go through two steps (1) selection and (2) crossover. In the selection step, we select half of the best individuals with respect to the score (i.e., summation of the score of three variables *J(S)*). For the second step, a crossover among the selected individuals occurs when two parents create two new children by exchanging some ensemble members. A new population is generated with $N_{pop}/2$ parents and $N_{pop}/2$ children. This process is repeated until a criterion is satisfied. In this case the criterion is the number of iterations that we specified. This algorithm will provide at the end a population of individuals with a better

rank histogram score than the initial population. Out of all those individuals we choose the sub-ensemble with the best score for the three variables (i.e., wind speed, wind direction and PBLH) and with a smaller bias than the large ensemble.

3.2.5 Parameterization of the Selection Algorithms

Various inputs are required to guide the selection algorithms. For example, we typically need to choose the initial and final temperature (T_0 and T_f) for the SA and its schedule, the best population size (N_{pop}) for the GA and the number of iterations for each algorithm. The temperature of the SA, the N_{pop} of the GA and the iterations were chosen by running the algorithms multiple times and confirming that the system reached similar solutions with independent minimization runs. If similar solutions were not achieved within multiple SA or GA runs, the algorithm parameters were altered to increase the breadth of the search. For the SA we found that 20,000 iterations yielded similar solutions after multiple runs of the algorithm. For the GA, 30 to 50 iterations were sufficient as long as the ensemble was smaller than 8-members. For an ensemble of 10-members we needed to increase to 100 iterations. Another factor that was important in the SA was the initial temperature used in the algorithm and the temperature decrease for each iteration. While the temperature is high, the algorithm will accept with more frequency the poorer solutions, as the temperature is reduced, the acceptance of poorer solutions is reduced too. Therefore, we need to provide an initial (T_0) and final (T_f) temperature that allows the system to reduce gradually and allowing it to search more combinations of members to identify the best solution or sub-ensemble. We were able to explore more solutions with T_0 equal to 20 and T_f equal to 1e-3. We need to specify the GA the size of the population that we want to create to find the calibrated ensemble. The larger the population, the more we can explore the space to find the optimal solution. In this case we

chose N_{pop} of 280 individuals, the smaller the population the harder it was to find similar solution (sub-ensembles) after multiple runs.

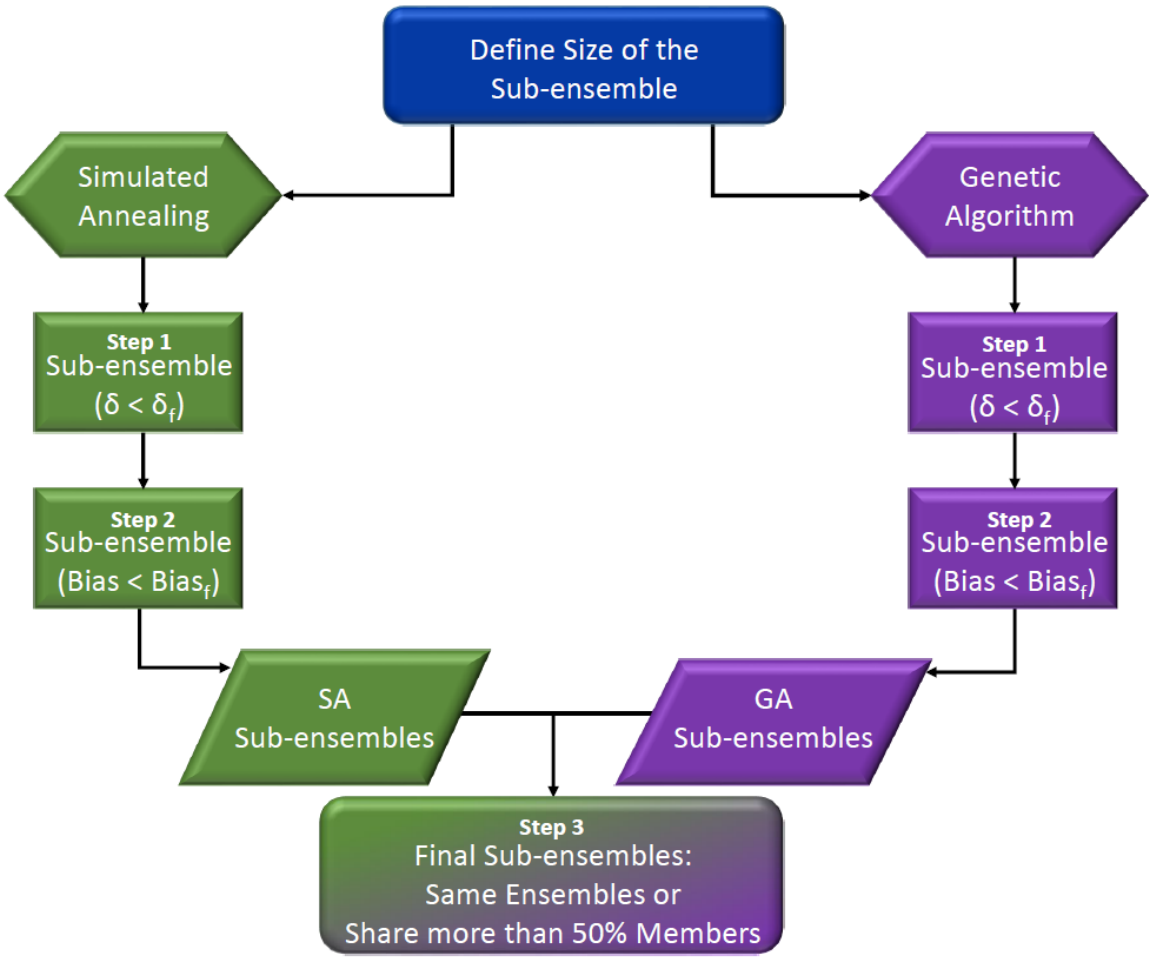

Figure 3-3. Diagram of the process of selection of reduced-sized ensembles.

3.2.6 Selection of the Optimal Reduced-sized Ensembles

The selection process is performed in three distinct steps to ensure that the final calibrated ensembles will be the optimal combinations of model configurations (Figure 3-3). First, the flatness of the rank histograms will control the acceptance of the calibrated sub-ensembles by the selection algorithms. The flatness is defined by equation (1) for the single-variable calibration and equation (3) for the calibration of the three variables simultaneously. The algorithms select multiple sub-

ensembles corresponding to a low score as defined by the lower percentile (here lowest 25%), as shown in Figure 3-4. To simplify the selection process, we accepted ensembles with a score smaller than six for each individual meteorological variable. As a second step, sub-ensembles accepted by SA and GA algorithms with a bias larger than the full ensemble are filtered out. The bias corresponds to the averaged model-data mismatch of the ensemble. Finally, the remaining calibrated ensembles are compared among SA and GA techniques to identify if both algorithms provide a common solution. If no solution was found by both techniques, the final sub-ensemble corresponds to the smallest score among the different solutions that shares more than half of the model configurations. This down-selection process was able to identify common ensembles in the majority of the experiments.

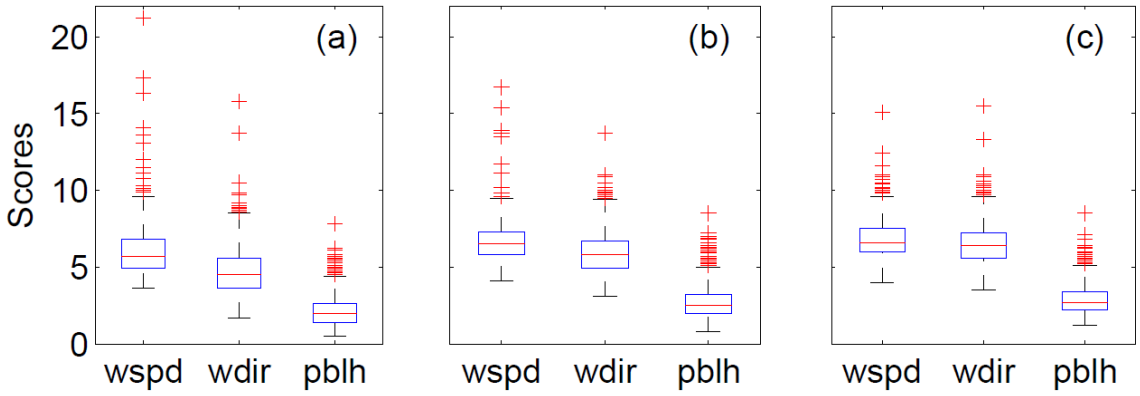

Figure 3-4. Box plot of the rank histogram scores of the different sub-ensembles of 10 (a), 8(b), and 5 (c) members accepted by the SA. Each figure shows the rank histograms scores for the different variables wind speed (wspd), wind direction (wdir) and PBLH. The outliers are plotted using the + symbol.

3.3 RESULTS

3.3.1 Evaluation of the Large Ensemble

In this section, we evaluate the performance of the different models and the large ensemble. Our goal is to test the skill (how close the models are to the observations) and the spread (how well the models represent the uncertainty) from the ensemble. We will evaluate the skill and the spread for wind speed, wind direction and PBLH across the region of study at 0000 UTC.

3.3.1.2 Model Skill

We evaluate the performance of the different models of the 45-member ensemble by computing the normalized standard deviation, center root mean square and the correlation coefficient for wind speed (Figure 3-5a), wind direction (Figure 3-5b) and PBLH (Figure 3-5c) (Taylor, 2001). Both wind speed and wind direction show the majority of the models with higher standard deviation (more variability) than the observations, whereas the simulations can over- or under-estimate PBLH variability depending on the model configuration. The models show correlations with wind speed and wind direction between 0.4 and 0.7, whereas the PBLH shows a smaller correlation, between 0.3 and 0.6. The range of variability in PBLH height will provide a wide range of alternatives for the calibrated sub-ensemble. Wind speed and wind direction do not show much variability among the different models. This reduces the potential selection of the models to produce a sub-ensemble that has sufficient spread.

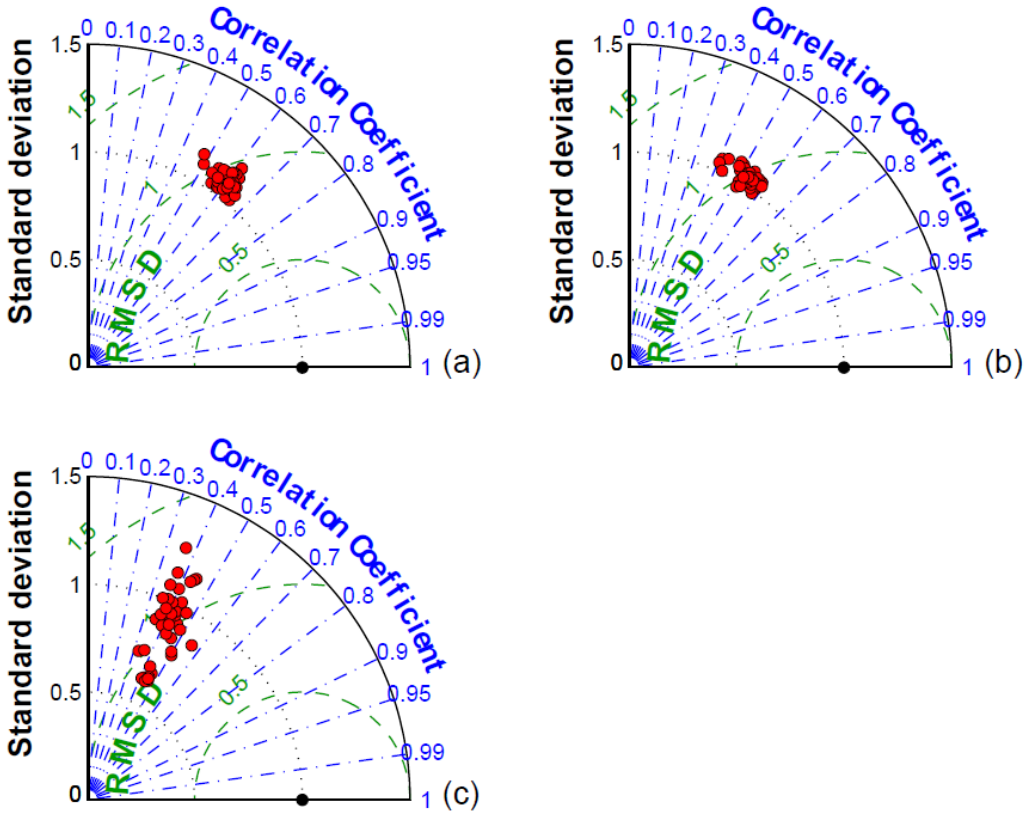

Figure 3-5. Taylor diagram comparing the 0000 UTC rawinsonde observations (300 m wind speed (a), 300 m wind direction (b) and PBL height (c)) to the 45 model configurations (red circles).

3.3.1.3 Reliability and Spread of the Ensemble

We illustrate the ensemble spread and how this ensemble captures the observations using the time series of the simulated and observed meteorological variables at 0000 UTC. Figure 3-6 shows the time series of the ensemble spread for wind speed, wind direction and PBLH at the GRB (Figure 3-6 a,c,e) and TOP (Figure 3-6 b,d,f) sites. The time series show qualitatively that simulated wind speed (Figure 3-6 a-b) and wind direction (Figure 3-6 c-d) have a smaller spread compared to PBLH (Figure 3-6 e-f). Figure 3-6 shows how the ensemble can have a small spread and still encompass the observations (i.e., DOY 183 Figure 3-6c); and have a large spread and not encompass the observation (i.e., DOY 174 Figure 3-6e). Similar to the Taylor Diagrams, the time

series of the large ensemble shows that more simulations have a similar behavior for wind speed and wind direction, whereas there is more variability in the performance of the model for the PBLH.

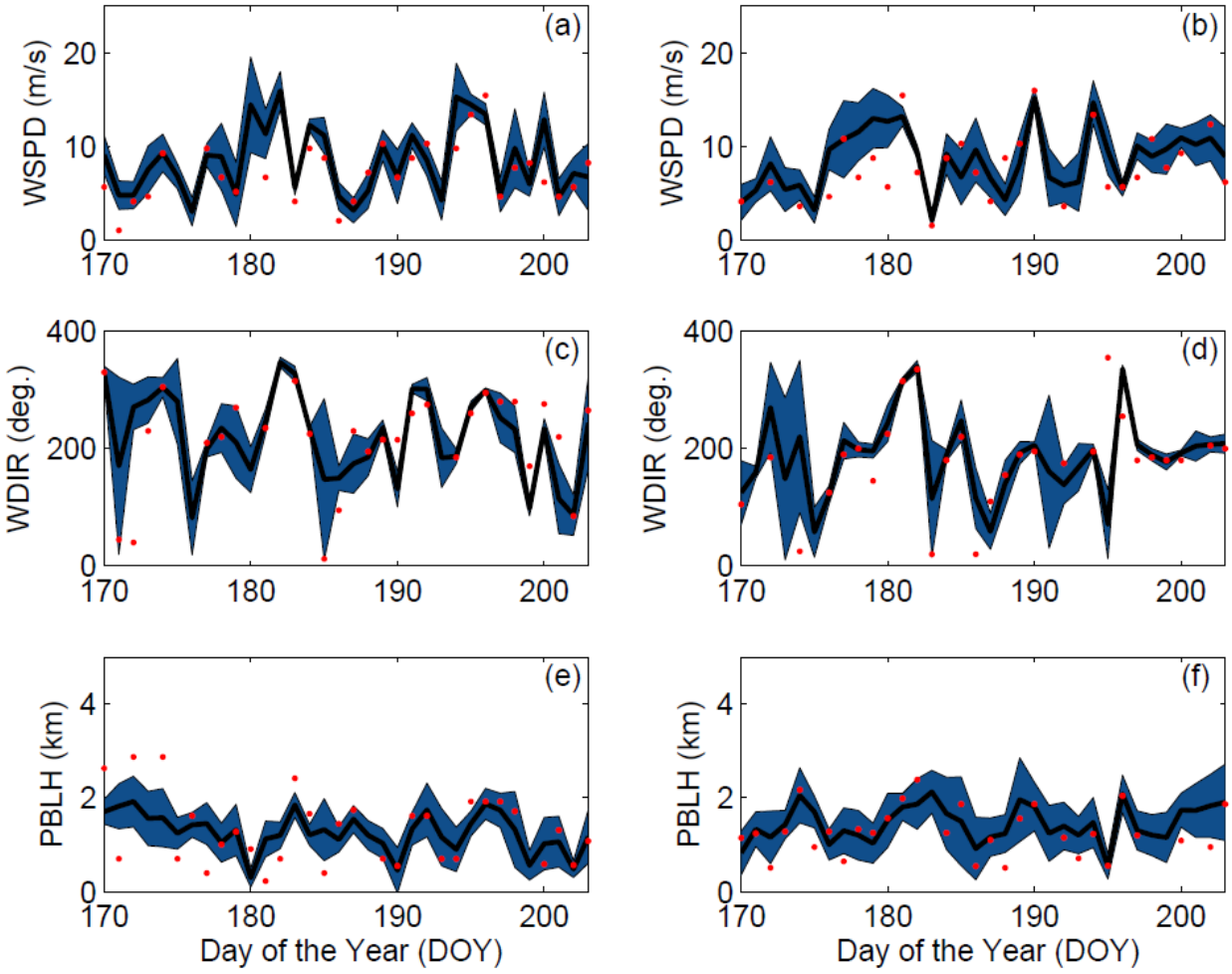

Figure 3-6. Time series of the large ensemble for wind speed (a-b), wind direction (c-d) and PBL height (e-f) at GRB (a,c,e) and TOP (b,d,f) sites. The shaded blue area represents the spread (i.e. RMSD) of the ensemble, the solid line the ensemble mean and the red dots the observations at 0000UTC.

Figure 3-7 shows the rank histograms of the 45-member ensemble for each of the meteorological variables that we are going to use to calibrate the ensemble (i.e., wind speed, wind direction and PBLH). In this rank histogram we include all the 14 rawinsonde sites. All the rank

histograms have a U-shape, where the highest frequencies are at the extremes of the histogram, indicating that the ensemble is under-dispersive, and that for many days the ensemble does not encompass the observations. Each rank histogram has the first rank as the highest frequency, indicating that observations are most frequently below the envelope of the ensemble. The rank histogram score for each of the variables is greater than one, confirming that we do not have optimal spread in our ensemble. Table 3-2 shows that both wind speed and wind direction have a higher rank histogram score (i.e., ≥ 6), compared to PBLH that has a score of 3.2. Wind speed and PBLH show a positive bias in the model-data mismatch across the regions, whereas wind direction has negative bias.

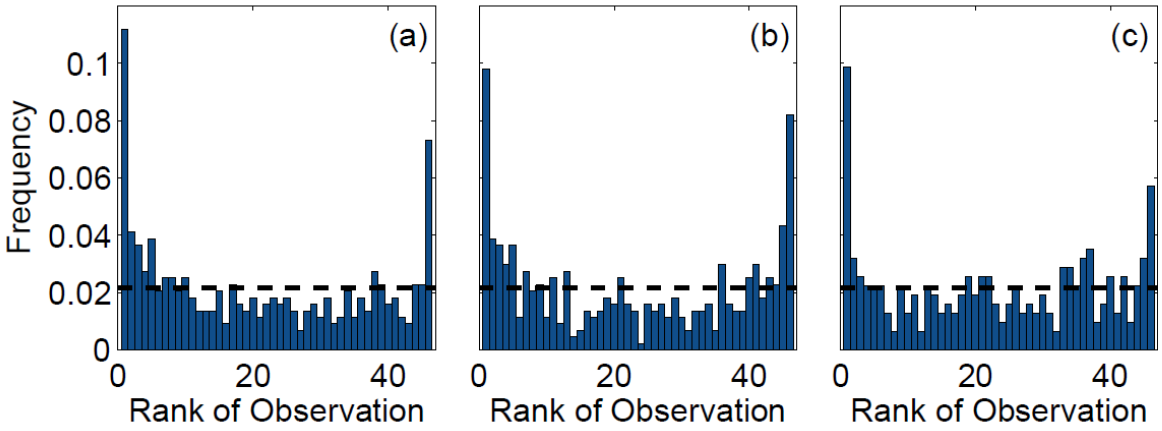

Figure 3-7. Rank histogram of the 45-member ensemble for wind speed (a), wind direction (b) and PBLH (c) using the 14 rawinsonde sites available over the region.

Table 3-2. Rank histogram score (δ), biases and standard deviation (σ) of the 45-member ensemble for wind speed, wind direction and PBL height computed across 14 rawindsonde sites using daily 0000 UTC observations for June 18 to July 21 of 2008 in the upper Midwest of the U.S.

Variables	δ	Bias	σ
Wind Speed	6.1	0.7 m/s	3.5 m/s
Wind Direction	6.2	-0.6 degrees	55.7 degrees
PBLH	3.2	98.2 m	787.5 m

Figure 3-8 shows the spread-skill relationship, another method that we use to examine the representation of errors of the ensemble. Wind direction (Figure 3-8b) shows a higher correlation between the spread and the skill compared to the PBLH (Figure 3-8c) and the wind speed (Figure 3-8a). The PBLH and wind speed show consistently larger skill than spread. This supports the conclusion that the large ensemble is under-dispersive for these variables. However, none of these variables shows a correlation equal to one; this implies that our ensemble spread does not match exactly the atmospheric transport errors on a day-to-day basis. This feature is common among ensemble prediction systems (Wilks et al., 2006) and should not impair the ability to identify the optimal reduced-size ensembles. However, linear coefficients are all between 0.7 and 1.1 which indicates that the ensemble captures a significant fraction of the daily model-data differences.

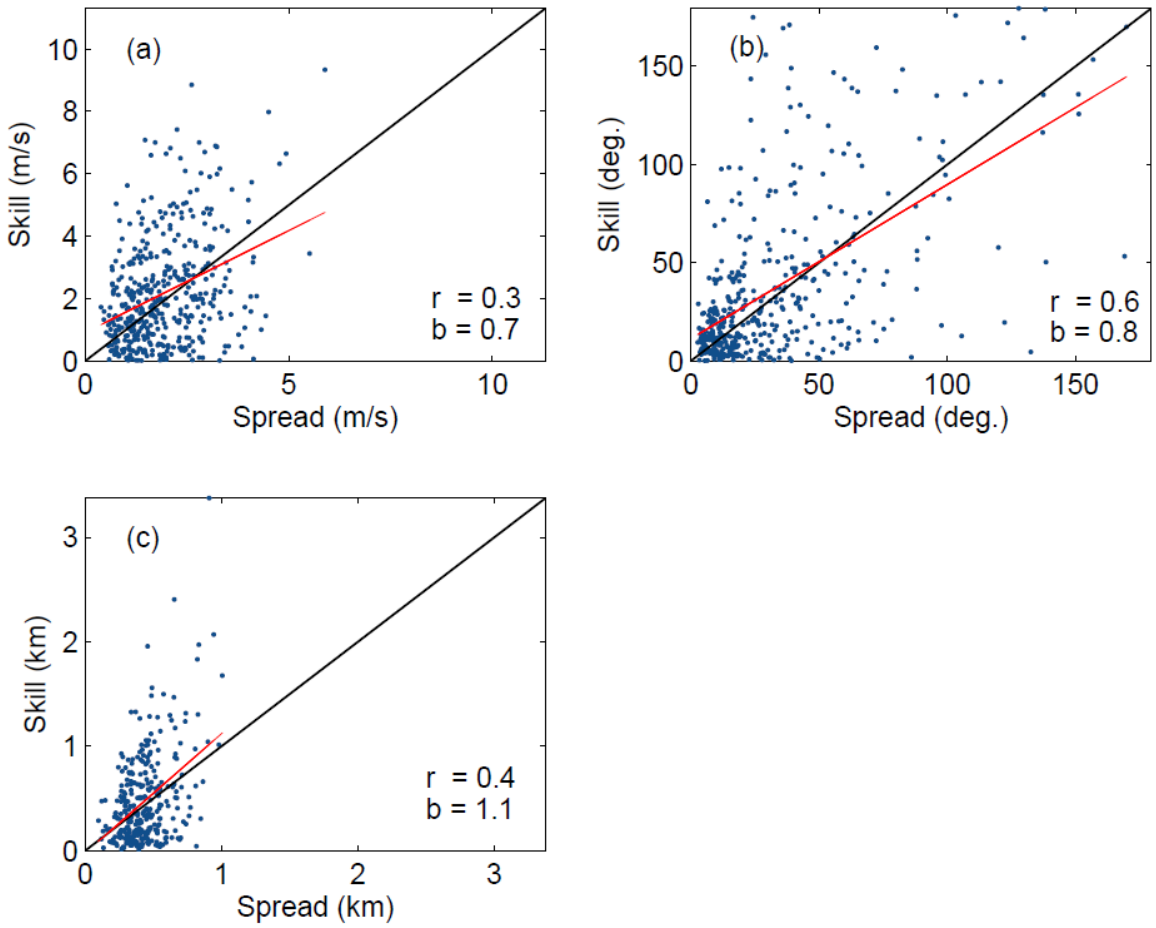

Figure 3-8. Spread-skill for (a) wind speed, (b) wind direction and (c) PBL height using the 14 rawinsonde sites available over the region. Each point represents the spread-skill of the different sites at different times. A one-to-one line is plotted in black and a line of best fit is plotted in red. Correlation (r) and slope (b) of the line of best fit of the spread-skill relationship.

3.3.2 Calibrated Ensemble

In this section, we will show the results of the calibrated ensembles generated with both SA and GA. Each calibration was performed for three different sub-ensemble sizes; the size of the ensembles is determined using the technique explained in Section 3.2.4. To compute the size of the sub-ensemble we need to use the maximum frequency of the rank histogram, in this case the maximum frequency is the left bar (r_0) of every rank histogram. This technique yields the

result that the calibrated ensemble should have about 8 to 10 members depending in the variable that to be used. Therefore, for this study we will generate three different ensemble sizes 10, 8 and 5-member ensembles using the two calibration techniques.

3.3.2.1 Individual Variable Calibration

Table 3-3 shows that both techniques (i.e., SA and GA) were able to find similar combinations of configurations (i.e., an ensemble that shares more than half of the members) when each variable was used separately. The final set of simulations chosen for each sub-ensemble varies significantly for the different variables, with the exception of the calibrated 10-member ensemble estimated using wind speed and wind direction. Although calibrated ensembles share less than 30% of the same configurations among the different variables, the majority of the ensembles include model configuration 14. This model configuration, as shown in (Chapter 2), presents large errors for both wind speed and wind direction. This implies that this model configuration contributes significantly to the spread of the three variables. The final scores of the calibrated ensembles for each variable show that finding a calibrated sub-ensemble that reaches a score of one is not possible for wind speed and wind direction. A sub-ensemble with a score less than or equal to one can be found for PBLH. Figure 3-9 shows the rank histograms of the different calibrated ensembles (i.e., 10, 8 and 5-member) for each meteorological variables shown in Table 3-3. The calibrated ensembles of PBLH (Figure 3-9 c, f, i) are nearly flat for all ensemble sizes, whereas the 10- and 8- member sub-ensembles keep a U-shape for wind speed and wind direction. The smallest score for wind speed and wind direction are obtained with a 5-member ensemble (Figure 3-9 g-h).

Table 3-3. Calibrated ensembles generated by SA and a GA and their rank histograms scores and bias for each variable.

N	Variable	Sub-Ensemble	δ	Bias
10	WSPD	[5 13 14 16 17 29 33 35 39 45]	3.8	0.4 m/s
	WDIR	[5 13 14 16 17 20 31 33 34 37]	3.4	-0.6 deg.
	PBLH	[2 11 14 23 27 31 35 37 43 44]	0.4	58 m
8	WSPD	[11 14 16 31 35 37 39 45]	3.7	0.5 m/s
	WDIR	[14 15 17 20 23 33 34 37]	3.9	-1 deg.
	PBLH	[12 13 14 23 26 28 37 44]	0.8	75.5 m
5	WSPD	[5 14 29 36 39]	3	0.4 m/s
	WDIR	[14 23 33 34 37]	1.9	0.3 deg.
	PBLH	[2 5 13 31 44]	0.1	69 m

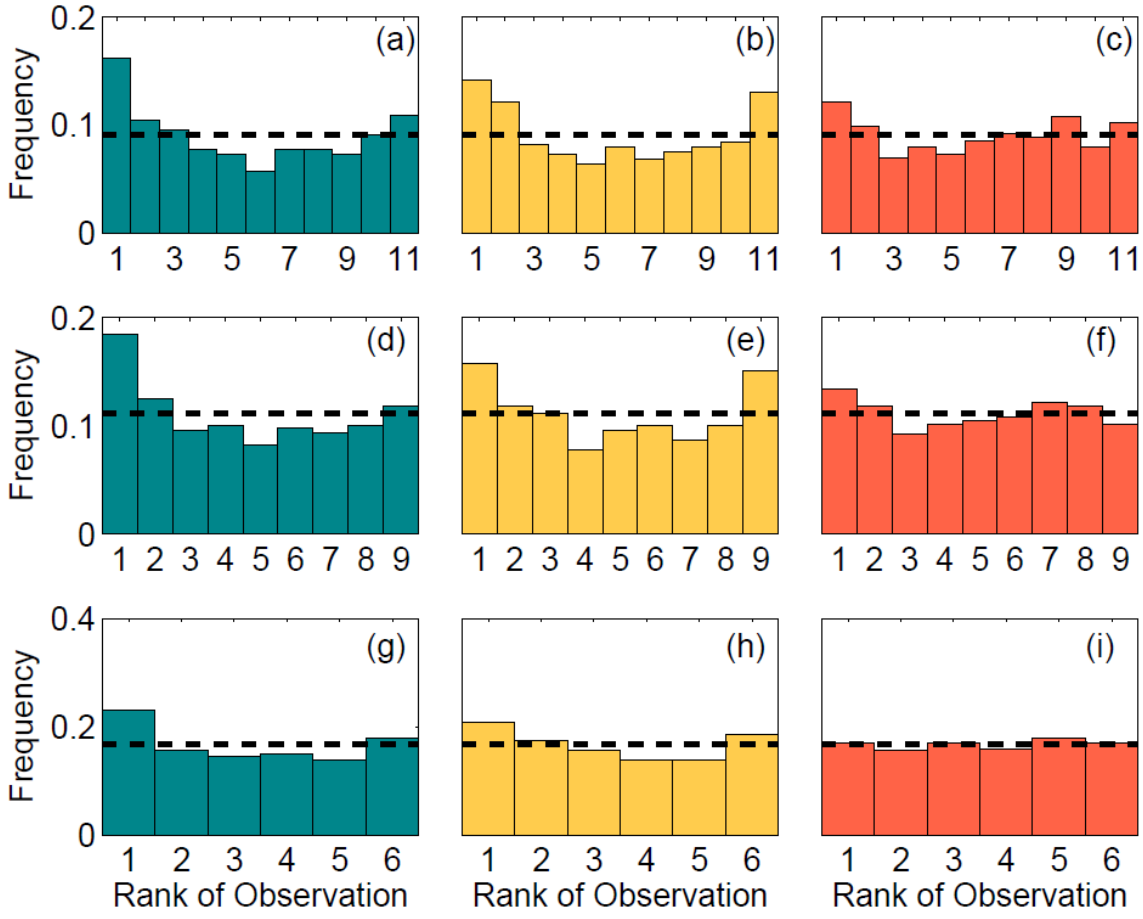

Figure 3-9. Rank histograms of the calibrated ensembles found for wind speed (a, d, g), wind direction (b, e, h) and PBL height (c, f, i) for each of the ensemble size. The upper, middle and lower panels correspond to the ensemble with 10, 8, and 5 members, respectively.

3.3.2.2 Multiple Variable Calibration

Table 3-4 shows the sub-ensembles selected by SA. Each of the sub-ensembles have two simulations in common (i.e., 17 and 33), implying that these models are crucial to build an ensemble that best represents the transport errors for the three variables. Figure 3-10 shows the rank histograms of the sub-ensembles shown in Table 3-4. These rank histograms show that we were able to flatten the histogram relative to the 45-member ensemble for each of the three meteorological variables. Similar to the individual variable calibration, the rank histogram for wind speed (Figure 3-10a, d) and wind direction (Figure 3-10b, e) still show a U-shape which is minimized for the smallest (i.e., 5-member) sub-ensemble (Figure 3-10g-h). The rank histograms are flatter for the PBLH (Figure 3-10c, f, i) and the score of each histogram is closer to one compared to wind speed and wind direction (Table 3-4). The rank histogram scores for all variables are greater than those for one-variable optimization (see Table 3-4), but are lower than the scores for the 45-member ensemble. In addition, all these calibrated sub-ensembles have biases smaller in magnitude than the 45-member ensemble. Both wind speed and PBLH retain an overall positive bias, and wind direction a negative bias. The standard deviations of these three calibrated ensembles are larger than those of the large ensemble, consistent with the effort to increase the ensemble spread.

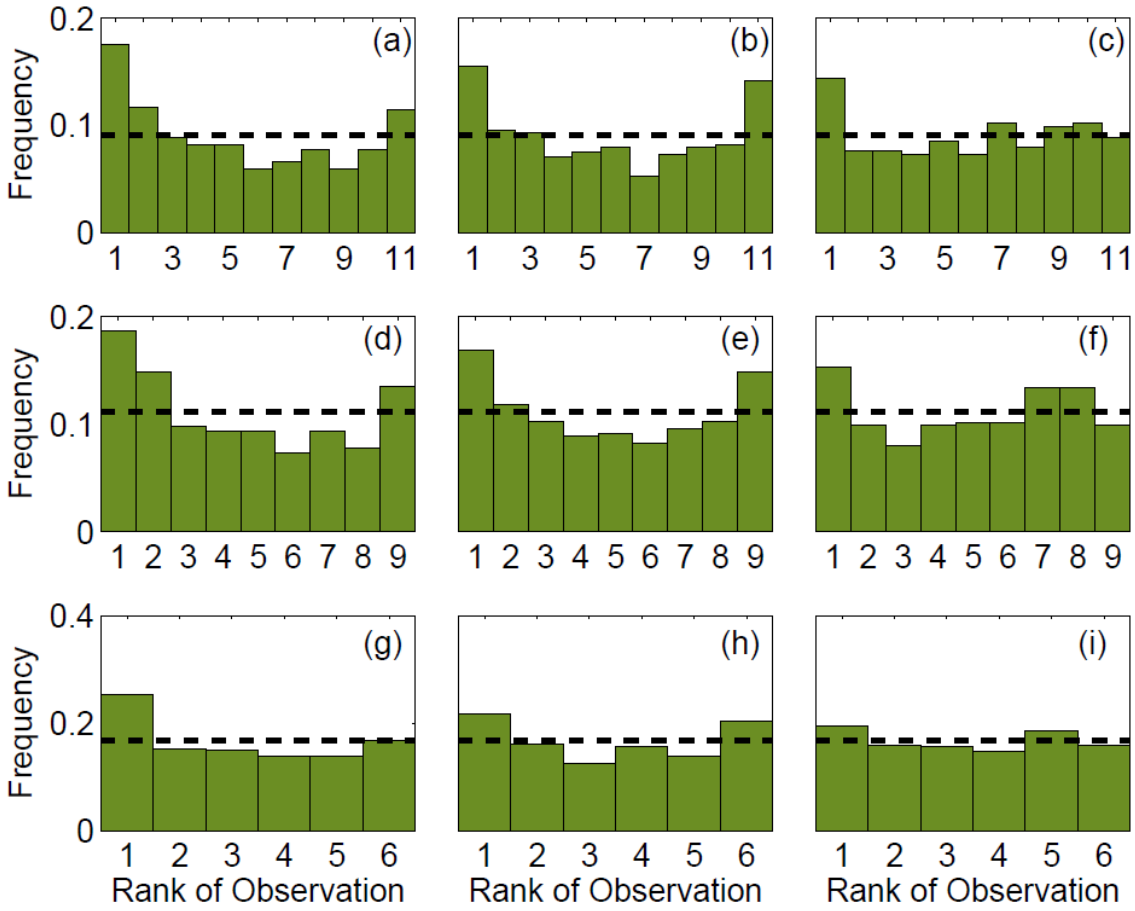

Figure 3-10. Rank histograms of wind speed (a, d, g), wind direction (b, e, h) and PBL height (c, f, i) using the calibrated ensembles found with simulated annealing. The upper, middle and lower panels correspond to the ensemble with 10, 8, and 5 members, respectively.

Table 3-4. Ensemble members, Rank histogram scores (δ), bias, and standard deviation (σ) for wind speed, wind direction and PBLH for the calibrated sub-ensembles generated with SA.

N	Sub-ensemble	Wind Speed			Wind Direction			PBL Height		
		δ	Bias m/s	σ m/s	δ	Bias Deg.	σ Deg.	δ	Bias m	σ m
10	[14 17 23 26 28 33 34 35 37 45]	5.5	0.6	3.6	4.6	-0.6	58	1.5	79.7	817.4
8	[5 6 14 17 26 33 34 37]	5.6	0.6	3.6	3.4	-0.7	58.5	1.6	71.8	823.4
5	[16 17 23 33 35]	5	0.5	3.6	3.4	-0.7	59	0.6	76.2	810.7

Table 35 shows the different simulations chosen by the GA to generate the different calibrated sub-ensembles of 10, 8 and 5-members. The different calibrated ensembles accepted by

the GA have two simulations in common (i.e., 17 and 33) and these two simulations are repeated in the SA calibrated ensembles. Figure 3-11 shows the rank histogram of the different sub-ensembles calibrated using the GA. Similar to SA, GA is capable of reducing the U-shape of the rank histogram. The GA also has difficulty generating a flat rank histogram for wind speed and wind direction, however it was capable of finding a sub-ensemble for both variables with a smaller rank histogram score compared to the large ensemble. Similar to the SA technique, the sub-ensembles generated for the PBLH have a flatter rank histogram and a smaller score compare to wind speed and wind direction. Compared to the SA the GA produces rank histogram with a higher score for the PBLH. Nevertheless, the GA was able to find calibrated sub-ensembles that fulfill the criteria of Section 3.2.3, a score smaller than six and an ensemble with a smaller bias than the large ensemble (see Table 3-2).

Table 3-5. Ensemble members, Rank histogram scores (δ), bias, and standard deviation (σ) for wind speed, wind direction and PBL height for the calibrated sub-ensembles generated with GA.

N	Sub-ensemble	Wind Speed			Wind Direction			PBL Height		
		δ	Bias m/s	σ m/s	δ	Bias Deg.	σ Deg.	δ	Bias m	σ m
10	[5 14 17 26 31 33 34 35 37 45]	5.6	0.6	3.6	4.1	-0.8	57.9	2.1	83.2	823.7
8	[5 14 15 17 33 34 37 38]	5.7	0.5	3.5	3.7	-0.4	58.1	2.5	99.3	828.3
5	[16 17 23 33 35]	5	0.5	3.6	3.4	-0.7	59	0.6	76.2	810.7

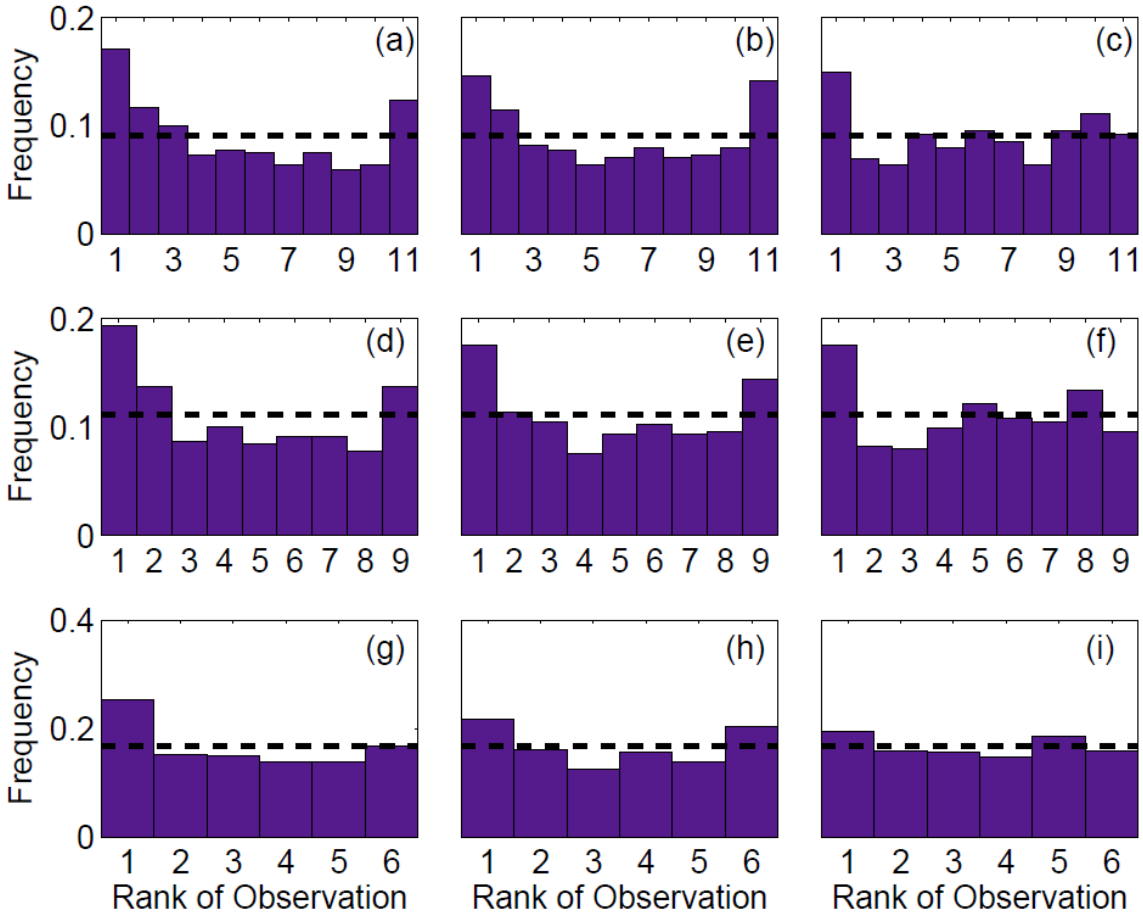

Figure 3-11. Rank histograms of wind speed (a, d, g), wind direction (b, e, h) and PBL height (c, f, i) using the calibrated ensembles found with genetic algorithm. The upper, middle and lower panels correspond to the ensemble with 10, 8, and 5 members, respectively.

Using SA and GA techniques and the selection criteria detailed in Section 3.2.6, we defined an optimal 5-member sub-ensemble (the optimal solution using both techniques) and nearly identical combinations of members for 10-and 8-member sub-ensembles, with only two model configurations not being shared by both algorithms. Both SA and GA techniques selected repeatedly two model configurations (i.e. 17 and 33) across the final sub-ensembles. However, statistical metrics (e.g. RMSE, ME) for these configurations are similar to other model configurations. When comparing the sub-ensembles from the multiple-variable calibration to the

single-variable calibrations, we found that both test generate sub-ensembles of 10- and 8-mebers that include simulation 14. This simulation shows larger errors in wind speed and wind direction compared to other configurations.

3.3.2.3 Evaluation of the Multiple Variable Calibrated Ensemble

Both optimization techniques were able to generate sub-ensembles that reduce the U-shape of the rank histograms while significantly reducing the number of members in the ensemble. The rank histograms tend to get flatter as we reduce the number of members, which indicates that the ensemble is more reliable and has a greater spread. However, as we mentioned in Section 3.2.3.1 a flat rank histogram does not always imply that the ensemble is reliable (Hamill, 2001). Therefore, we evaluate the spread-skill of the calibrated ensembles (Figure 3-12) as we did for the large ensemble (Figure 3-8). Figure 3-12 shows the spread-skill relationship of wind speed (Figure 3-12a), wind direction (Figure 3-12b) and PBLH (Figure 3-12c) for the 5-member ensemble generated by both SA and GA optimization techniques. This spread-skill relationship shows that the spread of the sub-ensemble generated by the optimization techniques is too small when the model errors are small, but too large when the model errors are large (i.e., linear coefficient is less than 1). The lack of correlation between ensemble spread and skill means that, at the daily time scale, ensembles do not represent correctly the transport errors. This result is a common deficiency in model ensembles which are primarily designed to represent the statistics of the model errors rather than the actual spatial and temporal variations of the errors at high frequency. The correlation for the wind direction increased while wind speed and PBLH remain similar. Therefore, we conclude that the calibrated sub-ensembles are equivalent or even better than the full ensemble to represent the daily model errors. This behavior is also characteristic for the 10- and 8-member sub-ensembles generated with SA and GA techniques.

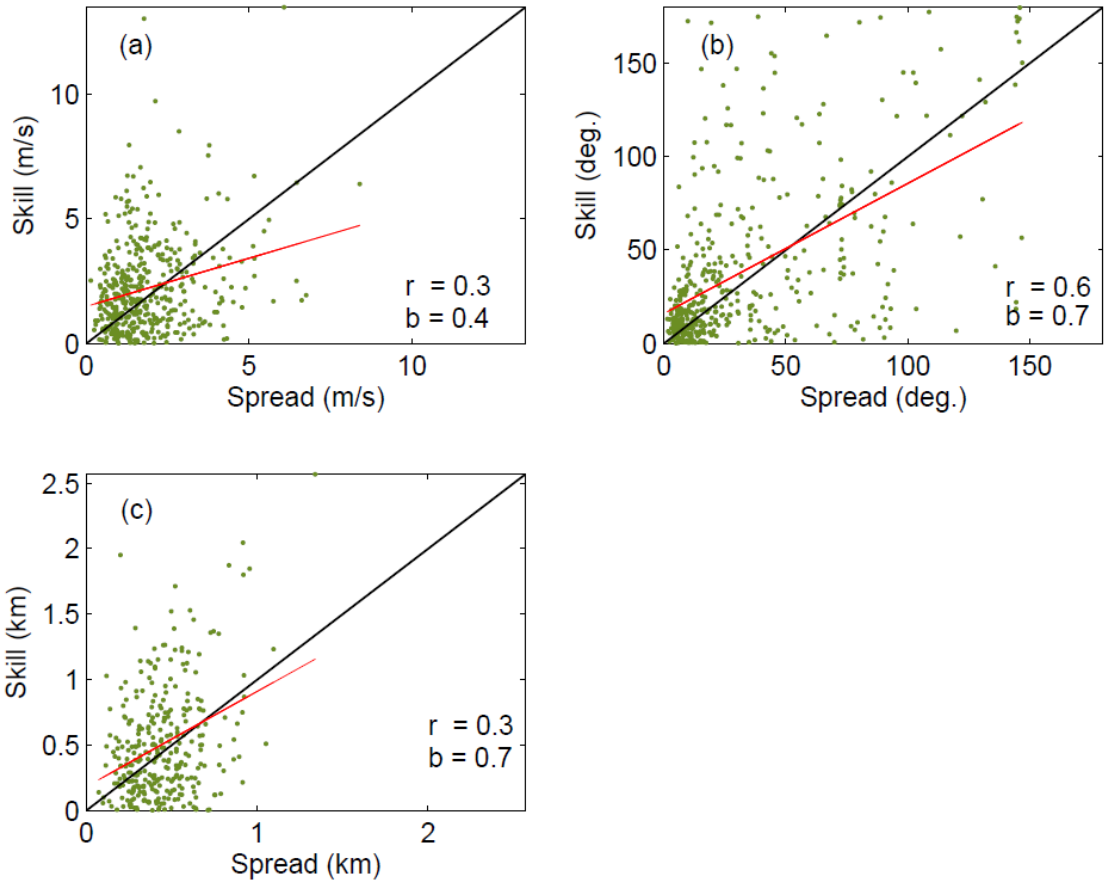

Figure 3-12. Spread-skill for (a) wind speed, (b) wind direction and (c) PBL height using the 14 rawinsonde sites available over the region using the 5-member calibrated ensemble. Each point represents the spread-skill of the different sites at different times. A one-to-one line is plotted in black and line of best fit is plotted in red. Correlation (r) and slope (b) of the line of best fit of the spread-skill relationship.

Figure 3-13 shows the time series of the different calibrated ensembles generated by the SA algorithm at TOP site. In general there are no major differences among 5- (Figure 3-13a,d,g), 8- (Figure 3-13,e,h) and 10-member (Figure 3-13c,f,i) ensembles. Some of the new sub-ensembles shows how the calibration can increase the spread of the ensemble to the extent of encompassing the observations (e.g., DOY 179 Figure 3-13 b-c) compared to the full ensemble (Figure 3-6b). Occasionally, the ensemble spread was reduced after calibration for a few specific points in space

and time. The reduced spread may be an improvement if model errors were actually smaller than the initial 45-member ensemble spread or a degradation of the spread if model errors of the initial ensemble were similar or larger, possibly due to the fewer ensemble members. Compared to the large ensemble, PBLH in the calibrated ensembles has a larger spread compared to the rest of the variables. Overall, ensemble spread in the calibrated ensembles is as large as the large ensemble for the three variables but using a smaller size ensemble.

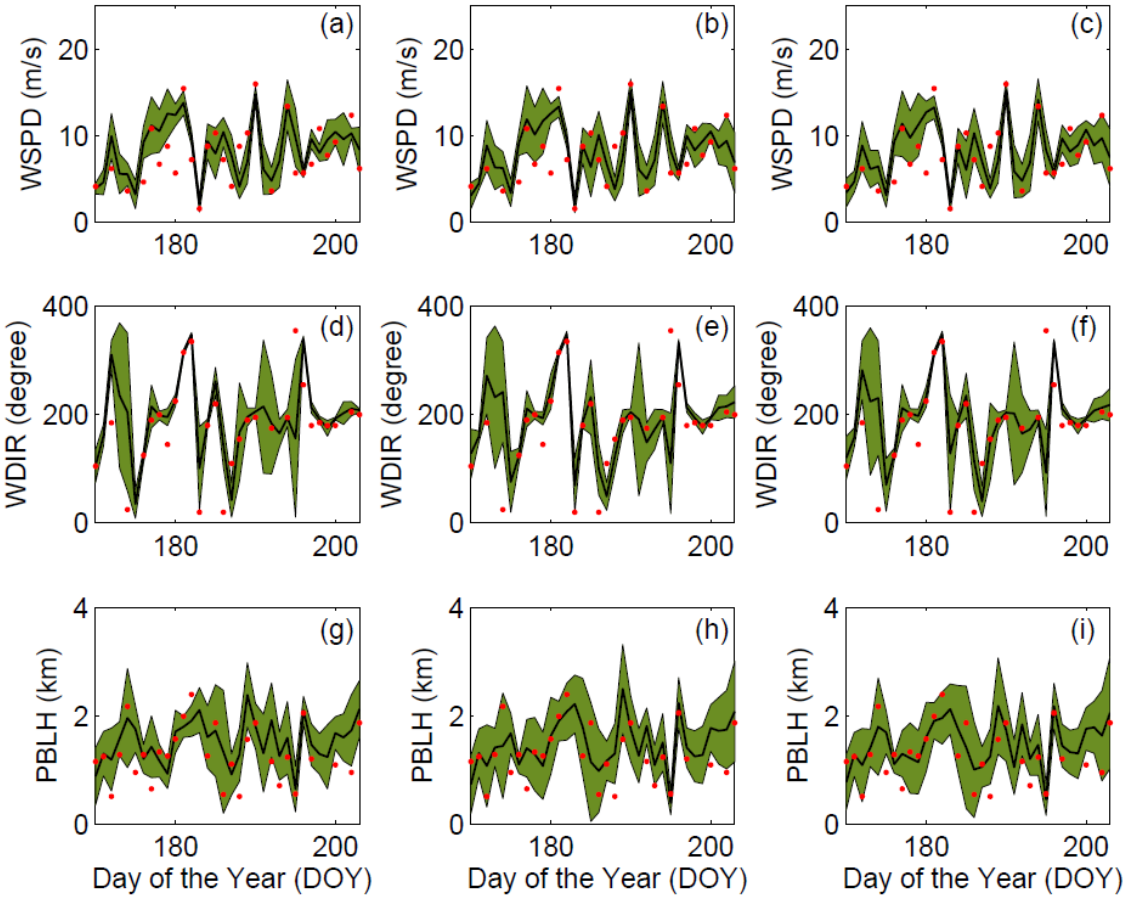

Figure 3-13. Time series of simulated wind speed (a-c), wind direction (d-f) and PBL height (g-i) using the 5-, 8- and 10-member calibrated ensembles at the TOP rawinsonde site. The green shaded area represents the spread (i.e., Root Mean Square Deviation) of the ensemble, the black line is the mean of the ensemble and the red dots are the observations at 0000UTC.

As explained in Section 3.3.2.2, both selection algorithms find similar calibrated sub-ensembles for 8- and 10-member ensembles, and identical configurations for 5-member sub-ensembles. The LSM, PBL, CP, and MP scheme, and reanalysis choice varies across all of the sub-ensemble members; no single parameterization is retained for all members in any of these categories. The selection of configurations is purely relative to the other configurations in the selected ensemble which limits the interpretation of any given scheme. We describe hereafter the statistics of the model configurations in the selection process. We find, however, that the calibrated ensembles rely upon certain physics parameterizations more than others. Figure 3-14 shows that most of the simulations in the calibrated ensemble use RUC and Thermal Diffusion (T-D) LSMs in preference to the Noah LSM. In addition, more simulations use MYJ PBL scheme than the other PBL schemes, and Kain-Fritsch as well as no Cumulus Parameterization more often than the Grell-3D CP. The physics parameterizations shown with a higher percentage in Figure 3-14 seem to contribute more to the spread of the ensemble than the other parameterizations.

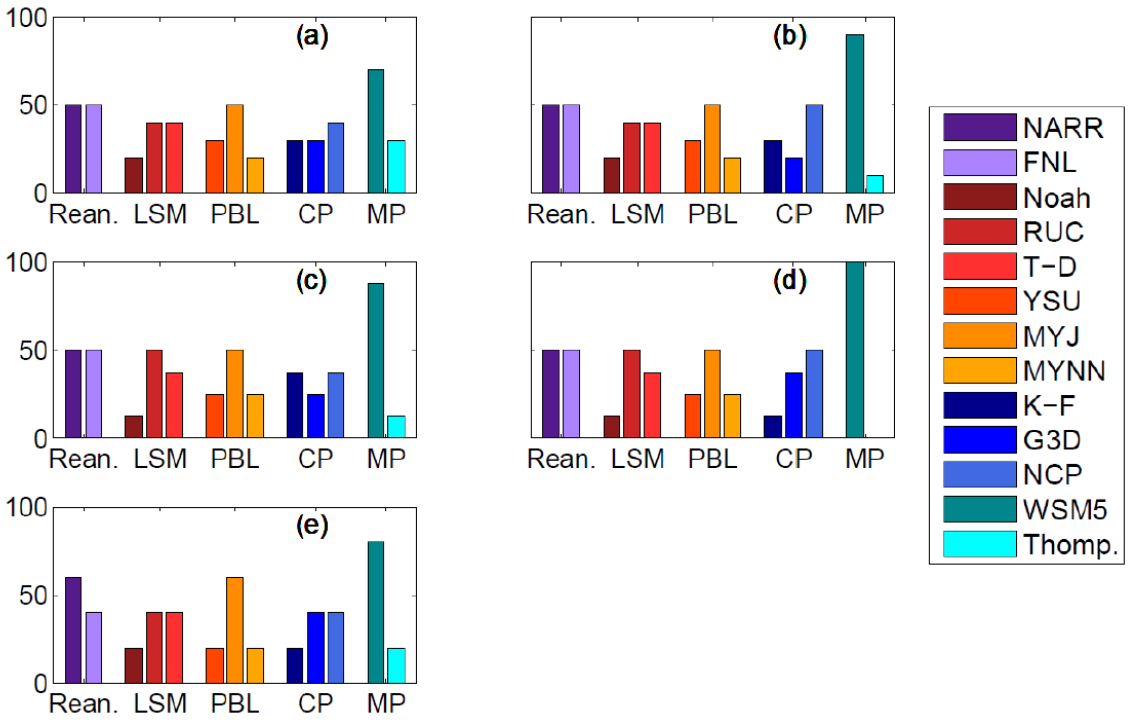

Figure 3-14. Frequency of the physics schemes used for each of the calibrated ensemble of 10 members (a-b), 8-members (c-d) and 5-members (e).

We next explore the characteristics of the individual ensemble members that are retained in an effort to understand what member characteristics are important to increase the spread of the ensemble. Figure 3-15 shows the mean and standard deviation of the residuals for each simulation included in the 5-member ensemble of SA and GA. Ensembles appear to need at least one member with a larger standard deviation to improve the spread for wind speed and wind directions (see member 23 from Figure 3-15a-b). Additionally, a member that has a large PBLH bias (see member 16 from Figure 3-15c) appears to be needed to generate a sub-ensemble that follows the criteria indicated in each of the optimization algorithms. It may be that the lack of an ensemble member with a persistently negative wind speed bias is a key factor in preventing us from finding a sub-ensemble with a flat rank histogram for wind speed, in contrast to PBLH. The sub-ensemble could

potentially be substantially improved by adding such a member. This behavior is repeated across the different calibrated ensemble sizes and with each of the techniques (not shown).

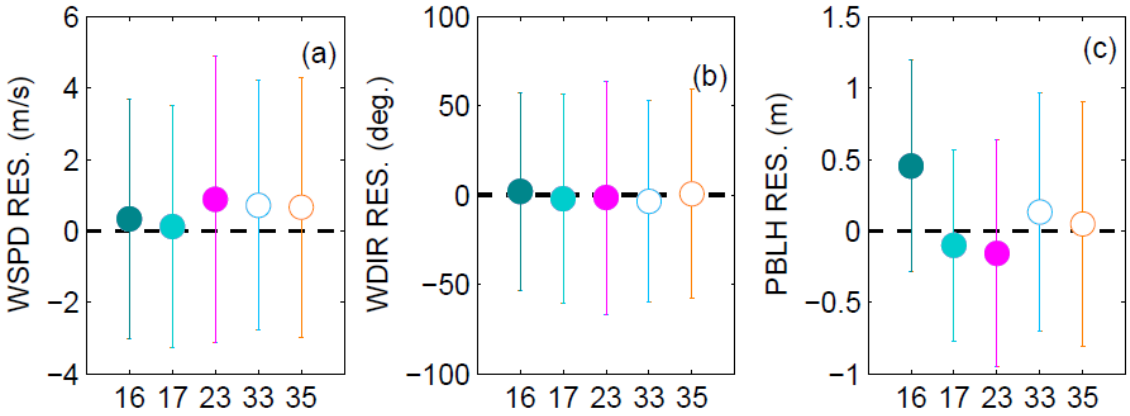

Figure 3-15. Bias of individual members for wind speed (a), wind direction (b), PBL height (c) using the SA and GA calibrated sub-ensemble of five members.

3.3.3 Propagation of Transport Errors into CO_2 Concentrations

The calibrated ensembles found in this study were chosen based on the meteorological variables and not on the CO_2 concentrations to avoid the propagation of CO_2 flux biases into the solution. We can now propagate these errors, represented by the ensemble spread, into the CO_2 concentration space. This straightforward calculation is possible because every model simulation included CO_2 concentration fields uses identical CO_2 fluxes. We present here the transport errors in both time and space as the spread in CO_2 concentrations comparing the initial 45-member un-calibrated ensemble to the calibrated sub-ensembles.

3.3.3.1 CO_2 Error Variances

Figure 3-16 shows the spread of daily daytime average CO_2 concentrations across the different sub-ensemble sizes at Mead (Figure 3-16a,d,g,j), West Branch (Figure 3-16b,e,h,k) and

WLEF (Figure 3-16c,f,i,l). The spread of the DDA CO_2 concentrations of the large ensemble (Figure 3-16a-c) does not appear to differ in a systematic fashion from the spread of the calibrated ensembles of various sizes (Figure 3-16 d-l). While the calibration can produce both an increase or a decrease of the ensemble spread at different locations and times, none of the ensembles consistently encompasses the observations. The remaining difference between the models and the observations for those days may be caused by flux errors, another component impacting the modeled CO_2 concentrations. However, the exact cause of the total difference cannot be determined from the CO_2 data alone. We simply note here that the model-data mismatch is not entirely due to the transport errors as represented by our sub-ensembles. Additionally, we compute the sum of the variances across the domain to evaluate the impact of the calibration on the total variance at different time steps. Figure 16 shows that the calibrated ensembles created using SA (Figure 16a) and GA (Figure 16b) have a larger CO_2 variance compare to the 45-member ensemble. The increase in total variance is maximum with 8-member ensembles which seems to be an optimal number of members for the calibration. We discuss further the number of members in the Discussion section.

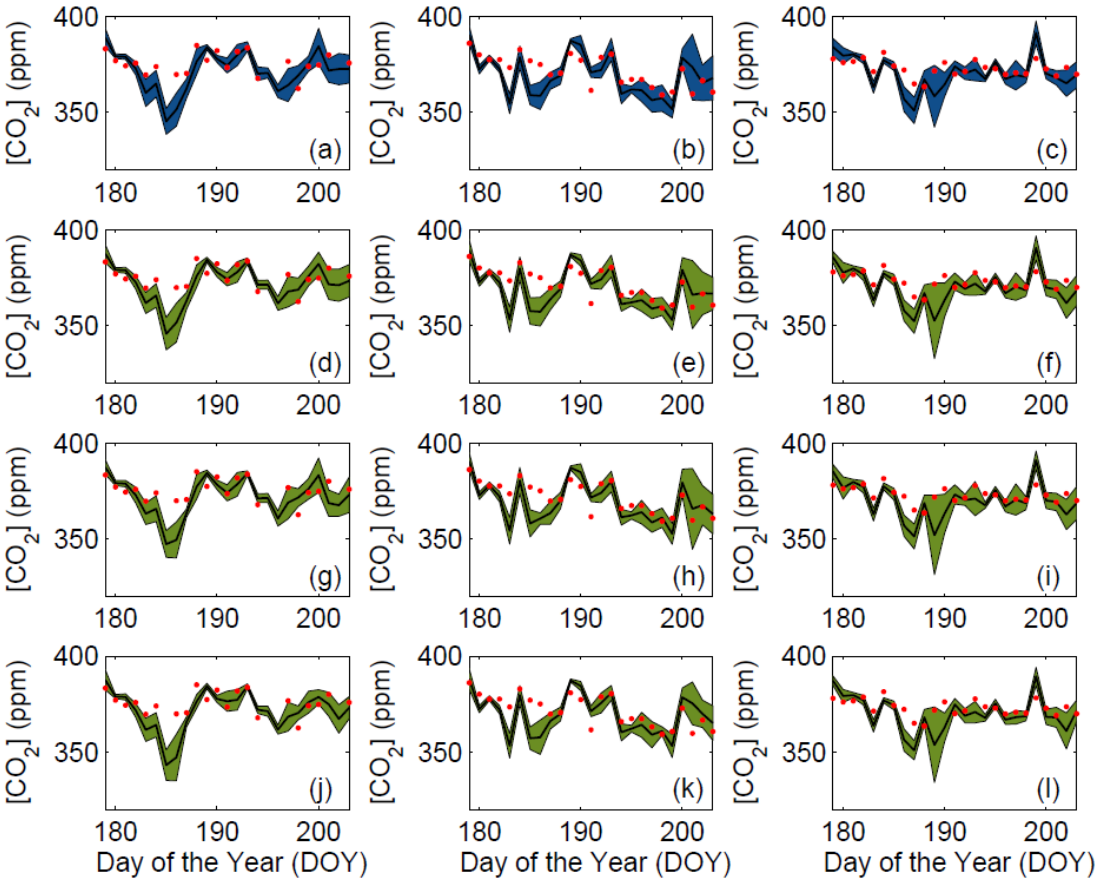

Figure 3-16. Spread (i.e., RMSD) of the DDA CO_2 concentrations of Mead (first column a,d,g,j), WBI (middle column b,e,h,k) and WLEF (last column c,f,i,l) using Simulated Annealing calibrated ensembles. Rows from top to bottom are 45, 10, 8 and 5 member ensembles. The blue area is the spread of the 45-member ensemble, green area is the spread is the spread of the calibrated (10-, 8- and 5-member) ensemble, the black line is the mean of the ensemble and the red dots are the observations.

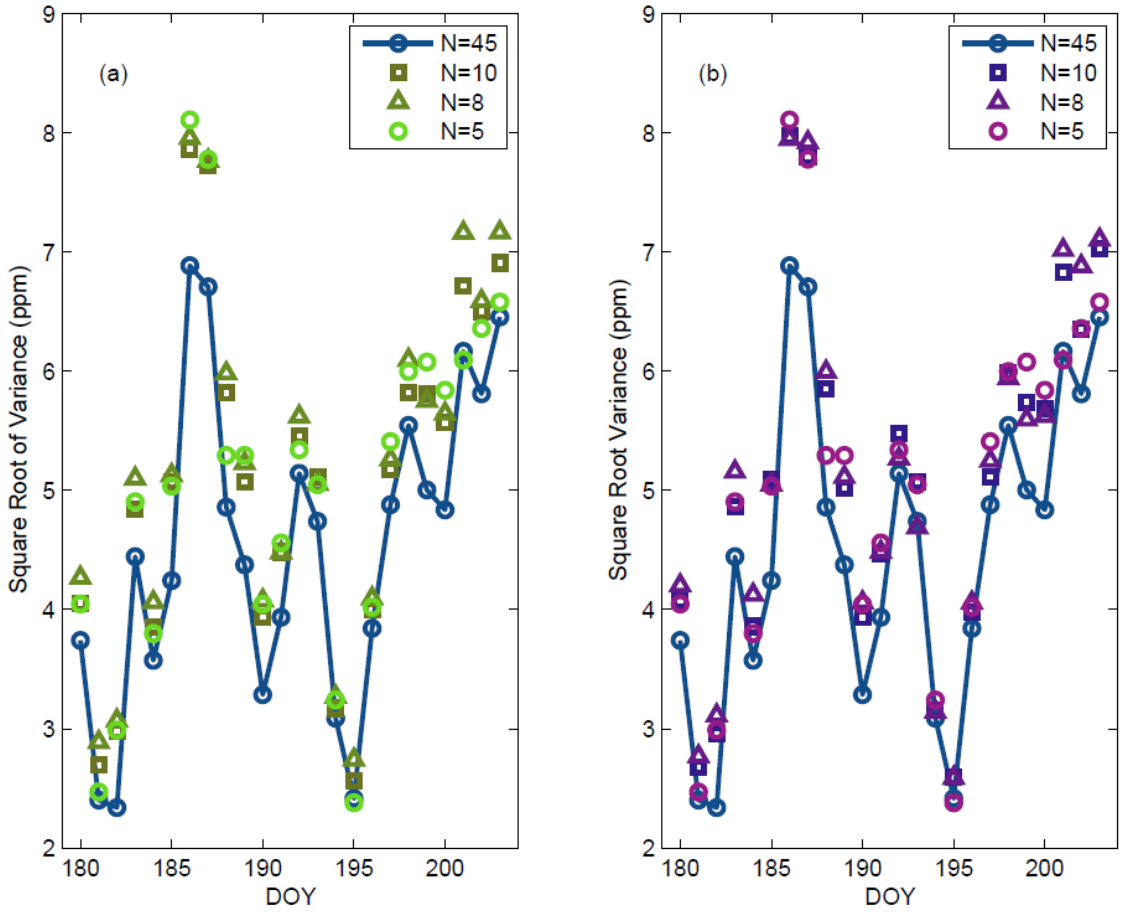

Figure 3-17. Sum of variance of the SA (a) and GA (b).

3.3.3.2 CO_2 Error Correlations

Figure 3-18 shows spatial correlation of the errors with respect to the Round Lake site on DOY 180. The structures in the errors are smoother using the large ensemble (Figure 3-18a) while smaller structures tend to appear as the number of members decreases (Figure 3-18 b,c,d). We suspect here that reduced-size ensembles are impacted by sampling noise which would require additional filtering (e.g., Ménétrier et al., 2015) or modeling (e.g., Lauvaux et al., 2009b). Concerning the magnitudes of the error correlation, the calibrated sub-ensembles exhibit a larger contrast in correlation values compared to the 45-member error correlations. Overall, the different

ensembles show similar flow-dependent spatial patterns which demonstrates that the calibration process, even if generating sampling noise, preserved the most important patterns in the error structures. Therefore, the calibrated ensemble is likely to provide a better representation of the variances and similar error structures for the construction of error covariance matrices in regional inversions.

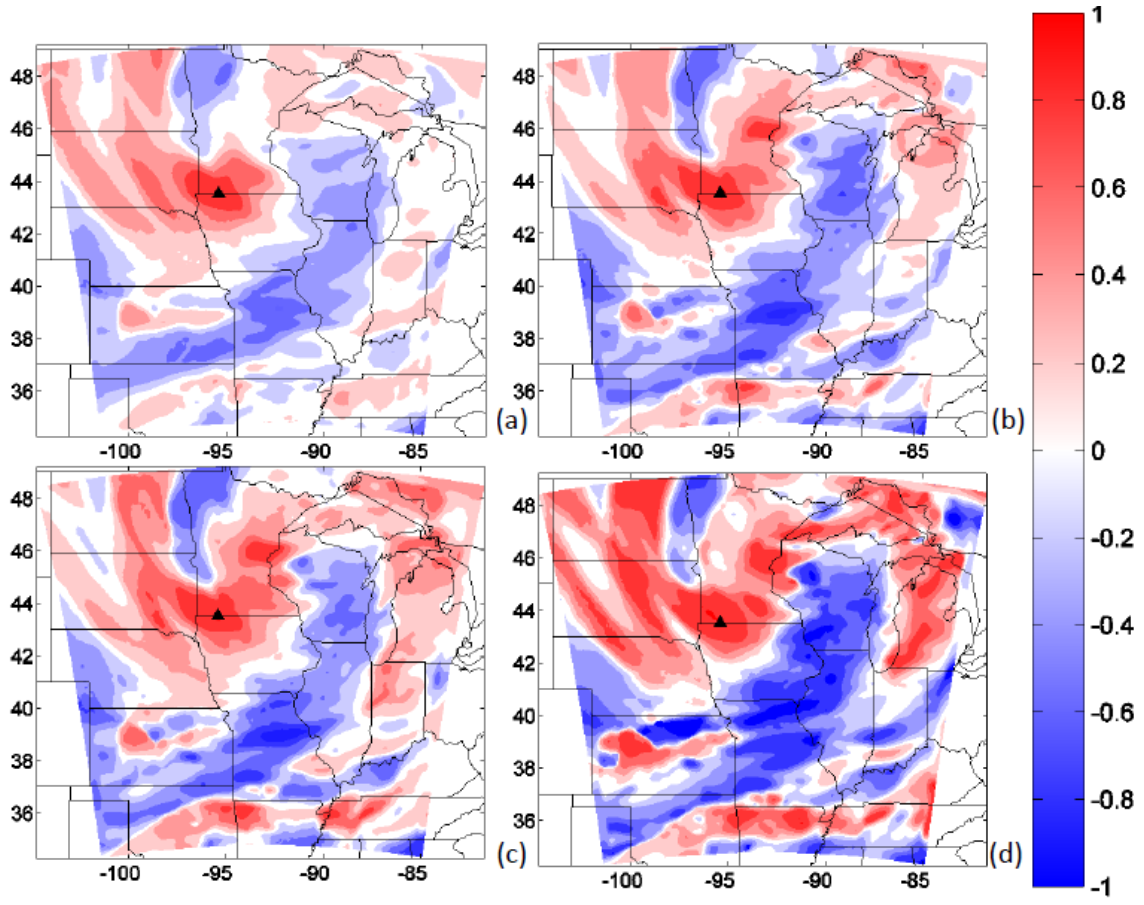

Figure 3-18. Spatial correlation of CO_2 across the 45- (a), 10-(b), 8-(c) and 5-members (d) ensembles with respect to the location of the Round Lake tower for DOY 180. This figure use calibrated ensembles of 10-, 8-, and 5-members found by SA technique.

3.4 DISCUSSION

3.4.1 Impact of Calibration on Ensemble Statistics

The calibration of the multi-physics/multi-analysis ensemble using SA and GA optimization techniques generated ensembles of 10-, 8- and 5- member with a better representation of the error statistics of the transport model than the initial 45-member ensemble. One of our goals here was to find a sub-ensemble that fulfill the criteria of Section 3.2.6, independent of the selection algorithm and for multiple meteorological variables. One can argue about the relative improvement after calibration remains relatively limited with potential remaining biases. For example, wind speed and wind direction only improve by a modest amount compared to the 45-member ensemble. Oppositely, PBLH shows a significant improvement in the calibrated ensembles. Overall, the variance increased relative to the 45-member ensemble but the improvement was limited by the spread in the initial ensemble. Stochastic perturbations (e.g. Berner et al., 2009) could increase the potential of the method if the initial ensemble offers a wider spread, which translates into additional combinations of model configurations better-suited to represent the model errors. Here, we limited the 45-member ensemble to mass-conserved, continuous flow (i.e., unpertubed) members that can be used in the regional inversion. Future work should address the problem of using an under-dispersive ensemble before the calibration of the ensemble.

3.4.2 Single-variable and Multiple-variable Ensembles

We first attempted to calibrate the ensemble for each meteorological variable (i.e., wind speed, wind direction and PBLH). Table 3-3 shows that the different sub-ensembles were able to follow the criteria presented on Section 3.2.6, but the calibration of the single-variable ensembles did not allow us to find a unique sub-ensemble that can be used to represent the errors of the three

variables. Therefore, the joint optimization of the three variables was required to identify an ensemble that best represents model errors across the three variables. By using the sum of the squared scores of the three variables (i.e., double-penalty with squared values), the selection algorithm found common solutions with higher scores than single-variable ensembles (cf. Table 3-4 and Table 3-5) but the results improved when considering the three variables together. We assumed that each variable was equally important to the problem, which could also be a limitation here. Future work on the relative importance of meteorological variables on CO_2 concentration errors would help weigh the scores in the selection algorithms.

3.4.3 Resolution and Reliability

The calibrated ensembles show the rank histogram score closer to one (Table 3-4 and Table 3-5), that is, flatter rank histograms (Figure 3-10 and Figure 3-11) compared to the 45-member ensemble (Table 3-2 and Figure 3-7). This implies that the sub-ensembles have a greater variance than the large ensemble (i.e., improved reliability). However, the spread-skill relationship (i.e., resolution) does not show any major improvement compared to the 45-member ensemble, implying that the spread of the ensemble does not represent the day-to-day transport errors. These results indicate that regardless of the application of a calibration technique the sub-ensembles are still incapable of representing the higher frequency of the transport errors (i.e., spatial and temporal patterns across the domain and summer time). The disagreement between the rank histogram and the spread-skill relationship suggests that using the score of the rank histogram alone may not be sufficient to measure the reliability of the ensemble (Hamill, 2001). Down-selection of ensembles has been implemented in other studies (e.g., Garaud and Mallet, 2011; Lee et al., 2016) but resolution is usually excluded from the calibration process. To represent daily model errors, additional metrics should be introduced and the initial ensemble should offer a sufficient spread,

possibly with additional physic parameterizations, additional random perturbations, or modifying the error distribution of the ensemble (Roulston and Smith, 2003).

3.4.4 Error Correlations

Rank histograms, as explained in Section 3.2.3, evaluate the ensemble by ranking observations in a relative sense, and therefore can be difficult to interpret. Although the calibrated ensembles show flatter rank histograms, the ensemble may be representing correctly the variances over the region but not the covariances (Hamill, 2001). In this study, the calibrated ensembles show an increase in the variance of CO_2 concentrations, but spatial structures of the errors (i.e., correlations) are impacted by sampling noise. Previous studies have suggested objective methods to filter the noise in small-size ensembles (i.e., Ménétrier et al., 2015) or modeling the error structures using the diffusion equation (e.g., Lauvaux et al., 2009b). Future work should address the impact of the calibration on the error structures as this information is critical in the observation error covariance to assess the inverse fluxes. Therefore, any misrepresentation of the errors could be directly detrimental to the improvement in the CO_2 variances.

3.5 CONCLUSION

We applied a calibration (or down-selection) process to a multi-physics/multi-analysis ensemble of 45 members. In this calibration process, two optimization techniques were used to extract a sub-set of members from the initial ensemble to optimally represent the transport model errors in CO_2 inversion modeling. We used purely meteorological criteria to calibrate the ensemble and avoid contaminating the calibration with CO_2 flux errors. The two optimization techniques were SA and GA. Both techniques control the acceptance of the calibrated ensemble using the

flatness of the rank histogram. We generated different calibrated ensembles for three meteorological variables; wind speed, wind direction and PBLH. With these techniques, we identified sub-ensembles by calibrating the three variables jointly. Both techniques show that calibrated small-size ensembles can reduce the score of the rank histogram flatness and therefore improve the representation of the model error variances with few members (between 5 and 10 members).

The calibration techniques improved the error representation of the ensembles (i.e., distribution of the model errors), except for the daily atmospheric transport errors as shown by the spread-skill relationship. We assessed how the calibrated ensemble errors propagate into the CO_2 concentrations simulated with identical CO_2 fluxes (i.e., independent of the atmospheric conditions). The spread from the new calibrated ensembles represented about 55% of the model-data CO_2 mismatches for summer 2008. These results suggest that additional errors in CO_2 fluxes and/or large-scale boundary conditions represent a large fraction of the differences between modeled and observed CO_2. Error correlations of the calibrated ensembles were compared to the large ensemble to identify any impact of the calibration. Compared to the initial error structures, the calibrated ensembles revealed sampling noise across the region which indicates that additional filtering or modeling of the errors would be required in order to construct the error covariance matrix for regional CO_2 inversion.

Chapter 4

CONCLUSION

In the present dissertation, I evaluated and quantified the atmospheric transport errors across a highly instrumented area, the Mid-Continental Intensive region of the Midwest U.S., for the period June 18 to July 21 of 2008. I isolated transport errors and propagated them into CO_2 mixing ratios using a multi-physics and multi-analysis ensemble of 45-members created with the Weather Research and Forecasting (WRF) mesoscale model. The transport differences of this ensemble come exclusively from the different physical parameterization (e.g., LSMs, PBL schemes, cumulus parameterizations and microphysics parameterizations) and different meteorological data sets. Each model configuration was coupled to the same CO_2 surface fluxes from CarbonTracker, which allowed me to generate simulated CO_2 mixing ratios. To evaluate the transport, I used a model-data comparison of three meteorological variables: wind speed, wind direction and PBL height. I assumed that these variables are the most important meteorological variables to understand the impact of atmospheric transport errors on in situ atmospheric CO_2 mixing ratios.

The first goal of this research was to identify which physics parameterizations beside the PBL schemes contributed the most to the errors in CO_2 mixing ratios. In Chapter 2, I found that all physics parameterizations except for microphysics have an impact on both CO_2 mixing ratios and meteorological variables. This analysis also showed that PBL height and CO_2 mixing ratios have similar sensitivities to the different physics schemes. The relationship between the two variables is reinforced by the high correlations found between PBL height errors and CO_2 mixing ratio errors. I note here that this result is supported by the physical relation between in situ mixing ratios and

mixing depths (i.e., volume of air and number of CO_2 molecules in the PBL). This result was expected based on previous studies, but this study was the first attempt to estimate more rigorously the relationship between the two, and compared it to other meteorological variables such as wind speed and wind direction.

Among the multiple configurations evaluated in this research, I intended to identify if any of them was best suited to represent the atmospheric transport over the region. Using a set of statistical metrics, I found that the performances of the model configurations vary widely across the radiosonde sites and for the different meteorological variables. However, some specific configurations were affected by systematic errors in PBL height across the different radiosonde sites. I performed a domain-average analysis to obtain more information from the different statistics. I found that the majority of the simulations showed a systematic positive (overestimation) bias for wind speed. This result seems fairly robust to the configurations and could be caused by a more general problem in the mesoscale model. I also noted that for the PBL height, systematic patterns were controlled by the choice of certain schemes. For example, the YSU PBL scheme generated high biases in PBL height while MYJ tends to under-estimate the PBL height. I found that the absence of cumulus parameterization tends to increase the PBL height, creating large positive biases in our analysis.

I evaluated the simulated CO_2 mixing ratios using the different model configurations and compared them to the observations. I am aware that model-data mismatch is due to both flux errors and transport errors, and avoided any direct comparison to absolute mixing ratios. However, daily variability and temporal correlations can be used to estimate the modeling performances of CO_2 mixing ratios. As a first result, I found that WRF can underestimate or overestimate the daily variability in CO_2 concentrations for the majority of sites except three of them affected by a systematic overestimation. In general, the model correlation with the observation range between 0.8 and 0.1 and only Round Lake showed all the models with correlations down to 0.5. In general,

I did not find any patterns in any model for the different statistics presented in Chapter 2. Additionally, I evaluated if any spatial correlation existed among the model meteorological variables errors and the atmospheric CO_2 mixing ratio errors. Correlations between errors were not related to the distance, that is, transport errors estimated from a rawinsonde near one of the CO_2 towers had similar correlations with CO_2 errors at the tower than any other locations across the domain. I conclude that the local component of the transport errors is not the primary driver of CO_2 errors, while flux errors and regional transport errors are likely to contribute significantly to the total CO_2 mixing ratio errors. Although I had a general idea of how the different members impacted the atmospheric CO_2 mixing ratios, the ensemble constituted by the different model configurations needed to be calibrated to represent the atmospheric transport errors.

For the second goal of this work, I addressed two problems related to transport errors: i) I decreased the number of simulations to reduce significantly the computational cost of our ensemble-based error assessment, and ii) I calibrated the ensemble to represent the actual model-data mismatches in WRF. Prior to the calibration, I evaluated if the 45-member ensemble (or a calibrated sub-ensemble extracted from the 45 simulations) would be sufficient to represent the atmospheric transport errors over the region. This verification was crucial because the calibration process can only improve the initial ensemble to a certain extent. I used wind speed, wind direction and PBL height to evaluate the performance of the ensemble and how it represents the atmospheric transport errors. I found that the ensemble was under-dispersive for the three meteorological variables. To evaluate further the initial ensemble, I considered the imbalance between low and high biases (i.e., the first and last bins of the rank histograms) and the distribution of errors (i.e., the bins in-between). The ratio was about 5-6 which indicates that the low and high biases could be significantly reduced (if not removed) using a calibration algorithm. Similarly, the spread-skill relationship did not show a 1:1 ratio between the ensemble spread and ensemble mean error. By reducing the number of model configurations (and therefore increase the variance), I was confident

that this ensemble would improve significantly with a calibration that can optimize the variance of the ensemble.

I used two optimization techniques to calibrate the ensemble: Simulated Annealing (SA) and Genetic Algorithm (GA). The calibration technique has two purposes: to improve the variance of the ensemble, and down-select the number of members to generate an ensemble that can represent errors similar to a large ensemble. Using the flatness of the rank histogram as a criterion caused some limitations because the selection became highly dependent on the choice of meteorological variables. Nevertheless, both techniques were able to generate calibrated sub-ensembles of 10-, 8- and 5-members that improved both the shape and score of the rank histogram for the three meteorological variables. The spread-skill relationship of these ensembles was not improved significantly for any of the variables, which indicated that despite the better statistics of the ensembles (i.e., better distributions), transport errors were not represented correctly on a daily basis. The calibration process improved the spread of the ensemble and the final ensembles were better able to reproduce transport errors with a smaller number of members.

After performing the calibration of the ensemble, I evaluated the impact of transport errors, now represented by our calibrated sub-ensembles, on CO_2 mixing ratios. I found that the spread of the calibrated ensembles represented about 50% of the total CO_2 model-data differences. This result confirmed that transport errors represent a significant fraction of the atmospheric signals, but also that the remaining mismatch associated with flux errors and background conditions is large enough to be optimized in future regional CO_2 inversions. I note here that atmospheric inversions use the observation error covariance matrix to represent the transport errors within the assimilation system. However, the calibration of the ensemble is mostly focusing on the ensemble variance. Thus, its impact on the covariance needs to be further studied. I estimated the final CO_2 error variances and covariances over the region, and showed that variances increased significantly with the calibration to better represent transport errors, while the error correlations indirectly representing the error

covariances, were amplified in magnitude, which is likely due to sampling noise in the calibrated reduced-size ensembles.

The research performed in this dissertation afforded me an insight into the magnitudes of the atmospheric transport errors that can be encountered over this region in summertime. Additionally, I showed how these transport errors can contribute to the atmospheric CO_2 mixing ratio errors and more specifically to the model-data mismatches. I was also able to use a multi-physics and multi-analysis ensemble to represent the atmospheric transport errors by calibrating a small-size ensemble. Nevertheless, some challenges and limitations still exist in the study. The ensemble used in this study shows all the members with a positive bias regardless of the model configuration. For future atmospheric inversions, these systematic biases need to be corrected or they will be propagated into the inverse fluxes. This ensemble of 45-members is still slightly under-dispersive regardless of the different physics parameterization used, therefore one would need to think about other methods to improve the spread of the ensemble. The calibration method used in this study is also highly dependent of the meteorological variables. Therefore, to improve the calibration system, each of the variables should be weighted to select a calibrated ensemble that represent the transport errors for the most important variables.

I propose here some directions for future work based on the different conclusions of this dissertation.

(a)Initial ensemble:

In this research, the ensemble was tested for a limited period of time, over a small region and with a limited amount of members. Therefore, new ensembles should be run over a longer period of time and a larger region. This will enable one to understand the errors during different seasons over the year and how these different models behave over other regions besides the U.S. Midwest. Additionally, the ensemble should expand the number of members by adding other parameterization schemes, including other models, or by adding stochastic perturbations. Ideally,

the addition of these components to the ensemble can potentially improve the spread of the initial ensemble and help obtain a better representation of the spread and the skills of the calibrated ensemble. Although the initial ensemble should be increased to generate more spread, a calibration should be applied to better represent the transport errors, and possibly with a smaller number of members.

(b) Calibration criteria:

The calibration process was highly constrained by the meteorological variables, thus weighting the different meteorological variables will help the system to optimize the ensemble based on the exact level of impact the meteorological variable has on CO_2 mixing ratios. Here, I simply used an equally weighted approach, assuming that wind speed, wind direction, and PBL height all contribute in a similar way to the transport model errors. The criterion used for the acceptance of the ensemble depends in the flatness of the rank histogram. I show that this does not necessarily represent a better skill in general. Other statistics that can measure the reliability of the ensemble such as Reliability Diagram or Discrete Rank Probability Score should be used in the future to permit better evaluation of the ensemble performance. To use these new statistics a threshold that can affect the CO_2 errors should be defined first to make these metrics useful. I did not use these statistics because currently these thresholds are unknown for this problem. Additionally, future calibration should include a criterion that can control the covariance of the ensemble; the current calibration only evaluates the variance. Therefore, the changes applied to the covariances remain unknown because this technique change only the variance of the ensemble.

(c) Covariances:

To measure the impact of these uncertainties on the inverse CO_2 fluxes, an error covariance matrix needs to be generated but the limited amount of members means that the covariance matrix becomes noisier. Therefore, a filtering technique should be applied to remove the noise from this matrix and then apply it in an atmospheric inversion. Future work should address this important

limitation in future inversions to fully utilize the information provided by the calibrated ensembles. Finally, it would be ideal to use the different model simulations from the calibrated ensemble and the filtered covariance matrix to perform multiple atmospheric inversions and quantify the impact of the atmospheric transport model errors on CO_2 fluxes. This approach will not only include a robust description of random transport errors but also propagate biases from individual ensemble members, allowing for the first time to characterize the impact of transport biases on the inverse CO_2 fluxes.

Bibliography

Ahmadov, R., C. Gerbig, R. Kretschmer, S. Körner, C. Rödenbeck, P. Bousquet, and M. Ramonet (2009), Comparing high resolution WRF-VPRM simulations and two global CO 2 transport models with coastal tower measurements of CO2, *Biogeosciences*, *6*(5), 807–817, doi:10.5194/bg-6-807-2009.

Ahmadov, R., C. Gerbig, R. Kretschmer, S. Koerner, B. Neininger, A. J. Dolman, and C. Sarrat (2007), Mesoscale covariance of transport and CO2 fluxes: Evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model, *J. Geophys. Res. Atmos.*, *112*(22), doi:10.1029/2007JD008552.

Alhamed, A., S. Lakshmivarahan, and D. Stensrud (2001), Cluster Analysis of Multimodel Ensemble Data from SAMEX, *Mon. Weather Rev.*, *130*, 226–256.

Anderson, J. L. (1996), A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *J. Clim.*, *9*(7), 1518–1530, doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.

Andrews, a E. et al. (2013), CO2, CO and CH4 measurements from the NOAA Earth System Research Laboratory's Tall Tower Greenhouse Gas Observing Network: instrumentation, uncertainty analysis and recommendations for future high-accuracy greenhouse gas, *Atmos. Meas. Tech. Discuss.*, *6*, 1461–1553, doi:10.5194/amtd-6-1461-2013.

Baker, D. F. et al. (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988-2003, *Global Biogeochem. Cycles*, *20*(1), doi:10.1029/2004GB002439.

Bakwin, P. S., P. P. Tans, D. F. Hurst, and C. Zhao (1998), Measurements of carbon dioxide on very tall towers: Results of the NOAA/CMDL program, *Tellus, Ser. B Chem. Phys. Meteorol.*, *50*(5), 401–415, doi:10.1034/j.1600-0889.1998.t01-4-00001.x.

Battle, M., M. L. Bender, P. P. Tans, J. W. C. White, J. T. Ellis, T. Conway, and R. J. Francey (2000), Global carbon sinks and their variability inferred from atmospheric O-2 and delta C-13, *Science (80-.).*, *287*(5462), 2467–2470, doi:DOI 10.1126/science.287.5462.2467.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer (2009), A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System, *J. Atmos. Sci.*, *66*(3), 603–626, doi:10.1175/2008JAS2677.1.

Bocquet, M. (2005), Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory, *Q. J. R. Meteorol. Soc.*, *131,*(No. 610), 2191–2208, doi:doi: 10.1256/qj.04.67.

Boden, T. a, G. Marland, and R. J. Andres (2009), Global, Regional, and National Fossil-Fuel CO2 Emissions, *Carbon Dioxide Inf. Anal. Cent. Oak Ridge Natl. Lab. USA Oak Ridge TN Dep. Energy*, U.S. Department of Energy, Oak Ridge, Tennessee, doi:10.3334/CDIAC/00001.

Bousquet, P. (2000), Regional Changes in Carbon Dioxide Fluxes of Land and Oceans Since 1980, *Science (80-.).*, *290*(5495), 1342–1346, doi:10.1126/science.290.5495.1342.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei (2005), A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, *Mon. Weather Rev.*, *133*(5), 1076–1097, doi:10.1175/MWR2905.1.

Butler, M. P., K. J. Davis, A. S. Denning, and S. R. Kawa (2010), Using continental observations in global atmospheric inversions of CO2: North American carbon sources and sinks, *Tellus, Ser. B Chem. Phys. Meteorol.*, *62*(5), 550–572, doi:10.1111/j.1600-0889.2010.00501.x.

Canadell, J. G., C. Le Quere, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, N. P. Gillett, R. A. Houghton, and G. Marland (2007), Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks, Proc. Natl. Acad. Sci. U.S.A., 104(47), 18,866–18,870, doi:10.1073/pnas.0702737104.

Candille, G., and O. Talagrand (2005), Evaluation of probabilistic prediction systems for a scalar variable, *Q. J. R. Meteorol. Soc.*, *131*(609), 2131–2150, doi:10.1256/qj.04.71.

Černý, V. (1985), Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, *J. Optim. Theory Appl.*, *45*(1), 41–51, doi:10.1007/BF00940812.

Chen, F., and J. Dudhia (2001), Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, *Mon. Weather Rev.*, *129*(4), 569–585, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Cheng, W. Y. Y., and W. J. Steenburgh (2005), Evaluation of Surface Sensible Weather Forecasts by the WRF and the Eta Models over the Western United States, *Weather Forecast.*, *20*(5), 812–821, doi:10.1175/WAF885.1.

Chevallier, F. et al. (2010), CO2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res. Atmos.*, *115*(21), doi:10.1029/2010JD013887.

Ciais, P., P. P. Tans, M. Trolier, J. W. C. White, and R. J. Francey (1995), A Large Northern Hemisphere Terrestrial CO2 Sink Indicated by the 13C/12C Ratio of Atmospheric CO2, *Science (80-.).*, *269*(5227), 1098–1102, doi:10.1126/science.269.5227.1098.

Ciais, P. et al. (2013), The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change, *Chang. IPCC Clim.*, 465–570, doi:10.1017/CBO9781107415324.015.

Coniglio, M. C., J. Correia, P. T. Marsh, and F. Kong (2013), Verification of Convection-Allowing WRF Model Forecasts of the Planetary Boundary Layer Using Sounding Observations, *Weather Forecast.*, *28*, 842–862, doi:10.1175/WAF-D-12-00103.1.

Corbin, K. D., A. S. Denning, E. Y. Lokupitiya, A. E. Schuh, N. L. Miles, K. J. Davis, S. Richardson, and I. T. Baker (2010), Assessing the impact of crops on regional CO2 fluxes and atmospheric concentrations, *Tellus, Ser. B Chem. Phys. Meteorol.*, *62*(5), 521–532, doi:10.1111/j.1600-0889.2010.00485.x.

Crosby, J. L. (1973), *Computer Simulation in Genetics*, John Wiley, Hoboken, N. J.

Davis, K. J., P. S. Bakwin, C. Yi, B. W. Berger, C. Zhao, R. M. Teclaw, and J. G. Isebrands (2003), The annual cycles of CO2 and H2O exchange over a northern mixed forest as observed from a very tall tower, *Glob. Chang. Biol.*, *9*(9), 1278–1293, doi:10.1046/j.1365-2486.2003.00672.x.

Denning, a. S., I. Y. Fung, and D. Randall (1995), Latitudinal gradient of atmospheric CO2 due to seasonal exchange with land biota, *Nature*, *376*(6537), 240–243, doi:10.1038/376240a0.

Díaz Isaac, L. I., T. Lauvaux, K. J. Davis, N. L. Miles, S. J. Richardson, A. R. Jacobson, and A. E. Andrews (2014), Model-data comparison of MCI field campaign atmospheric CO2 mole fractions, *J. Geophys. Res. Atmos.*, *119*(17), 10536–10551, doi:10.1002/2014JD021593.

Dlugokencky, E., and P. Tans (2015), *Trends in Atmospheric Carbon Dioxide*.

Dolman, A. J. et al. (2006), The CarboEurope regional experiment strategy, *Bull. Am. Meteorol. Soc.*, *87*(10), 1367–1379, doi:10.1175/BAMS-87-10-1367.

Dudhia, J. (1996), A multi-layer soil temperature model for MM5, *Prepr. from Sixth PSU/NCAR Mesoscale Model Users' Work. 22-24 July 1996, Boulder, Color.*, 49–50.

Dudhia, J. et al. (2000), A multi-layer soil temperature model for MM5, *Mon. Weather Rev.*, *134*(1), 4077, doi:10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Enting, I. G. (1993), Inverse problems in atmospheric constituent studies: III. Estimating errors in surface sources, *Inverse Probl.*, *9*(6), 349–362.

Evensen, G. (1994), Inverse Methods and Data Assimilation in Nonlinear Ocean Models, *Phys. D Nonlinear Phenom.*, *77*(1), 108–129, doi:10.1016/0167-2789(94)90130-9.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10143, doi:10.1029/94JC00572.

Fang, Y., A. M. Michalak, Y. P. Shiga, and V. Yadav (2014), Using atmospheric observations to evaluate the spatiotemporal variability of CO2 fluxes simulated by terrestrial biospheric models, *Biogeosciences*, *11*(23), 6985–6997, doi:10.5194/bg-11-6985-2014.

Feng, S. et al. (2016), Los Angeles megacity: A high-resolution land-atmosphere modelling system for urban CO2 emissions, *Atmos. Chem. Phys.*, *16*(14), 9019–9045, doi:10.5194/acp-16-9019-2016.

Fraser, A., and D. Burnell (1970), *COMPUTER MODELS IN GENETICS*.

Fung, I., K. Prentice, E. Matthews, J. Lerner, and G. Russell (1983), Three-dimensional tracer model study of atmospheric CO 2 : Response to seasonal exchanges with the terrestrial biosphere, *J. Geophys. Res.*, *88*(C2), 1281, doi:10.1029/JC088iC02p01281.

Garaud, D., and V. Mallet (2011), Automatic calibration of an ensemble for uncertainty estimation and probabilistic forecast: Application to air quality, *J. Geophys. Res. Atmos.*, *116*(19), doi:10.1029/2011JD015780.

Geels, C. et al. (2007), Comparing atmospheric transport models for future regional inversions over Europe—Part 1: Mapping the atmospheric CO_2 signals, *Atmos. Chem. Phys.*, *7*, 3461–3479, doi:10.5194/acp-7-3461-2007.

Gerbig, C., S. Körner, and J. C. Lin (2008), Vertical mixing in atmospheric tracer transport models: error characterization and propagation, *Atmos. Chem. Phys. Atmos. Chem. Phys.*, *8*, 591–602, doi:10.5194/acp-8-591-2008.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, a. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. a. Grainger (2003), Toward constraining regional-scale fluxes of CO 2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res. Atmos.*, *108*(D24), n/a-n/a, doi:10.1029/2002JD003018.

Giglio, L., I. Csiszar, and C. O. Justice (2006), Global distribution and seasonality of active fires as observed with the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, *J. Geophys. Res. Biogeosciences*, *111*(2), doi:10.1029/2005JG000142.

Gloor, M., N. Gruber, J. Sarmiento, C. L. Sabine, R. A. Feely, and C. Rödenbeck (2003), A first estimate of present and preindustrial air-sea CO2 flux patterns based on ocean interior carbon measurements and models, *Geophys. Res. Lett.*, *30*(1), 1010, doi:10.1029/2002GL015594.

Göckede, M., A. M. Michalak, D. Vickers, D. P. Turner, and B. E. Law (2010), Atmospheric inverse modeling to constrain regional-scale CO2 budgets at high spatial and temporal resolution, *J. Geophys. Res.*, *115*(D15), D15113, doi:10.1029/2009JD012257.

Gourdji, S. M., A. I. Hirsch, K. L. Mueller, V. Yadav, A. E. Andrews, and A. M. Michalak (2010), Regional-scale geostatistical inverse modeling of North American CO_2 fluxes: a synthetic data study, *Atmos. Chem. Phys.*, *10*(13), 6151–6167, doi:10.5194/acp-10-6151-2010.

Grell, G. a. (2002), A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geophys. Res. Lett.*, *29*(14), 10–13, doi:10.1029/2002GL015311.

Grell, G. a. (1993), Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations, *Mon. Weather Rev.*, *121*(3), 764–787, doi:10.1175/1520-0493(1993)121<0764:PEOAUB>2.0.CO;2.

Gruber, N., J. L. Sarmiento, and T. F. Stocker (1996), An Improved Method for Detecting Anthropogenic CO 2 in the Oceans, *Global Biogeochem. Cycles*, *10*(4), 809–837, doi:10.1029/96GB01608.

Gurney, K. R., D. Baker, P. Rayner, and S. Denning (2008), Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO2 inversions for the period 1980 to 2005, *Global Biogeochem. Cycles*, *22*(3), doi:10.1029/2007GB003082.

Gurney, K. R. et al. (2002), Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models., *Nature*, *415*(February), 626–630, doi:10.1038/415626a.

Hamill, T. M. (2001), Interpretation of Rank Histograms for Verifying Ensemble Forecasts, *Mon. Weather Rev.*, *129*(3), 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Hamill, T. M., and S. J. Colucci (1997), Verification of Eta–RSM Short-Range Ensemble Forecasts, *Mon. Weather Rev.*, *125*(6), 1312–1327, doi:10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2.

Hilton, T. W., K. J. Davis, K. Keller, and N. M. Urban (2013), Improving North American terrestrial CO2 flux diagnosis using spatial structure in land surface model residuals, *Biogeosciences*, *10*(7), 4607–4625, doi:10.5194/bg-10-4607-2013.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems*. *Ann Arbor MI University of Michigan Press* (Vol. Ann Arbor). https://doi.org/10.1137/1018105

Hong, S.-Y., J. Dudhia, and S.-H. Chen (2004), A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, *Mon. Weather Rev.*, *132*(1), 103–120, doi:10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2.

Hong, S.-Y., Y. Noh, and J. Dudhia (2006), A new vertical diffusion package with an explicit treatment of entrainment processes., *Mon. Weather Rev.*, *134*(9), 2318–2341, doi:10.1175/MWR3199.1.

Houghton, R. A., J. L. Hackler, and K. T. Lawrence (1999), The U.S. Carbon Budget: Contributions from Land-Use Change, *Science (80-.).*, *285*(5427), 574–578, doi:10.1126/science.285.5427.574.

Houghton, R. A. (2003), Why are estimates of the terrestrial carbon balance so different?, *Glob. Chang. Biol.*, *9*(4), 500–509, doi:10.1046/j.1365-2486.2003.00620.x.

Houtekamer, P. L., and H. L. Mitchell (2001), A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Mon. Weather Rev.*, *129*(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houweling, S. et al. (2010), The importance of transport model uncertainties for the estimation of CO2 sources and sinks using satellite measurements, *Atmos. Chem. Phys.*, *10*(20), 9981–9992, doi:10.5194/acp-10-9981-2010.

Hu, X. M., P. M. Klein, and M. Xue (2013), Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments, *J. Geophys. Res. Atmos.*, *118*(18), 10490–10505, doi:10.1002/jgrd.50823.

Hu, X. M., J. W. Nielsen-Gammon, and F. Zhang (2010), Evaluation of three planetary boundary layer schemes in the WRF model, *J. Appl. Meteorol. Climatol.*, *49*(9), 1831–1844, doi:10.1175/2010JAMC2432.1.

Huntzinger, D. N. et al. (2012), North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison, *Ecol. Modell.*, *232*, 144–157, doi:10.1016/j.ecolmodel.2012.02.004.

Jacobson, A. R., S. E. M. Fletcher, N. Gruber, J. L. Sarmiento, and M. Gloor (2007), A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, *Global Biogeochem. Cycles*, *21*(1), doi:10.1029/2005GB002556.

Janjic, Z. (2002), Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model, *NCEP Off. Note*, *437*, 61.

Janssens, I. A. et al. (2003), Europe's terrestrial biosphere absorbs 7 to 12% of European anthropogenic CO2 emissions, *Science (80-.).*, *300*(5625), 1538–1542, doi:10.1126/science.1083592.

Johnson, A., X. Wang, M. Xue, and F. Kong (2011), Hierarchical Cluster Analysis of a Convection-Allowing Ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part II: Ensemble Clustering over the Whole Experiment Period, *Mon. Weather Rev.*, *139*(12), 3694–3710, doi:10.1175/MWR-D-11-00016.1.

Kain, J. S. (2004), The Kain–Fritsch Convective Parameterization: An Update, *J. Appl. Meteorol.*, *43*(1), 170–181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Keeling, C. D., S. C. Piper, and M. Heimann (1989), *Aspects of Climate Variability in the Pacific and the Western Americas*.

King, A. W., L. Dilling, G. P. Zimmerman, D. M. Fairman, R. A. Houghton, G. Marland, A. Z. Rose, and T. J. Wilbanks (2007), The first state of the carbon cycle report (SOCCR): The North American carbon budget and implications for the global carbon cycle, *first state carbon cycle Rep. North Am. carbon Budg. Implic. Glob. carbon cycle*, (November), 242.

King, A. W. et al. (2015), North America's net terrestrial CO2 exchange with the atmosphere 1990-2009, *Biogeosciences*, *12*(2), 399–414, doi:10.5194/bg-12-399-2015.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983), Optimization by Simulated Annealing, *Science (80-.).*, *220*(4598), 671–680, doi:10.1126/science.220.4598.671.

Kretschmer, R., C. Gerbig, U. Karstens, G. Biavati, A. Vermeulen, F. Vogel, S. Hammer, and K. U. Totsche (2014), Impact of optimized mixing heights on simulated regional atmospheric transport of CO2, *Atmos. Chem. Phys.*, *14*(14), 7149–7172, doi:10.5194/acp-14-7149-2014.

Kretschmer, R., C. Gerbig, U. Karstens, and F. T. Koch (2012), Error characterization of CO2 vertical mixing in the atmospheric transport model WRF-VPRM, *Atmos. Chem. Phys.*, *12*(5), 2441–2458, doi:10.5194/acp-12-2441-2012.

Krol, M., S. Houweling, B. Bregman, M. van den Broek, A. Segers, P. van Velthoven, W. Peters, F. Dentener, and P. Bergamaschi (2005), The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, *Atmos. Chem. Phys.*, *5*, 417–432, doi:10.5194/acpd-4-3975-2004.

Lauvaux, T., and K. J. Davis (2014), Planetary boundary layer errors in mesoscale inversions of column-integrated CO2 measurements, *J. Geophys. Res. Atmos.*, *119*(2), 490–508, doi:10.1002/2013JD020175.

Lauvaux, T. et al. (2009), Bridging the gap between atmospheric concentrations and local ecosystem measurements, *Geophys. Res. Lett.*, *36*(19), doi:10.1029/2009GL039574.

Lauvaux, T., O. Pannekoucke, C. Sarrat, F. Chevallier, P. Ciais, J. Noilhan, and P. J. Rayner (2009), Structure of the transport uncertainty in mesoscale inversions of CO2 sources and sinks using ensemble model simulations, *Biogeosciences*, *6*(6), 1089–1102, doi:10.5194/bg-6-1089-2009.

Lauvaux, T. et al. (2012), Constraining the CO2 budget of the corn belt: Exploring uncertainties from the assumptions in a mesoscale inverse system, *Atmos. Chem. Phys.*, *12*(1), 337–354, doi:10.5194/acp-12-337-2012.

Law, R. M. et al. (2008), TransCom model simulations of hourly atmospheric CO2: Experimental overview and diurnal cycle results for 2002, *Global Biogeochem. Cycles*, *22*(3), doi:10.1029/2007GB003050.

Law, R. M., P. J. Rayner, L. P. Steele, and I. G. Enting (2003), Data and modelling requirements for CO2 inversions using high-frequency data, *Tellus, Ser. B Chem. Phys. Meteorol.*, *55*(2), 512–521, doi:10.1034/j.1600-0889.2003.00029.x.

Le Quéré, C. et al. (2015), Global Carbon Budget 2015, *Earth Syst. Sci. Data*, *7*(2), 349–396, doi:10.5194/essd-7-349-2015.

Lee, J. (2012), Techniques for Down-Selecting Numerical Weather Prediction Ensembles, The Pennsylvania State University.

Lee, J. a., W. C. Kolczynski, T. C. McCandless, and S. E. Haupt (2012), An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Wind Prediction, *Mon. Weather Rev.*, *140*(7), 2270–2286, doi:10.1175/MWR-D-11-00065.1.

Lee, J. A., H. SE, and Y. GS (2016), Down-Selecting Numerical Weather Prediction Multi-Physics Ensembles with Hierarchical Cluster Analysis, *J. Climatol. Weather Forecast.*, *4*(1), 1–16, doi:10.4172/2332-2594.1000156.

Lin, J. C., and C. Gerbig (2005), Accounting for the effect of transport errors on tracer inversions, *Geophys. Res. Lett.*, *32*(1), 1–5, doi:10.1029/2004GL021127.

Lokupitiya, E., S. Denning, K. Paustian, I. Baker, K. Schaefer, S. Verma, T. Meyers, C. Bernacchi, A. Suyker, and M. Fischer (2009), Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands, *Biogeosciences*, *6*(1), 969–986, doi:10.5194/bgd-6-1903-2009.

Marland, G., and R. M. Rotty (1984), Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950???1982, *Tellus B*, *36 B*(4), 232–261, doi:10.1111/j.1600-0889.1984.tb00245.x.

Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre (2015), Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization, *Mon. Weather Rev.*, *143*(5), 1622–1643, doi:10.1175/MWR-D-14-00157.1.

Mesinger, F. et al. (2006), North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, *87*(3), 343–360, doi:10.1175/BAMS-87-3-343.

Miles, N. L., S. J. Richardson, K. J. Davis, T. Lauvaux, A. E. Andrews, T. O. West, V. Bandaru, and E. R. Crosson (2012), Large amplitude spatial and temporal gradients in atmospheric boundary layer CO2 mole fractions detected with a tower-based network in the U.S. upper Midwest, *J. Geophys. Res. Biogeosciences*, *117*(1), doi:10.1029/2011JG001781.

Miller, S. M., M. N. Hayek, A. E. Andrews, I. Fung, and J. Liu (2015), Biases in atmospheric CO_2 estimates from correlated meteorology modeling errors, *Atmos. Chem. Phys.*, *15*(5), 2903–2914, doi:10.5194/acp-15-2903-2015.

Molod, A., H. Salmun, and M. Dempsey (2015), Estimating planetary boundary layer heights from NOAA Profiler Network wind profiler data, *J. Atmos. Ocean. Technol.*, *32*(9), 1545–1561, doi:10.1175/JTECH-D-14-00155.1.

Nakanishi, M., and H. Niino (2004), An improved Mellor-Yamada Level-3 model with condensation physics: Its design and verification, *Boundary-Layer Meteorol.*, *112*(1), 1–31, doi:10.1023/B:BOUN.0000020164.04146.98.

Ogle, S. M. et al. (2015), An approach for verifying biogenic greenhouse gas emissions inventories with atmospheric CO 2 concentration data, *Environ. Res. Lett.*, *10*(3), 34012, doi:10.1088/1748-9326/10/3/034012.

Ogle, S., K. J. Davis, A. E. Andrews, K. R. Gurney, T. O. West, R. B. Cooke, T. Parkin, J. Morissette, S. Verma, and S. Wofsy (2006), *Science plan: Mid-Continent Intensive Campaign of the North American Carbon Program, NACP*, Greenbel, MD.

Pacala, S. W. et al. (2001), Consistent Land- and Atmosphere-Based U.S. Carbon Sink Estimates, *Science (80-.).*, *292*(5525), 2316–2320, doi:10.1126/science.1057320.

Pan, Y. et al. (2011), A large and persistent carbon sink in the world's forests., *Science*, *333*(6045), 988–93, doi:10.1126/science.1201609.

Patra, P. K. et al. (2008), TransCom model simulations of hourly atmospheric CO2: Analysis of synoptic-scale variations for the period 2002-2003, *Global Biogeochem. Cycles*, *22*(4), doi:10.1029/2007GB003081.

Pei, L., N. Moore, S. Zhong, L. Luo, D. W. Hyndman, W. E. Heilman, and Z. Gao (2014), WRF model sensitivity to land surface model and cumulus parameterization under short-term climate extremes over the Southern great plains of the United States, *J. Clim.*, *27*(20), 7703–7724, doi:10.1175/JCLI-D-14-00015.1.

Pérez-Landa, G., P. Ciais, M. J. Sanz, B. Gioli, F. Miglietta, J. L. Palau, G. Gangoiti, and M. M. Millán (2006), Mesoscale circulations over complex terrain in the Valencia coastal region, Spain, Part 1: simulation of diurnal circulation regimes, *Atmos. Chem. Phys. Discuss.*, *6*(2), 2809–2852, doi:10.5194/acpd-6-2809-2006.

Peters, W. et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker., *Proc. Natl. Acad. Sci. U. S. A.*, *104*(48), 18925–18930, doi:10.1073/pnas.0708986104.

Peylin, P. et al. (2013), Global atmospheric carbon budget: Results from an ensemble of atmospheric CO2 inversions, *Biogeosciences*, *10*(10), 6699–6720, doi:10.5194/bg-10-6699-2013.

Pickett-Heaps, C. A. et al. (2011), Atmospheric CO2 inversion validation using vertical profile measurements: Analysis of four independent inversion models, *J. Geophys. Res. Atmos.*, *116*(D12), n/a--n/a, doi:10.1029/2010JD014887.

Riccio, A., A. Ciaramella, G. Giunta, S. Galmarini, E. Solazzo, and S. Potempski (2012), On the systematic reduction of data complexity in multimodel atmospheric dispersion ensemble modeling, *J. Geophys. Res. Atmos.*, *117*(5), doi:10.1029/2011JD016503.

Richardson, S. J., N. L. Miles, K. J. Davis, E. R. Crosson, C. W. Rella, and A. E. Andrews (2012), Field testing of cavity ring-down spectroscopy analyzers measuring carbon dioxide and water vapor, *J. Atmos. Ocean. Technol.*, *29*(3), 397–406, doi:10.1175/JTECH-D-11-00063.1.

Román-Cascón, C., C. Ya Ue, M. Sastre, G. Maqueda, F. Salamanca, and S. Viana (2012), Observations and WRF simulations of fog events at the Spanish Northern Plateau, *Adv. Sci. Res*, *8*, 11–18, doi:10.5194/asr-8-11-2012.

Roulston, M. S., and L. A. Smith (2003), Combining dynamical and statistical ensembles, *Tellus, Ser. A Dyn. Meteorol. Oceanogr.*, *55*(1), 16–30, doi:10.1034/j.1600-0870.2003.201378.x.

Sarmiento, J. L., M. Gloor, N. Gruber, C. Beaulieu, A. R. Jacobson, S. E. M. Fletcher, S. Pacala, and K. Rodgers (2010), Trends and regional distributions of land and ocean carbon sinks, *Biogeosciences*, *7*(8), 2351–2367, doi:10.5194/bg-7-2351-2010.

Sarrat, C., J. Noilhan, A. Dolman, and C. Gerbig (2007), Atmospheric CO< sub> 2</sub> modeling at the regional scale: an intercomparison of 5 meso-scale atmospheric models, *Biogeosciences*.

Schuh, A. E., A. S. Denning, K. D. Corbin, I. T. Baker, M. Uliasz, N. Parazoo, A. E. Andrews, and D. E. J. Worthy (2010), A regional high-resolution carbon flux inversion of North America for 2004, *Biogeosciences*, *7*(5), 1625–1644, doi:10.5194/bg-7-1625-2010.

Schuh, A. E. et al. (2013), Evaluating atmospheric CO2 inversions at multiple scales over a highly inventoried agricultural landscape, *Glob. Chang. Biol.*, *19*(5), 1424–1439, doi:10.1111/gcb.12141.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., Powers, J. G. (2005), A description of the advanced research WRF version 2, , 100.

Smallman, T. L., M. Williams, and J. B. Moncrieff (2014), Can seasonal and interannual variation in landscape CO2 fluxes be detected by atmospheric observations of CO2 concentrations made at a tall tower?, *Biogeosciences*, *11*(3), 735–747, doi:10.5194/bg-11-735-2014.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and D. Kim (2000), Parameterization of cold-season processes in the MAPS land-surface scheme, *J. Geophys. Res.*, *105*(D3), 4077, doi:10.1029/1999JD901047.

Solazzo, E., and S. Galmarini (2015), The Fukushima-^{137}Cs deposition case study: Properties of the multi-model ensemble, *J. Environ. Radioact.*, *139*, doi:10.1016/j.jenvrad.2014.02.017.

Stephens, B. B. et al. (2007), Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO2., *Science*, *316*(5832), 1732–5, doi:10.1126/science.1137004.

Stull, R. B. (1988), An Introduction to Boundary Layer Met Meteorology, 666 pp., Kluwer Acad., Dordrecht, Netherlands.

Talagrand, O., R. Vautard, and B. Strauss (1999), Evaluation of Probabilistic Prediction System, in *Workshop on Predictability*, ECMWF, Reading, U. K.

Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the Global Atmospheric CO2 budget, *Science (80-.).*, *247*, 1431–1438, doi:10.1126/science.247.4949.1431.

Tarantola, A. (2004), Inverse Problem Theory and Methods for Model Parameter Estimation, *Book*, doi:10.1137/1.9780898717921.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192, doi:10.1029/2000JD900719.

Thompson, G., R. M. Rasmussen, and K. Manning (2004), Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part I: Description and Sensitivity Analysis, *Mon. Weather Rev.*, *132*(2), 519–542, doi:10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2.

van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, and J. Arellano, A. F. (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, *6*(11), 3423–3441, doi:10.5194/acpd-6-3175-2006.

Van Der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, M. Mu, P. S. Kasibhatla, D. C. Morton, R. S. Defries, Y. Jin, and T. T. Van Leeuwen (2010), Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009), *Atmos. Chem. Phys.*, *10*(23), 11707–11735, doi:10.5194/acp-10-11707-2010.

Wang, W., K. J. Davis, B. D. Cook, C. Yi, M. P. Butler, D. M. Ricciuto, and P. S. Bakwin (2007), Estimating daytime CO2 fluxes over a mixed forest from tall tower mixing ratio measurements, *J. Geophys. Res. Atmos.*, *112*(10), doi:10.1029/2006JD007770.

Whitaker, J. S., and A. F. Loughe (1998), The Relationship between Ensemble Spread and Ensemble Mean Skill, *Mon. Weather Rev.*, *126*(12), 3292–3302, doi:10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2.

Wilks, D. S. (Department of E. and A. S. C. U. (2006), *Statistical Methods in the Atmospheric Sciences*.

Williams, I. N., W. J. Riley, M. S. Torn, J. A. Berry, and S. C. Biraud (2011), Using boundary layer equilibrium to reduce uncertainties in transport models and CO2 flux inversions, *Atmos. Chem. Phys.*, *11*(18), 9631–9641, doi:10.5194/acp-11-9631-2011.

Yerramilli, A., V. S. Challa, V. B. R. Dodla, H. P. Dasari, J. H. Young, C. Patrick, J. M. Baham, R. L. Hughes, M. G. Hardy, and S. J. Swanier (2010), Simulation of Surface Ozone Pollution in the Central Gulf Coast Region Using WRF/Chem Model: Sensitivity to PBL and Land Surface Physics, *Adv. Meteorol.*, *2010*, 1–24, doi:10.1155/2010/319138.

Yi, C., K. J. Davis, P. S. Bakwin, A.S. Denning, N. Zhang, A. Desai, J.Ch.-H. Lin, and C. Gerbig, (2004), The observed covariance between ecosystem carbon exchange and atmospheric boundary layer dynamics in North Wisconsin, *Journal of Geophysical Research*, **109**(D08302): doi10.1029/2003JD004164.

Yussouf, N., D. J. Stensrud, and S. Lakshmivarahan (2004), Cluster analysis of multimodel ensemble data over New England, *Mon. Weather Rev.*, *132*(10), 2452–2462, doi:10.1175/1520-0493(2004)132<2452:CAOMED>2.0.CO;2.

Yver, C. E., H. D. Graven, D. D. Lucas, P. J. Cameron-Smith, R. F. Keeling, and R. F. Weiss (2013), Geoscientific Instrumentation Methods and Data Systems Evaluating transport in the

WRF model along the California coast, *Atmos. Chem. Phys*, *13*, 1837–1852, doi:10.5194/acp-13-1837-2013.

Zhang, Y., M. K. Dubey, and S. C. Olsen (2009), Comparisons of WRF/Chem simulations in Mexico City with ground-based RAMA measurements during the MILAGRO-2006 period, *Atmos. Chem. Phys. Discuss.*, *9*(1), 1329–1376, doi:10.5194/acpd-9-1329-2009.

VITA

Liza Ivelisse Díaz
503 Walker Building
Department of Meteorology
University Park, PA 16802
e-mail: lzd120@psu.edu

EDUCATION

The Pennsylvania State University — University Park, PA
Ph.D. Candidate, Meteorology — 2010 – 2017

The Pennsylvania State University — University Park, PA
M.S. Meteorology — 2007 – 2010

University of Puerto Rico – Río Piedras — San Juan, PR
B.A. Geography — 2003 – 2007

RESEARCH EXPERIENCE

The Pennsylvania State University — University Park, PA
Research Assistant; Advisor: Dr. Kenneth J. Davis, Co-Advisor: Thomas Lauvaux — 2010-2017
Dissertation Title: *"Quantification of Transport Errors and their Impacts on CO_2 inverse fluxes over the Mid-Continental intensive (MCI) region"*

The Pennsylvania State University — University Park, PA
Research Assistant; Advisor: Dr. Kenneth J. Davis — 2008 – 2010
Thesis Title: "*Model-Data Comparison of Mid-Continental Intensive Field Campaign Atmospheric CO_2 Mixing Ratios"*

TEACHING EXPERIENCE

The Pennsylvania State University — University Park, PA
Teaching Assistant, Fundamentals of Atmospheric Science — Spring 2008

PUBLICATIONS

Díaz-Isaac L. I., T. Lauvaux, K. J. Davis. *Sensitivity and Uncertainty Analysis of Physical Parameterization and Initial Conditions on Meteorological Variables and CO_2 Mole Fractions*. (in prep.)

Díaz-Isaac L. I., T. Lauvaux, M. Bocquet. K. J. Davis, *Calibration of a Multi-Physics Ensemble for Greenhouse Gas Atmospheric Transport Model Uncertainty Estimation*. (in prep.)

Díaz Isaac, L. I., K. J. Davis, T. Lauvaux, N. Miles, S. Richardson, A. Jacobson, and A. Andrews, 2014. *Model-Data Comparison of Mid-Continental Intensive (MCI) Field Campaign Atmospheric CO_2 Mole Fractions.* J. Geophys. Res. Atmos., 119, 10,536-10,551, doi:10.1002/2014JD021593