An Examination of the Downscaled Regional Climate Model (RCM) Scenarios for Assessing Air Quality Response to Climate Variability and Change

An Internal EPA Report Prepared By

Ellen Cooter, Robert Gilliam, Alice Gilliland and Jenise Swall
U.S. Department of Commerce, National Oceanic and Atmospheric Administration,
Air Resources Laboratory, Atmospheric Sciences Modeling Division
Research Triangle Park, North Carolina 27711

In partnership with the U.S. Environmental Protection Agency, Office of Research and Development

Prepared For

Brooke Hemming, Air Quality Team Lead, NCEA-IO
Anne Grambsch, Acting Global Staff Director, NCEA
Joel Scheraga, National Program Director
U.S. EPA Global Change Program, Office of Research and Development

23 September, 2005

TABLE OF CONTENTS

LIST OF FIGURES ..... **II**

LIST OF TABLES ..... **VII**

EXECUTIVE SUMMARY ..... **IX**

1.0 BACKGROUND AND OVERVIEW **1**

2.0 REGIONAL CLIMATE MODEL SCENARIO OVERVIEW **5**

3.0 SPATIAL ANALYSIS **9**

3.1 METHODOLOGY 10
3.1.1 Climate variable selection, data preparation and quality assurance *10*
3.1.2 Statistical Method Selection and Application Development *12*
3.1.3 Method quality assurance *14*
3.2 SEASONAL ANALYSIS 16
3.2.1 Spring *21*
3.2.2 Summer *29*
3.2.3 Autumn *40*
3.2.4 Winter *46*

4.0 DISTRIBUTIONAL AND TIME SERIES ANALYSIS **54**

4.1 RESEARCH METHODS AND DATASETS 54
4.1.1 Climate Observations *54*
4.1.2 Time series extraction, Distributions Generation, and Mean Patterns *56*
4.1.3 Time Series Filter *58*
4.1.4 Analysis Overview *60*
4.2 ASSESSMENT OF MM5 REGIONAL CLIMATE MODEL BASE CASE CLIMATE 61
4.2.1 Sea-Level Pressure and Surface Wind Flow *61*
4.2.2 Temperature Evaluation *80*
4.2.3 Precipitation Evaluation *96*
4.3 POINT COMPARISON OF CURRENT AND FUTURE REGIONAL CLIMATE SIMULATIONS 100
4.3.1 Temperature *101*
4.3.2. Wind patterns *112*
4.3.3 Precipitation *116*

5.0 SUMMARY **120**

5.1 BASE CASE CLIMATE SIMULATION 120
5.2 FUTURE CLIMATE CONDITIONS 122

6.0 REFERENCES **124**

APPENDIX A. LIST OF NATIONAL WEATHER SERVICE (NWS) SITES USED IN CLIMATE ANALYSIS 129

APPENDIX B. DATA MANAGEMENT AND QUALITY ASSURANCE 134

B.1 METEOROLOGICAL DATA PROCESSING 134
B.2 SPECIFIC DATA PROCESSING ISSUES 136
B.2.1 MCIP Settings 136
3.2.2 Data Extraction for Various Climate Analyses 137
B.3 QUALITY ASSURANCE PROCEDURES 137
B.3.1 File existence and integrity – Layer I Quality Control 138
B.3.2 Data Scan – Spatial and Temporal Range Checks – Layer II Quality Control 139
B.3.3 Incidental Data Integrity Checks –Layer III Quality Control 144
B.4 MCIP PROCESSING 145
B.4.1 Climate Data Extraction and Analysis 145
B.4.2 700 mb Wind Extraction for Cluster Analysis 145
B.5 QUALITY ASSURANCE PROJECT PLAN FOR THE CLIMATE IMPACT ON AIR QUALITY(CIRAQ) PROJECT (PHASE I, 2005-2007) 147

List of Figures

Figure 1. Box plot of relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank. 21
Figure 2. Box plot of relative frequency distributions for base case (B) and future (F) RCM grouped by relative frequency rank..... 21
Figure 3. 700 mb NCEP-R1 and AMIP-R2 spring season transport pattern differences representative of Table 4 comparisons a, c and d. Black arrows are NCEP-R1 directions. Red arrows are AMIP-R2 directions. Velocity differences (m/s) are computed as (R2 - R1). 26
Figure 4. 700 mb NCEP-R1 and AMIP-R2 Spring season transport pattern differences representative of Table 4 comparisons b and e. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2 - R1). 26
Figure 5. 700 mb reanalysis and base case RCM spring season transport pattern differences representative of Table 4 comparison a. Black arrows are base case RCM wind directions. Red arrows are reanalysis wind directions. Velocity difference (m/s) is computed as (reanalysis – base case). 27
Figure 6. 700 mb reanalysis and base case RCM spring season transport pattern differences representative of Table 4 comparisons b through e. Black arrows are RCM wind directions. Red arrows are reanalysis wind directions. Velocity differences are computed as (reanalysis – base case). 27
Figure 7. 700 mb base case and future RCM spring season transport pattern differences representative of Table 4 comparisons b and c. Black arrows are base case wind directions. Red arrows are future wind directions. Velocity difference (m/s) is computed as (future - base case). 28
Figure 8. 700 mb base case and future RCM spring season transport pattern differences representative of Table 4 comparisons a, d and e. Black arrows are base case wind directions. Red arrows are future wind directions. Velocity difference (m/s) is computed as (future - base case). 28
Figure 9. Box plot of synthetic spring season relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank. 29
Figure 10. Box plot of synthetic spring season relative frequency distributions for base case (B) and future (F) RCM scenarios grouped by relative frequency rank..... 29
Figure 11. 700 mb NCEP-R1 and AMIP-R2 summer season transport pattern differences representative of Table 6 comparisons a through d. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1). 35
Figure 12. 700 mb NCEP-R1 and AMIP-R2 summer season transport pattern differences representative of Table 6 comparison e. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1). 35
Figure 13. RCM base case summer season 700 mb wind speed (m/s) and direction illustrating the persistent western trough feature present in all base case average dominant transport patterns. 36

Figure 14. 700 mb RCM base case and reanalysis summer season transport pattern differences representative of all reanalysis and base case comparisons in Table 6. Note the location of the reanalysis anti-cyclone centered over Arkansas. Black arrows are RCM base case wind directions. Red arrows are reanalysis wind directions. Velocity differences (m/s) are computed as (reanalysis-base case). 36
Figure 15. 700 mb RCM base case and future summer season transport pattern differences representative of Table 6 comparisons a through d. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future-base). 37
Figure 16. 700 mb RCM base case and future summer season pattern differences representative of Table 6 comparison e. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future- base case). 37
Figure 17. Map of RCM future summer season wind speed (m/s) and direction illustrating weak transport and lack of active synoptic features. 38
Figure 18. Box plot of synthetic summer season relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank. 39
Figure 19. Box plot of synthetic summer season relative frequency distributions for base case (B) and future (F) RCM grouped by relative frequency rank. 39
Figure 20. 700 mb NCEP-R1 and AMIP-R2 autumn season transport pattern differences representative of Table 8 comparisons b, c and e. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1). 44
Figure 21. 700 mb NCEP-R1 and AMIP-R2 autumn season transport pattern differences representative of Table 8 comparison d. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1). 44
Figure 22. 700 mb RCM base case and reanalysis autumn season transport pattern differences representative of all Table 8 comparisons. Black arrows are RCM base case wind directions. Red arrows are reanalysis wind directions. Velocity differences (m/s) are computed as (reanalysis- base case). 45
Figure 23. 700 mb RCM base case and future autumn season transport pattern differences representative of all Table 8 comparisons. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future-base case). 45
Figure 24. Box plot of synthetic autumn season relative frequency distributions for NCEP-R1(R1), base case RCM (GM), and AMIP-R2(R2) grouped by relative frequency rank. 46
Figure 25. Box plot of synthetic autumn season relative frequency distributions for base case (B) and future (F) RCM scenarios grouped by relative frequency rank..... 46
Figure 26. 700 mb NCEP-R1 and AMIP-R2 winter season transport pattern differences representative of Table 10 comparisons a through d. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1). 51
Figure 27. 700 mb NCEP-R1 and AMIP-R2 winter season transport pattern differences for Table 10 comparison e. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-

R2 wind directions. Velocity differences (m/s) are computed as (R2-R1). 51
Figure 28. 700 mb RCM base case and reanalysis transport pattern differences representative of all Table 10 comparisons. Black arrows are RCM base case wind directions. Red arrows are reanalysis wind directions. Velocity differences (m/s) are computed as (reanalysis- base case)..... 52
Figure 29. 700 mb RCM base case and future winter season transport pattern difference representative of Table 10 comparisons a through d. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future-base case). 52
Figure 30. Box plots of synthetic winter season relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank. 53
Figure 31. Box plots of synthetic winter season relative frequency distribution for base case (B) and future (F) RCM grouped relative frequency rank. 53
Figure 32. Weather observation sited used in the climate analysis 55
Figure 33. The 10-year mean sea-level pressure pattern in the wintertime (DJF). The top panels is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale. Note that the sea-level pressure in the western U.S. is not reliable. 63
Figure 34. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the wintertime (DJF). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction is within +/-40° of the observed (light green)..... 66
Figure 35. Sea surface temperature patterns during the winter. Upper panel is an example of the sea surface temperature from the regional climate simulation on a particular day in February. The lower panel is the observed 14 km scale sea surface temperature in February 2002. Both datasets are plotted using the same color scale. 68
Figure 36. The 10-year mean sea-level pressure pattern in the springtime (MAM). The top panel is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale. Note that the sea-level pressure in the western U.S. is not reliable. 72
Figure 37. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the springtime (MAM). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction mode is within +/- 40° of the observed (light green). 73
Figure 38. The 10-year mean sea-level pressure pattern in the summertime (JJA). The top panel is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale, and simulated sea-level pressure over the western U.S. is not reliable. 75
Figure 39. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the summertime (JJA). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction mode is within +/- 40 ° of the observed (light green). 76
Figure 40. The 10-year mean sea-level pressure pattern in the fall (SON). The top panel is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale. Note that the sea-level pressure in the

western U.S. is not reliable. 79
Figure 41. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the fall (SON). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction mode is within +/- 40° of the observed (light green). 80
Figure 42. Mean 2 m temperature (10-year) difference (K) between the climate model simula- 81
Figure 43. The seasonal mean 2-m temperature (K) difference between the regional climate simulation and observations (model-observation). 81
Figure 44. Mean (10-year) seasonal temperature perturbation and interannual seasonal variability of the regional climate simulation (blue) and observation (red). 87
Figure 45. The percentage difference of the model and observed temperature 2 m temperature variability for different time scales (diurnal and synoptic) and the different seasons. The percent difference is scaled by the total variability of the non-filtered time series. 91
Figure 46. The percentage difference of the model and observed temperature 2-m temperature variability for the intraday, seasonal and interannual time scales. The percent difference is scaled by the total variability of the non-filtered time series. 95
Figure 47. Mean seasonal precipitation (cm) calculated from 10-years of data. The background shading represents the observed precipitation, and the foreground dots represent the simulated precipitation at the climate reference sites. 97
Figure 48. Seasonal precipitation variability (cm) calculated from 10-years of data. The background shading represents the observed precipitation variability, and the foreground dots represent the simulated precipitation variability at the climate reference sites. 97
Figure 49. The difference in annual mean of the hourly temperature (K) for the future versus current regional climate simulations. 103
Figure 50. Same as Figure 49, except that the mean temperature difference is calculated separately for each season. 103
Figure 51. The hourly baseline average component of the 2m temperature (K) for select NWS sites from the Eastern and Midwestern U.S., including Atlanta, GA (KATL), Raleigh, NC (RDU), Roanoke, VA (KROA), Albany, NY (KALB), Springfield, MO (KSGF), and Greenbay, WI (KGRB). Each plot includes the median (horizontal line), the upper 75th and lower 25th quartiles (box), outer fences (75th quartile plus 1.5 times the interquartile range; 25th quartile minus 1.5 times the interquartile range), and outliers (solid dots outside of the fences). The baseline component represents the long-term moving average mean over the ten year simulation with seasonal, synoptic, diurnal, and intra-day variability removed. ... 105
Figure 52. The hourly baseline average component of the 2m temperature (K) for select NWS sites from the Southwestern and Western U.S., including Dallas-Ft. Worth, TX (KDFW), Alburquerque, NM (KABQ), Los Angeles, CA (KLAX), Ukiah, CA (KUKI), Portland, OR (KPDX), and Seattle, WA (KSEA). Each plot includes the median (horizontal line), the upper 75th and lower 25th quartiles (box), outer fences (75th quartile plus 1.5 times the interquartile range; 25th quartile minus 1.5 times the interquartile range), and outliers (solid dots outside of the fences). The baseline component represents the long-term moving average mean over the ten year simulation with seasonal, synoptic, diurnal, and intra-day variability removed. 106
Figure 53. Difference in future versus current number of hourly-modeled temperature values above 100°F (311 K). 110

Figure 54. The mean length (days) of heat episodes: Future – Current Scenarios. The numbers on above (below) each marker represent the mean number of days in a heat wave episode in the future (current) scenario. Heat episodes are defined as the number of consecutive days with at least one simulated hourly temperature greater than 90°F (305 K). 111
Figure 55. The seasonal component of the current and future modeled temperatures at the NWS site in Ukiah, CA. (a) shows the seasonal temperature component for each year separately to assess interannual variability, and (b) shows the average seasonal temperature component for all ten years and the ±1 σ standard deviation around that mean for all ten years of current and future simulations. 111
Figure 56. The seasonal component of the current and future modeled temperatures at the NWS site in Portland, OR. (a) shows the seasonal temperature component for each year separately to assess interannual variability, and (b) shows the average seasonal temperature component for all ten years and the ±1 σ standard deviation around that mean for all ten years of current and future simulations. 112
Figure 57. Current and future RCM predicted dominant 10-m wind directions at the climate reference sites for the winter and summer. The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the future (dark green) wind direction is within 40° of the current (light green). If the wind directions are not within 40° of each other, the wind vectors follow a different color patter of blue (current) and future (red). 114
Figure 58. Current (top) and future (bottom) MM5-RCM mean sea-level pressure patterns for the winter (left) and summer (right). 115
Figure 59. Average rainfall (cm) for the future and current scenarios and the different between Future-Current Scenario. 118
Figure 60. The summer average rainfall (cm) for the future and base case scenarios and the different between Future-Base Case scenario. 119

List of Tables

Table 1. Regional Climate Model (RCM) specifications for CIRAQ and New York Climate and Health Projects (NYCHP). 8
Table 2. Best pattern matches for example NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (RCM - reanalysis) and RCM (future - base case). 20
Table 3. Example statistical summaries for synthetic relative frequency distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap. 20
Table 4. Best visual matches for spring season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - base case) or RCM (future - base case). 22
Table 5. Statistical summaries of spring season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap. 24
Table 6. Best visual matches for summer season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - RCM) or RCM (future - base case). 30
Table 7. Statistical summaries of summer season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap. 33
Table 8. Best visual matches for autumn season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - RCM) or RCM (future - base case). 40
Table 9. Statistical summaries of autumn season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap. 42
Table 10. Best visual matches for winter season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - RCM) or RCM (future - base case). 47
Table 11. Statistical summaries of winter season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap. 49

DISCLAIMER

The research presented here was performed under the Memorandum of Understanding between the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to NOAA's Air Quality and Climate Programs. Although it has been reviewed by EPA and NOAA and has been approved for publication, it does not necessarily reflect their policies or views.

ACKNOWLEDGEMENTS

Thanks are extended to Dr. Brian Eder, NOAA/ARL/AMD in Research Triangle Park, NC and Dr. Richard Cohn, Constella Health Sciences Team, Constella Group, Inc., Durham, NC for many helpful discussions and thoughtful suggestions on the adaptation and application of their spatial analysis techniques.

Executive Summary

The United States EPA has established ambient air quality standards for criteria and other air pollutants to protect human health and the environment. There is increasing recognition that climate and air quality are inter-related through atmospheric chemical, radiative, and dynamic processes. There are, however, significant gaps in our understanding of many climate-chemistry linkages and the potential impacts from future human actions on these linkages. The Climate Impact on Regional Air Quality (CIRAQ) project, a collaborative research effort involving multiple Federal Agencies and academic institutions, is designed to examine global climate change scenarios and their potential effects on ozone and particulate matter (PM) concentrations throughout North America. Global climate simulations have been derived from the NASA Goddard Institute for Space Studies (GISS) version II'(two prime) and the IPCC Special Report on Emission Scenarios (SRES) A1B "business as usual" emission scenario. Scientists with the Department of Energy (DOE) Pacific Northwest National Laboratory have used these scenarios to provide boundary and initial conditions to a regional climate model (RCM) based on the Fifth Generation Pennsylvania State/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5). The MM5-RCM was then used to generate 10 years each of present (~2000) and future (~2050) hourly climate scenarios for the continental United States with horizontal grid cell size of 36 km. This report summarizes the analysis of RCM-derived steering level winds and surface temperature, pressure, wind and precipitation on various time scales (e.g., seasonal, annual, inter-annual). MM5-RCM simulations of base case (current) conditions have been compared with historical point and gridded reanalysis datasets. Comparisons between base

case and future modeled conditions have also been completed. These comparisons are used to identify model biases and uncertainties on temporal and spatial scales that are relevant to regional and national air quality assessment. Results of the CIRAQ project contribute to the U.S. Climate Change Science Program (CCSP) assessments addressing the effects of global change on human health and welfare and human systems.

The analyses presented in this report establish that the base case MM5-RCM scenarios reflect observed regional climate conditions of particular relevance for air quality for some, but not all seasons and geographic regions (e.g. New England, Central Great Plains) of the continental United States. Air quality assessments driven by this set of RCM scenarios should focus on regional-scale events that occur on daily or longer time scales. Comparisons of base case to future model scenarios suggest that changes in atmospheric conditions may occur over the next 50 years (e.g., warmer surface temperatures and weaker surface and upper level atmospheric transport) that could impact the way we manage air quality for ozone and PM. This issue will be addressed in detail in the next CIRAQ report in FY07. Several of the analyses reported here mark the first time these climate variables have undergone rigorous study with an eye towards climate change assessment application. We find these results to be interesting and strongly recommend that similar studies should be performed for alternative GCM and RCM model configurations and emissions scenarios so that a better description of current and future climate variability and change can be made available to the air quality research and regulatory community. This is being done to some extent under the US EPA Science to Achieve Results (STAR) grant program. We look forward to including those findings as they become available, but also see the need to continue to build strong collaborations with other Federal Agencies conducting

assessment research dealing with climate variability and change (e.g., NOAA, NASA and the DOE).

1.0 Background and Overview

Improving air quality is a primary mission of the U.S. Environmental Protection Agency (EPA) and there is increasing recognition that climate and air quality are closely coupled through atmospheric chemical, radiative, and dynamic processes [*NRC*, 2001]. Our understanding of such climate-chemistry linkages and the potential impacts of future human actions on these linkages is, however, limited. The importance of this limitation is acknowledged in the U.S. Climate Change Science Program Research Strategy [*Climate Change Science Program (CCSP)*, 2003] through the following key research question:

> **Question 3.4:** ***What are the effects of regional pollution on the global atmosphere and the effects of global climate and chemical change on regional air quality and atmospheric chemical inputs to the ecosystems?***

The EPA's Climate Impact on Regional Air Quality (CIRAQ) Project directly responds to this research challenge. The CCSP question breaks logically into two parts: 1) "What are the effects of regional pollution on the global atmosphere?" and 2) "What are the effects of global climate and chemical change on regional air quality and atmospheric chemical inputs to the ecosystems?" CIRAQ addresses part two of this question.

The CIRAQ project formally began in December, 2001 with a workshop in Research Triangle Park, North Carolina to solicit input to a sound science research strategy that would meet EPA's mandated responsibilities as well as contribute to answering CCSP question 3.4. The workshop included representatives from EPA, academia, other research organizations, and EPA. Two major consensus items were identified: 1) The effect of climate change on air quality

must be considered separately from future emission scenarios so that the air quality impacts from climate change can be isolated from impacts of future air quality emission estimates; and 2) In comparing model results from current and future air quality simulations, multiple year simulations are needed in order to separate differences due to interannual variability from climate trend impacts on air quality. These recommendations are reflected in three CIRAQ activities: 1) the acquisition and quality control of 10 years each of base case and future (~2050) regional climate model (RCM) scenarios; 2) analysis of these scenarios to better understand biases in climatological means, distributions and variability that could impact air quality model performance and the interpretation of air quality model results; and 3) the generation, analysis and interpretation of base case and future CMAQ simulation results. This report summarizes progress related to activities 1 and 2 performed from FY03 through FY05.

A number of studies have explicitly investigated the effects of climate change on air quality, raising concerns that it could adversely affect air quality e.g., [*Hogrefe et al.*, 2004a; *Hogrefe et al.*, 2004b; *Mickley et al.*, 2004]. Global climate changes will likely result in changes in regional and local weather conditions and their variability. Changes in meteorology will affect air pollution levels by altering: 1) Rates of atmospheric chemical reactions and transport processes; 2) Anthropogenic emissions, including adaptive responses involving increased fuel combustion from power generation; and 3) Biogenic emissions rates from natural sources. The CIRAQ project examines potential climate change impacts on ozone (O_3) and particulate matter (PM) using EPA's regional scale Community Multiscale Air Quality (CMAQ) model linked to global scale climate and chemical transport models.

The scope of the present report is limited given the opportunities offered by the regional

climate model (RCM) scenarios described in Chapter 2. Simply put, our goal is to analyze and evaluate elements of the RCM scenarios, selected on the basis of their potential importance to CMAQ's performance for base case and future year air quality conditions with regard to: 1) Their appropriate temporal and spatial characterization relative to present-day observations; and 2) Their description of RCM projected changes in these elements and characteristics 50 years into the future (~2050). Our analysis must focus on atmospheric phenomena having spatial and temporal scales commensurate with those of the RCM. A conservative estimate of appropriate spatial scale is four times the model grid cell dimension or, in our case ~ 150km. This scale includes synoptic features such as cyclones (e.g., low pressure systems, hurricanes, tropical storms, etc.), anti-cyclones (e.g., high pressure systems, Bermuda Highs, etc.), and fronts that are largely responsible for day-to-day weather changes and persist on the order of days to weeks. Given these requirements as well as time and resource constraints, we have limited our initial analyses to temperature, pressure, dominant wind direction and annual precipitation at the surface and steering level winds at ~3000 meters (700 mb). If the present scope were expanded to encompass a complete climatology for air quality applications, then any number of additional derived measures such as stagnation indices, ventilation indices, storm track analyses and precipitation measures such as frequency, duration and intensity could be added. This report provides a detailed analysis for several previously unstudied or under-studied elements, e.g., dominant surface and steering level winds, and paves the way for the addition of more meteorological elements and derived variables in the future. Through this approach, the present report also contributes to CCSP Question 4.2, "***How can predictions of climate variability and projections of climate change be improved, and what are the limits of their predictability?"***

A brief summary of the regional climate scenarios provided to CIRAQ by the Pacific Northwest National Laboratory (PNNL) immediately follows the present background description. RCM analysis methods and results are summarized in Chapters 3 and 4. Methods employed include distributional and time series analyses at individual locations or grid positions and spatial analyses across the contiguous United States. A synthesis of these findings is provided in Chapter 5. A description and summary report of in-house quality assurance, quality control and processing to model-ready format of these scenarios is presented in Appendix B.

2.0 Regional Climate Model Scenario Overview

RCM scenarios have been developed by Dr. Lai-Yung Leung of the PNNL, in Richland, Washington, in cooperation with EPA/NCEA (Anne Grambsch, Project Officer). In this work, a regional climate model based on the Penn State/NCAR Mesoscale Model MM5[*Grell et al.*, 1994], hereafter referred to as the MM5-RCM, was used to downscale global climate simulations of current and future (mid-century) climate conditions for the continental U.S., i.e., dynamical downscaling (IAG DW-89-93963401). The global simulations, which are used to provide boundary and initial conditions for the regional model, were derived from the National Aeronautic and Space Administration (NASA) GISS version II' (two prime) Global Climate Model (GCM) [*Rind et al.*, 1999]. They were produced and archived at 6-hourly intervals by Dr. Loretta Mickley and Dr. Daniel Jacob at the Harvard University while performing climate simulations using the GISS model with embedded tropospheric chemistry (U.S. EPA Science to Achieve Results, STAR, grant R83095). These global-scale climate change scenarios follow the Special Report on Emission Scenario (SRES) "business as usual" scenario (A1B) for greenhouse gases and aerosols. All regional simulations were performed at 36 km spatial resolution and 23 vertical levels with outputs archived at hourly intervals for the U.S. Further details regarding the MM5-RCM and GISS models and downscaling may be found elsewhere [*Leung and Gustafson*, 2005; *Mickley et al.*, 2004]. Dynamical regional downscaling was selected for this application because air quality assessment requires a full suite of dynamically-consistent meteorological data (e.g., 3-D structures of temperature, humidity, and winds) that cannot be produced by statistical downscaling, and it is not yet feasible to perform long-term global climate simulations at the

smaller spatial scale required for regional air quality assessments. These MM5-RCM outputs later undergo post-processing at the U.S. EPA National Exposure Research Laboratory (NERL) in Research Triangle Park, North Carolina, for input to the air quality assessment. Post-processing descriptions and quality assurance/quality control summaries are provided in Appendix B.

A first task performed by Dr. Leung included sensitivity experiments to select an appropriate set of model configuration and physics parameterizations. The importance of these selections to the RCM scenario result has been thoroughly explored for one GCM/RCM combination [*Lynn et al.*, 2004; *Lynn et al.*, submitted]. Major elements of the final configuration are summarized in Table 1. Choices were made such that, to as great an extent as possible, the regional scenarios maintain the information provided by the GISS model. That is, choices were made that produce regional scenarios that most closely reflect the large-scale GCM patterns as opposed to choices that would more closely reproduce present-day observation or reanalysis datasets. The impact of these choices with regard to bias and uncertainty is highlighted throughout the analyses that follow.

The lower and lateral MM5 boundary conditions were updated every 6 hours based on large-scale conditions of the global simulations. Large-scale variables used in the boundary conditions include temperature, wind, water vapor mixing ratio, geopotential height, sea level pressure and sea surface temperature from the GISS simulations. In addition to providing meteorological boundary conditions for the regional model, the GISS model also provides the climate fields for the global scale chemical transport model that will be used as chemical boundary conditions for the regional air quality assessment. Regional simulations driven by the

GISS scenarios will, therefore, provide meteorological changes that are consistent with the chemical boundary conditions used in the air quality assessment.

The only other comprehensive air quality assessment that included a significant dynamic downscaling component was the New York Climate and Health Project (NYHCP) STAR grant study [*Hogrefe et al.*, 2004a; *Hogrefe et al.*, 2004b]. Table 1 facilitates comparison of critical model configuration attributes across both projects. Attributes that differentiate the PNNL downscaling completed for CIRAQ and the NYHCP project include downscaling performed throughout the entire continental U.S. and the generation of continuous decades of hourly current and future meteorology. In addition, the maintenance of the GCM scenario integrity in the downscaled results and the use of an alternative IPCC emissions scenario will result in RCM outcome differences, but may also present opportunities for future model intercomparisons.

Table 1. Regional Climate Model (RCM) specifications for CIRAQ and New York Climate and Health Projects (NYCHP).

Specification	CIRAQ	NYCHP
Global Climate Model	GISS II' [*Rind et al.*, 1999]	GISS AOM [*Russell et al.*, 1995]
Regional Climate Model (RCM) domain	36 x 36km, 115 x 169 grid Continental U.S., southern Canada and northern Mexico	108 x 108 km for U.S.; 36 x 36 km for Eastern U.S.
Dynamical Downscaling Model	MM5	MM5
Number of vertical layers	23	35
lowest model layer	40 m above the surface	35 m above the surface
boundary layer scheme	Medium Range Forecast Model (MRF)	Medium Range Forecast Model (MRF)
Convection Scheme	Grell [*Grell*, 1993; *Grell et al.*, 1994]	Grell [*Grell*, 1993; *Grell et al.*, 1994]
Radiation Scheme	Rapid Radiative Transfer Model (RRTM) for shortwave radiation [*Mlawer et al.*, 1997]; Dudhia scheme for shortwave radiation [*Dudhia*, 1989]	Rapid Radiative Transfer Model (RRTM) [*Mlawer et al.*, 1997]
Microphysics	Mixed phase microphysics, Reisner1 [*Reisner et al.*, 1998]	Mixed phase microphysics, Reisner2 [*Reisner et al.*, 1998]
Duration of simulation	10 years present, 10 years mid-century (continuous, hourly)	1993-1997, 2053-2057 (May1 through September 1)
Initial and boundary condition source	GISS II'	GISS AOM

3.0 Spatial Analysis

The goal of the spatial analysis is to identify regional to large scale patterns in climatological variables that are relevant to air quality status and prediction, and the comparison of those patterns across multiple present and future scenarios. Climate variables of interest for air quality include those that have been shown to control or predict rates of atmospheric chemical reactions and transport processes; anthropogenic emissions, and biogenic emissions rates from natural sources. A characteristic of these variables that is important to expanding our ability to predict air quality now and under future conditions is the identification of a dominant (in space or time) condition or state. Analysis of climatological patterns and characteristics in present-day observations help us to understand current relationships between air quality (e.g., ozone and PM) and climate means, extremes and variability as well supporting quality assurance for methods development. Comparisons of present-day observations to base case (current) RCM scenarios help us to better understand model-related biases in these scenarios that might impact CMAQ performance and predicted outcomes. Finally, comparison of these characteristics across base case and future RCM scenarios allows us to make statements regarding changes (or lack thereof) in air-quality relevant climate variables.

Details regarding the climatological variables selected for study, data preparation and associated quality assurance information is provided in section 3.1.1, spatial analysis method selection and application development are described in section 3.1.2 and method quality assurance is described in section 3.1.3. These discussions are followed by a detailed seasonal analysis presented in section 3.2.

3.1 Methodology

3.1.1 Climate variable selection, data preparation and quality assurance

The meteorological variables selected for analysis are the 1800 UTC 700 mb zonal (*u*) and meridional (*v*) component winds. Atmospheric conditions at about 3000 meters (700 mb) have long been established as reflecting steering flow controlling the movement of synoptic scale features (cyclones and anti-cyclones) which, in turn, control large-scale surface temperature and precipitation patterns, e.g. [*Kasahara*, 1959]. Previous studies linking meteorological condition and air quality have utilized conditions at lower altitudes e.g., [*Eder et al.*, 1994], but generally focus on the Eastern U.S. which lacks significant regional-scale orographic influence. The CIRAQ project addresses the conterminous United States, including the mountainous West, and so we must look slightly higher in the atmosphere to avoid the influence of local surface features. 700 mb transport patterns are commonly included in air quality forecast guidance and evaluation and have been shown to be strongly associated with surface temperature, humidity and extinction coefficient, the latter which is used as a surrogate for visibility [*Cohn et al.*, 2001].

MM5-RCM scenarios have been processed through the data management and quality control tool as described in Appendix B. 700 mb 1800 UTC *u* and *v* wind component data were extracted from the model-ready Meteorology-Chemistry Interface Processor (MCIP) files. MCIP files are used in this analysis since these are the data on which the air quality assessment will be based.

Ten years (1985 through 1994) of data from two widely used gridded reanalysis datasets

were acquired; National Center for Environmental Prediction (NCEP) [*Kalnay and Coauthors*, 1996] hereafter referred to as NCEP-R1 or R1 and the NCEP-DOE AMIP-II [*Kanamitsu et al.*, 2002] hereafter referred to as AMIP-R2 or R2. Both datasets characterize observed conditions across the globe at a resolution of 2.5 degrees through a combination of assimilated observation and physically-based models. R2 follows R1. Its purpose is to correct known problems in R1 and to serve as a basic verification dataset for the Second Atmospheric Model Intercomparison Project (AMIP) [*Kistler et al.*, 2001]. The R2 global analyses are made using an updated (relative to R1) forecast model, updated data assimilation system, improved diagnostic outputs, and corrections of known R1 processing problems. Notable differences include improved shortwave fluxes by introduction of a new shortwave parameterization, improved precipitation estimates through changes in convective parameterization, boundary layer physics, and moisture diffusion and more realistic interannual variability and better fit to observation through use of observed pentad precipitation in the soil wetness assimilation.

The choices of data to be assimilated and specific model physics result in slightly different interpretations of the same atmospheric state in space and time. Each dataset has its own strengths and weaknesses, but neither is considered by the atmospheric community to be more or less correct than the other. This “difference of opinion” is used in the present context to provide a sense of current levels of uncertainty regarding the observed climate. Variables included in the datasets are classified into three types [*Kalnay and Coauthors*, 1996]. Type **A** variables are generally strongly influenced by the available observations and are therefore the most reliable product of the reanalysis. Type **B** variables are influenced by both the observations and by the model, and are therefore less reliable. Type **C** variables are completely determined

by the model and should be used with caution and, wherever possible, compared with model-independent estimates. The zonally averaged *u* component of the wind is considered type a type **A** variable except in the Tropics, where the model influence is larger because of much sparser observational data making it a type **B** variable. The zonally averaged meridional velocity, *v*, corresponds to divergent flow and is considered a **B** variable [*Kistler et al.*, 2001]. Taken together, the *u* and *v* wind components allow us to express our results in the more familiar terms of wind speed and direction.

3.1.2 Statistical Method Selection and Application Development

The CIRAQ assessment goals suggest that a general numerical technique is needed that 1) identifies large-scale or persistent spatial patterns among multiple climate variables, and 2) provides a uniform or consistent measure that can be compared across datasets. We would like to identify patterns of climatological condition that are known to be indicative of air quality conditions of interest. These linkages can be complex, requiring consideration of several simultaneous atmospheric measures whose analysis requires the application of multivariate data analysis techniques. The distinguishing attribute of these methods is that both the joint behavior of the multiple simultaneous observations and the variations of the individual data elements are considered [*Wilks*, 1995].

Cluster analysis is a common multivariate technique that separates data into groups whose identities are not known in advance. The degree of similarity and difference between individual observations is used to define the groups, and to assign group membership. Numerical rules are established that quantify data "similarity" and "difference." Examples of cluster

analysis applications in the climatological literature include grouping daily weather observations into synoptic types [*Kalkstein et al.*, 1987], defining weather regimes from upper-air flow patterns [*Mo and Ghil*, 1988], definition of a synoptic climatology of air mass movement [*Fernau and Samson*, 1990], and identification of dominant synoptic patterns associated with meteorological conditions leading to poor air quality [*Cohn et al.*, 2001].

Ward's Minimum Variance method [*Ward*, 1963], was selected to facilitate development of dominant patterns of 700 mb transport. Ward's method merges pairs of clusters that will minimize the within-cluster sum of squares summed over the variable, but tends to produce compact clusters which may have poor separation properties [*Anderberg*, 1973] and gives priority to merging clusters with fewer observations, encouraging similar-sized clusters. Another well known limitation of clustering techniques is that the assignment of specific observations to specific clusters and subsequent cluster frequencies can be unstable. This behavior can confound our ability to explore changes in the frequency with which dominant transport patterns occur, and so synthetic cluster assignment distributions were developed. Synthetic distributions were generated by randomly removing 25% of observations within a season (e.g., spring) and repeating the cluster analysis for the reduced dataset. This process of removal and re-clustering was repeated to generate a distribution of cluster assignment frequencies made up of 75 samples. Box plots were produced to summarize these data. The ranges of the middle 95% of frequencies (95% data interval) as well as the ranges between maximum and minimum values were noted and hypothesis testing was explored to support method validation (i.e., reproducibility of result) and cross-dataset comparison. The synthetic distributions only rarely meet the assumptions for application of standard parametric statistics, and so non-

parametric measures are employed. Kruskal-Wallis rank and median-based scores are both considered when testing hypotheses regarding distribution location (i.e., central tendency). A non-parametric two sample Kolmogorov-Smirnov (K-S) test is used to test hypotheses regarding the distribution shape. At times the non-parametric test of location suggests the null hypothesis of equal location should not be accepted, even though there is clear overlap of the 95% data intervals (e.g. Cluster 4 in Figure 2 and Table 3). This result, along with other apparent inconsistencies noted during the analysis point to the need for further consideration of issues such as the influence of serial correlation on test statistic performance before unambiguous statements of frequency similarity or difference can be made.

3.1.3 Method quality assurance

The analysis to be performed follows closely that reported in [*Cohn et al.*, 2001]. Comparison to this source is the primary means of method quality assurance for the present application. We began by extracting the reported time period and domain. In addition to spatial sampling of the full reanalysis domain, temporal sampling was also performed. In the interest of computational efficiency, consecutive 5-day periods from 1984 through 1992 were constructed and the first, third and fifth days of each 5-day period were declared to be a single observation. This results in a single observation (or 5-day event) consisting of 2016 elements (the 2 u and v components x 336 grid nodes x 3 days considered per observation). Later analyses demonstrated that results were insensitive to the choice of consecutive versus overlapping periods and so running pentads (5-day periods) were used in the analysis reported in sections 3.2.2 through

3.2.5.

The original domain was selected to minimize the impact of off-shore patterns that are less relevant to CONUS air quality (Richard Cohn, personal communication). It is also reported that retention of the most frequent 5 clusters per season resulted in an improved ability to explain variation in surface level visibility *via* extinction coefficient and surface temperature and surface humidity; together these are good indicators of photochemical pollution potential.

Average dominant pattern cluster maps matching those that were published were developed relatively quickly. A pilot study making use of higher resolution meteorological model output, the Eta Data Assimilation System (EDAS [*Rolph*, 2002]) was explored to test the stability of these patterns as a function of model scale and essentially identical maps were produced. This, in combination with discussions with Ruby Leung (DOE/PNNL) indicate that by working at 700 mb, a height at which surface or local-scale influences should be minimal, model scale should not unduly influence our results. As with R1 and R2 reanalysis datasets, EDAS employs yet another set of physics and observation assimilation protocols that may result in dominant pattern differences. Future studies could expand the present reanalysis uncertainty portion of this study to include these data, but resource limitations preclude their inclusion here.

In addition to average cluster patterns, [*Cohn et al.*, 2001] present relative frequency rankings of clusters. Reproduction of these results proved more elusive. Even when the same re-analysis dataset (R1) and the same years were used, reported rankings could not be reproduced beyond the most frequent patterns (e.g., patterns ranked 1, 2 or 3). Following discussions with Cohn and other applied statisticians, the instability described in Section 3.1.2, combined with post-clustering assignment of remaining observations by Cohn, appear to be the primary

source of disagreement. When the Cohn results are compared to the synthetic distributions described previously, the Cohn relative frequencies for similarly ranked clusters fell within the inter-quartile range of the distributions.

3.2 Seasonal Analysis

The sections that follow contain seasonal analyses for observed (reanalysis), MM5-RCM base case and MM5-RCM future scenarios of 700 mb *u* and *v* component winds. Summaries are presented in terms of 700 mb wind speed and wind direction. Wind speed is computed from the *u* and *v* wind components as $|V| = \sqrt{u^2 + v^2)}$ and wind direction is computed as arc tan(*u*/*v*). Ward's method is used to assign each observation to one of five clusters (see Section 3.1.2). For each cluster, a mean value is computed at each domain grid point. In the case of reanalysis datasets R1 and R2 (see section 3.1.1), a one-to-one comparison for a 2.5°×2.5° degrees (~280 km × 280 km) grid domain is performed. The MM5-RCM data are produced at a higher grid resolution (36 km × 36 km) and so the MM5-RCM grid is sampled at an interval of once every 6 grids to ensure a consistent spatial scale for comparison. This sampling should not affect analysis results for large-scale, smoothly varying fields such as those expected at ~3000 meters (700 mb).

The primary analysis of interest is the identification of dominant (in space or time) patterns of atmospheric transport, and the description of similarities and differences between various datasets. This is done through a consistent set of questions and formatted responses.

The first question is, *"Do transport patterns look alike across datasets?"* To answer this question, maps of mean cluster wind speed and direction are examined and a set of difference maps is constructed. "Best-fit" matches are identified through a visual survey assessing the range and texture of wind speed differences and the degree to which arrows indicating wind direction are superimposed. Sometimes a best fit choice is not obvious (i.e., multiple pairings appear to be equally good or poor). In these cases domain-wide wind speed bias, computed as the sum of the point difference of the mean cluster values, is also considered. Such domain-wide statistics should be considered only after preliminary visual examination strongly suggests a good match because large positive and negative deviations from a central value can produce the same bias value as a more uniform field of small deviations.

The output of this analysis is a discussion of visual similarities and differences and a table summarizing the results. An example of a summary table of reanalysis datasets (an estimate of uncertainty) and downscaled base case and future MM5-RCM patterns is presented in Table 2. Columns A, B and C summarize results for the reanalysis comparison. Columns A and B provide the relative frequency rank for the reanalysis clusters showing the closest visual agreement and Column C contains the domain-wide wind speed bias for the comparison. Columns D, E and F summarize the comparison of base case MM5-RCM patterns to the two reanalysis datasets. Column D contains the MM5-RCM cluster rank. Column E contains the reanalysis dataset (R1 or R2) and cluster rank that provides the best visual agreement. Column E contains the domain-wide wind speed bias for the comparison. Columns G, H and I provide similar information for comparison between MM5-RCM base case and MM5-RCM future conditions. In addition to the tabled information, examples illustrating the range of wind speed

and direction differences are provided for each seasonal analysis.

The second analysis question is, *"Do transport patterns occur with the same frequency across datasets?"* This question assesses similarity with regard to the relative frequency of clusters across datasets using the synthetic distributions described in Section 3.2.3, non-parametric tests of central tendency and distribution shape and comparison of 95% data intervals and data ranges. First, the distributions are summarized in two box and whisker displays (e.g. Figure 1 and Figure 2). Each distribution summary contains the median (red dot), the upper and lower quartiles (solid box), upper and lower inner fences (25th or 75th percentile ± 1.5 times the interquartile range), and outliers (black dots).

Figure 1 contains R1 (R1), R2 (R2) and base case MM5-RCM (GM) synthetic distributions for each of five seasonal clusters. Figure 2 contains similar information for MM5-RCM base case (B) and future (F) synthetic distributions. Table 3 provides a summary of non-parametric statistical test results as well as comparisons of 95% intervals (is there overlap?) and ranges. Agreement across relative frequency parent distributions but not location suggests there is a significant difference in the frequency of the cluster pattern across datasets. A lack of agreement in parent distribution implies a fundamental difference in spatial (extent) and/or temporal (persistence) structure that is used to define a "dominant" transport pattern. In this case, i.e., distributions are not similar; a lack of agreement regarding location (expected frequency) is not particularly informative. An overlap of 95% data intervals suggests that apparent differences in relative frequency central tendency may not be statistically significant. Distribution ranges are compared by ratio. Ratio values close to 1.0 suggested the distributions being compared have similar variability.

Finally, information contained in the tables and figures is combined to describe the relationship among the various datasets being compared. For example, Table 2 and Table 3 tell us that reanalysis (R1 and R2) clusters ranked 1 and 3 look similar and appear to be derived from similar relative frequency parent distributions with similar central tendency, but the relative frequency range for cluster 1 dataset R1 is only 60% of the R2 range. For the base case MM5-RCM and reanalysis comparison, Table 2 and Table 3 suggest that base case patterns appear similar to present day patterns, have similar internal structure and occur with similar relative frequencies. For the base case to future MM5-RCM comparison, Table 2 indicates substantial visual similarity. There is substantial overlap of the base case and future 95% data intervals (Table 3), but data ranges and tests of location and distribution suggest the presence of substantial differences that should be explored further.

Table 2. Best pattern matches for example NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (RCM - reanalysis) and RCM (future - base case).

R1 cluster Rank (A)	R2 cluster Rank (B)	Total wind speed bias (m/s) (C)	RCM Base case Cluster Rank (D)	Reanalysis Cluster Rank (E)	Total wind speed bias (m/s) (F)	RCM Future Cluster Rank (G)	RCM Base case Cluster Rank (H)	Total wind speed Bias (m/s) (I)
2	4	+0.10	3	R2-5	+0.86	1	3	-0.30
3	3	+0.15	2	R2-2	+1.61	2	2	-1.04
1	1	+0.77	1	R1-1	+1.21	3	1	+0.48
5	2	-0.15	4	R1-3	-0.08	4	4	+0.43
4	5	-1.45	5	R1-4	-1.18	5	5	-0.44

Table 3. Example statistical summaries for synthetic relative frequency distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap.

Cluster Rank	Reanalysis				RCM and Reanalysis				RCM Base Case and Future			
	95% Interval Overlap	Loc	Range Ratio\1	Dist	95% Interval Overlap (R1/R2)	Loc	Range Ratio \2 (R1/R2)	Dist	95% Interval Overlap	Loc	Range Ratio\3	Dist
1	Y	*	0.59	*	Y/Y	* (R1,R2)	1.08/0.64	*(R1,R2)	Y		1.18	
2	Y	*	1.00	*	Y/Y	*(R1, R2)	1.21/1.21	* (R1, R2)	Y	*	0.84	*
3	Y	*	1.20	*	Y/Y	*(R1, R2)	1.11/1.33	* (R1, R2)	Y	*	0.56	*
4	Y	*	1.21	*	Y/Y	* (R1,R2)	0.79/0.96	* (R1,R2)	Y		1.00	
5	Y	*	1.19	*	Y/Y	* (R1,R2)	0.82/0.98	* (R1,R2)	Y		1.11	

\1 reanalysis range ratio is computed as R1/R2
\2 MM5-RCM (RCM) and reanalysis range ratio is computed as reanalysis/RCM.
\3 MM5-RCM (RCM) base case and future range ratio is computed as base/future.

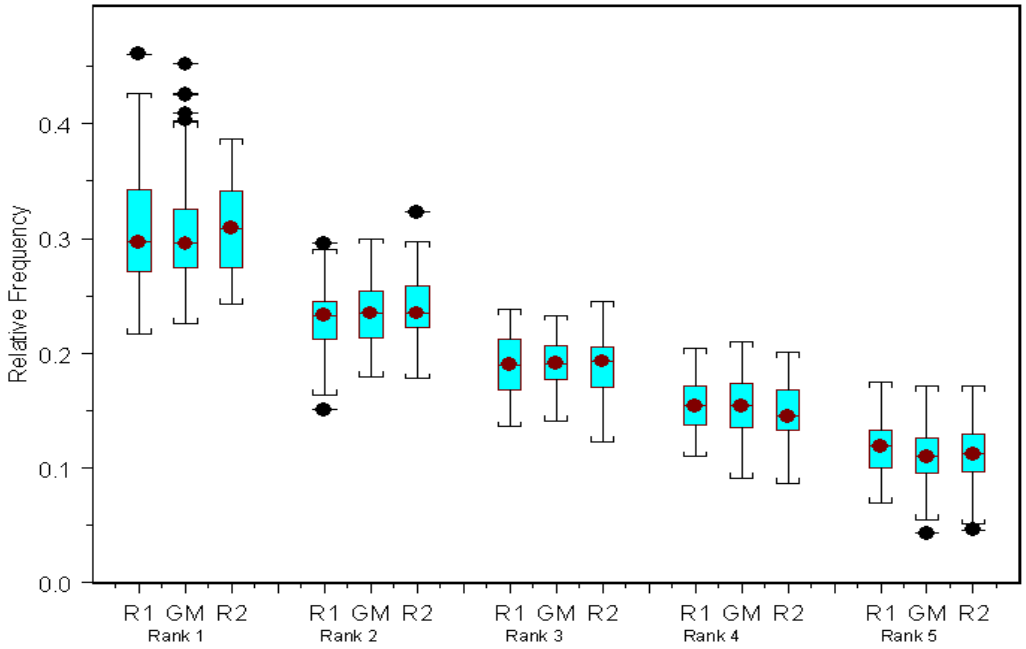


Figure 1. Box plot of relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank.

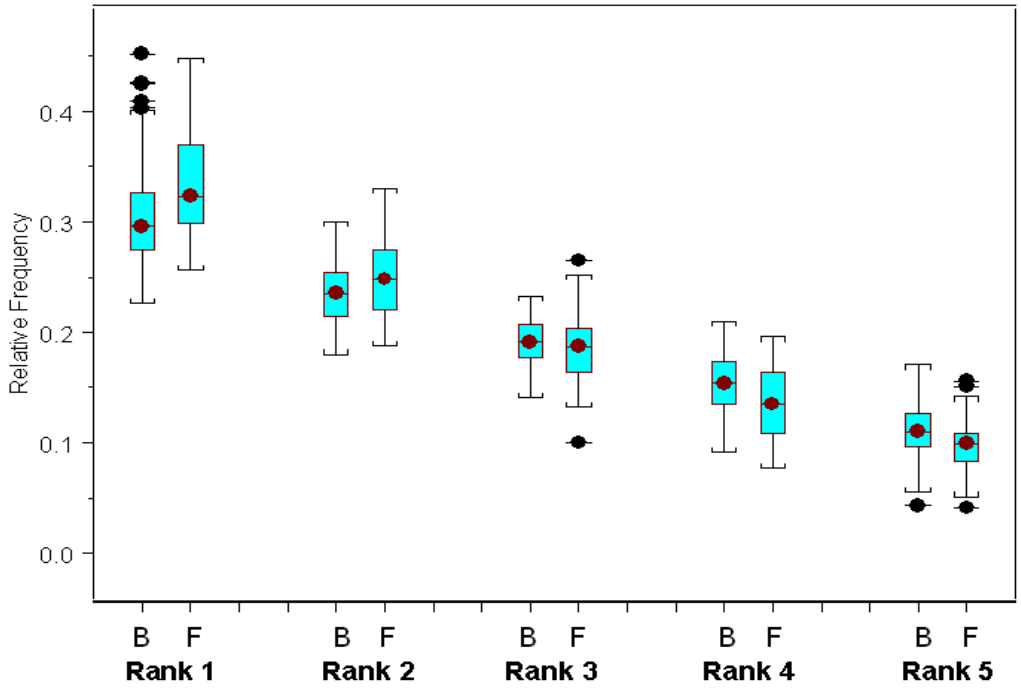


Figure 2. Box plot of relative frequency distributions for base case (B) and future (F) RCM grouped by relative frequency rank.

3.2.1 Spring

This analysis summarizes comparisons of 10 years (1985-1994) of spring season (March, April, May) reanalysis results (NCEP-R1 and AMIP-R2), a climatological decade of MM5-RCM base case (~2000) and a decade of MM5-RCM future (~2050) 700 mb *u* and *v* wind component data. All cluster ranks refer to relative frequency, where rank 1 is assigned to the most frequently occurring pattern.

"Do transport patterns look alike across datasets?"

Table 4. Best visual matches for spring season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - base case) or RCM (future - base case).

	R1 cluster Rank	R2 cluster Rank	Total wind speed bias (m/s)	RCM Base case Cluster Rank	Reanalysis Cluster Rank	Total wind speed bias (m/s)	RCM Future Cluster Rank	RCM Base case Cluster Rank	Total wind speed Bias (m/s)
(a)	2	4	+0.10	3	R2-5	+0.86	1	3	-0.30
(b)	3	3	+0.15	2	R2-2	+1.61	2	2	-1.04
(c)	1	1	+0.77	1	R1-1	+1.21	3	1	+0.48
(d)	5	2	-0.15	4	R1-3	-0.08	4	4	+0.43
(e)	4	5	-1.45	5	R1-4	-1.18	5	5	-0.44

There is generally good visual wind speed and wind direction agreement across R1 and R2 mean dominant cluster patterns, but cluster ranks suggest that in most cases these patterns occur with different relative frequency. Figure 3 is representative of Table 4 comparisons **a**, **c** and **d**. Figure 4 is representative of Table 4 comparisons **b** and **e**.

Base case MM5-RCM visual similarity to reanalysis data is good with regard to wind

direction, and there do not appear to be major differences with regard to transport feature identification or placement at 700 mb. This differs from Chapter 4 findings for the southeastern U.S. and should be explored further at a later time. Domain-wide wind speed bias tends to be quite large; especially for comparisons **b** and **c**. In these cases the reanalysis scenario places weaker transport off the west coast and generally higher wind speeds elsewhere, especially along the northern and southern boundaries of the domain. For cases **d** and **e**, weaker reanalysis transport spans the entire U.S. mid-section, producing negative as opposed to positive domain-wide bias. The range of base case MM5-RCM and reanalysis pattern agreement is illustrated by Figure 5 and Figure 6. Figure 5 illustrates Table 4 comparison **a**. All other comparison results (**b** through **e**) are similar to those shown in Figure 6.

Visual similarity between similarly ranked MM5-RCM base case and future scenarios is moderately good, with rank agreement in 3 of 5 cases. While domain-wide wind speed bias remains larger than that of the reanalysis (R1 to R2) comparison, it is generally lower than that for base case to reanalysis comparisons. The range of wind speed and direction future to base case pattern differences is illustrated in Figure 7 and Figure 8. Figure 7 is representative of Table 4 comparisons **b** and **c**. Figure 8 is representative of Table 4 comparisons **a**, **d** and **e**. There is some indication of significant changes in synoptic feature placement and behavior across these scenarios. For instance, the primary feature of comparison **a**, is an area of closed anti-cyclonic circulation centered over the southeastern U.S. This feature is present in both base case and future patterns, but Table 4 suggests that in the future, it may become the most dominant seasonal feature. The dominant feature of comparisons **d** and **e** is the position of the northern jet stream. During the spring season the jet retreats northward. Visual comparison

suggests that its future dominant position may be further north than current base case projections, but the frequency of these jet-focused patterns remains about the same.

Do transport patterns occur with the same frequency across datasets?

Figure 9 compares synthetic distributions of ranked relative frequency across reanalysis (R1 and R2) and MM5-RCM base case scenarios. Figure 10 compares MM5-RCM base case and future scenarios. Table 5 summarizes statistical tests of distribution and location and 95% data interval overlap for all distribution comparisons. In this analysis, differences in parent distribution can, among other things, indicate internal correlation structure (spatial pattern extent or pattern persistence) differences.

Table 5. Statistical summaries of spring season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap.

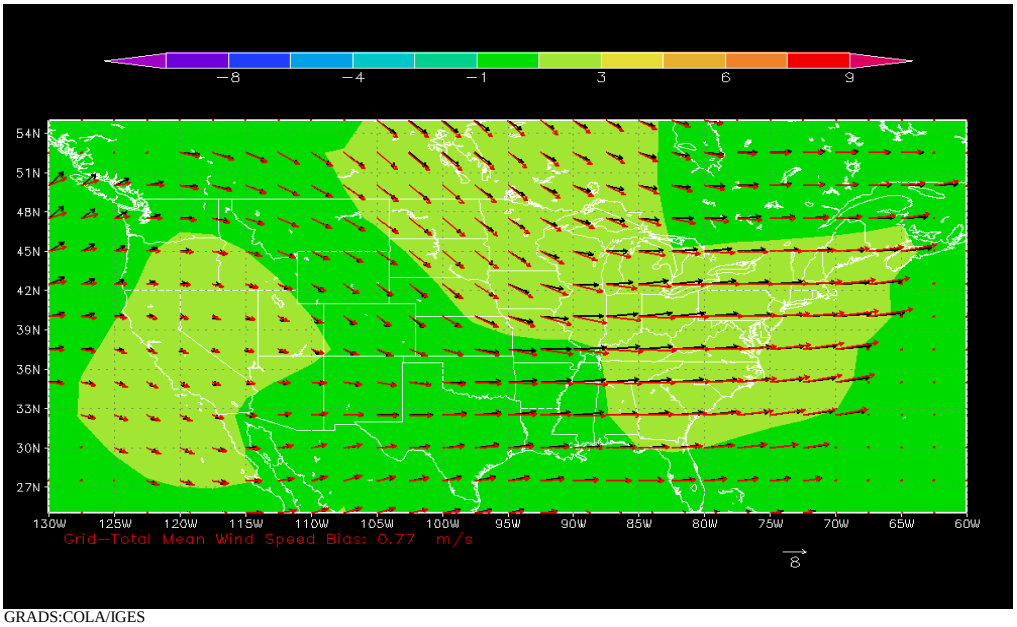
Cluster Rank	Reanalysis				RCM and Reanalysis				RCM Base Case and Future			
	95% Interval Overlap	Loc	Range Ratio\1	Dist	95% Interval Overlap (R1/R2)	Loc	Range Ratio \2 (R1/R2)	Dist	95% Interval Overlap	Loc	Range Ratio\3	Dist
1	Y	*	0.59	*	Y/Y	* (R1,R2)	1.08/0.64	*(R1,R2)	Y		1.18	
2	Y	*	1.00	*	Y/Y	*(R1, R2)	1.21/1.21	* (R1, R2)	Y	*	0.84	*
3	Y	*	1.20	*	Y/Y	*(R1, R2)	1.11/1.33	* (R1, R2)	Y	*	0.56	*
4	Y	*	1.21	*	Y/Y	* (R1,R2)	0.79/0.96	* (R1,R2)	Y		1.00	
5	Y	*	1.19	*	Y/Y	* (R1,R2)	0.82/0.98	* (R1,R2)	Y		1.11	

\1 reanalysis range ratio is computed as R1/R2
\2 MM5-RCM (RCM) and reanalysis range ratio is computed as reanalysis/RCM.
\3 MM5-RCM (RCM) base case and future range ratio is computed as base/future.

All similarly ranked reanalysis clusters appear to have been drawn from similar parent populations (distribution shape) having similar central tendency and there is substantial overlap of all 95% data intervals. Similarly, all MM5-RCM base case ranked clusters agree both in terms of parent distribution and central tendency and 95% data interval with the reanalysis datasets. There is little agreement among similarly ranked base case and future MM5-RCM clusters suggesting potentially significant frequency and correlation structure differences between the two sets of scenarios. This result is not surprising given the synoptic changes noted previously in the visual comparison.

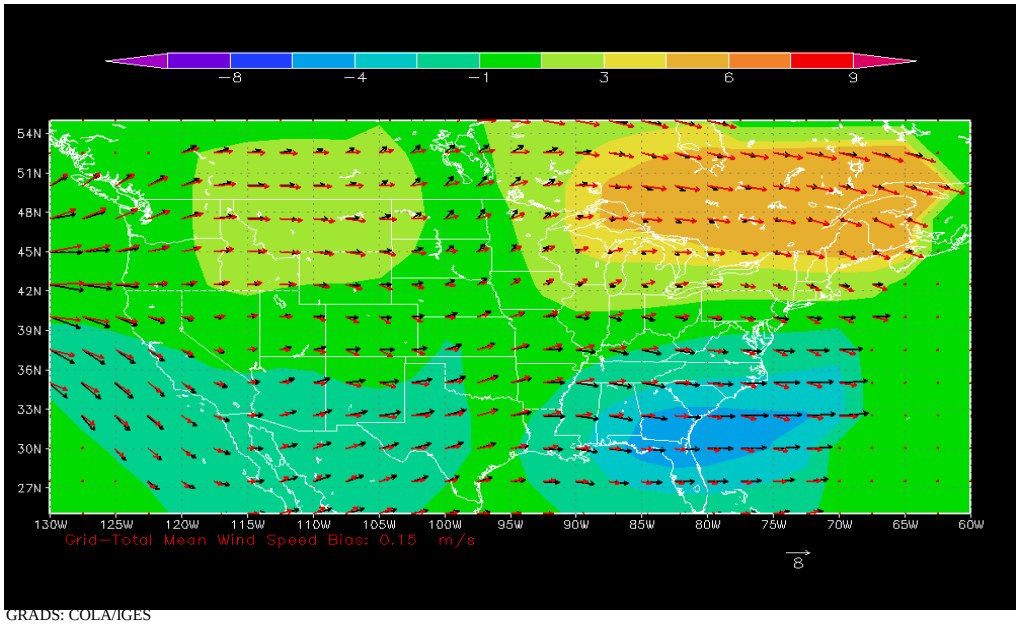
Conclusion

The reanalysis study indicates that there is good visual agreement and frequency of occurrence among present day dominant patterns of steering level atmospheric transport. Visual comparison of MM5-RCM base case clusters to reanalysis yield differences regarding wind speed that exceed the range defined by the reanalysis study. We conclude that the MM5-RCM base case represents present day 700 mb transport patterns reasonably well. There is moderate visual similarity between base case and future MM5-RCM scenarios, but measures of within-rank distribution suggest significant differences in relative frequency parent distribution and location in spite of substantial 95% data interval overlap. Subtle visual changes in the dominant seasonal position of the northern jet and persistence of dominant southeastern circulation suggest the potential for significant regional changes in projected spring season air quality.



GRADS:COLA/IGES

Figure 3. 700 mb NCEP-R1 and AMIP-R2 spring season transport pattern differences representative of Table 4 comparisons **a**, **c** and **d**. Black arrows are NCEP-R1 directions. Red arrows are AMIP-R2 directions. Velocity differences (m/s) are computed as (R2 - R1).



GRADS: COLA/IGES

Figure 4. 700 mb NCEP-R1 and AMIP-R2 Spring season transport pattern differences representative of Table 4 comparisons **b** and **e**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2 - R1).

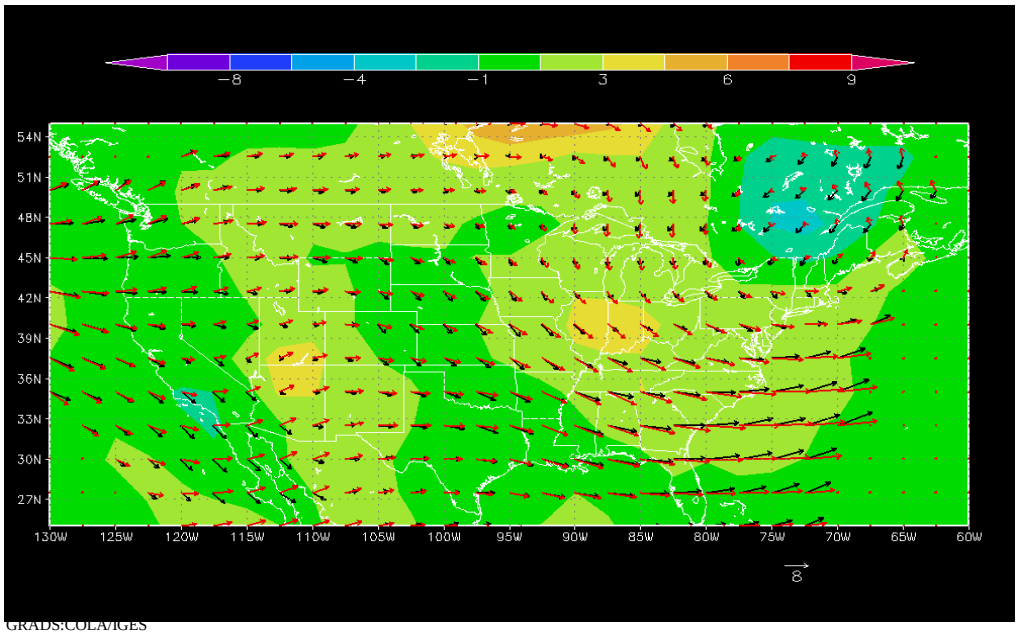


Figure 5. 700 mb reanalysis and base case RCM spring season transport pattern differences representative of Table 4 comparison **a**. Black arrows are base case RCM wind directions. Red arrows are reanalysis wind directions. Velocity difference (m/s) is computed as (reanalysis – base case).

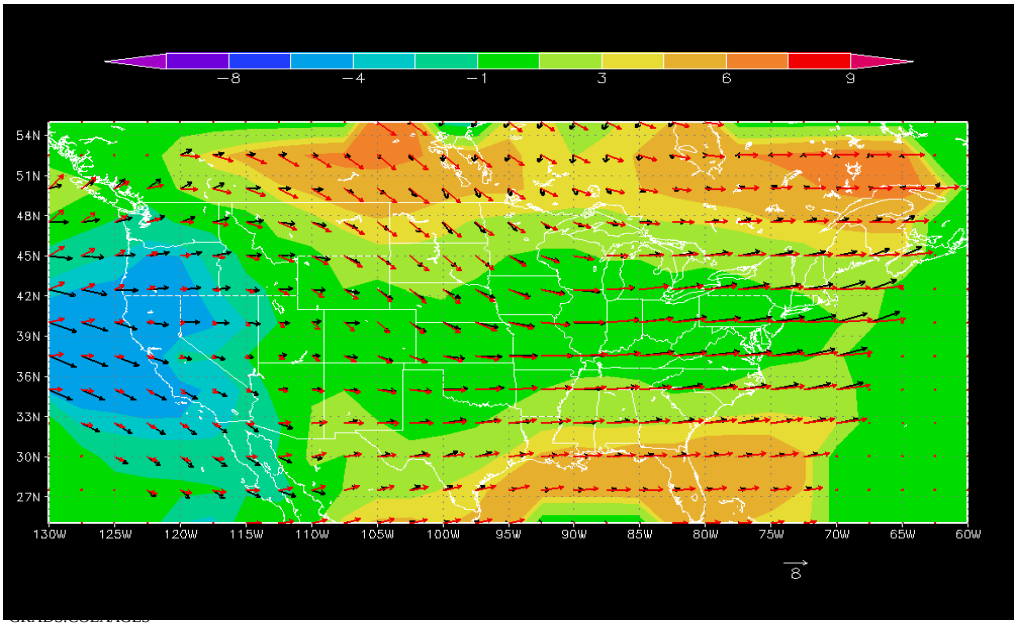


Figure 6. 700 mb reanalysis and base case RCM spring season transport pattern differences representative of Table 4 comparisons **b** through **e**. Black arrows are RCM wind directions. Red arrows are reanalysis wind directions. Velocity differences are computed as (reanalysis – base case).

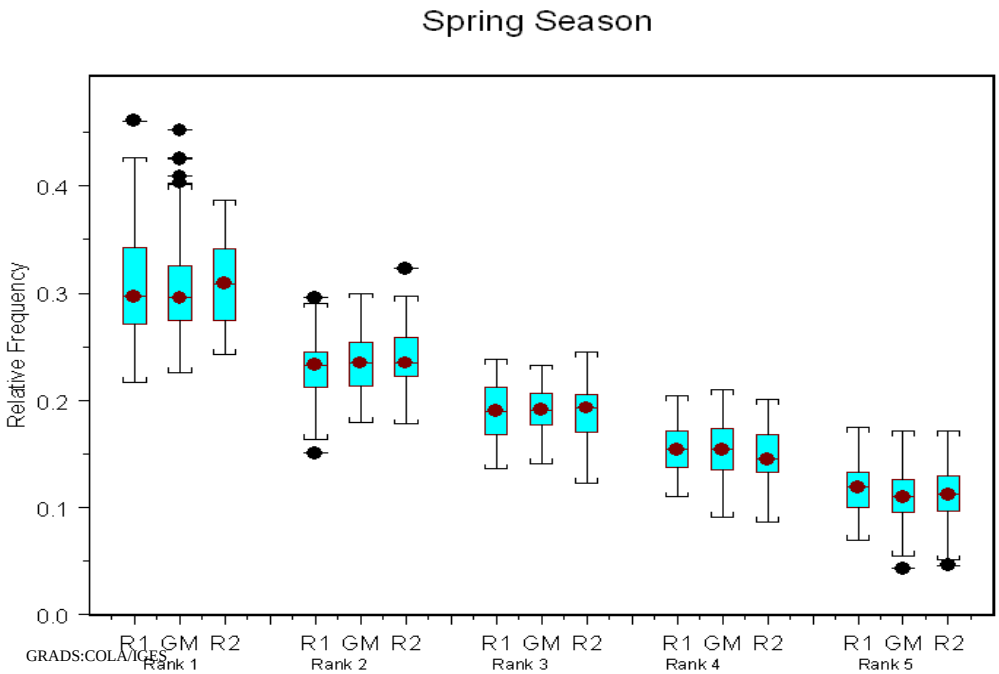


Figure 7. 700 mb base case and future RCM spring season transport pattern differences representative of Table 4 comparisons **b** and **c**. Black arrows are base case wind directions. Red arrows are future wind directions. Velocity difference (m/s) is computed as (future - base case).

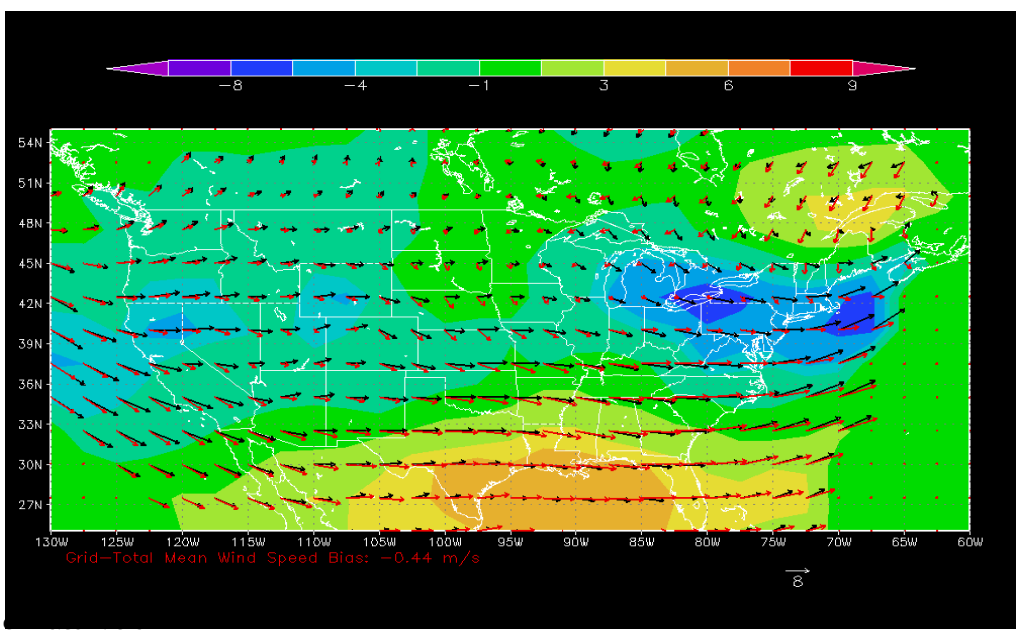


Figure 8. 700 mb base case and future RCM spring season transport pattern differences representative of Table 4 comparisons **a**, **d** and **e**. Black arrows are base case wind directions. Red arrows are future wind directions. Velocity difference (m/s) is computed as (future - base case).

Figure 9. Box plot of synthetic spring season relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank.

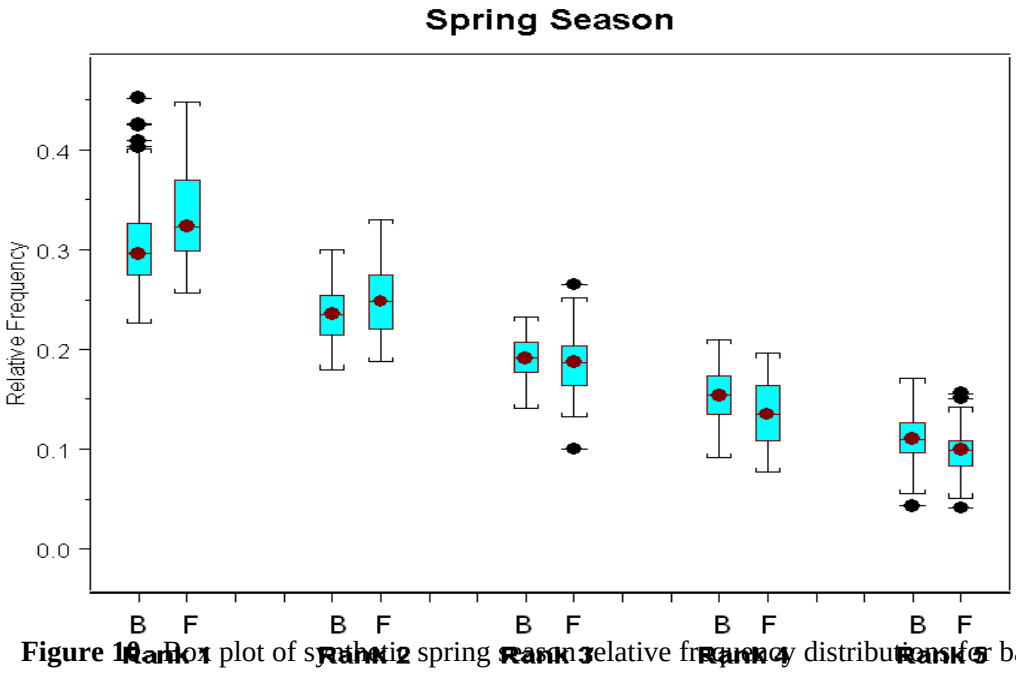


Figure 10. Box plot of synthetic spring season relative frequency distributions for base case (B) and future (F) RCM scenarios grouped by relative frequency rank.

3.2.2 Summer

This analysis summarizes comparisons of 10 years (1985-1994) of summer season (June, July, August) reanalysis results (NCEP-R1 and AMIP-R2), a climatological decade of MM5-RCM base case (~2000) and a decade of MM5-RCM future (~2050) 700mb *u* and *v* wind component data. All cluster ranks refer to relative frequency, where rank 1 is assigned to the most frequently occurring pattern.

"Do transport patterns look alike across datasets?"

Table 6. Best visual matches for summer season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - RCM) or RCM (future - base case).

	R1 cluster Rank	R2 cluster Rank	Total wind speed bias (m/s)	RCM Base case Cluster Rank	Reanalysis Cluster Rank	Total wind speed bias (m/s)	RCM Future Cluster Rank	RCM Base case Cluster Rank	Total wind speed Bias (m/s)
(a)	1	4	-0.27	1	R1-3	+0.30	1	1	-0.17
(b)	3	2	+0.31	4	R1-3	+0.92	2	2	+0.01
(c)	3	1	-0.32	2	R2-1	+0.62	3	3	-0.46
(d)	5	3	-0.34	3	R2-3	+0.44	4	4	-0.06
(e)	4	5	+0.20	5	R2-3	+0.15	5	5	-1.86

Visual comparison of NCEP-R1 and AMIP-R2 mean dominant cluster patterns shows excellent agreement in wind speed as well as direction in most cases, but agreement among patterns of similar rank is poor (Table 6). That is, patterns across reanalysis datasets look similar, but the frequency with which they occur is quite different. The range of R1 to R2 wind speed and direction differences is illustrated by Figure 11 and Figure 12. Figure 11 is representative of comparisons **a** through **d**. Figure 12 illustrates pattern differences for Table 6 compari-

son **e**.

Figure 14 is representative of difference patterns for all reanalysis and MM5-RCM comparisons listed in Table 6. Comparison of MM5-RCM base case to reanalysis patterns reveals relatively poor visual agreement with wind speed differences generally in excess of the reanalysis range. Base case MM5-RCM scenarios place a persistent trough off the U.S. West Coast (Figure 13). This feature is clearly present in every MM5-RCM base case pattern. This steering level MM5-RCM feature does not appear to link directly to the surface analysis presented in Chapter 4, which indicates good agreement between model and observed sea surface pressure and temperature. This apparent disagreement could reflect confounding surface local effects or model domain edge effects resulting in a region of greater scenario uncertainty requiring additional future study. In addition to the western trough, most MM5-RCM base case patterns indicate a persistent, "cut-off" (in most cases) subtropical high pressure area, i.e., closed anti-cyclonic circulation in the southeastern quadrant of the U.S. and weaker zonal (east-west) transport across the northern U.S. Reanalysis wind speeds throughout the South are weaker than the base case, except where the base case pattern is under the influence of the anticyclonic circulation mentioned previously. Finally, the northeastern quadrant of Figure 14 shows dramatic disagreement between reanalysis and RCM base case flow patterns. Chapter 4 discussion highlights a tendency for this particular RCM configuration to place the Bermuda high well northeast of its climatological position. The reflection of this RCM bias at 700 mb should be explored further in later analyses.

In stark contrast to these poor reanalysis and base case results, there is remarkable visual agreement between base case and future MM5-RCM scenarios. This is in general agreement

with the Chapter 4 surface analysis. Figure 15 is representative of wind speed and direction differences for Table 6 comparisons **a** through **d**. Figure 16 is illustrative of Table 6 comparison **e**. Review of this latter case shows that, with the exception of the west coast, weak transport dominates throughout the U.S. and few, if any, well-organized synoptic features can be distinguished (Figure 17).

Do transport patterns occur with the same frequency across datasets?

Figure 18 shows synthetic distribution comparisons of ranked relative frequency across reanalysis (R1 and R2) and MM5-RCM base case scenarios. Figure 19 compares MM5-RCM base case and future scenarios. Table 7 summarizes statistical tests of distribution and location and 95% data interval overlap for all distribution comparisons. Although there is substantial 95% data interval overlap, all dataset comparisons show only moderate distribution and location agreement. Table 7 indicates there is parent distribution as well as location agreement between base case and future clusters of frequency rank 2 and 3 and 5 suggesting most future change is predicted to occur for patterns ranked 1 and 4. Future variability in rank 5 cluster frequency, however, may be reduced by one-half.

Table 7. Statistical summaries of summer season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap.

Cluster Rank	Reanalysis				RCM and Reanalysis				RCM Base Case and Future			
	95% Interval Overlap	Loc	Range Ratio\1	Dist	95% Interval Overlap (R1/R2)	Loc	Range Ratio \2 (R1/R2)	Dist	95% Interval Overlap	Loc	Range Ratio\3	Dist
1	Y	*	1.36	*	Y/Y		0.70/0.95		Y		1.27	*
2	Y	*	1.25	*	Y/Y	*(R1)	0.67/0.84	* (R1)	Y	*	1.02	*
3	Y		1.31		Y/Y	*(R2)	0.66/0.86	* (R2)	Y	*	0.96	*
4	Y	*	1.01	*	Y/Y		1.33/1.34		Y		0.83	
5	Y		1.03		Y/Y	* (R1)	0.87/0.89	* (R1)	Y	*	1.91	

\1 reanalysis range ratio is computed as R1/R2
\2 MM5-RCM (RCM) and reanalysis range ratio is computed as reanalysis/RCM.
\3 MM5-RCM (RCM) base case and future range ratio is computed as base/future.

Conclusion

The reanalysis study indicates that there is good visual agreement regarding present day dominant patterns of 700 mb atmospheric transport, but only moderate agreement regarding the frequency with which these patterns occur within the 10-year time period. This is not particularly surprising since persistence is generally the rule during the summer season and transport regime transitions tend to be weak and difficult to predict.

Of far greater interest is the poor visual as well as frequency distribution similarity between base case MM5-RCM transport patterns and either of the reanalysis datasets. As noted previously, there are fundamental differences between reanalysis and base case transport scenarios that are quite likely to impact the CMAQ scenarios produced. There are clear indications of weakened 700 mb transport patterns (Table 6 domain-wide bias) and possible increased persistence of dominant patterns (Figure 19).

A comparison of MM5-RCM base case and future patterns suggests predominantly stable transport patterns through time. In most cases there is very good visual agreement and there is agreement regarding parent distribution and location in three of five cases. We conclude that the MM5-RCM model does not do a very good job reproducing present day 700 mb transport patterns. There is little agreement regarding pattern frequency as well, but our understanding of present-day frequencies as represented by our two reanalysis datasets is uncertain as well. Base case and future summer season MM5-RCM scenario patterns suggest little change over time.

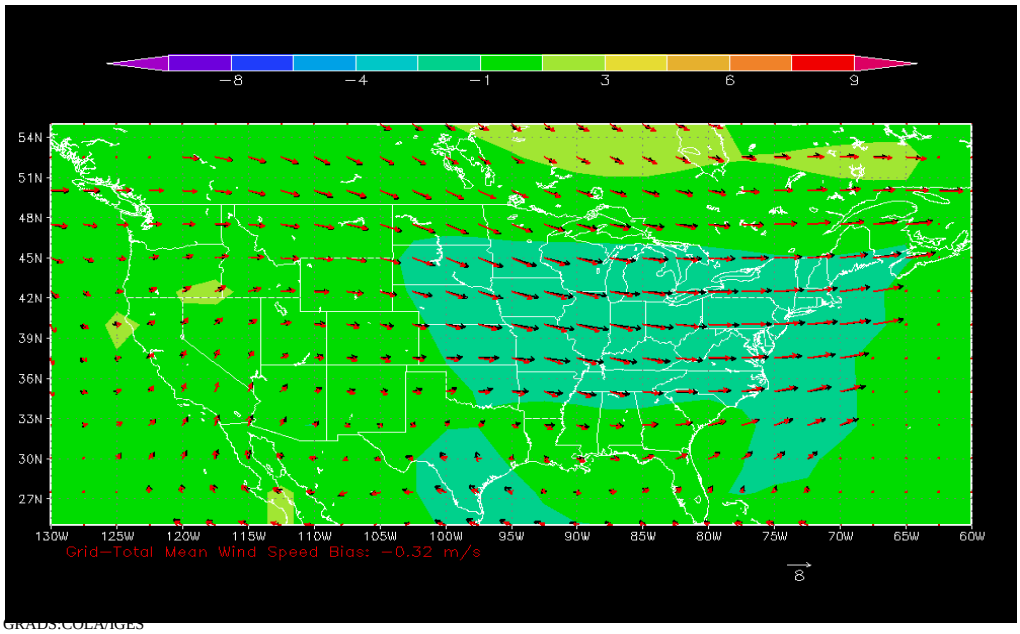


Figure 11. 700 mb NCEP-R1 and AMIP-R2 summer season transport pattern differences representative of Table 6 comparisons **a** through **d**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1).

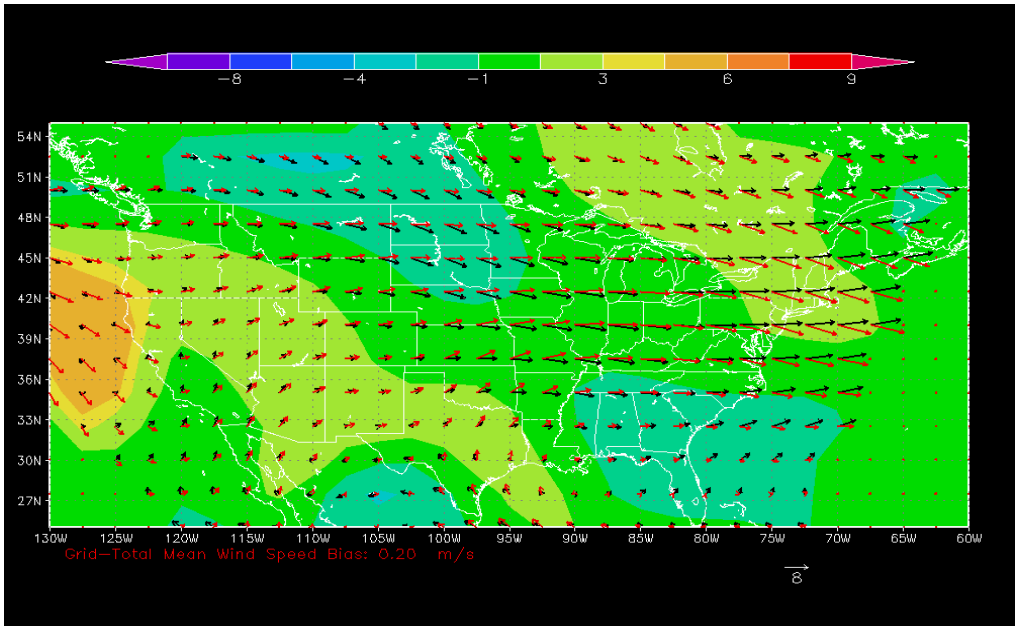


Figure 12. 700 mb NCEP-R1 and AMIP-R2 summer season transport pattern differences representative of Table 6 comparison **e**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1).

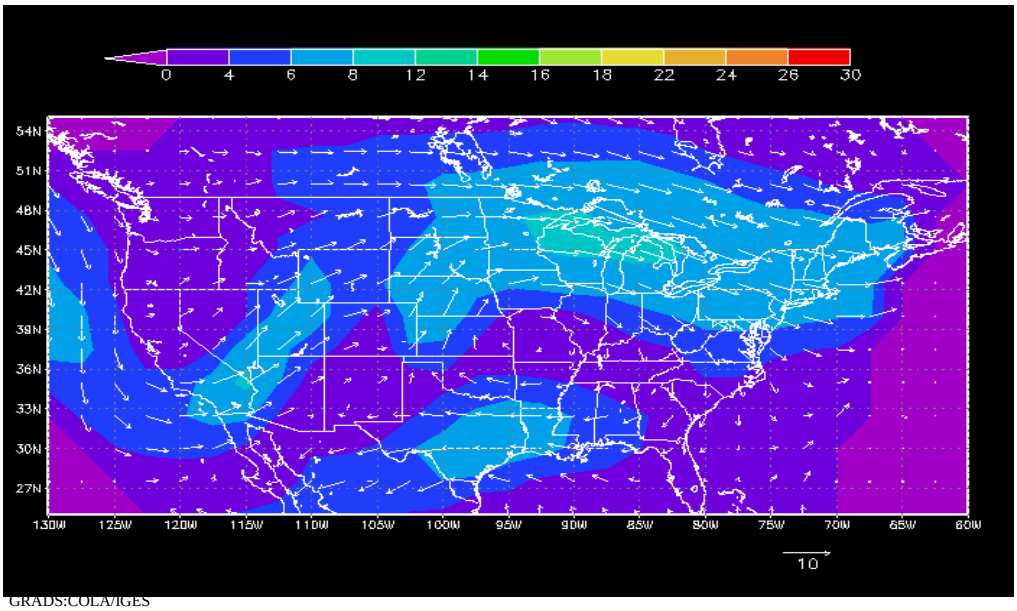


Figure 13. RCM base case summer season 700 mb wind speed (m/s) and direction illustrating the persistent western trough feature present in all base case average dominant transport patterns.

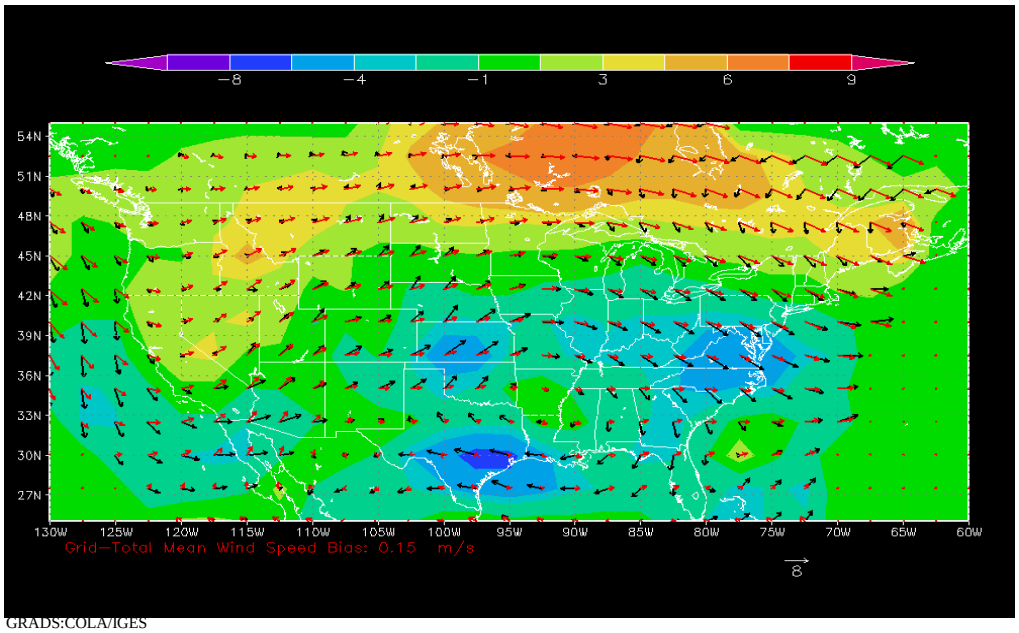
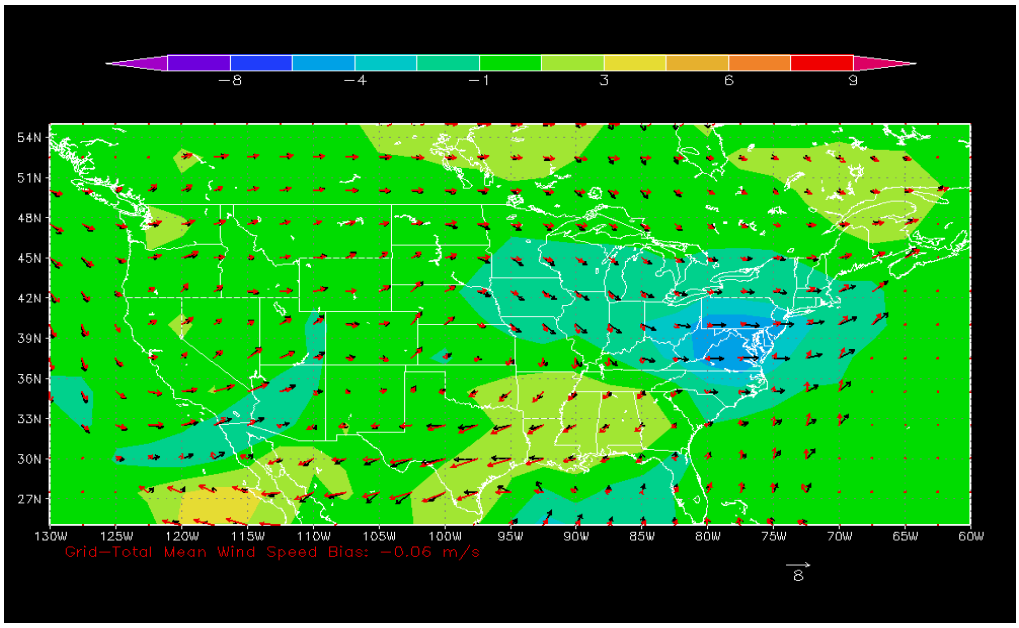


Figure 14. 700 mb RCM base case and reanalysis summer season transport pattern differences representative of all reanalysis and base case comparisons in Table 6. Note the location of the reanalysis anti-cyclone centered over Arkansas. Black arrows are RCM base case wind directions. Red arrows are reanalysis wind directions. Velocity differences (m/s) are computed as (reanalysis-base case).



GRADS:COLA/IGES

Figure 15. 700 mb RCM base case and future summer season transport pattern differences representative of Table 6 comparisons **a** through **d**. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future-base).

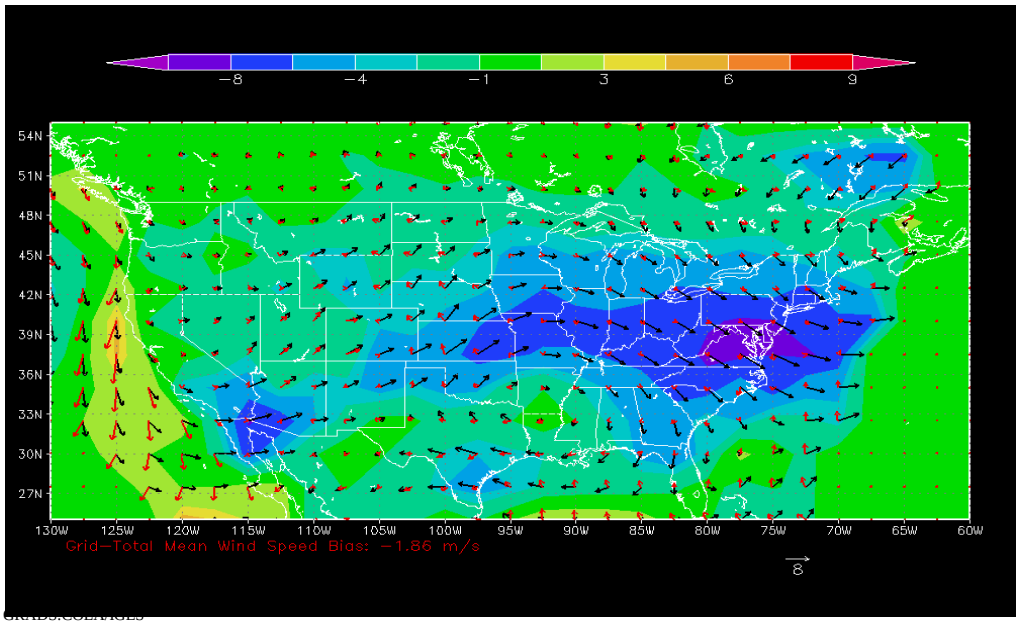
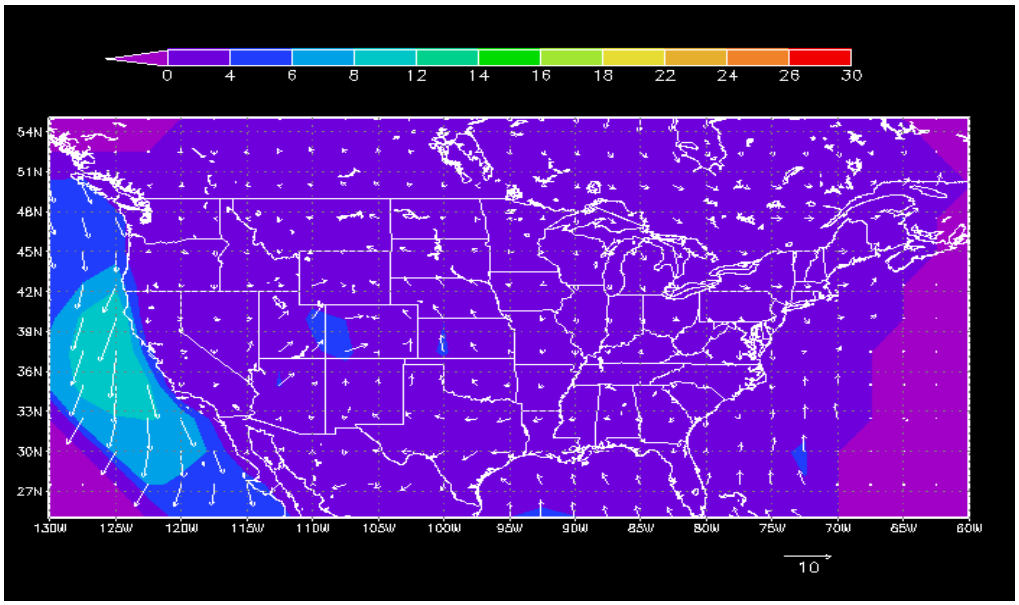


Figure 16. 700 mb RCM base case and future summer season pattern differences representative of Table 6 comparison **e**. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future- base case).



GRADS:COLA/IGES

Figure 17. Map of RCM future summer season wind speed (m/s) and direction illustrating weak transport and lack of active synoptic features.

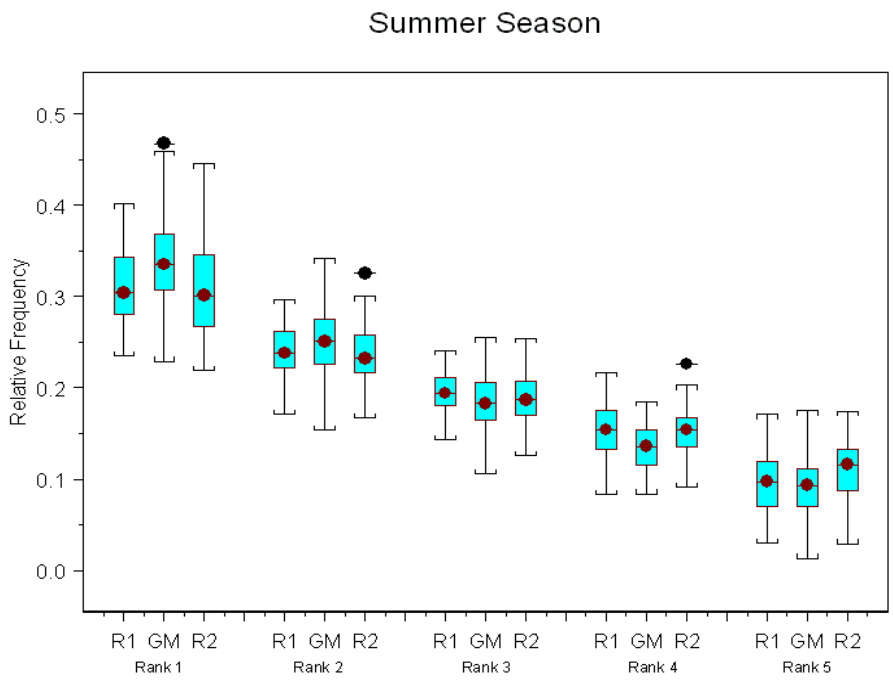


Figure 18. Box plot of synthetic summer season relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank.

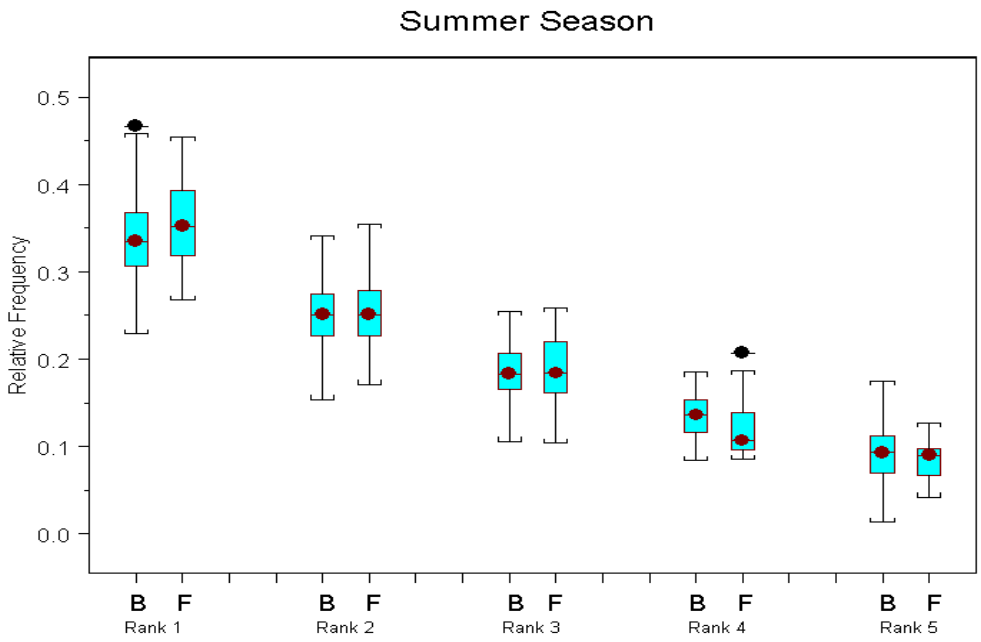


Figure 19. Box plot of synthetic summer season relative frequency distributions for base case (B) and future (F) RCM grouped by relative frequency rank.

3.2.3 Autumn

This analysis summarizes comparisons of 10 years (1985-1994) of autumn season (September, October, November) reanalysis results (NCEP-R1 and AMIP-R2), a climatological decade of MM5-RCM base case (~2000) and a decade of MM5-RCM future (~2050) 700mb *u* and *v* wind component data. All cluster ranks refer to relative frequency, where rank 1 is assigned to the most frequently occurring pattern.

"Do transport patterns look alike across datasets?"

Table 8. Best visual matches for autumn season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - RCM) or RCM (future - base case).

	R1 cluster Rank	R2 cluster Rank	Total wind speed bias (m/s)	RCM Base case Cluster Rank	Reanalysis Cluster Rank	Total wind speed bias (m/s)	RCM Future Cluster Rank	RCM Base case Cluster Rank	Total wind speed Bias (m/s)
(a)	1	3	+1.26	1	R2-2	-0.30	1	1	+0.05
(b)	4	1	-1.50	4	R2-5	+0.10	2	2	+0.29
(c)	2	5	-0.10	2	R1-2	+1.62	3	3	-0.05
(d)	5	2	-.019	3	R1-2	+0.39	4	1	+0.47
(e)	3	4	+0.19	5	R1-3	+2.07	5	1	-0.15

There is good visual agreement between Table 8 reanalysis comparisons **b**, **c** and **e** whose difference patterns are represented in Figure 20. Table 8 comparison **a** is also quite good, but R1 has a larger area of cyclonic circulation in the Southwest resulting in more of the west coast experiencing light southerly winds than R2. In the case of Table 8 comparison **d** (Figure 21), R1 indicates an area of cyclonic circulation in the southwest located further inland than does R2. This results in northerly flow off the west coast that gains a strong southerly component (base of

trough) over the Baja peninsula and then flows strongly from the southwest across the central plains. The range of reanalysis uncertainty is illustrated by Figure 20 and Figure 21.

The MM5-RCM and reanalysis comparison suggests that wind direction agreement across the western two-thirds of the model domain is reasonably good, but wind speed agreement is poor throughout. Figure 22 illustrates the best agreement for this comparison, but is illustrative of all the differences listed in Table 8. In most cases, the reanalysis indicates lower wind speeds off the west coast and higher wind speeds in the northwestern portions of the domain than does the MM5-RCM base case (Table 8 positive bias). There is generally good agreement regarding light winds in the southeast, but some disagreement regarding the location of the center of anti-cyclonic circulation. In most cases, wind speed and direction differences exceed those estimated across the two reanalysis datasets. Comparison to reanalysis patterns for the summer season produce more similar results suggesting a seasonal shift that may warrant additional analysis.

The comparison of base case and future MM5-RCM poses some interesting similarities and differences. There is excellent visual agreement in both wind speed and direction between the three most frequently identified patterns across datasets. In addition, the 4th and 5th most frequent patterns also look remarkably like the top ranked future pattern. This result suggests a "collapsing" of patterns in the future, indicating reduced variability and increasing persistence. Figure 23 is illustrative of all MM5-RCM base case to future comparisons described in Table 8.

Do transport patterns occur with the same frequency across datasets?

Figure 24 compares synthetic distributions of ranked relative frequency across reanalysis

(R1 and R2) and MM5-RCM base case scenarios. Figure 25 compares MM5-RCM base case and future scenarios. Table 9 summarizes statistical tests of distribution and location and 95% data interval overlap for all distribution comparisons.

Table 9. Statistical summaries of autumn season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap.

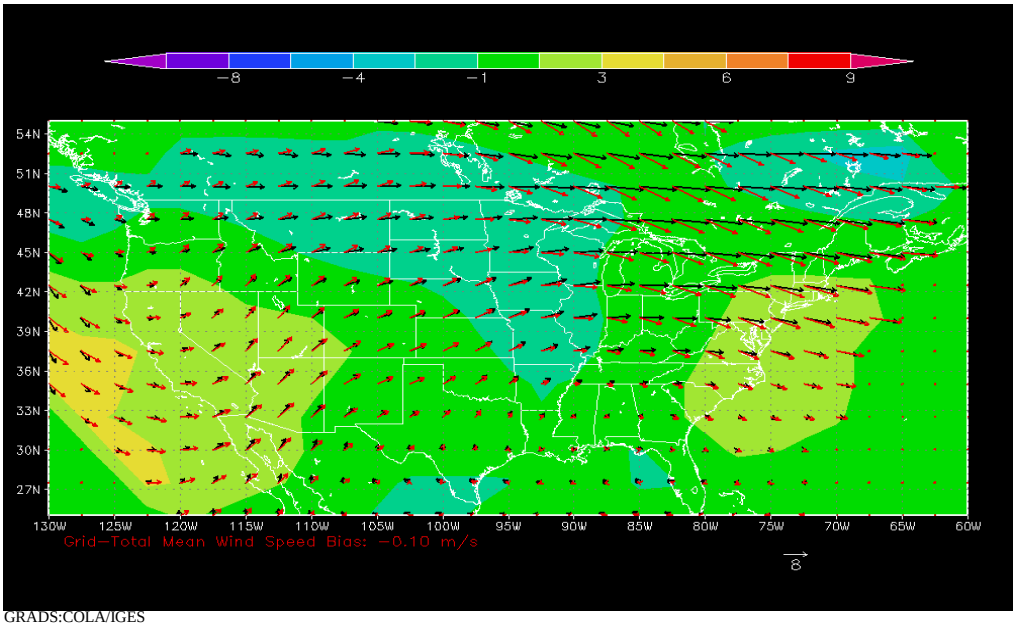
Clus-ter Rank	Reanalysis				RCM and Reanalysis				RCM Base Case and Future			
	95% Interval Overlap	Loc	Range Ratio\1	Dist	95% Interval Overlap (R1/R2)	Loc	Range Ratio \2 (R1/R2)	Dist	95% Interval Overlap	Loc	Range Ratio\3	Dist
1	Y	*	1.27	*	Y/Y	* (R2)	0.93/1.18	*(R2)	Y		0.93	
2	Y	*	1.47	*	Y/Y	*(R1, R2)	0.70/1.03	* (R1, R2)	Y		0.81	
3	Y	*	0.98	*	Y/Y	*(R1, R2)	0.90/0.88	* (R1, R2)	Y		1.01	
4	Y	*	1.22	*	Y/Y		0.78/0.95		Y		1.13	
5	Y	*	1.00	*	Y/Y	* (R1,R2)	0.89/0.89	* (R1,R2)	Y		1.60	

\1 reanalysis range ratio is computed as R1/R2
\2 MM5-RCM (RCM) and reanalysis range ratio is computed as reanalysis/RCM.
\3 MM5-RCM (RCM) base case and future range ratio is computed as base/future.

Table 9 and Figure 24 indicate that reanalysis data sets occur in a similar fashion in terms of central tendency and distribution shape. There is agreement between MM5-RCM base case and at least one similarly ranked reanalysis dataset in four of five clusters. Although there is some degree of 95% data interval overlap for all five patterns, no similarly ranked base case and future comparisons appear to be statistically similar. The lack of distribution agreement is immediately clear upon inspection of Figure 25. The increased persistence of base case patterns ranked 1 in the future, primarily deriving from the lower ranked base cases, is confirmed here and suggests fundamental changes in the temporal characteristics of transport patterns in the future.

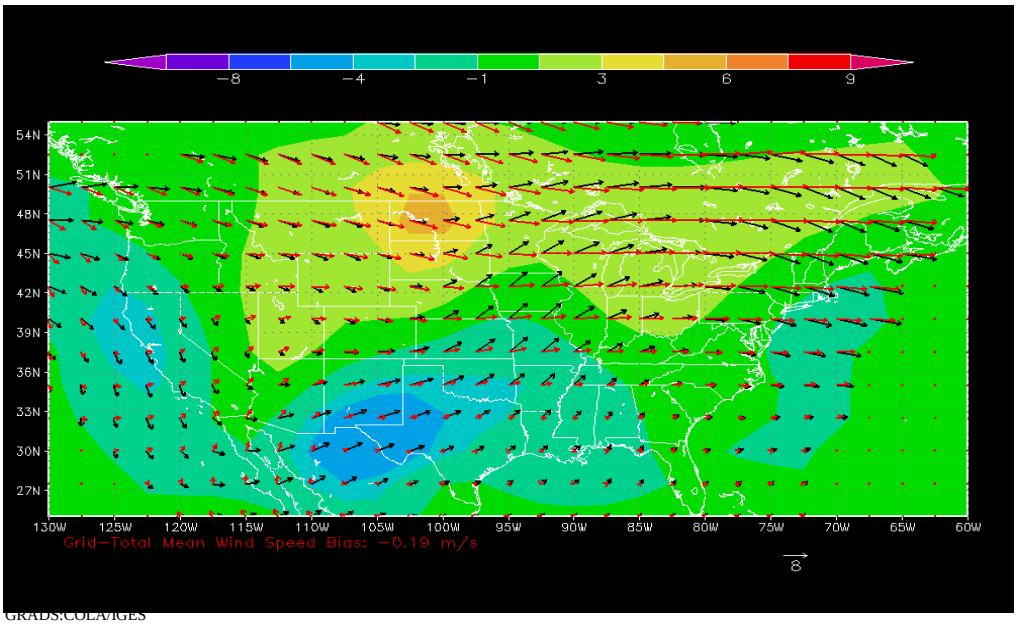
Conclusion

There is generally good visual and relative frequency reanalysis agreement, with the primary source of visual disagreement being in the location of the center of cyclonic circulation in the far west. Wind speed and direction differences are generally large for base case to reanalysis comparisons, but they appear to be within the uncertainty range defined by the R1 and R2. The comparison of base case to future MM5-RCM projections suggests little expected change in transport speed and direction, but a substantial change with respect to persistence of the most frequently expected base case transport patterns.



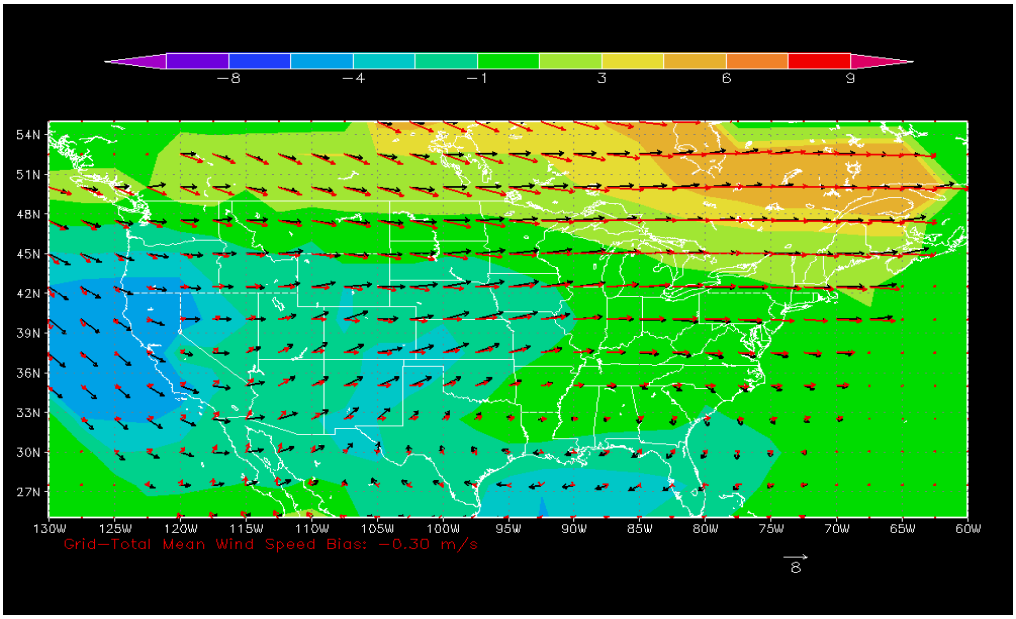
GRADS:COLA/IGES

Figure 20. 700 mb NCEP-R1 and AMIP-R2 autumn season transport pattern differences representative of Table 8 comparisons **b**, **c** and **e**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1).



GRADS:COLA/IGES

Figure 21. 700 mb NCEP-R1 and AMIP-R2 autumn season transport pattern differences representative of Table 8 comparison **d**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1).



GRADS:COLA/IGES

Figure 22. 700 mb RCM base case and reanalysis autumn season transport pattern differences representative of all Table 8 comparisons. Black arrows are RCM base case wind directions. Red arrows are reanalysis wind directions. Velocity differences (m/s) are computed as (reanalysis- base case).

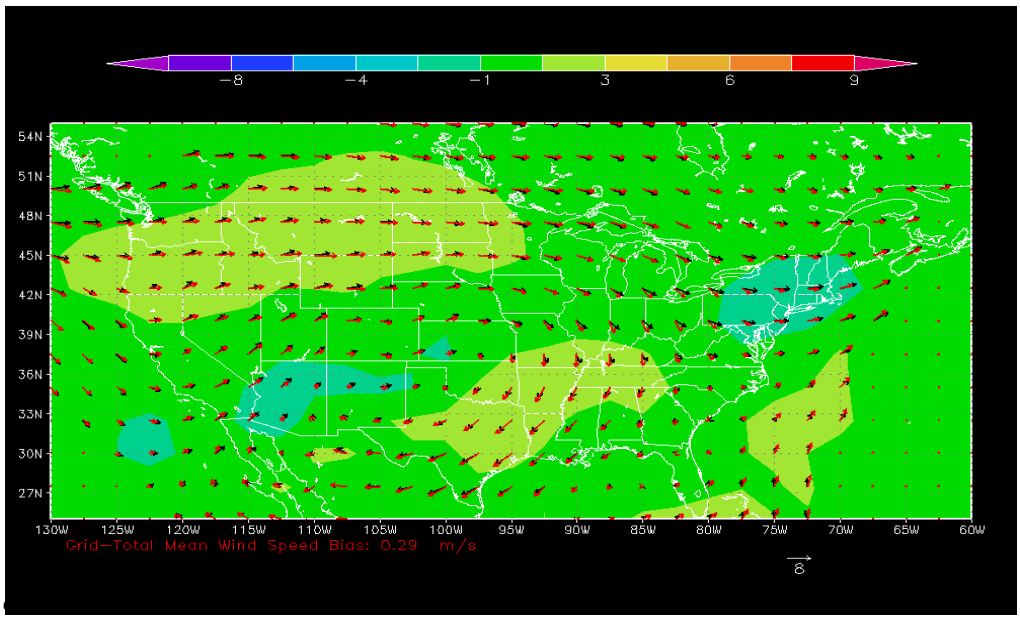


Figure 23. 700 mb RCM base case and future autumn season transport pattern differences representative of all Table 8 comparisons. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future-base case).

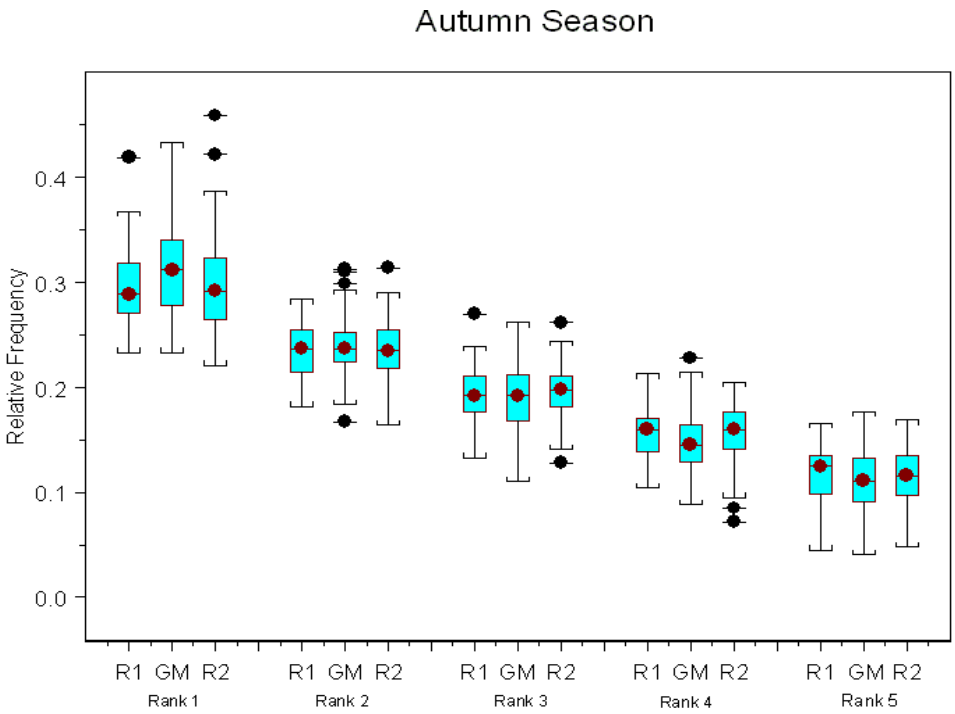


Figure 24. Box plot of synthetic autumn season relative frequency distributions for NCEP-R1(R1), base case RCM (GM), and AMIP-R2(R2) grouped by relative frequency rank.

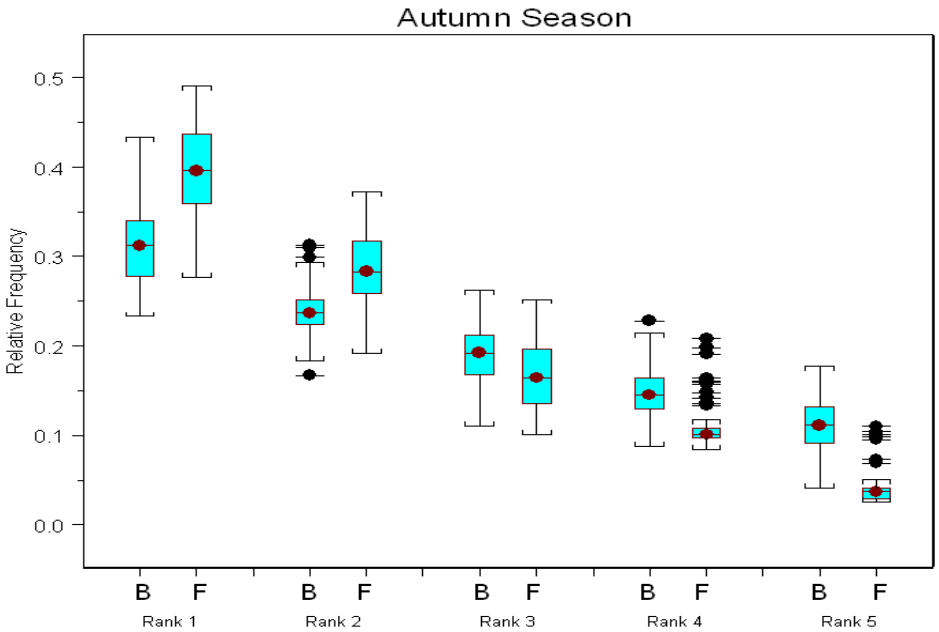


Figure 25. Box plot of synthetic autumn season relative frequency distributions for base case (B) and future (F) RCM scenarios grouped by relative frequency rank.

3.2.4 Winter

This analysis summarizes comparisons of 10 years (1985-1994) of winter season (December, January, February) reanalysis results (NCEP-R1 and AMIP-R2), a climatological decade of MM5-RCM base case (~2000) and a decade of MM5-RCM future (~2050) 700mb *u* and *v* wind component data. All cluster ranks refer to relative frequency, where rank 1 is assigned to the most frequently occurring pattern.

"Do transport patterns look alike across datasets?"

Table 10. Best visual matches for winter season NCEP-R1, AMIP-R2, base case RCM and future RCM. Total wind speed bias is computed as (R2-R1), (reanalysis - RCM) or RCM (future - base case).

	R1 cluster Rank	R2 cluster Rank	Total wind speed bias (m/s)	RCM Base case Cluster Rank	Reanalysis Cluster Rank	Total wind speed bias (m/s)	RCM Future Cluster Rank	RCM Base case Cluster Rank	Total wind speed Bias (m/s)
(a)	3	3	-0.14	1	R1-4	+1.81	1	2	+0.05
(b)	2	5	+0.37	2	R2-1	+1.29	2	5	-0.50
(c)	4	2	+0.16	4	R1-4	+1.78	3	3	+0.45
(d)	1	1	+.01	5	R1-1	+0.85	4	4	-0.17
(e)	5	4	+0.41	3	R1-5	+1.50	5	1	-4.13

There is good visual wind speed and wind direction agreement between Table 10 comparisons **a** through **d** whose difference patterns are represented in Figure 26. Although the domain-wide wind speed bias shown in Table 10 is not large, visual agreement for comparison **e** is not very good (Figure 27). This result, in combination with the lack of relative frequency rank agreement indicates substantial uncertainty in the frequency of winter season dominant transport patterns. The principal source of uncertainty at the steering level appears to be the location and

spatial extent of an area of low pressure located in the southwestern portion of the domain. Table 10 suggests R2 domain-wide mean dominant patterns wind speeds are higher overall than R1.

There is little visual similarity in either wind speed or direction across base case MM5-RCM and reanalysis datasets as illustrated by Figure 28. The details of other comparisons vary, but the level of agreement does not improve. The primary issues appear to be the treatment of cyclonic circulation at 700 mb and variability of on-shore Jet Stream flow along the West Coast. Only 2 of 5 MM5-RCM dominant patterns include an area of weak cyclonic circulation at 700 mb which is present in nearly every dominant reanalysis pattern. In the two MM5-RCM cases in which it is present, the center is placed farther inland than the reanalysis, near the Four-Corners area of the Southwest suggesting that the model mean Jet Stream path is located south of the reanalysis position. In addition, MM5-RCM patterns suggest less variability in the direction of on-shore flow from the Pacific. In most cases, the MM5-RCM base case patterns appear more similar to those in reanalysis R1, but contain lower domain-wide wind speeds (Table 10, large positive bias). The discussion in Chapter 4 further confirms that these differences in the west result in simulation of the major winter storm track too far south and reduced variability leading to persistent Canadian high-pressure over the central U.S.

Wind speed and direction pattern agreement is much better between base case and future than between base case and reanalysis with patterns ranked 3 and 4 suggesting relative stability through time. Table 10 comparisons **a** through **d** are represented by Figure 29. The poorest agreement reported in Table 10 is comparison **e**. Further consideration of these base case and future patterns suggest much weaker, less well organized future flow for these cases as compared

to base case patterns, in agreement with findings reported in Chapter 4.

Do transport patterns occur with the same frequency across datasets?

Figure 30 compares synthetic distributions of ranked relative frequency across reanalysis (R1 and R2) and MM5-RCM base case scenarios. Figure 31 compares MM5-RCM base case and future scenarios. Table 11 summarizes statistical tests of distribution, location and 95% data interval overlap for all distribution comparisons. There is good agreement across winter season reanalysis relative frequency distributions. There is moderate relative frequency agreement between the MM5-RCM base case and at least one reanalysis dataset, and poor agreement between base case and future MM5-RCM scenarios. This latter finding is in agreement with the surface wind flow analysis presented in Chapter 4.

Table 11. Statistical summaries of winter season synthetic distributions. An "*" indicates that the null hypothesis of similar central tendency (Loc) or distribution shape (Dist) cannot be rejected. "Y" indicates that the 95% data intervals overlap.

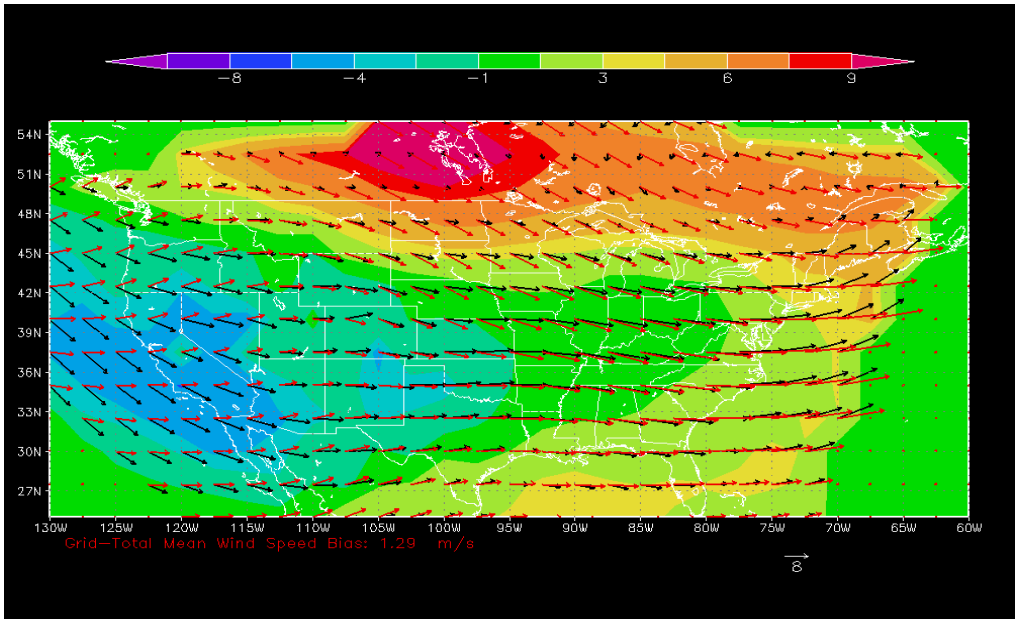
Cluster Rank	Reanalysis				RCM and Reanalysis				RCM Base Case and Future			
	95% Interval Overlap	Loc	Range Ratio\1	Dist	95% Interval Overlap (R1/R2)	Loc	Range Ratio \2 (R1/R2)	Dist	95% Interval Overlap	Loc	Range Ratio\3	Dist
1	Y	*	0.59	*	Y/Y	* (R1,R2)	1.05/1.35	*(R1,R2)	Y		0.76	
2	Y	*	1.00	*	Y/Y	*(R1)	0.81/0.77	* (R1)	Y	*	0.97	
3	Y	*	1.20	*	Y/Y	*(R1, R2)	0.76/0.79	* (R1, R2)	Y	*	1.22	*
4	Y	*	1.21	*	Y/Y	* (R1)	1.09/0.96	* (R1)	Y		1.01	
5	Y	*	1.19	*	Y/Y		1.45/1.28		Y		1.19	

\1 reanalysis range ratio is computed as R1/R2
\2 MM5-RCM (RCM) and reanalysis range ratio is computed as reanalysis/RCM.
\3 MM5-RCM (RCM) base case and future range ratio is computed as base/future.

Conclusion

Although the reanalysis comparison suggests that there is good visual agreement regarding the appearance of dominant present day winter 700 mb transport patterns there is a lack of agreement regarding strength of on-shore transport in the West. There is some disagreement regarding the frequency with which these patterns occur, i.e., the most visually similar are not of the same rank, but the rank location (central tendency) differences do not appear to be statistically significant and there is substantial overlap in the 95% data intervals.

Comparison of MM5-RCM base case to reanalysis datasets suggests weak agreement both in terms of visual appearance, variability and relative frequency primarily in the western U.S. and is associated with the position of the major winter storm track (Jet Stream). Base case and future MM5-RCM comparison suggest little future change with regard to patterns of transport is expected, particularly for Table 10 comparison **c** (rank 3). A notable exception are patterns with future relative frequency rank 5, for which both visual and relative frequency analysis suggest a marked departure from MM5-RCM base case conditions.



GRADS:COLA/IGES

Figure 26. 700 mb NCEP-R1 and AMIP-R2 winter season transport pattern differences representative of Table 10 comparisons **a** through **d**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1).

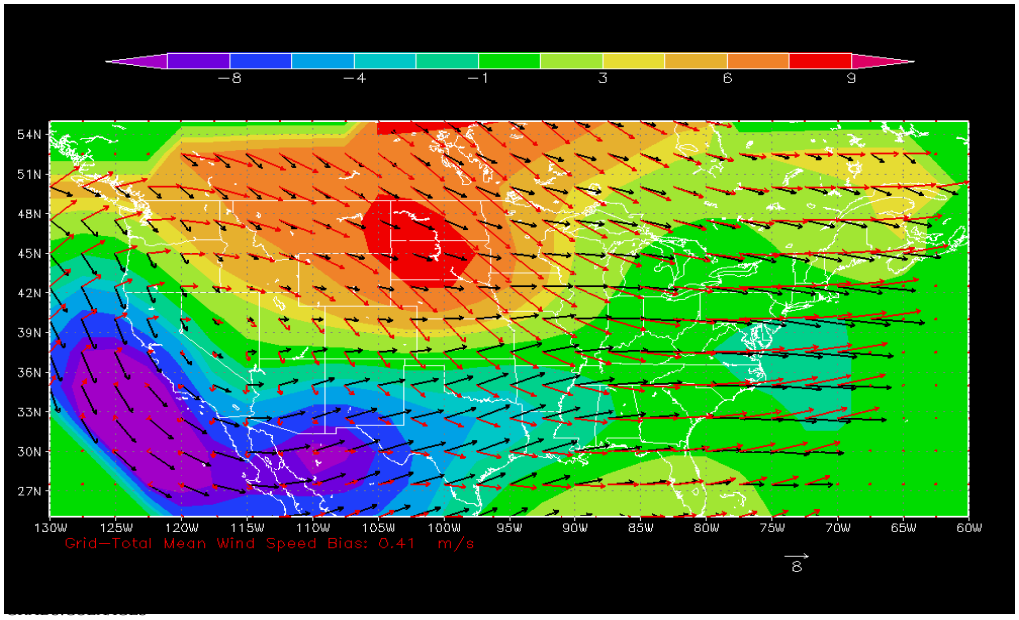


Figure 27. 700 mb NCEP-R1 and AMIP-R2 winter season transport pattern differences for Table 10 comparison **e**. Black arrows are NCEP-R1 wind directions. Red arrows are AMIP-R2 wind directions. Velocity differences (m/s) are computed as (R2-R1).

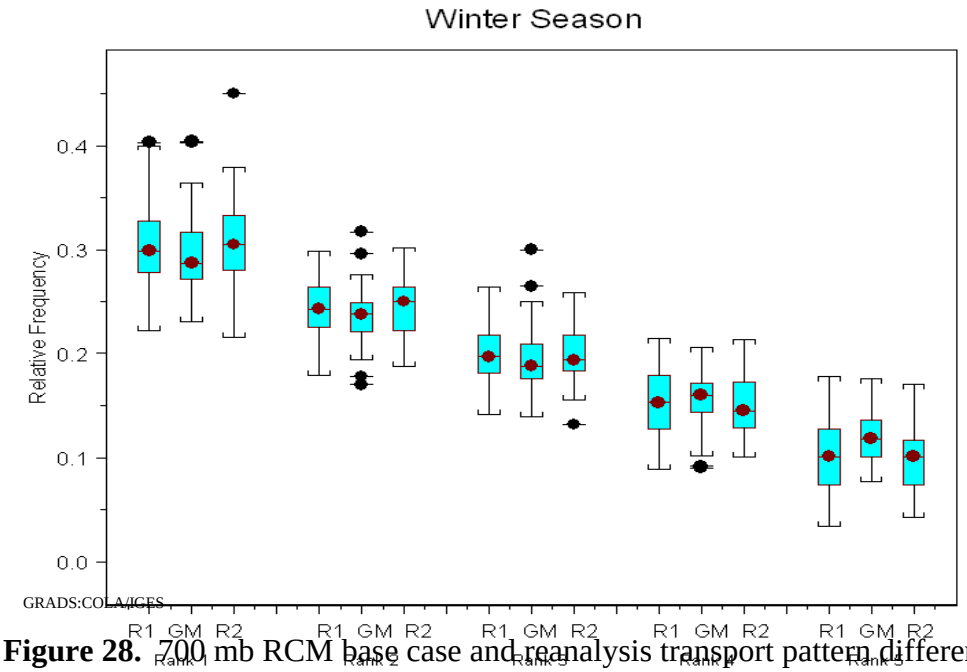


Figure 28. 700 mb RCM base case and reanalysis transport pattern differences representative of all Table 10 comparisons. Black arrows are RCM base case wind directions. Red arrows are reanalysis wind directions. Velocity differences (m/s) are computed as (reanalysis- base case).

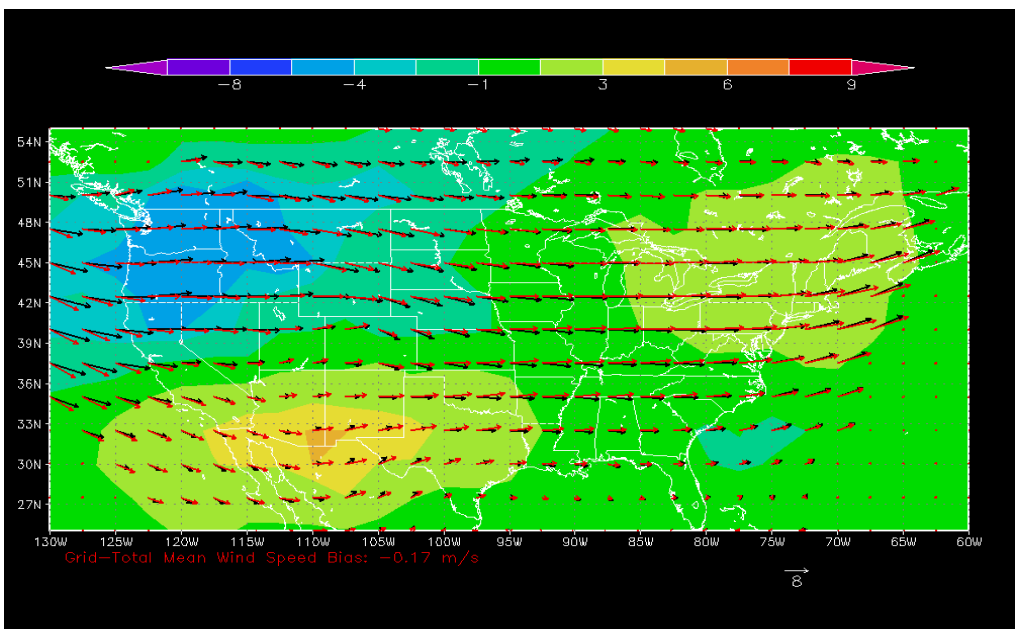


Figure 29. 700 mb RCM base case and future winter season transport pattern difference representative of Table 10 comparisons **a** through **d**. Black arrows are RCM base case wind directions. Red arrows are RCM future wind directions. Velocity differences (m/s) are computed as (future-base case).

Figure 30. Box plots of synthetic winter season relative frequency distributions for NCEP-R1 (R1), base case RCM (GM), and AMIP-R2 (R2) grouped by relative frequency rank.

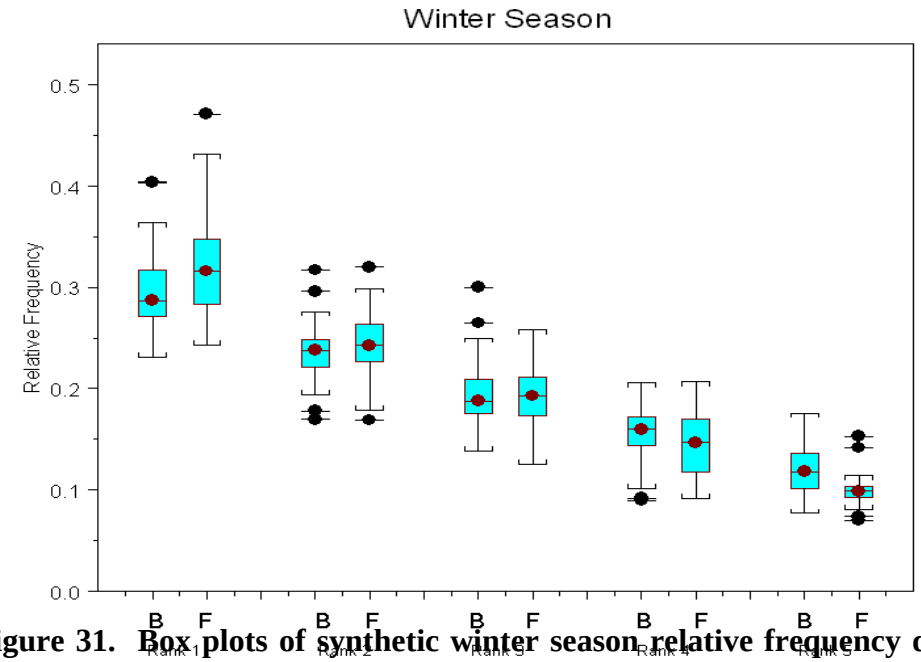


Figure 31. Box plots of synthetic winter season relative frequency distribution for base case (B) and future (F) RCM grouped relative frequency rank.

4.0 Distributional and Time Series Analysis

4.1 Research Methods and Datasets

4.1.1 Climate Observations

Several observational and analysis data sets were used to evaluate the base case GISS/MM5 (i.e., RCM) simulation. National Weather Service (NWS) Techniques Development Laboratory (TDL) observations, provided by the Data Support Section at the National Center for Atmospheric Research (NCAR-DSS) spanning 1992 through 2001 were used for the hourly surface meteorology. All surface observations in this data set have undergone a quality control process (TDL website, http://www.nws.noaa.gov/tdl) to remove anomalous records. The specific variables that were used in the evaluation are 2-m temperature, sea-level pressure and wind. The observations from 194 National Weather Service (NWS) sites were extracted for the comparison with the regional climate simulation; the observation locations are shown in Figure 32. **Wea**. Appendix A provides the city, state, latitude, and longitude for each of these NWS sites. Throughout the analyses, the sites may be referred to as climate reference sites, observation sites, or observation locations.

A daily United States precipitation analysis available through NOAA's Climate Prediction Center (CPC, ftp://ftpprd.ncep.noaa.gov/pub/precip/wd52ws/us_daily) was also used in the climate evaluation. The daily gridded precipitation values are derived from radar and rain gauge data (resolved on a 0.25° by 0.25° grid). The daily precipitation files were collected from 1996 to 2004. For each season, the daily precipitation was summed and then averaged (1996-2004) to

derive a seasonal mean precipitation. Additionally, seasonal variability (standard deviation) was calculated using the individual nine years of mean seasonal precipitation.

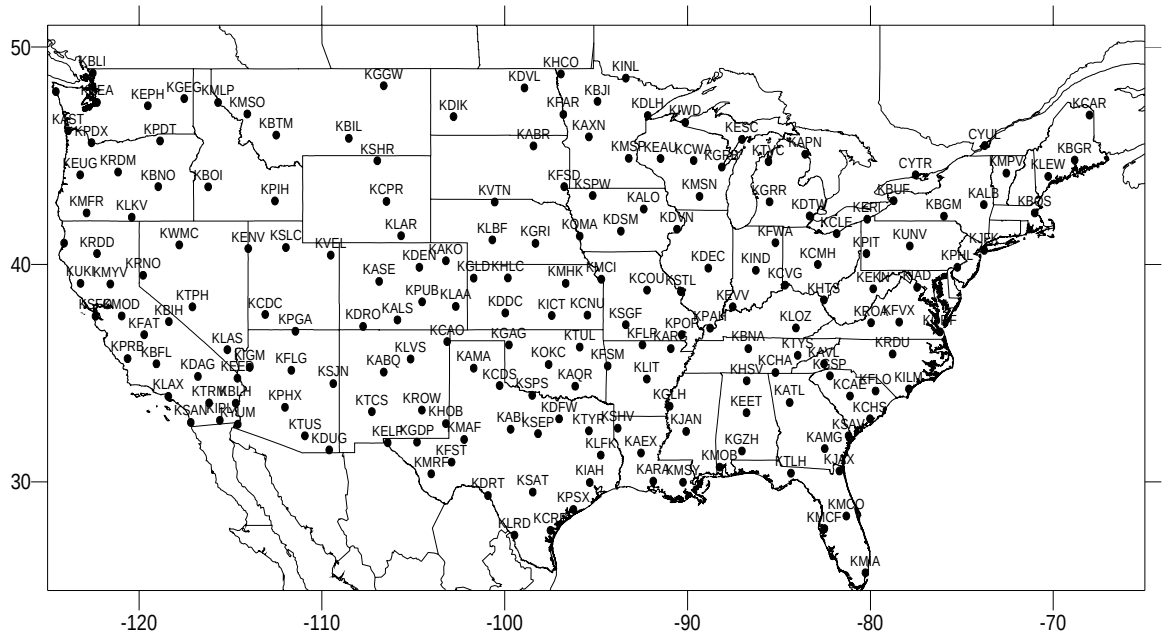


Figure 32. Weather observation sited used in the climate analysis

The two other data sets compared with the climate simulation are the NCEP-DOE AMIP-II[Kanamitsu et al., 2002] reanalysis monthly-mean sea-level pressure and a 14 km satellite derived sea-surface temperature analysis. The sea-level temperature analysis is derived from the Polar Operational Environmental Satellites (POES) platform (8 km) and blended with *in situ* data (http://www.class.noaa.gov). The sea-level pressure from the reanalysis is compared with the mean RCM pressure patterns. Sea-surface temperature differences are used to explain discrepancies in the synoptic patterns.

4.1.2 Time series extraction, Distributions Generation, and Mean Patterns

Time series are constructed from the 10-year base case RCM simulation to represent the meteorology (2-m temperature, hourly precipitation, first-level wind speed, and first i.e., lowest atmospheric level wind direction) at the 194 climate reference sites (Figure 32). This is done by choosing the model grid point closest to the observation site. The first-level wind speed and direction represent conditions approximately 35 meters above the surface. The observed variable time series are similarly constructed, and both RCM and observed time series are filtered into different temporal components (see Section 4.1.3).

Several analyses are performed with the observed and simulated time series. The primary analysis compares the seasonal mean sea-level pressure, temperature, and precipitation patterns of the model and observations. A second analysis examines the simulated and observed frequency distributions. This approach is appropriate because the climate simulation is a representation of the meteorology driven by a global climate model, and not a direct space-time replication.

The distributions are presented in several formats. The first format presents the 10 m wind distributions as a function of wind speed and direction. These distributions are used to compare observed and RCM wind roses in order to establish the way in which the climate simulation represents the surface wind-flow patterns over the United States. The synoptic-scale component of the near-surface wind flow is examined by determining and then plotting the three most dominant (frequent) modeled and observed wind directions for wind speeds greater than 2.5 $m \cdot s^{-1}$ (e.g.,Figure 34). The observed wind direction modes (i.e., the three wind directions occurring most frequently in the observations) are indicated by the red vectors. Blue vectors indicate the simulated wind direction modes. Each set of vectors have different lengths, which

correspond to the frequency of occurrence (i.e., longest vector indicates most dominant mode). Green vectors are plotted for instances where the observed and simulated wind modes have a difference of less than 40°. 40° was used as a cut-off because it is a typical mean wind direction error of meteorological models, and it is twice the bin size of the wind direction distribution. In these instances, the observed vector is plotted light green and the simulated vector dark green. A more in-depth discussion is provided for the winter and summer seasons since these are considered most important in terms of air quality.

Two other distributional methods are applied for this report; both are used to examine the filtered time series. The first uses the 10-year seasonally filtered time series separated into 10 annual time series. A mean and standard deviation for each hour of the year is calculated across the 10 annual time series. This is repeated for each variable, at each of the 194 weather observation sites, for the observations, the current, and the future climate simulations. In the base case climate simulation *versus* observation analysis, the 10-year mean hourly values from the simulation and observations are plotted, and the standard deviation is used to define variability bounds about the mean value. The net result is a plot of the mean seasonal times series over the 10-year period with the seasonal variability. The simulated and observed plots are overlaid to identify clearly the difference in the characteristics of the seasonal distribution. This same method is applied to the diurnal filtered time series, but rather than generating the mean and standard deviation for each hour of the year, it is done for each hour of the day. Diurnal distribution analyses are generated for the entire year, and for each season to provide insight on how well the climate simulation represents the amplitude and variability of the diurnal component of a particular variable (i.e., how well the temperature or wind speed range is simulated).

A second time series method is applied to examine the variability of the various filtered components of the time series. For the observed, base case simulation, and future simulation, the standard deviation is calculated for each temporal component. The difference of variability between the simulated and observed time series is calculated then normalized by the variability of the non-filtered time series to provide a percentage difference relative to the total variability. This information is plotted spatially for the different seasons (diurnal and synoptic components) and for the entire year (intraday, seasonal and interannual components). These variability plots are used to determine spatial patterns of the variability differences.

The full (non-filtered) 10-year observed and simulated time series (base case and future) are used to generate the mean seasonal value of each variable at each meteorological observation site. Throughout this document, the seasons are defined as winter (December, January, February), spring (March, April, May), summer (June, July, August), and fall (September, October, November). The annual means are also calculated and used to determine the standard deviation, which is a measure of inter-annual variability. Differences between the observations and the base case climate simulation as well between the base case and future climate simulations are also plotted and discussed.

4.1.3 Time Series Filter

Filtering is a valuable diagnostic method when applied to a variable time series that is largely composed of naturally occurring cycles (such as those representing physical processes). A filter can separate small-scale, random variability (“noise”) from the larger-scale phenomena,

as well as separate the components on strategic time scales that relate to physical processes [*Rao et al.*, 1997]. All observed and simulated climate time series were filtered using the Kolmogorov-Zurbenko (KZ) filter [*Hogrefe et al.*, 2000; *Rao et al.*, 1997] .

The user must first determine the frequencies that are of most interest, based on an understanding of the physical processes. For instance, if annual cycles are of interest in a collection of daily meteorological observations, the user would seek to remove all cycles which occur at a frequency greater than 1/365 (1 cycle per 365 days). Based on the identified frequency, the user determines the window length and the number of applications for the filtering algorithm. The KZ filter is then applied to separate the patterns that occur with a frequency greater than that of the cutoff level; the filtered result will then only contain components with cycles of period length 365 or longer. Repeated applications of the method can be used to isolate the portion of the time series associated with each time scale.

The major advantage of the KZ filter is its ease of use. The method does not require an advanced understanding of statistical methods, such as those used for parameter estimation in traditional time series analysis. It also does not require the assumptions that underlie many commonly used statistical methods and may not be appropriate in many applications. Compared with more traditional spectral analysis techniques, the method handles missing observations well.

The method is not based on a true stochastic model, and it does not have parameters to estimate, as do many traditional statistical methods. The user will not be able to make statistical inferences, i.e., form confidence intervals, or test hypotheses in the same way as for other time series procedures. The absence of stochastic assumptions makes it difficult to quantify the likely

amount or type of error associated with the component separation process. Finally, since the method is based on a simple linear filter, it is possible that any jumps or discontinuities in the time series may be oversmoothed, and their presence or importance overlooked by the user [*Eskridge et al.*, 1997]. In practice, the user should carefully examine and justify each filtered component based on the physical processes known to be at work.

In the present analysis, a KZ filter is applied to decompose the time series into intraday (variations with period of 12 hours or less), diurnal (1 days), synoptic (2-21 days), seasonal (~ 90 days) and interannual or baseline average (365 days) components. The various simulated and observed components of the time series are then compared to elucidate the strengths and weakness of the climate simulation in representing the climate.

4.1.4 Analysis Overview

One of the most important phases in the application of a climate model, especially when dealing with policy-related issues, is to asses whether the model is capable of replicating the spatial and temporal variability of the current climate. This is done by comparing the observed climate, using various observational datasets, to the simulated base case climate scenario.

The focus of this analysis is on climate and weather features that occur on scales of hundreds of kilometers and persist on the order of days or weeks. This criterion is clearly met by patterns of sea-level pressure. It is also generally accepted that other aspects of weather (e.g., dominant near-surface wind flow, temperature and regional precipitation) directly reflect the influence of these large-scale patterns of high and low pressure. Thus, this analysis begins with an examination of the simulated mean seasonal sea-level pressure patterns and the dominant

near-surface flow that is strongly influenced by these patterns.

Next, differences in the mean seasonal 2-m temperature are examined to provide a sense of the overall representativeness of the climate model in terms of near-surface temperature. Then, the filtered components of temperature are examined to provide more insight into processes that may contribute to differences in the temperature. Among these temporal scales are the diurnal, synoptic and seasonal variations in 2-m temperature. These temperature analyses will be connected to major conclusions from the pressure and wind flow analysis. The evaluation concludes by comparing climate-scale precipitation patterns and variability, and relating the results to conclusion drawn by the other analyses. The GISS/MM5 assessment concludes with a comparison of base case to future RCM scenario results.

4.2 Assessment of MM5 Regional Climate Model Base Case Climate

4.2.1 Sea-Level Pressure and Surface Wind Flow

The sea-level pressure and dependent near-surface flow patterns are critical aspects of the regional climate model simulation. Many qualities of the regional climate simulation including synoptic flow patterns, regional variability of the wind, boundary layer evolution, and stagnation events can be assessed by an in-depth examination of the sea-level pressure patterns and dependent near-surface wind flow. The discussions that follow consider observed (as represented by reanalysis data set) and modeled mean sea-level pressure and related dominant wind flow.

4.2.1.1 Winter

Sea-level pressure patterns are a standard meteorological proxy of synoptic-scale weather. Many of the standard surface-related variables are linked to the sea-level pressure pattern including wind, precipitation, temperature and moisture. Figure 33 is a comparison of the 10-year observed (reanalysis) and simulated mean wintertime (DJF) sea-level pressure pattern. The sea-level pressure over the higher elevation of the Rocky Mountains is not reliable, and so these areas are not considered in the sea-level pattern analysis. Important similarities and differences exist in the pressure pattern. In general, both model and observations indicate mean high pressure over the central and eastern United States, mean low pressure off the northwest and northeast United States, mean high-pressure off the southern California coast, and mean lower pressure around Baja California. The major difference, which has a dramatic influence on many of the other meteorological fields, is the configuration of the mean continental high-pressure area located in the central United States. The RCM simulation has a clearly defined zone of higher sea-level pressure from west-central Canada, down through the central United States. This is a typical track for wintertime Arctic high-pressure systems. Areas to the south and east of the zone of higher pressure will experience relatively cold temperatures. In contrast, the reanalysis field does not indicate a well-defined high-pressure zone in central Canada and the United States. A general area of relatively high pressure is depicted in the mean reanalysis pattern, but it is diffused over a large part of the central and eastern United States. Several important details can be deduced from these pattern differences. First, the well-defined high-pressure track in the model output indicates that a persistent Arctic high-pressure cell slides

down the eastern edge of the Rocky Mountains and takes residence in the central United States.

Conversely, the climate

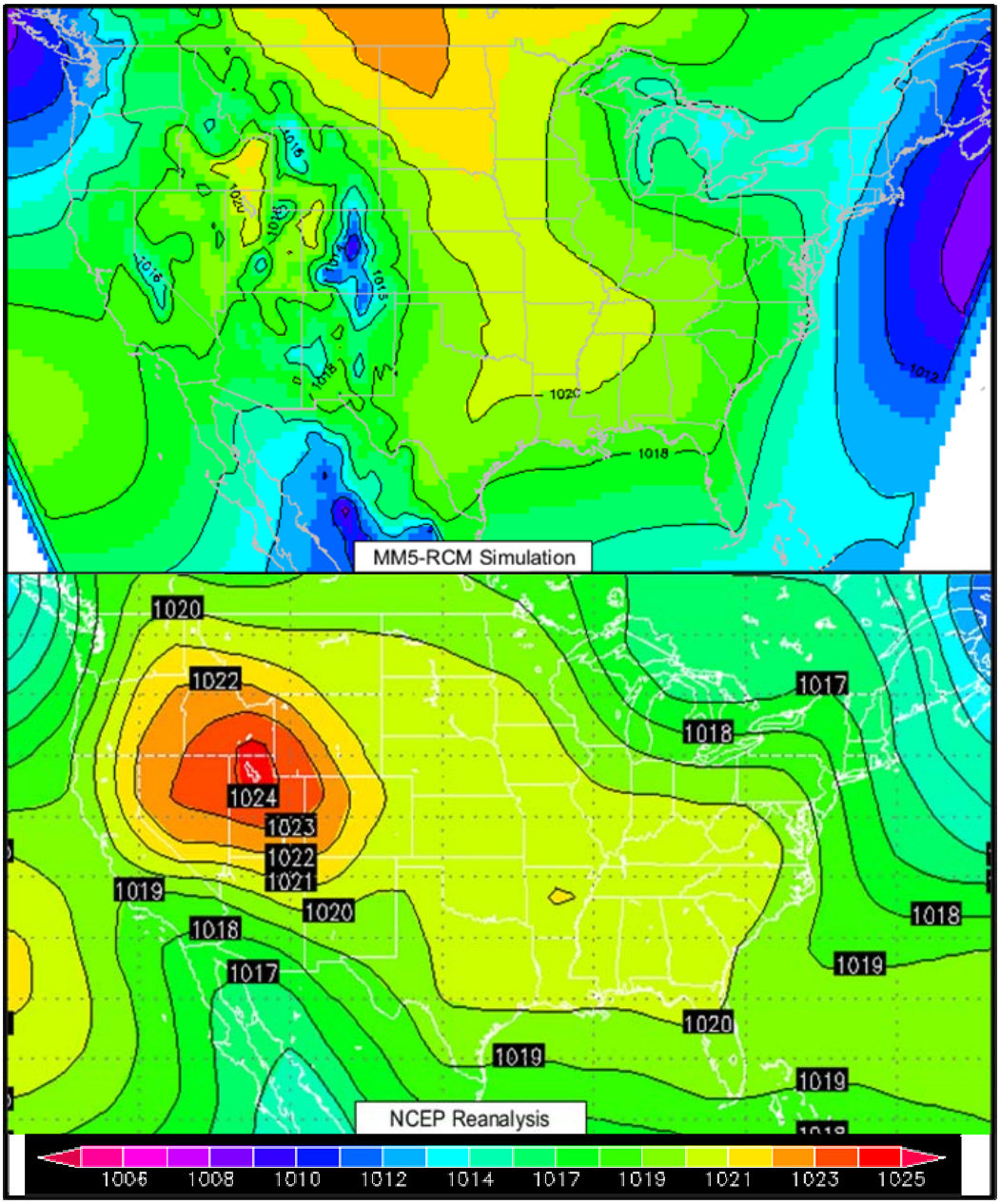


Figure 33. The 10-year mean sea-level pressure pattern in the wintertime (DJF). The top panels is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale. Note that the sea-level pressure in the western U.S. is not reliable.

analysis suggests that these high-pressure systems, in reality, are either less frequent, weaker, or there is more variability in the track as compared to the climate simulation. Another significant feature that is not well represented by the model scenario is the wintertime Canadian Clipper low-pressure system that swings on a southeastward arc from the Canadian Rockies to the northeast United States. This feature would tend to lower the mean sea-level pressure along its climatological path. This arc of lower pressure can be clearly seen in the reanalysis pressure field. However, the simulated pressure field does not indicate such a feature across the northern Plains and Midwest. These major synoptic pattern differences indicate that the RCM suppresses the wintertime storm track across the eastern United States. An analysis of 700 mb wind flow patterns in Section 3.2 also indicates major flow differences aloft during the winter seasons that are consistent with these differences in sea-level pressure. Specifically, the 700 mb flow supports the idea that the RCM simulates the major winter storm track (Jet Stream) too far south, and lacks variability in the synoptic weather pattern, which leads to the persistent Canadian high-pressure over the central United States.

Another significant difference in the sea-level pressure pattern is the location of the low-pressure zone off the northeast United States coast. The RCM simulates the Atlantic low-pressure east of Cape Cod and south of Nova Scotia while the reanalysis data indicate this low should be further north and east. Another difference is the low-pressure trough that extends to the south and west of the low-pressure minimum off the East Coast. The reanalysis data indicate the mean axis of this feature lies just off the East Coast while the RCM has this feature further offshore. The evolution of the western Atlantic coastal low along the East Coast is a critical component of the climate. These mean pressure pattern features are the climate footprint of the

cyclone's evolution. Thus, the differences indicate the RCM is not properly representing the evolution of these coastal cyclones off the East Coast. With these exceptions, the wintertime RCM pattern is comparable to the climate analysis over much of the United States.

Next, consider how these pressure pattern similarities and differences are reflected in the dominant near-surface wind flow. The observed and simulated dominant near-surface wind directions for the winter season are plotted in Figure 34. In the northeastern United States, the principal simulated wind flow (blue and dark green where agreement exists) is northwesterly (arrows indicate the direction from which the wind is blowing). Although the observed wind has a major northwest wind component, the climate model does not simulate a frequent southerly wind component at many of the sites. Nonetheless, there is generally good agreement between the observed and simulated wind directions and, in some cases, regional variability. This is in agreement with the comparative sea-level pressure pattern over the Great Lakes and Northeast (Figure 33). The winter wind-flow comparison farther south and east of the Appalachian Mountains, from the Mid-Atlantic to areas of the Southeastern United States, indicates the climate model does not replicate a frequent synoptic flow pattern. An observed southerly wind component is consistent over the area; the climate model does not simulate this frequent flow direction properly. One plausible explanation is the winter climate of the Mid-Atlantic and the Southeast is largely influenced by the warm Gulf Stream off the East Coast. The regional climate model uses sea-surface temperature information from the courser coupled ocean-atmosphere global climate model. It is well known that the warm Gulf Stream influences many aspects of the evolution of wintertime low-pressure systems [Dirks et al., 1988; Hirsch et al., 2001], and if major errors exist

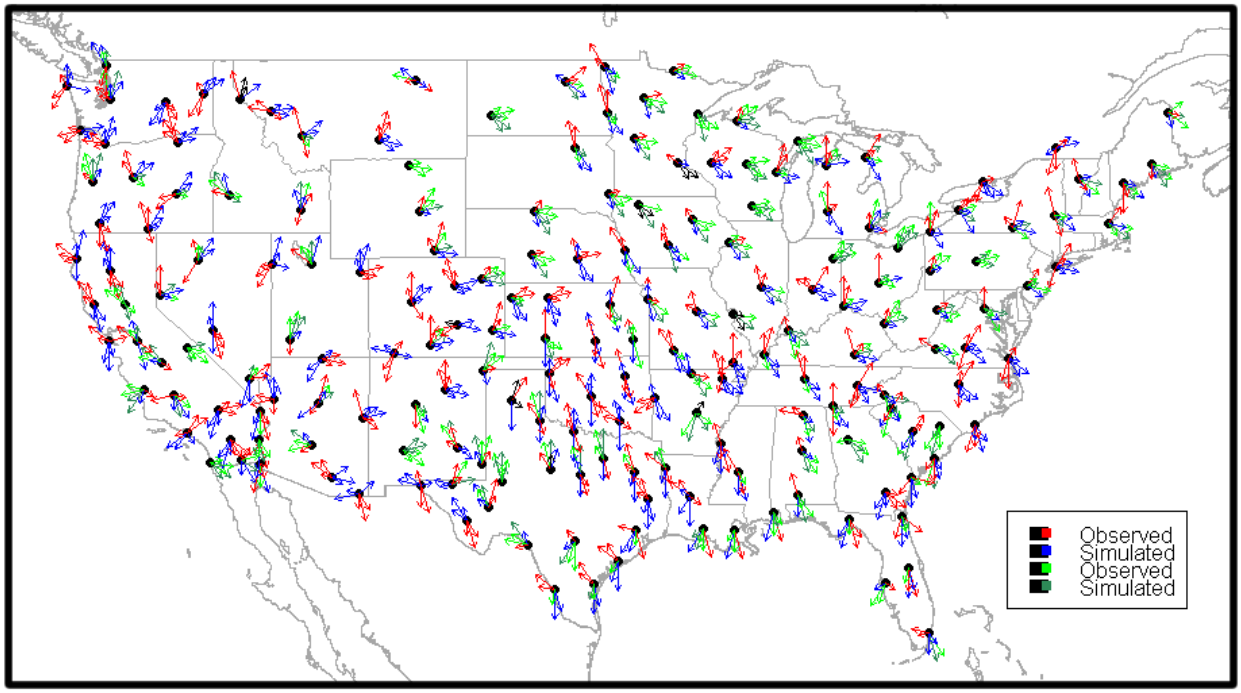


Figure 34. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the wintertime (DJF). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction is within +/-40° of the observed (light green).

in the model's sea-surface temperature, the climate model would likely be affected [Hurrell and Trenberth, 1999]. To explore this further, the climate model's sea-surface temperature was examined and compared to a typical pattern that is observed in the wintertime (Figure 35). As illustrated in the Figure, the main differences are that the Gulf Stream represented by the climate simulation is further offshore and suppressed to the south. Additionally, the water temperatures off the Northeast and Mid-Atlantic are warmer in the climate simulation than what observations indicate. The Gulf Stream displacement (south and east) in the simulation would tend to cause a south and eastward displacement of the simulated wintertime storm track along the Atlantic Coast. In addition, coastal storms prefer to develop, strengthen, and track along the stronger

observed sea surface temperature gradients. The strong observed sea temperature gradient noted in the observations is substantially smoother in the climate simulation. The net result, in theory, would be fewer southerly wind instances, as the wintertime storm track would remain south and east of the southeastern United States. This is supported by the comparison of the RCM and reanalysis mean sea-level pressure patterns (Figure 33). The reanalysis indicates an area of mean lower sea-level pressure just off the southeast coast over the Gulf Stream that is not simulated by the climate model. The RCM also has the main northwest Atlantic low, which results from coastal lows deepening off the Eastern Seaboard, much further south (well east of Cape Cod) than the reanalysis (over Newfoundland).

Although large observed and RCM differences in the southerly wind flow regimes exist across the Southeast, some of the wind flow modes are partially captured by the climate model. For instance, with the exception of Jacksonville (FL), Savannah (GA) and Miami (FL), the dominant north and northeasterly wind modes along coastal sections of the Carolinas and Florida are simulated well by the regional climate model. In the Miami case and the three sites along the south Texas coast, the observed southeasterly wind mode is a wintertime sea breeze signal the simulation will not capture because of the spatial resolution issues discussed previously. The suppressed storm track idea is further solidified by the wind flow regimes along and just inland of the Gulf coast and even coastal areas of Texas, where no southerly wind is simulated, but is

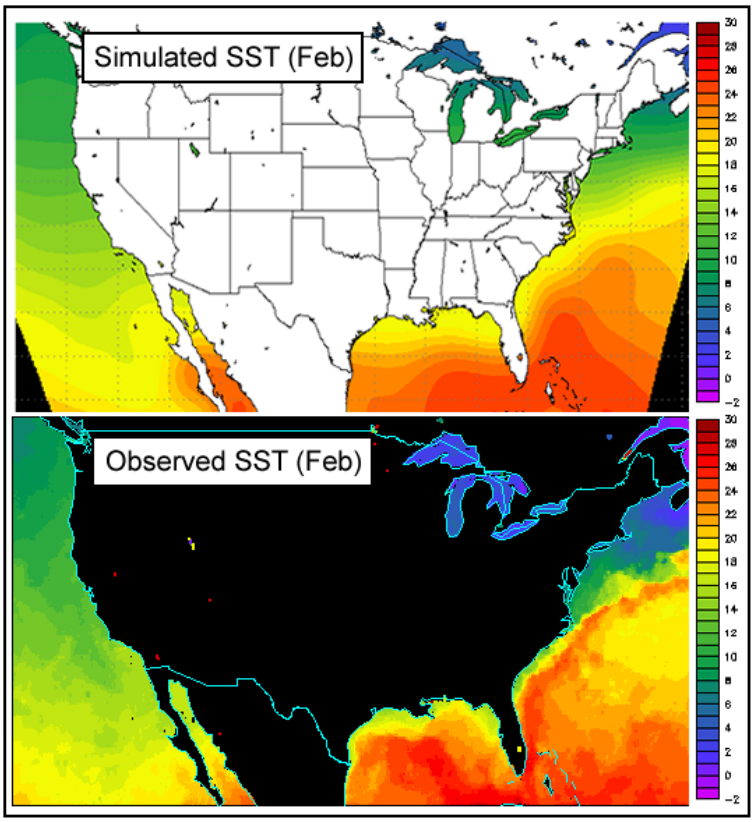


Figure 35. Sea surface temperature patterns during the winter. Upper panel is an example of the sea surface temperature from the regional climate simulation on a particular day in February. The lower panel is the observed 14 km scale sea surface temperature in February 2002. Both datasets are plotted using the same color scale.

observed. This indicates the climate model does not allow winter low-pressure system to make the frequently observed inland track through the Deep South.

Areas of northern and western Texas and southern portions of the Great Plains are much better simulated by the climate model in terms of dominant wind flow patterns. The climate model represents both the northerly and southerly wind flow that is observed in North Texas, Oklahoma and Nebraska; even though the northerly component is more northwest in many instances and the rank of the wind direction modes are not exactly the same. This bipolar wind

flow is caused by the passing of developing wintertime low-pressure systems that have strong southerly flow ahead of the low-pressure system and strong northerly wind behind. This implies that at least the developmental stages and frequency of the southern Rockies wintertime low-pressure systems are reasonably simulated by the climate model, just not the subsequent track over the eastern United States. Also in agreement is the mean winter sea-level pressure pattern (Figure 33) where both the model and observed pressure patterns indicate a lower pressure zone on average over this area and higher pressures to the east.

Wind flow patterns in the Ohio Valley, Great Lakes and the Midwest United States during the wintertime are not well simulated by the climate model. Again, the observations indicate there is southerly wind component at many of the sites, while the climate model has almost a pure northwesterly wind mode. The predominant simulated northwest flow over the Ohio Valley and Midwest, coupled with the southerly component over the southern Great Plains and Texas, implies that the RCM frequently simulates the migration of Arctic high-pressure systems from central Canada to Arkansas, western Tennessee, and northern Mississippi. However, climatology has frequent clipper-type low-pressure systems that arc from northwest to southeast across the northern Plains and Midwest. Again, this idea is well supported by the mean sea-level pressure pattern (Figure 33), which shows a much stronger high-pressure signal from central Canada to the central United States in the RCM compared to the reanalysis. The mean sea-level pressure pattern of the reanalysis indicates that the high-pressure area over the central United States is cut-off from the central Canadian high because the frequently observed clipper-lows lessen the mean pressure over the northern Plains.

The near-surface wind flow patterns over the western United States are more difficult to

evaluate because the RCM is optimized to simulate regional climate not local scale wind variations that are a result of small-scale topographical variability. However, the high plains areas of Montana, Wyoming, and Colorado, which are relatively flat in open, should be simulated well if the climate model is representative. This is not necessarily the case though, and differences are related to the poor performance over the north-central United States that has already been discussed. The model, as in other parts of the United States, has a predominant northwesterly component while the observations indicate a high frequency of southwesterly wind. Again, this discrepancy is a reflection of the climate model simulating an over-active southerly storm track and few Canadian Clippers low-pressure systems.

Along the West Coast of the United States, the climate model performance is mixed in terms of wind flow regimes. Areas in far Southern California, central and northern Arizona and New Mexico show fair agreement between the observed and simulated wind flow at some sites, and poor agreement at others. Additionally, the dominant wind directions over inland valley areas of central California are well represented at several sites. However, along a majority of coastal California, much of Oregon and Washington state the observed wind flow regimes are not well represented. Much of this discrepancy in the western United States could be due to the influence that the coastal and interior mountain ranges have on the local wind flow[Leung et al., 2003; Mass et al., 2003].

4.2.1.2 Spring

Springtime is a transition season; so, synoptic weather patterns will be a mixture of warm and cool season signatures. The March, April and May (MAM) 10-year mean sea-level pressures

are plotted in Figure 36. As with the winter comparison, there are several notable similarities and differences in the sea-level pressure patterns that relate to the placement and frequency of high and low pressure systems. The RCM well simulates the area of mean high pressure off the California coast, even the high-pressure tongue that exists inland across the Pacific Northwest. This agreement is reflected in the dominant wind-flow comparison shown in Figure 37. RCM directions are shown in blue and observations are shown in red. Areas where the arrows are green indicate reasonable agreement of the dominant wind flow. Along the Pacific Coast and inland valleys is agreement between of some of the dominant wind directions, which supports the agreement of the sea-level pressure pattern. A 700 mb wind flow analysis for spring, presented in Section 3.2, illustrates general agreement in the upper-level flow over this region.

The mean sea-level pressure comparison in Figure 36 indicates good agreement in the pattern over the central part of the United States. In both pressure patterns an area of relatively high-pressure, over the eastern United States, results in southerly flow from Texas northward to the upper Great Plains. This agreement is clearly shown in the comparison of the dominant wind flow in Figure 37. On the other hand, over much of the eastern United States, the pressure and wind flow patterns do not agree. The mean sea-level pressure patterns hint at a fundamental difference in the simulated synoptic weather when compared to the observed climate. The observed mean pressure pattern indicates the existence of a warm, sub-tropical high (the Bermuda High) being established over the southeastern United States in the spring. The RCM does not simulate this feature as a dominant synoptic pattern. The RCM indicates the main synoptic pattern across the eastern United States is high-pressure of Canadian origin (cool season). This suggests that the RCM is slow to transition from cold to warm season weather systems. It should

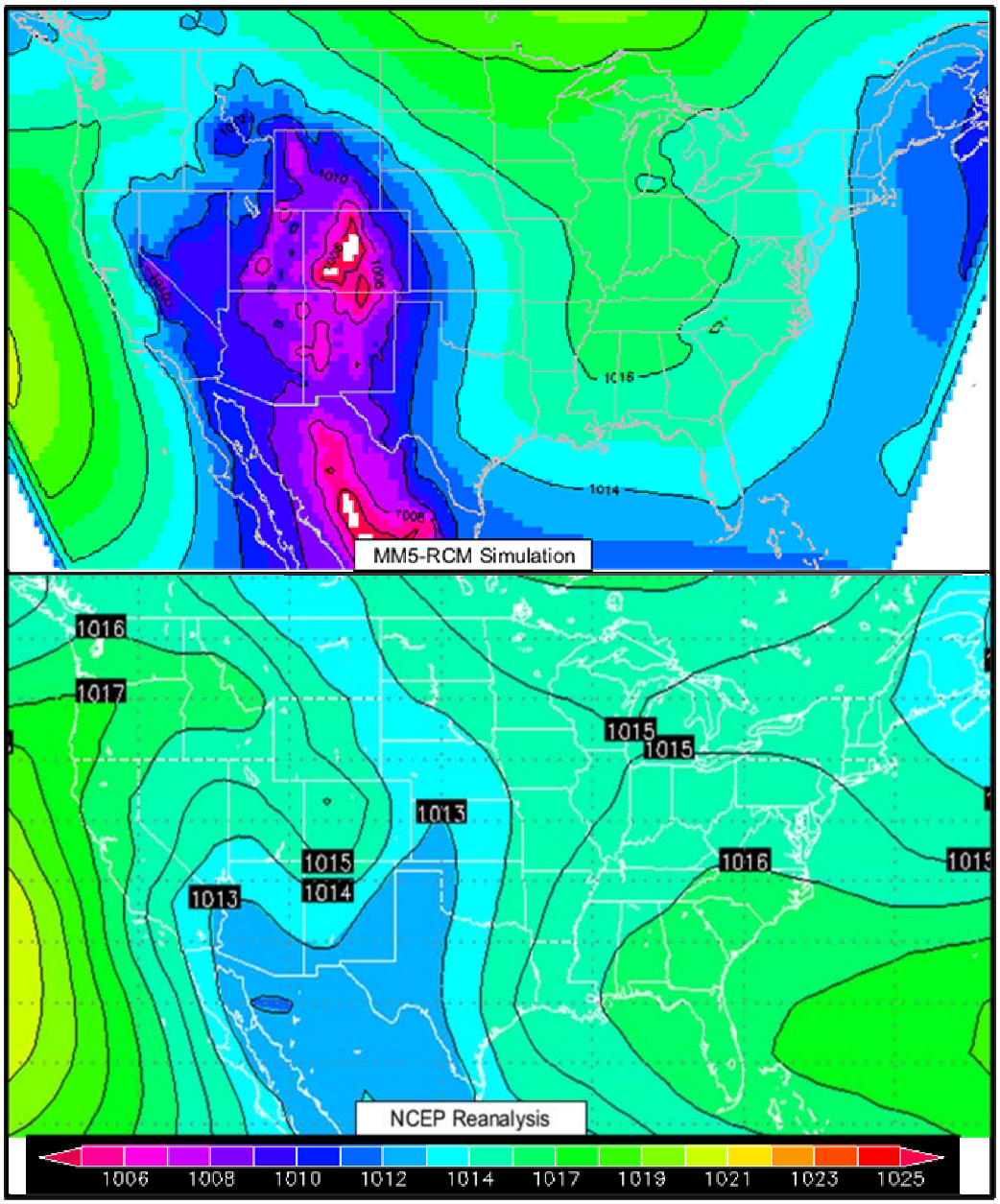


Figure 36. The 10-year mean sea-level pressure pattern in the springtime (MAM). The top panel is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale. Note that the sea-level pressure in the western U.S. is not reliable.

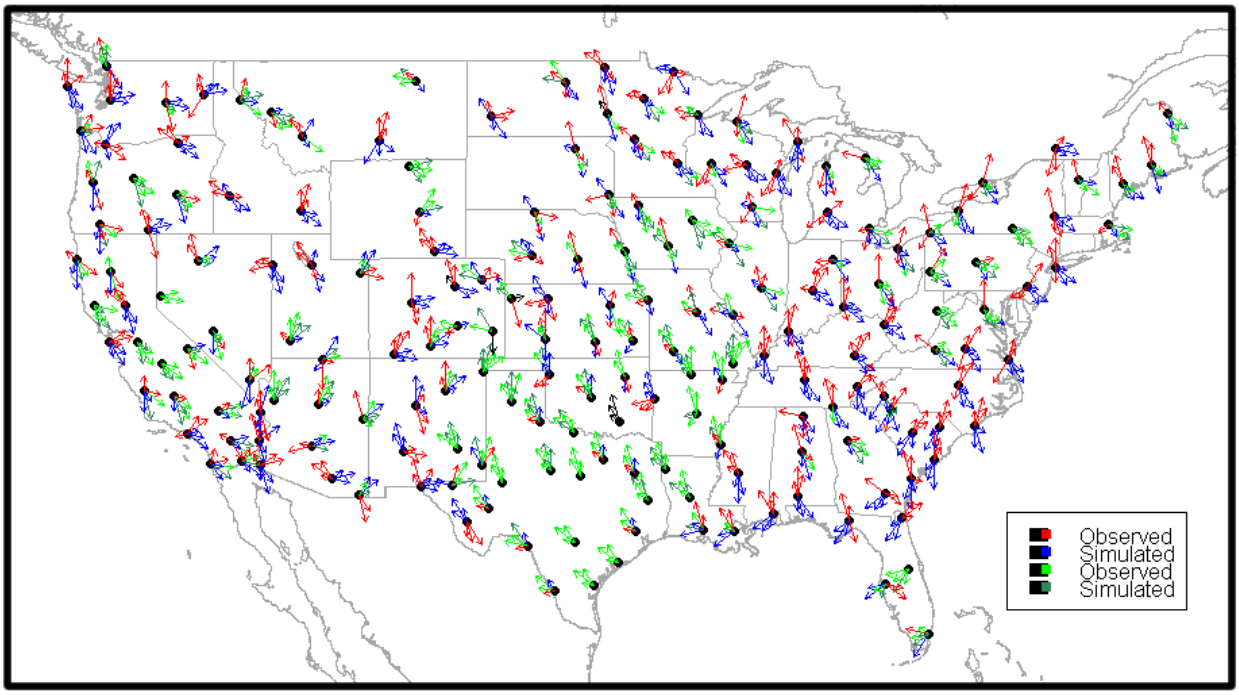


Figure 37. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the springtime (MAM). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction mode is within +/- 40° of the observed (light green).

be mentioned that there is a climatological cold-season high-pressure mode (higher-pressure area to the north of the Great Lakes in lower panel of Figure 36), but it is not as persistent as the RCM suggests. The wind flow analysis in Figure 37, again, clearly illustrates these pattern differences. With the exception of New England, many sites in the eastern United States indicate poor wind flow agreement between the climate simulation and the observations. The areas of the Northeast and New England are in fair agreement because the simulated and observed mean sea-level pressure pattern has a gradient of higher to lower pressure from west to east, which induces a general northerly wind. These pattern similarities and differences will directly affect the temperature and precipitation patterns, which will be discussed in more detail in Section 4.2.2.

4.2.1.3 Summer

Summertime is traditionally a season where air quality concerns are the greatest, especially with species such as ozone and particulate matter. During the summer, it is generally known that wind flow and synoptic patterns have a strong correlation with the regional air quality [*Hurrell and Trenberth*, 1999]. It is therefore imperative that the climate model represents the range of weather patterns that exist across the United States.

Figure 38 presents the 10-year mean summertime sea-level pressure pattern of the RCM and climate analysis. The most dramatic difference is in the eastern United States. It is well known that a persistent sub-tropical high-pressure (Bermuda high) area dominates the summertime climate along the middle and southern Atlantic coast of the United States. The classic pattern (lower panel, Figure 38) is clearly seen in the climate reanalysis, but not at all present in the RCM mean sea-level pressure field (top panel,Figure 38). In fact, the climate simulation has a low-pressure belt that extends from the Gulf to Southeast coast, and northeastward into the Atlantic Ocean. One plausible explanation for this dramatic synoptic pattern difference in the eastern United States is the fact that the driving global climate model does not force this pattern properly on the southeast boundary of the nested model domain. If the sub-tropical Bermuda high is not correctly simulated by the global climate model, along the eastern boundary of the model domain, the regional climate model cannot replicate this feature because the domain does not extend far enough across the Atlantic Ocean to resolve explicitly the entire high-pressure system. The observed and simulated mean sea-level pressure patterns shown in Figure 38 indicate this is the case. A climate study [*Hurrell and Trenberth*, 1999] using the same GISS-GCM found that

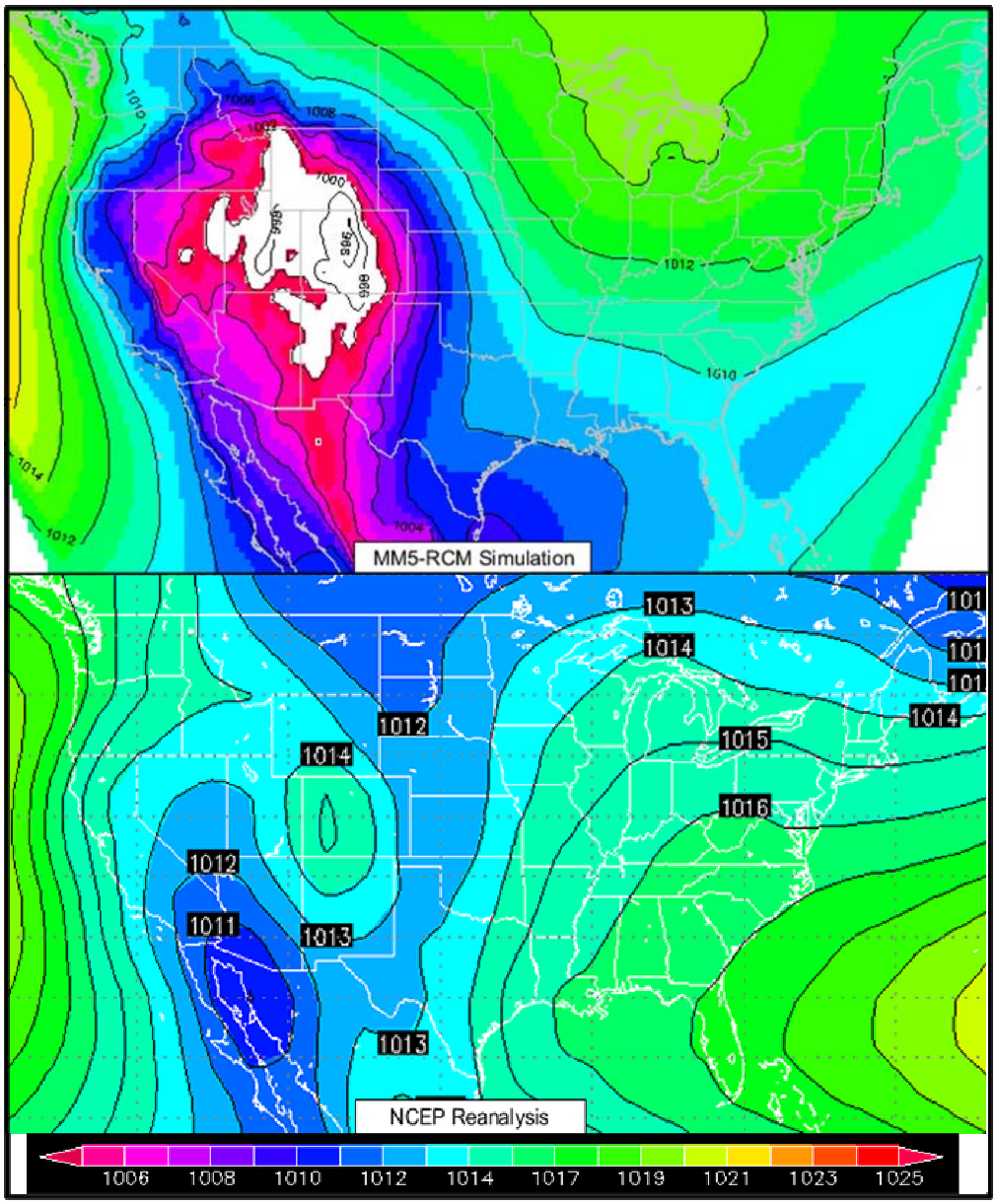


Figure 38. The 10-year mean sea-level pressure pattern in the summertime (JJA). The top panel is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale, and simulated sea-level pressure over the western U.S. is not reliable.

the climate model displaced the Bermuda high well northeast of its climatological position and this difference influenced chemical transport simulations, which used the global climate model meteorology. The dominant wind flow analysis for summertime is provided in Figure 39. This analysis reveals stark differences between the observed and simulated wind field over the entire Eastern United States that directly corresponds to the mean sea-level pressure differences. At almost every site east of the Mississippi River, the top three dominant wind directions from the model simulation are approximately 180° from the observations. The 700 mb wind flow analysis in Section 3.2 supports the differences in the sub-tropical high-pressure placement over the eastern United States, as does a upcoming comparison of temperature and precipitation.

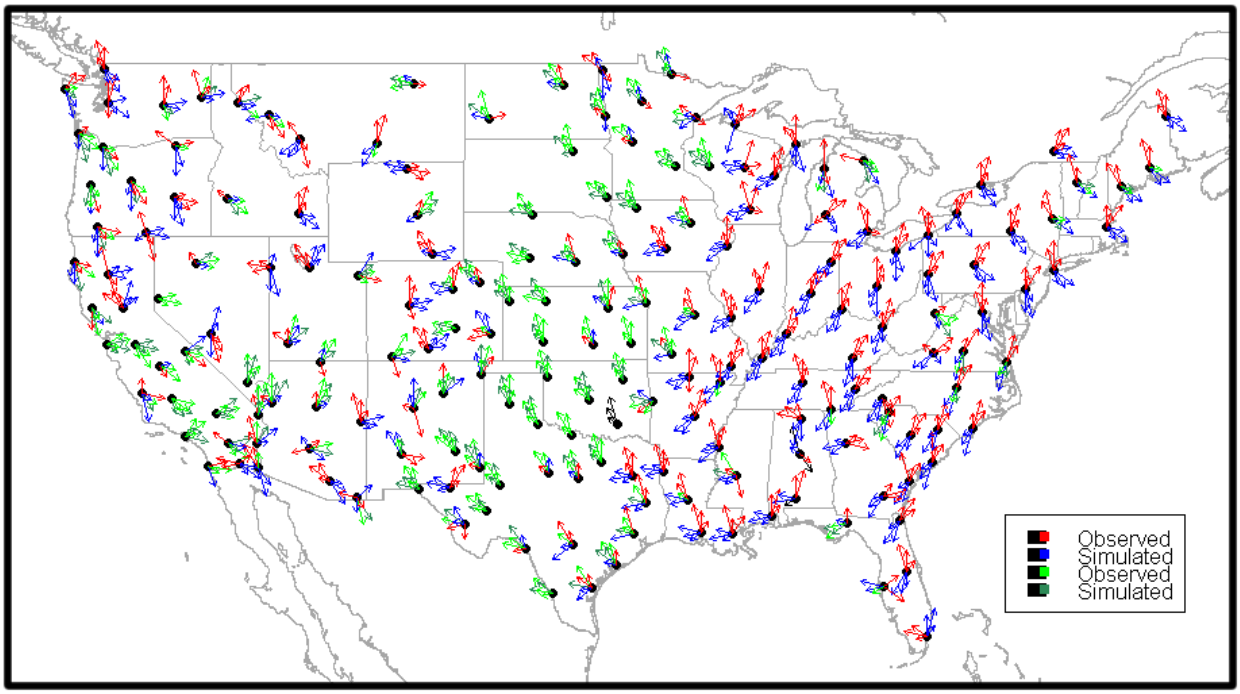


Figure 39. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the summertime (JJA). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction mode is within +/- 40 ° of the observed (light green).

On the other hand, west of the Mississippi River and east of the Rocky Mountains, the climate simulation is in good agreement with the climate analysis in terms of mean sea-level pressure (Figure 38). Although the eastern sub-tropical high-pressure area is not simulated at the correct geographic location, the fact that is it over the eastern United States causes a general south to east wind over the center of the country. The near-surface dominant wind direction in Figure 39, in particular over the central Plains, supports the claim that the RCM is performing well in the area.

Mean sea-level pressure patterns in Figure 38 suggest that the model captures the dominant synoptic weather patterns along the Pacific Coast of the United States. Both the RCM and climate analysis have a similar spatial configuration of high-pressure off the West Coast. In support of this, the climate model well represents the near-surface dominant wind direction modes over much of the central and southern portions of California and the Desert Southwest, as indicated by the numerous green wind vectors in Figure 39. The regional climate model also performs well in areas of Oregon, Idaho, Nevada and Utah.

4.2.1.4 Autumn

A comparison of the 10-year mean RCM and climate analysis sea-level pressure patterns is provided in Figure 40. The general sea-level pressure pattern off the west coast of the United States is simulated well by the RCM (i.e., relatively low pressure off the Pacific Northwest, high-pressure off the California coast and low-pressure over Baja California). The dominant wind flow analysis for the autumn (Figure 41) indicates some agreement between the RCM and climate, but the results are mixed. More agreement exists (green arrows) across inland portions

of California. There is little agreement in the top three dominant wind directions along the entire Pacific Coast. The autumn season was not simulated well in this region in terms of 700 mb wind (Section 3.2), as the RCM simulated a stronger northerly wind component. This is supported by the near-surface wind flow analysis, but it is still uncertain why the sea-level pressure pattern along and just off the coast is similar. The answer could lie in local scale effects that the RCM is not designed to simulate.

The RCM 10-year mean sea-level pressure pattern is similar to the climate reanalysis over parts of the Southeast, Deep South, Texas and the southern Plains (Figure 40). The pressure pattern leads to easterly winds across the Deep South and southerly winds across the lower and central Plains states. Figure 41 indicates that the dominant wind flow regimes from the RCM compare well with many of the climate reference sites, as indicated by the green vectors. However, a key feature in the climate analysis (Figure 40), the high-pressure area over the Mid-Atlantic, and the extension of this higher-pressure zone across the Atlantic, is not simulated by the RCM as a primary synoptic pattern. Rather, the RCM simulates a high-pressure over the Great Lakes region. The wind flow analysis in Figure 41 agrees well with these important differences in the mean synoptic flow pattern. There are very few of the climate reference sites that are similar to the RCM dominant wind flow in the northeastern quadrant of the United States. In fact, many of the sites indicate the dominant wind direction of the RCM is opposite of the climate observations, indicated by the lack of green vectors over the Ohio Valley, Mid-Atlantic, Northeast and Great Lakes. In the upcoming temperature and precipitation analyses, these key differences are linked to significant differences in the near-surface temperature and precipitation. Additionally, the 700 mb wind flow analysis shows major differences in the upper-

level flow patterns over these regions during the autumn.

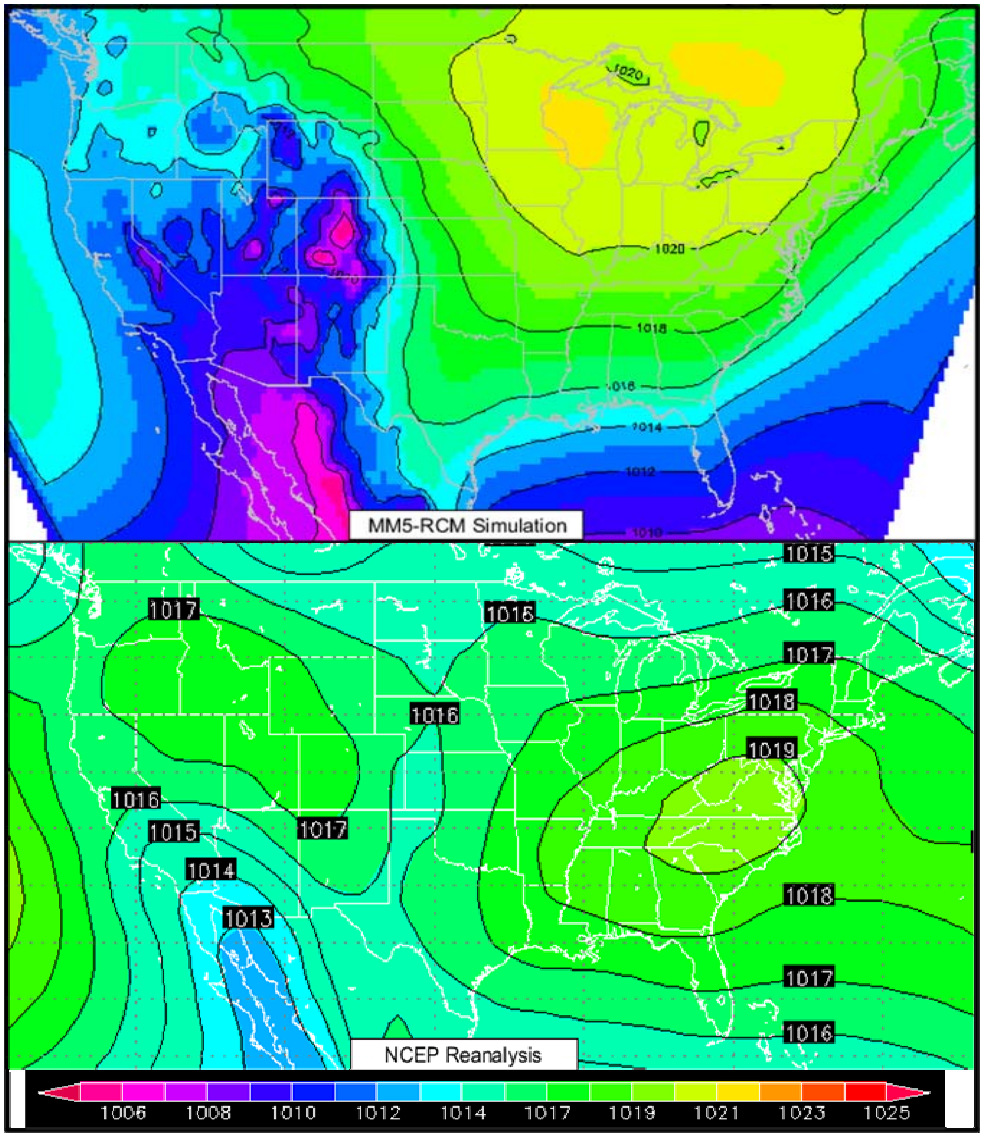


Figure 40. The 10-year mean sea-level pressure pattern in the fall (SON). The top panel is from the regional climate model and the bottom panel is from the NCAR Reanalysis II dataset. Both datasets are plotted using the same color scale. Note that the sea-level pressure in the western U.S. is not reliable.

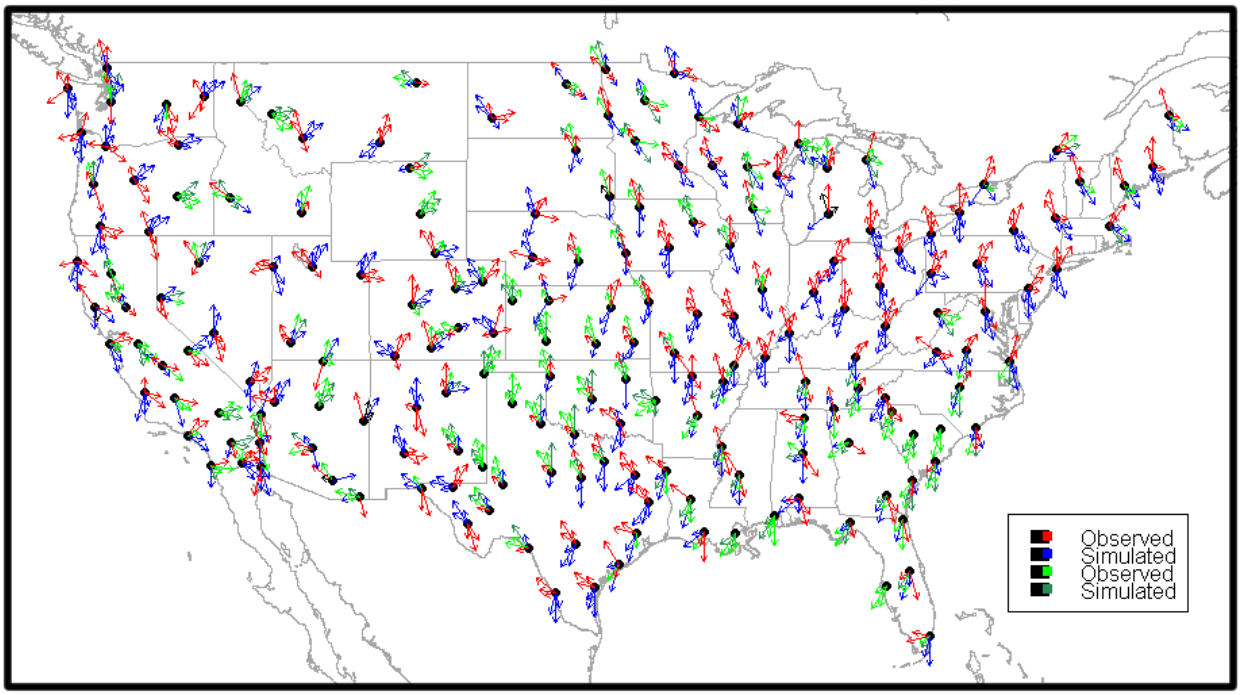


Figure 41. Observed (red) and simulated (blue) dominant 10-m wind directions at the climate reference sites for the fall (SON). The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the simulated (dark green) wind direction mode is within +/- 40° of the observed (light green).

4.2.2 Temperature Evaluation

Many properties of the temperature time series were generated. However, to first generalize regional performance of the climate model, mean annual temperature was calculated for the observed and simulated time series at each climate station. Figure 42 shows the difference or bias in the 10-year mean annual temperature (model-observation). Overall, the RCM is colder over most of the United States by 1 to 7 K. This is more evident in the western United States where typical cold biases are approximately 5 to 7 K. Over the eastern United States, the cold bias of the climate simulation is approximately 1 to 2 K. The central United States, including the Great Plains and Midwest areas are best simulated in terms of mean annual temperature with a

cold bias of 0 to 1.5 K. Alternatively, coastal areas of the United States, especially in the eastern United States and Great Lakes are biased warm (1 to 3 K). These annual temperature biases are important to the overall analysis, but the seasonal differences that were noted in the synoptic pattern analysis indicate that temperature biases likely have a seasonal variation.

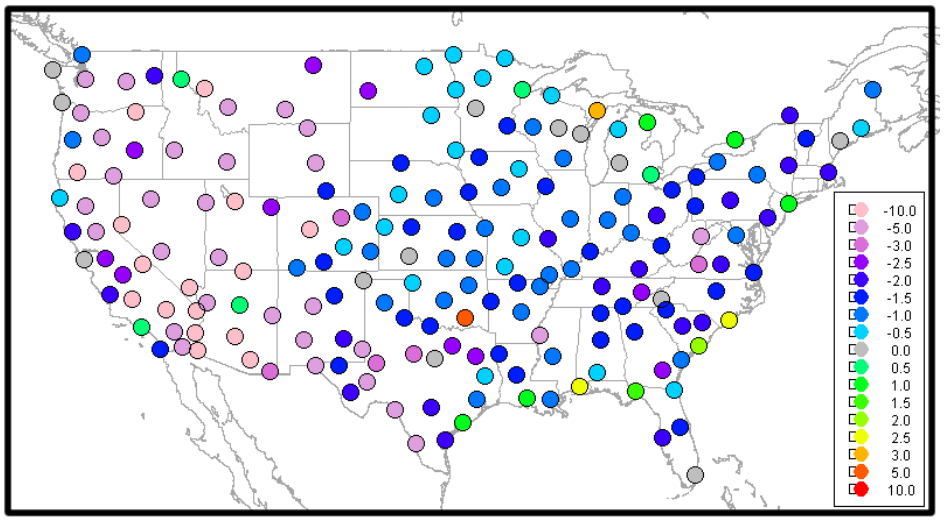


Figure 42. Mean 2 m temperature (10-year) difference (K) between the climate model simulation and observations.

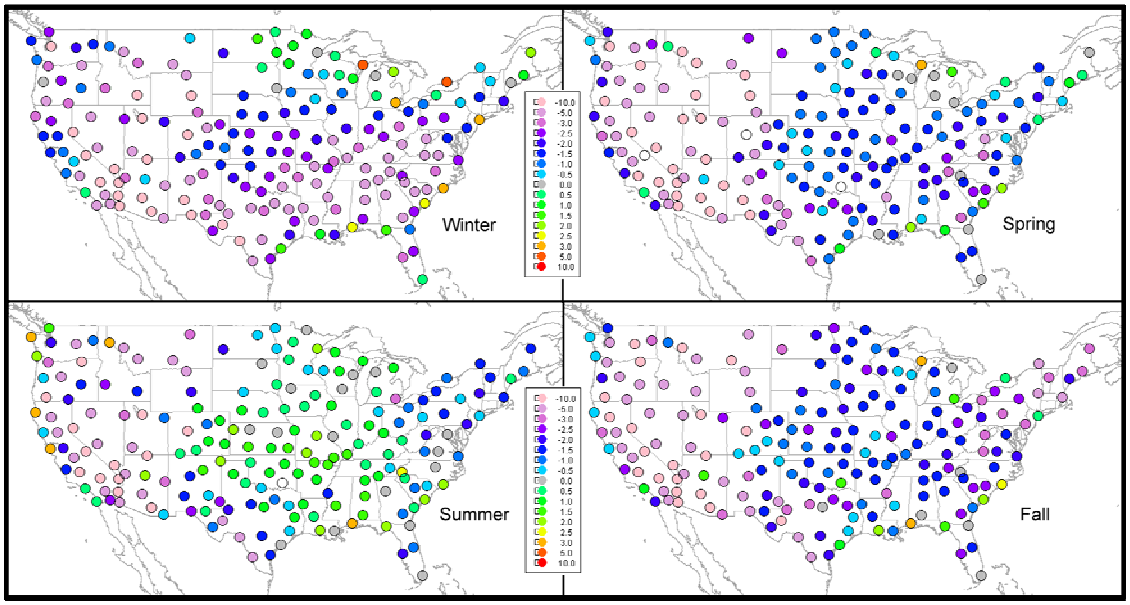


Figure 43. The seasonal mean 2-m temperature (K) difference between the regional climate simulation and observations (model-observation).

4.2.2.1 Winter

Seasonal biases in the mean 10-year temperatures were calculated to determine if there was a bias in simulated seasonal temperature across the United States. Winter biases shown in Figure 43 (top left panel), infer that the wintertime cold bias of the climate simulation is much more dramatic than the annual bias. Many of the cold bias values are on the order of 3 to 7 K. Inland areas of the middle-Atlantic, southeastern and southern United States are biased cold bias by 3 to 5 K, which is much greater than the annual mean cold bias of 1 to 2 K. This seasonal anomaly indicates the annual cold bias over much of the United States is a result of unrealistically cold winters simulated by the climate model. These colder simulated temperatures are consistent with the sea-level pressure and wind flow analyses that indicates the model frequently simulates Arctic high-pressure systems over the central United States.

Alternatively, areas of the upper sections of the northeastern U.S. and Great Lakes have only a slight warm bias, which corresponds to the general better agreement of the wind flow patterns of the climate model in this region. Also noted, the wintertime sea-surface temperature in Figure 35 suggests that the climate model has much warmer Great Lakes temperatures than the observations (~ 5 K warmer). These warmer water temperatures in the climate simulation are directly responsible for the approximate 5 K warm bias at the sites directly on the shores of the Great Lakes. Areas in the central portion of the United States have a colder bias relative to the annual mean, indicating the winter is also the main contributor to the overall cold bias. The 2 K to 3 K cold bias exists in north Texas and the southern Great Plains even though the wind flow patterns were fairly well simulated. The cold bias over most of the inter-mountain West is similar to the annual mean. Coastal areas of the northwestern United States and California as

well as valley areas of California have a lower relative cold bias compared to other parts of the country, approximately 1 to 2 K. Many of these areas were also better simulated in terms of wind flow patterns than other parts of the United States. Preliminary statistical analysis indicates that the reported biases are statistically significant.

4.2.2.2 Summer

Differences in the mean temperature during the summer portion of the 10-year time series are shown in Figure 43 (bottom-left panel). The overall temperatures are much better represented in the summer by the RCM, especially in the central, southeast, and Mid-Atlantic sections of the United States. Bias in the Southeast and mid-Atlantic regions improve from a 3 to 5 K cold bias in the winter to 0 to 1 K warm bias. Many sections of the upper Midwest and central United States have a 1 to 2 K warm temperature bias. The northeastern United States has a slight cold bias of 0 to 1 K. In the mountainous portions of the western United States, the temperature bias remains cold.

The general temperature bias pattern over the eastern United States correlates well with differences in the summer synoptic pattern (Figure 38) that was discussed in the sea-level pressure and wind flow analyses. It was shown that the climate model simulates a persistent subtropical high-pressure ridge that extends from the Great Lakes to the Ohio River valley, which would lead to warmer temperatures than have been observed over the past 10-years. This warm bias area in Figure 43 (lower-left panel) that results from the sub-tropical ridge is evident. Also consistent with the sub-tropical ridge is cooler than observed temperatures along the northeast quadrant of this simulated feature (i.e., Northeast). The climate model, because of the

sub-tropical ridge, has persistent cool northerly flow over the Northeast rather than simulating frequent warm southwesterly and westerly flow. This is also true (cold bias) along the East Coast including Florida where the primary simulated wind directions in the summer are northerly.

Along the west coast simulated temperature bias reverses from 2 K cold in the winter to 2 to 3 K warm in the summer. At least a few of the sites have summer wind-flow patterns that explain that the warm bias of the climate simulation is a result of the model not simulating the sea breeze and its affects (coastal fog). For example, Figure 39 indicates at Los Angeles two of the dominant wind flow directions are directly onshore, while the climate simulation has an along shore dominant wind. Similar differences exist at the other sites in northern California that have a large warm bias. This supports the idea that the major differences are from local scale circulations that the climate model cannot replicate. Other interior sections of the intermountain West have a large cold bias in the summer (3 to 5 K). Differences in the model grid-point elevation and the actual observation site elevation (not shown) are likely the cause for this large cold bias. On average, the model grid-point elevations are 200-600 m higher than the observation sites, which translates to 2 to 6 K cooler temperatures. Preliminary statistical analyses indicate that, with the exception of areas most likely to be impacts by local effects, e.g., long the east and west coasts and the mountainous west, most mean summer season model to observed temperature differences are not significant.

4.2.2.3 Filtered Temperature Time Series and Variability

Alternative methods were used to evaluate differences between the simulated and observed temperatures of the past 10 years on different time scales. First, the seasonal component

of the temperature time series was partitioned by hour of the year and then the mean and standard deviation was calculated. The mean was plotted as a function of hour of the year, and the standard deviation was used to place bounds around the mean value. This display of the data succinctly identifies how well the regional climate model replicates seasonal aspects of the temperature distribution. Plots are presented for six of the climate reference sites across the country. The second analyses will examine the variability of temperature on the diurnal, synoptic, and seasonal time scale. This was achieved by examining the variability (i.e., standard deviation) of the temperature on these time scales for both the simulation and observed time series.

Seasonal aspects of the 2-m temperature are illustrated in Figure 44 for six observation sites in different sections of the country. The first site (upper left panel) is Boston (MA). The mean temperature difference (Figure 43) indicates Boston's modeled temperatures are biased cool when compared to observations. The climate model is about 1 K cooler on average during all seasons except the autumn, where the cold bias is about 5 K. The observed seasonal signal in Figure 44 is compared to the climate simulation, and at first glance, the model seems to track the observations well, except during the later part of the year (Oct.-Dec.) where the model has a clear cold bias. For the first half of the year, the model's seasonal signal is actually slightly warmer than the observations, which is opposite of what the mean temperature comparison indicates (Figure 43). This implies the overall cold bias may not be a seasonal effect, but a result of differences on another time scale. The variability, which can be diagnosed by comparing the width of the shading about the mean value in Figure 44, is generally similar between the simulation and model. The climate model is most accurate in reproducing the seasonal temperature in

July where the interannual mean and variability are about the same. This is true even though the flow patterns, examined in other sections of this study, were not simulated well by the climate model. The interannual variability from January to March is also well reproduced, which does agree with the wind flow analysis, which suggest the climate simulation agreed very well with the observations in the winter. In addition, the climate simulation reaches the coldest part of the winter earlier (mid-January) than observations suggests.

The same information for the Raleigh (NC) site is presented in the top right panel of Figure 44. The mean temperature (from Figure 42 and Figure 43) is simulated 1.5 K cold overall and 5 K cold in the winter, 0.5 K in the spring, no bias in the summer, and -2.5 K in the fall. The seasonal, interannual mean and variability is shown in Figure 44. It is clear from the data that the winter and fall cold bias is largely a result of the seasonal temperature component as the model is colder over these entire seasons. Temperatures transition colder, more quickly in the climate simulations starting in November, and do not recover until April. The variability in winter temperature from year to year is greater in the observed time series (red shading). This implies that the fewer large-scale pattern changes from winter to winter, which is supported by the wind flow and sea-level pressure analysis over the Southeast. During the summer, the climate model is actually warmer in the seasonal time series. This is opposite of the mean temperature bias, which is close to zero in the summer. This implies that one of the other temporal components is offsetting the warm bias in the summer on the seasonal time scale. The interannual variability in the spring and summer is very similar between model and observations.

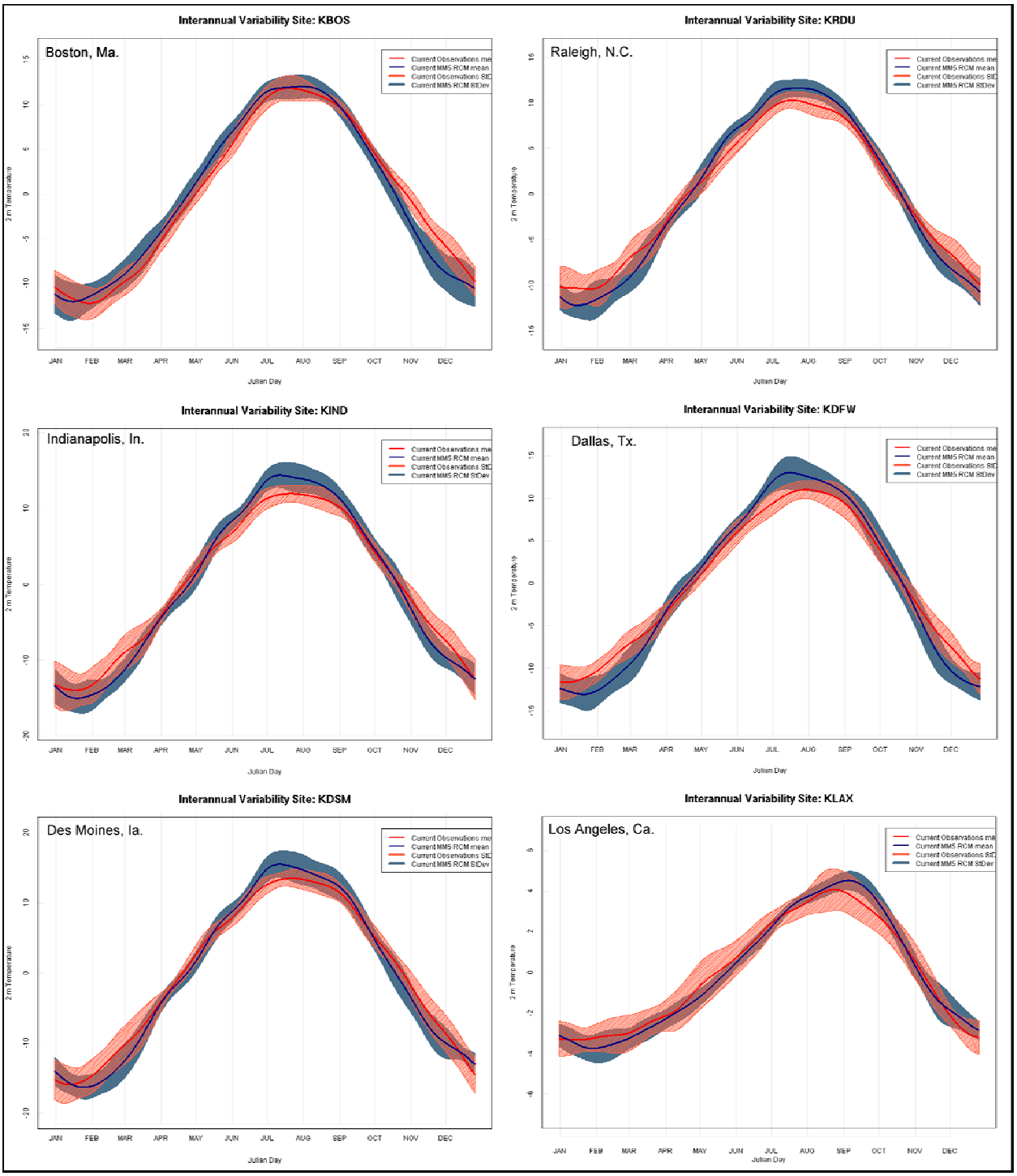


Figure 44. Mean (10-year) seasonal temperature perturbation and interannual seasonal variability of the regional climate simulation (blue) and observation (red).

The middle-left panel in Figure 44 is the same plot for Indianapolis (IN). The mean unfiltered temperature bias of the model was 1 K cold overall, 2 K cold in the winter, 0.5 K cold in the spring, 1 K warm in the summer and 1.5 K cold in the fall. This mean seasonal bias from the unfiltered time series is consistent with the model bias of the seasonal signal, i.e., the climate model cools more quickly in the fall and remains cooler through the winter. Again, this winter bias is consistent with the difference in winter synoptic patterns that indicates, of the top-three dominant wind direction modes in climate simulation (Figure 34), none were southwesterly (warmer return flow) and all were northwesterly (cold Canadian flow). Additionally, the warmer simulated summers as indicated by the seasonal time series, are in agreement with differences in the mean sea-level pressure pattern (Figure 38). The model has a mean subtropical high-pressure ridge over Indiana, while the observed sea-level pressure reveals a more active, cooler weather pattern. Interannual variability is similar in the spring, summer and fall, but slightly greater in the winter.

The fourth and fifth locations are Dallas (TX) and Des Moines (IA) (Figure 44, middle-right panel). Similar to the previous sites and consistent with the mean unfiltered temperature bias of the regional climate model; the seasonal component of the full time series at these sites is largely responsible for the dramatic cold model bias in the fall and winter and for the warm model bias in the late spring and summer. In these cases, unlike many of the other sites, the model performed relatively well in terms of wind flow patterns. This will be explored in a subsequent analysis that examines other temperature time scales.

The final interannual distribution shown in Figure 44 is for Los Angeles (CA) (lower-right panel). The seasonal temperature component is well simulated by the climate model, which

is similar to what the temperature bias of the unfiltered time series indicates (0 K to 0.5 K warm biases for all seasons). The interannual mean difference between the simulation and observations is consistent with the very good wind flow representation in the spring and summer, and less representative flow in the fall and winter. However, the observed interannual variability of the seasonal time series is double the climate simulation during the spring and summer, but similar in the fall and winter. Furthermore, the model tends to transition from summer to fall about a month later than the observations indicate. The results signify that more variations in the weather on time scales greater than a few weeks occur from year to year than what was simulated, even though the temperatures are generally well represented.

Temperature variability can also be expressed as the total variability of the time series. If the climate model represents the real climate, the variability of the complete time series will be similar. To examine this variability in more detail, the variability of the filtered time series components, which represent different physical processes that occur in the atmosphere, are compared.

Figure 45 provides the difference between the simulated and observed variability of temperature of the diurnal and synoptic filtered time scales. For these two time scales, the annual differences in variability, as well as the variability difference separated by winter and summer season are presented. It should be noted that the difference is expressed as the percentage difference (model-observation) relative to the variability of the unfiltered time series, so the variability of each time scale is normalized by the variability of the unfiltered time series.

Figure 45 (upper-left panel) indicates the annual variability of the diurnal temperature component is over-simulated (approximately 5 to 15%) by the climate model in the Desert Southwest.

This would imply that the range of diurnal filtered temperature is greater than the observations (on average) in these drier climates. Two processes may be contributing to the difference in the diurnal temperature variability. The first is the influence from synoptic weather, which will cause increase and decrease the diurnal temperature fluctuation with passing weather systems. The other process is related to errors in the model representing the diurnal temperature range because land-surface physics (e.g., model cools surface too much at night). In the case of the Desert Southwest, the larger model temperature variability is most likely related to model physics as the RCM simulated drastically cooler temperature at night, while the daytime high temperature were well reproduced. The net result is a larger diurnal temperature range. Alternatively, sites in the Midwest, upper Great Plains and Eastern United indicate the variability of the simulated diurnal temperature is 5 to 15% less than the observed. In this case, the major reason for the difference in variability is likely the lack of simulated synoptic variability, which will tend to lower variation in the diurnal temperature amplitude. This assumption is supported by the center-right panel in Figure 45, which indicated the simulated synoptic variability is lower over much of the eastern half of the United States. Other parts of the country are simulated fairly well in terms of diurnal temperature variability (annual).

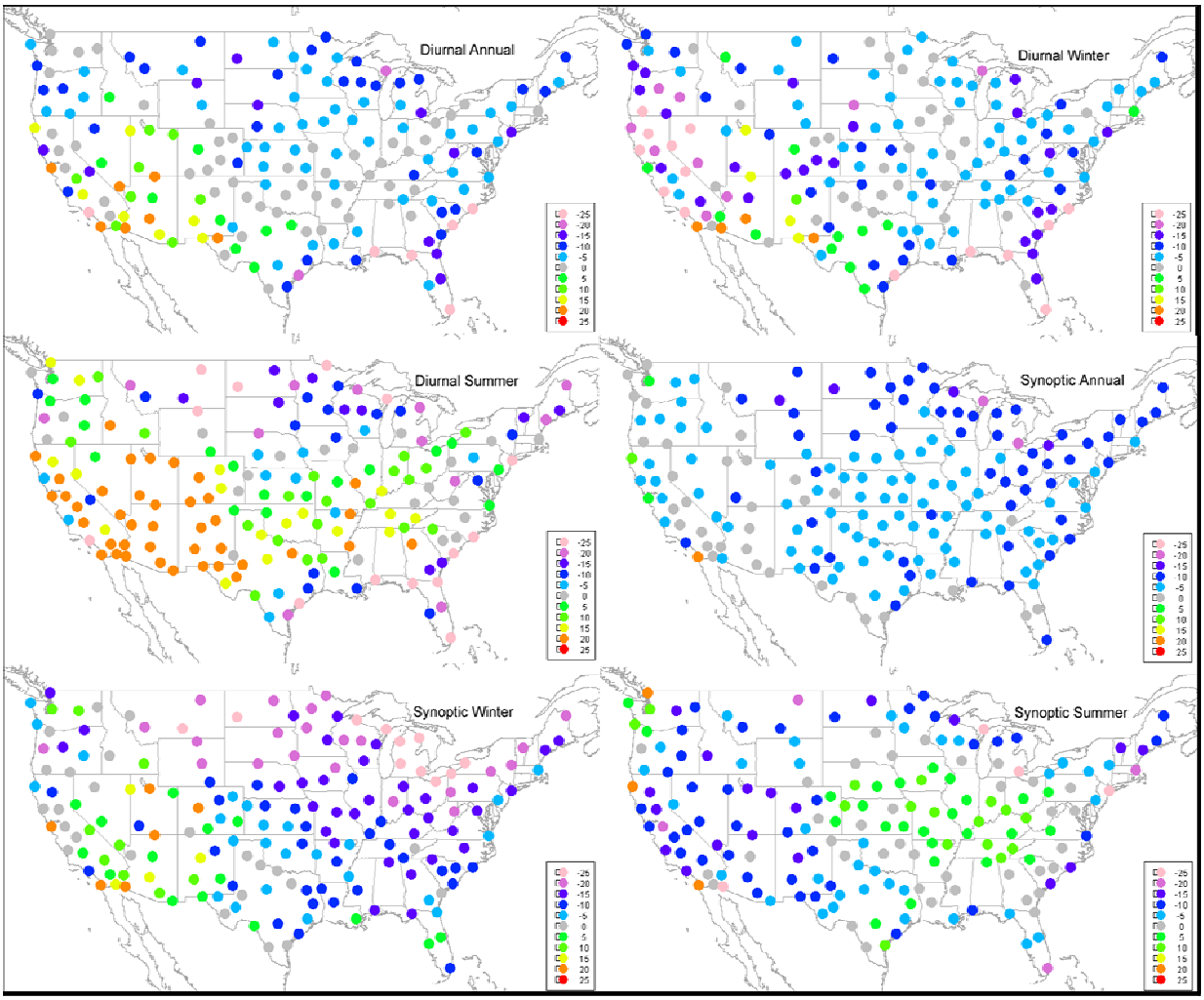


Figure 45. The percentage difference of the model and observed temperature 2 m temperature variability for different time scales (diurnal and synoptic) and the different seasons. The percent difference is scaled by the total variability of the non-filtered time series.

When the winter diurnal variability difference is isolated (Figure 45, upper-right panel), it is revealed that the simulated diurnal temperature variability is similar over much of the eastern United States. The climate model does not have as large a temperature fluctuation as the observations because of the lack of synoptic weather changes during the winter. At many sites in the center of the country, the diurnal variability is well simulated. However, the diurnal variability at sites near the west coast of the United States is significantly underestimated. When the synoptic variability is consider along with an examination of the time series, the lack of temperature variability on the diurnal scale is likely because of errors in local scale forcing, which is consistent with the sea-level pressure and wind flow analysis.

Differences in the variability of the diurnal time series for the summertime are presented in Figure 45 (middle-left panel). The most notable characteristic is the sharp spatial change from lower simulated variability in the Southeast (-5% to -20%), to greater simulated variability in central-eastern United States (10% to 20%), to lower simulated variability in the north-central United States (-10% to -20%). The time series were examine for a number of these sites and the extremely low variability of the diurnal temperature at many of the coastal sites is from the interpolation of the temperature that considers nearby gridpoints that are over water. The net result is a lower diurnal temperature range. Over the center of the country, the greater variability of the simulated diurnal temperature is a result of the inaccurate synoptic weather pattern as indicated by the sea-level pressure analysis. In this area, the climate model has a false preference to anchor a sub-tropical high pressure, which leads to much higher daytime high temperatures, thus a larger diurnal temperature range. These continental sub-tropical type high-pressure systems are associated with clear skies and drier air. This would tend to cause a large diurnal

temperature range in the climate model versus observations. A side effect of this false synoptic weather pattern is responsible for the lower diurnal variability in the Midwest and upper Plains. The model does not simulate frequent cool Canadian air masses in the summer. The time series indicates many periods of cool weather exist in the observations that are not simulated. In the western United States, the significant over-simulated diurnal temperature variability is from errors in the nighttime low temperatures that are much less than observations.

The synoptic variability difference for the whole year is presented in the middle-right panel of Figure 45. Differences indicate the simulated temperatures are generally less variable on the synoptic time scale (by approx. 5% to 15%), or in other words, the model has fewer changes in temperature because large-scale weather systems are more infrequent. Areas in the western United States are generally better simulated in terms of synoptic variability. The variability difference in the eastern United States is even more dramatic in the winter (Figure 45, lower-left panel). The large decrease in synoptic temperature variations over this region concur with previous analyses in this report, specifically the idea that the climate model simulates frequent, unrealistically cold air masses of the eastern United States. Furthermore, the climate simulates a winter storm track that is well south of where climatology suggests. The temperature variability on the synoptic time scale is 20% to 25% less over the north-central to northeastern United States, and 10% to 15% less over the central to southeastern United States. This directly implies that fewer oscillations in temperature, which result from the passage of synoptic systems, are simulated over these areas. On the other hand, the synoptic temperature variability is greater by 5% to 10% in the climate simulation over areas of the Southwest, south Texas, and Gulf of Mexico states, which further supports the idea that the wintertime storm track is simulated too

far south.

During the summer season, the difference in temperature variability on the synoptic time scale (Figure 45, lower-right panel) also has a clear spatial pattern. The states surrounding Illinois are simulated with much more temperature variability than the observations (approx. 15%), while areas to the south and east and areas to the north and west are simulated with similar, but less synoptic variability. This pattern is similar to the summertime diurnal variability (Figure 45, middle-left panel), in which it was suggested that the side effects from differences in the synoptic weather patterns was the root cause of the difference in variability.

Figure 46 provides the similar variability difference, but for the other time scales (intraday, seasonal, and interannual) and from the annual time series. The intraday variability differences are +/- 5% of the total temperature variability. This component is not nearly as significant to climate as the others are since these features are typically more chaotic and not generally well simulated by 36 km scale models [Hogrefe et al., 2004a]. On the other hand, the seasonal differences in variability clearly show a major problem of the climate simulation. The entire southeast quadrant of the United States has a 10-15% greater variability in the seasonal time series. This is the main indicator of areas where the simulated winter climate is too cold and summer climate is too warm. Areas of the upper Great Plains, Northeast and western United States, as well as areas along all coasts have 5% less seasonal temperature variability than what observations suggests. The differences in interannual variability or the mean temperature changes from year to year are low compared to some of the other time scales, but a distinct pattern is noticed across the United States. The climate model has lower (5%) variability in interannual temperature north and west of a line from Texas to New York, and virtually no

difference in the year-to-year temperature variations to the south and east. This relatively well-simulate aspect of the climate simulation may have implications on similar comparisons between the current climate simulation and future regional climate simulation.

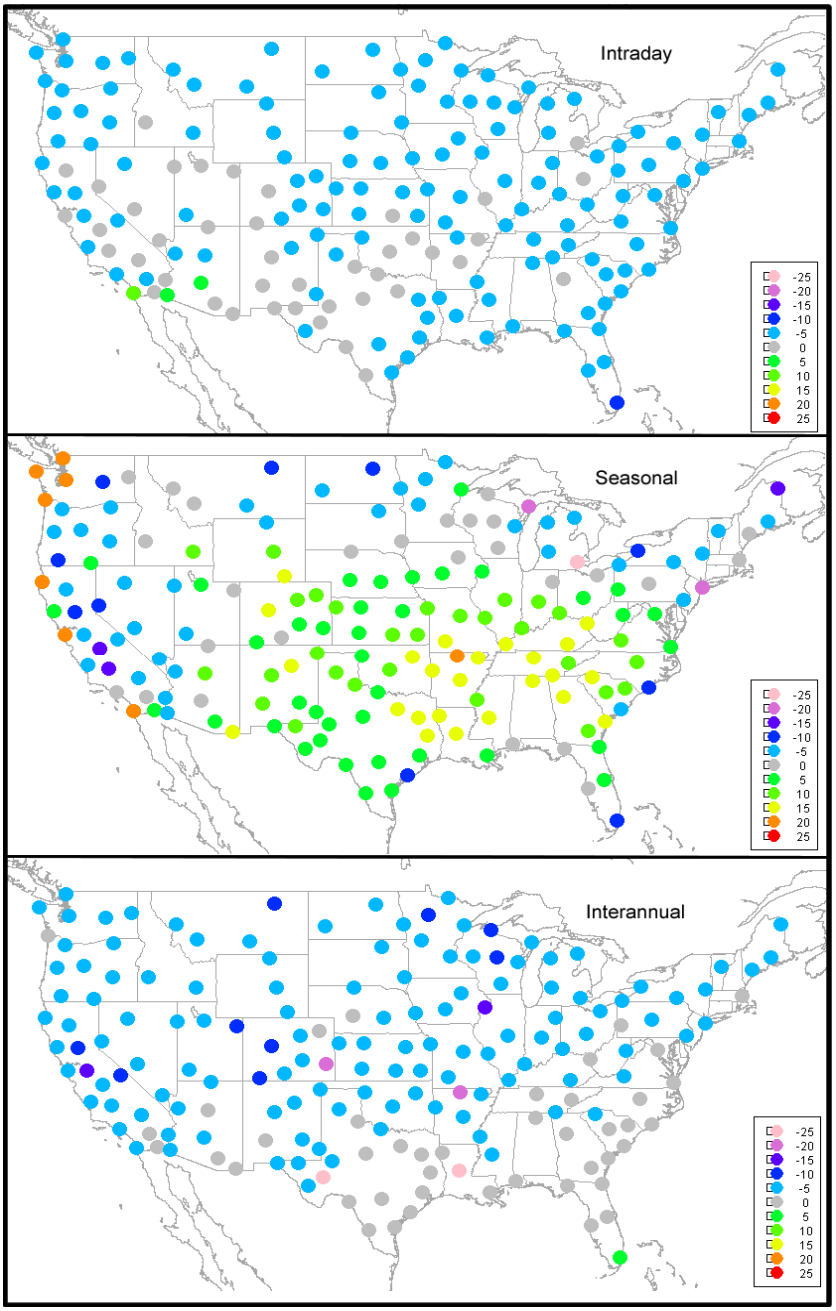


Figure 46. The percentage difference of the model and observed temperature 2-m temperature variability for the intraday, seasonal and interannual time scales. The percent difference is scaled by the total variability of the non-filtered time series.

4.2.3 Precipitation Evaluation

Precipitation is one of the most examined variables in climate studies. In terms of impact on future air-quality simulations, the precipitation will effect the wet deposition and concentrations of all pollutants. Precipitation patterns also provide clues as to the ability of the model to simulate observed synoptic patterns, since the two are intimately related. The following precipitation evaluation will be related to well-founded results from the examination of the wind flow and synoptic weather patterns of the climate model. Two sets of figures are presented; the first is the mean annual precipitation (separated by season) that was calculated from the simulated and observed 10-year datasets. The second is the interannual variability of the annual precipitation(also, separated by season). The background shading shows the observed precipitation, while the simulated precipitation is represented by the overlaid colored circles. The color scales are equivalent, so instances where the circles fade into the background shading, the model is the same as the observations, and vice-versa.

Figure 47 (upper left panel) indicates the winter precipitation is well simulated by the regional climate model over the entire western United States. Regional variations are also very well simulated. These similar precipitation totals are consistent with the well- simulated sea-level pressure and wind flow patterns, notwithstanding the poor agreement in certain situations that results from local scale effects. The spring, summer, and to a lesser degree, the autumn, are similarly well simulated in terms of mean precipitation (Figure 47). The model does produce less autumn precipitation than the observations suggest in the Pacific Northwest. In areas of central and northern California, the model produces more spring precipitation than the observations. This probably relates to differences in the large-scale weather transitions from winter to summer.

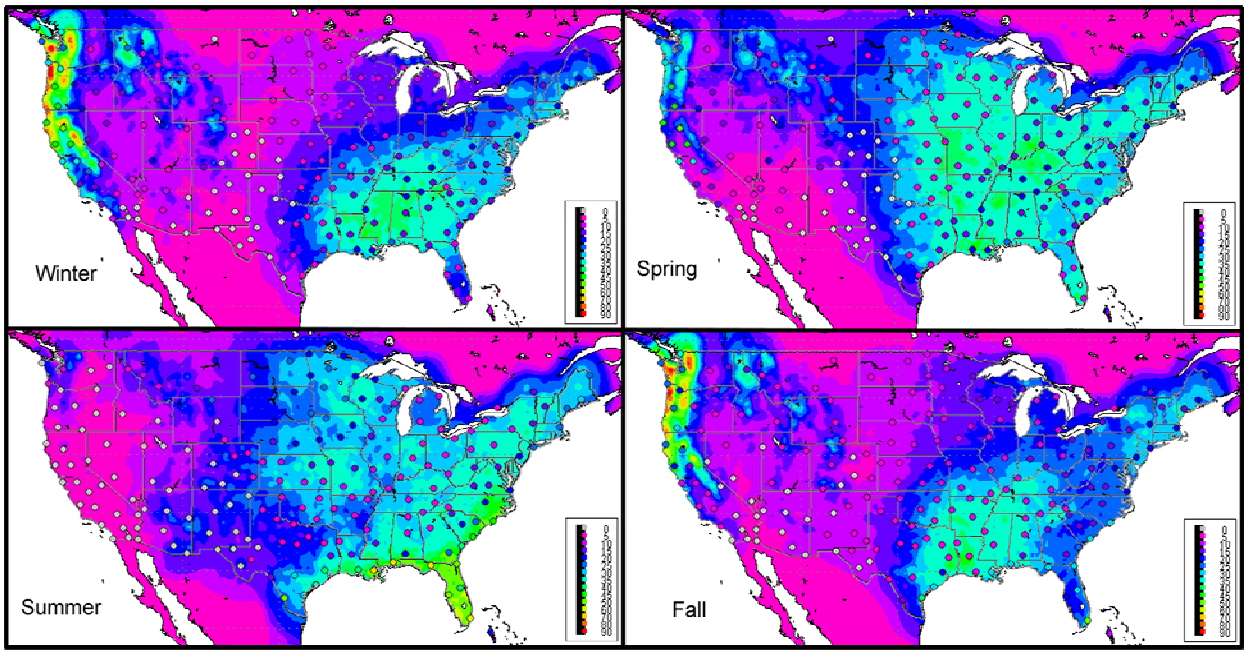


Figure 47. Mean seasonal precipitation (cm) calculated from 10-years of data. The background shading represents the observed precipitation, and the foreground dots represent the simulated precipitation at the climate reference sites.

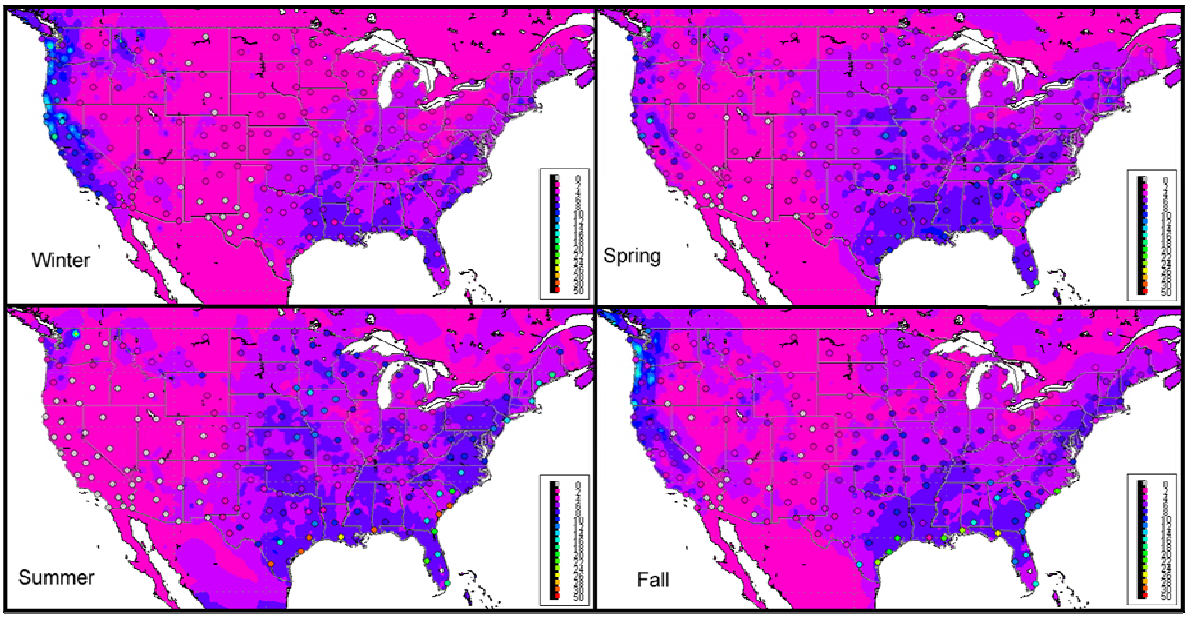


Figure 48. Seasonal precipitation variability (cm) calculated from 10-years of data. The background shading represents the observed precipitation variability, and the foreground dots represent the simulated precipitation variability at the climate reference sites.

Figure 48 presents the interannual variability of precipitation. Over the western United States, the model, as with the mean annual precipitation, seems to capture much of the year-to-year variability in total precipitation. During the autumn in the Pacific Northwest, the observations indicate more variability at a few sites than the observations suggest.

The eastern United States is not as well simulated in terms of annual precipitation patterns or annual variability. This has a clear connection to the poor wind flow, temperature and sea-level pressure evaluation results. During the winter (Figure 47, top-left panel) the climate model produces anywhere from 5 cm of precipitation in the central United States to a maximum of 20 cm over portions of the southern United States. However, observed precipitation totals

infer the mean annual precipitation in these areas is 10-15 cm (central U.S.) to 35-45 cm (southern U.S.). This is interconnected to colder air masses (Arctic high-pressure system) that were simulated over much of the eastern United States, and the simulated southern storm track, which suppresses subtropical moisture stream. The model does simulate a similar (approximately 20 cm) precipitation pattern over the Northeast and Great Lakes, which agrees with the better-simulated wind flow over this region.

During the spring (Figure 47, upper-right panel), similar to the winter, the climate model drastically under-simulates the amount of precipitation over much of the eastern United States. The widespread swath of 35-40 cm observed precipitation totals over a large portion of the central United States is about 2-3 times greater than the climate simulation (10-20 cm). Even the observed 30 cm area of precipitation along the East Coast is about twice the simulated amount. The climate simulation of fall precipitation over the Mississippi Valley (Figure 47, lower-right panel), as in the spring, is drastically under-done (by factor of three).

The climate model performs much better during the summer (Figure 47, bottom-left panel) along the immediate Gulf Coast, Florida, the Southeast and Northeast coasts, where the mean annual precipitation is almost identical to the observations. The climate model performs reasonably well in areas of the northern Great Plains. However, the summer annual precipitation is significantly under-simulated from Texas northeastward to areas of the Great Lakes as well as the Midwest. Observations indicate that many areas have summer average precipitation of about 30 to 35 cm, while the model only simulates 5 to 10 cm. This lack of precipitation is almost certainly related to other analyses, like the diurnal temperature variability difference, which has the same spatial pattern. Less simulated precipitation in this area causes less cloudiness, which increases the daily temperature range relative to observations that indicate more precipitation. As mention in the previous section, the air mass type (i.e., subtropical high-pressure) is also related since it is characterized by lower moisture and clear skies (allows larger amplitude of daily temperatures). These ideas are also supported by the summer sea-level pressure and wind flow analysis that revealed much different weather patterns between the model and observations.

Interannual variability of the observed seasonal mean precipitation in the eastern United States is much lower than the Pacific Northwest, generally less than 10 cm for all seasons (Figure 48, all panels). The winter simulation has little change in seasonal precipitation from year-to-year, which does agree with the observations. However, this agreement is likely, because the overall mean simulates winter precipitation is low. The spring comparison (Figure 48, upper-right panel) may indicate some skill by the climate model in simulating interannual variability, especially along the east coast and Appalachian mountain range. Many of the other areas in the eastern United States have similar variability between the model and observations. During the

summer, the simulation indicates many areas, most along the Gulf of Mexico and East Coast, have much larger interannual variability. This may be attributed to tropical activity and generally lower pressure zone along the Gulf Coast in the summer when compared to observations.

4.3 Point Comparison of Current and Future Regional Climate Simulations

This section presents comparisons of the current and future regional climate simulation predictions at select NWS sites on a range of temporal scales. To consider future predictions of climate change particularly relevant to air quality, the primary focus of the comparison is on surface temperature, precipitation patterns, and wind speeds. Analyses presented will primarily consider the long-term mean changes in these meteorological parameters to focus on climate scale changes in these meteorological conditions. The longer-term changes are assessed first by looking at the average differences between current and future simulations using all ten years of model predictions and by looking at these average differences separately for each season (e.g., a comparison of the average temperature for all future versus current predictions for the fall season). These long-term averages include higher frequency variability, such as synoptic scale changes associated with weather patterns; therefore, we use the KZ filter (see Section 4.1.3 for description of method) to then separate the seasonal and base case components of the meteorological variables. The base case component represents the long-term, multi-year average component of the meteorological variable of interest, while removing higher frequency changes that do not affect the long-term average. This base case component of the temperature is particularly

relevant to climate changes that occur over many decades. If a climate change signal is seen in the base case component of temperature and precipitation, for example, this could also suggest that we may see changes in background concentrations of ozone or particulate matter in the future simulations as well. However, it is also relevant to consider whether there are changes on higher frequencies (e.g., changes in synoptic weather patterns, extreme temperature events) that could affect the frequency of air pollution episodes. Previous studies have found an increase in ozone episodes and duration of episodes when modeling current versus future (similar mid-21st century) summer conditions, where they also compared out to 2050 [*Hogrefe et al.*, 2004b]. To consider the higher frequency changes, an analysis of the seasonal component of temperature and wind speed changes is included here.

The evaluation of the base case regional climate simulation was presented in Section 4.2. When key differences between the current and future regional climate simulations are identified, it is important consider these changes in light of how well the current regional climate simulation compares to observations. These modeled differences in current versus future climate could have an influence on the forthcoming air quality simulations, and the comparison to current meteorological observations is needed to establish confidence in the model simulations' ability to capture the meteorological processes of note.

4.3.1 Temperature

When comparing the future 10-year regional climate simulations to the base case 10-year simulation, the simplest comparison of temperature differences is first shown in Figure 49 . For the 10-year simulation, the annual mean of the hourly 2 m surface temperature is calculated for each NWS site based on all hourly temperature values predicted. When comparing the future to

the base case mean temperatures, a regional structure in the difference is evident. Temperature increases are simulated throughout the continental U.S. domain ranging between 1 and 2.5 K. Increases are smallest in the Eastern U.S., and some Midwestern states at about 1 K. Increases in the Midwest and western coastal U.S. are predominantly around 1.5 K. In the Southwest U.S., increases are the highest reaching about 2 K with a few extreme differences reaching 2.5 K.

When considering how these temperatures change for each season (Figure 50), we see some differences from the annual comparison in Figure 49. In the winter, there is very little temperature difference between current and future predictions on the west coast, while temperatures are up to 2.5 K higher in the future at some Midwestern sites. Opposing that pattern, summer shows very little increase in future temperatures at the Midwestern sites, while future temperatures reach up to 3 K higher at some west coast and Southwest sites. Southeastern sites appear to show the most consistency across seasons with temperature increases ranging between 1 and 2 K in the future at most sites. The autumn comparison in future versus base case temperatures shows the most striking increases, with future temperatures at many Southwestern sites that are up to 4 K higher and west coast sites between 2 and 3 K higher than current simulations. Preliminary t-tests were performed on these seasonal temperature differences and results suggest that the winter and autumn differences were significant across the domain while most of the spring and summer differences do not appear significant. This will be explored further in future analyses. While differences across the seasons are evident, it is notable that in some regions, fairly similar temperature changes are estimated for three out of four seasons, such as the uppermost Northern Midwest sites and Southeastern sites. This suggests the presence of long-term

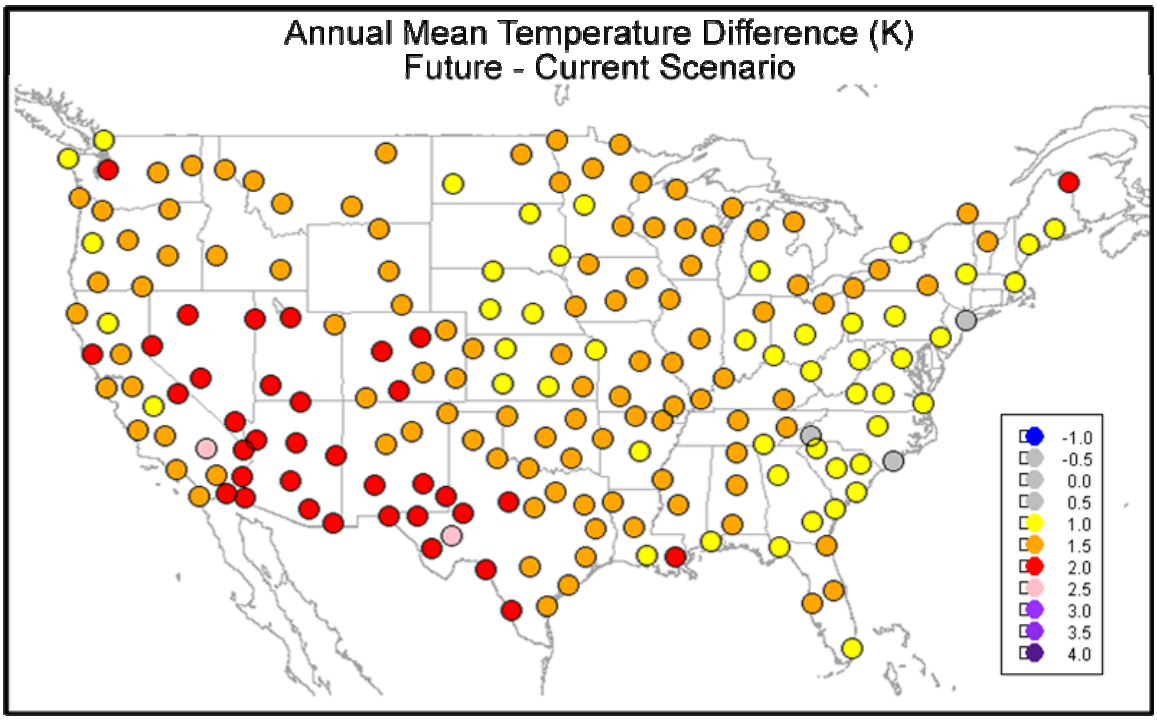


Figure 49. The difference in annual mean of the hourly temperature (K) for the future versus current regional climate simulations.

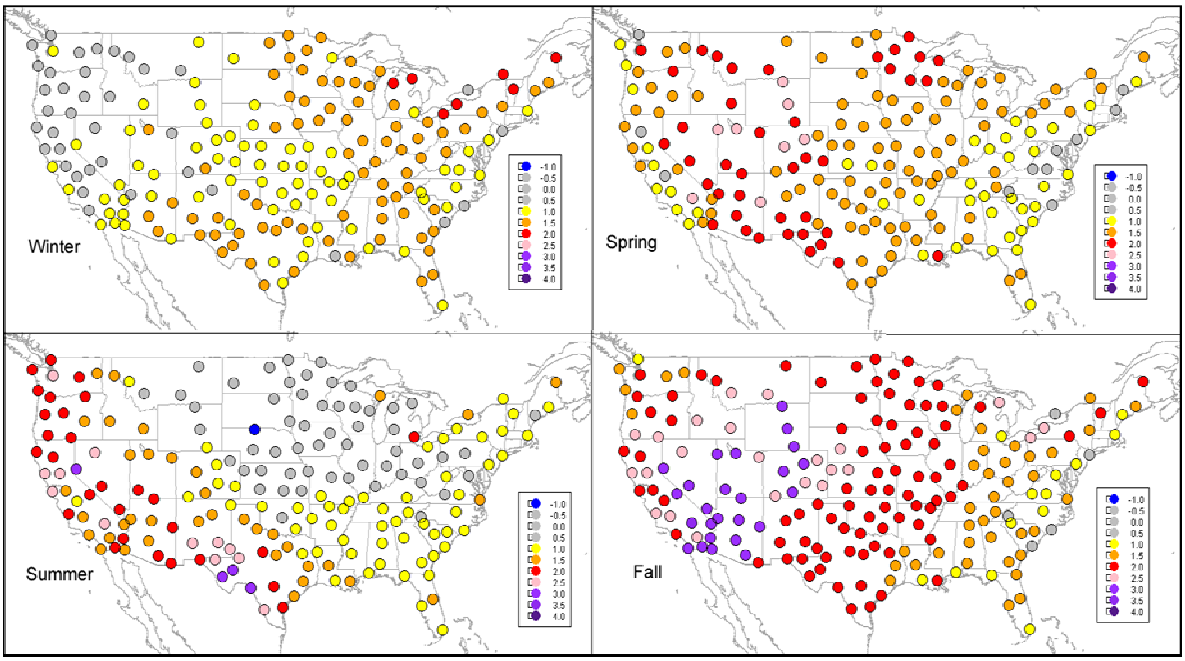


Figure 50. Same as **Figure 49**, except that the mean temperature difference is calculated separately for each season.

average temperature changes. To assess this, the baseline temperature component, which represents the hourly temperature values after higher frequency variability has been removed using the KZ filtering technique (Section 4.1.3), is considered. Since the baseline temperature component does not include seasonal, synoptic, or diurnal variability; differences in this average baseline component of the temperature are most relevant to the multi-decadal temporal scale of climate change.

The baseline temperature component is shown in Figure 51 and Figure 52 for select NWS sites. Note that as climatological simulations, the timing of the interannual variations should not be compared between the current and future simulations. Rather, we want to determine whether the range of temperatures without higher frequency (e.g., synoptic or seasonal) variations is significantly different between the current and future years for each site. Thus, a boxplot format was chosen to illustrate the difference in the range of base case temperature for the current and future simulations at these sites. It should be noted, however, that these data are serially correlated and this can reduce the range of the interquartile and outlying points. Nevertheless, the comparison of the interquartiles suggests that the interannual range of base case temperatures is not as large as differences between the base case and future temperatures. We see that at many sites, the interannual variability among the 10 years of base case or future simulation is smaller than the difference in the base case versus future temperature. In many cases the overlap between base case and future baseline temperature predictions is only in the outer range of the predicted base case temperatures.

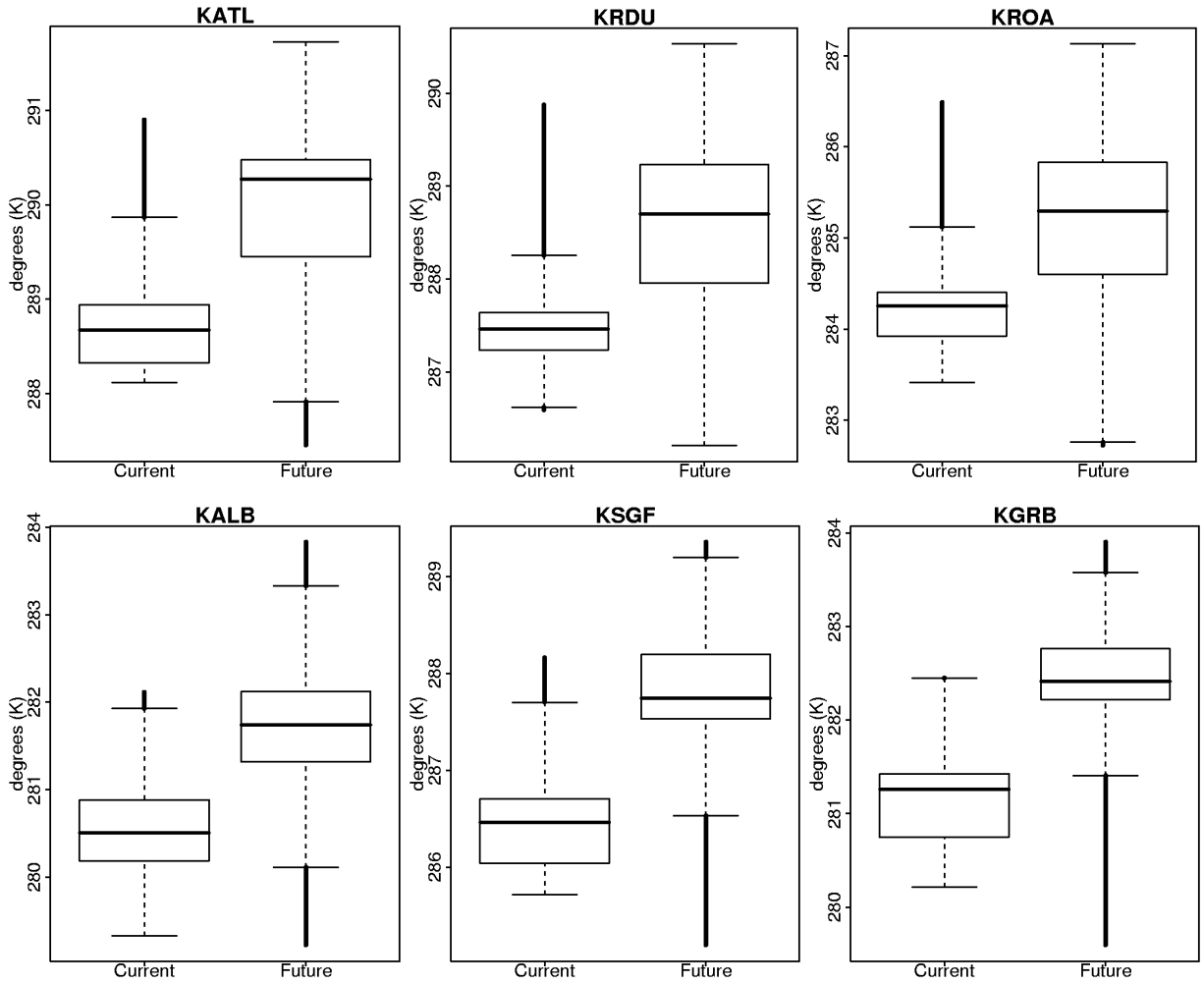


Figure 51. The hourly baseline average component of the 2m temperature (K) for select NWS sites from the Eastern and Midwestern U.S., including Atlanta, GA (KATL), Raleigh, NC (RDU), Roanoke, VA (KROA), Albany, NY (KALB), Springfield, MO (KSGF), and Greenbay, WI (KGRB). Each plot includes the median (horizontal line), the upper 75th and lower 25th quartiles (box), outer fences (75th quartile plus 1.5 times the interquartile range; 25th quartile minus 1.5 times the interquartile range), and outliers (solid dots outside of the fences). The baseline component represents the long-term moving average mean over the ten year simulation with seasonal, synoptic, diurnal, and intra-day variability removed.

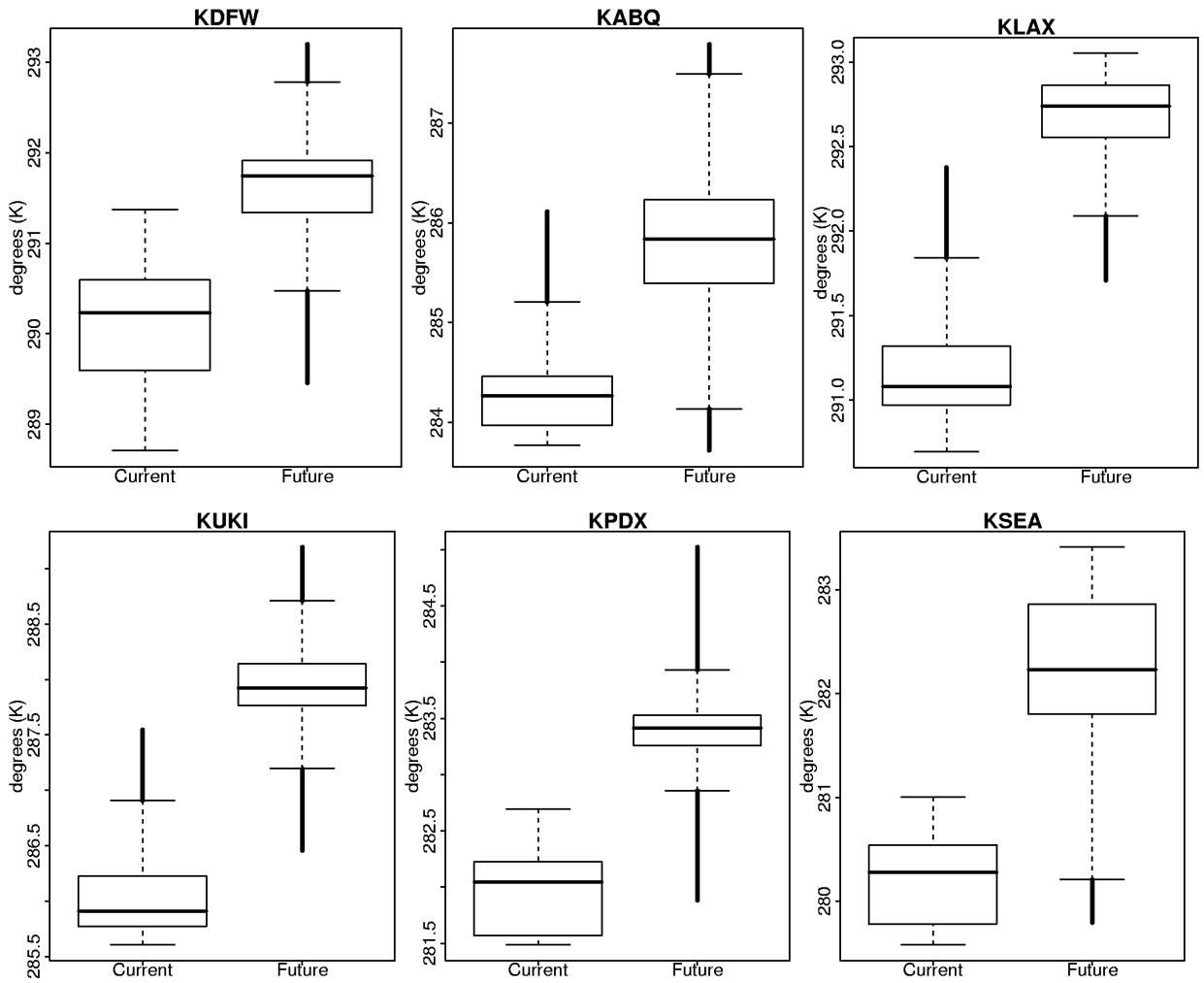


Figure 52. The hourly baseline average component of the 2m temperature (K) for select NWS sites from the Southwestern and Western U.S., including Dallas-Ft. Worth, TX (KDFW), Albuquerque, NM (KABQ), Los Angeles, CA (KLAX), Ukiah, CA (KUKI), Portland, OR (KPDX), and Seattle, WA (KSEA). Each plot includes the median (horizontal line), the upper 75th and lower 25th quartiles (box), outer fences (75th quartile plus 1.5 times the interquartile range; 25th quartile minus 1.5 times the interquartile range), and outliers (solid dots outside of the fences). The baseline component represents the long-term moving average mean over the ten year simulation with seasonal, synoptic, diurnal, and intra-day variability removed.

This is especially noticeable in the western sites in Figure 52, where the only overlap is often for the outlier values only. Based on these results, it is reasonable to conclude that the temperature changes predicted in the regional climate simulations presented here (Figure 49 and Figure 50) represent long-term climate forcing changes and can be distinguished from the interannual variability in the predictions (Figure 51 and Figure 52).

In Figure 50, where differences in mean seasonal temperatures are presented, we see some noticeable differences across seasons for some U.S. regions. Most notable are the differences among the seasons for the Southwest and West Coast. The changes are relatively consistent with the differences in average temperature that are shown in Figure 49, Figure 51 and Figure 52, but seasonal differences may be larger than what can be detected in the comparison of multi-year mean values (e.g., a cold winter bias may be compensated by a warm summer bias). To consider additional differences on the seasonal scale, the seasonal components of the surface temperatures at the NWS sites were analyzed. Differences in seasonal variability are important to consider for air quality. Aerosols formed secondarily in the atmosphere, such as sulfate and nitrate aerosols, have a very pronounced seasonal variation with sulfate concentrations highest (lowest) in the summer (winter) and nitrate concentrations highest (lowest) in the winter (summer). Also, more extreme temperatures in the summer could also affect ozone concentration levels that may be seen in the future CMAQ air quality simulations. We are interested in seeing whether future predictions suggest any substantial changes or shifts in the seasonal variability in temperature or whether we see any major differences in the interannual variability in the seasonal temperature structure.

The seasonal component of the temperature values for many sites in the Eastern and Midwestern U.S. show no notable differences in temperatures across the seasons (not shown). An

increase in hourly temperatures exceeding 311 K (100°F) is noted at some Eastern, Southwestern, and west coast sites (Figure 53), suggesting that synoptic scale extreme temperature events do contribute to the mean average temperature increases seen in these areas during the summer. We anticipate that summer temperature differences could have the most impact on ozone. Previous research has shown strong correlations with daily maximum temperatures and wind speeds [*Wise and Comrie*, (in press)]. Comparing Figure 53 to the summer difference in mean summer season temperatures in Figure 50, some similarities are evident. A decreased number of temperature exceedances are found in the upper Midwest in the future, where the average temperatures show little or no change from the current simulation. There is also a decrease in the mean length of extreme heat episodes in the same area (Figure 54). An increased number of temperatures above 100°F are found in the Southwest and west coast in the future simulation where Figure 53 also shows average temperature increases in the future up to 3 K. The Southwest also shows an increase in the mean length of extreme heat episodes as well (Figure 54). Areas in the Eastern U.S that show a more modest increase in future temperatures between 1 and 2 K show little or no increase in temperatures above 100°F or in mean length of heat episodes.

Where we see an increase in these short-term temperature extremes at many western sites along the California coast, Oregon, and Washington in Figure 53, we also see increased seasonal temperatures in the future. Several examples are shown here. In Ukiah, CA (Figure 55), the summer component of future temperature shows increases up to about 2.5 K in addition to differences in the base case temperature values in Figure 52. Other sites along the west coast, such as Portland, OR (Figure 56), also show increases in the summer temperatures that are slightly larger than one standard deviation of the current seasonal temperature values during some periods.

Similar types of comparisons are evident at many other sites along the west coast such as Seattle, WA and Reno, NV (not shown). For many of these sites, the observed seasonal temperatures show a similar range of difference to the current model simulations (see section 4.2.1) suggesting that the differences are within the range of uncertainty for the model predictions.

In Figure 50, the temperature increases in the fall are very noticeable in the Southwest U.S. The seasonal component of the temperature predictions does not show increased seasonal temperatures at some sites in the area, such as Reno, NV (not shown). Higher-frequency synoptic changes in temperature may be contributing to these large temperature increases.

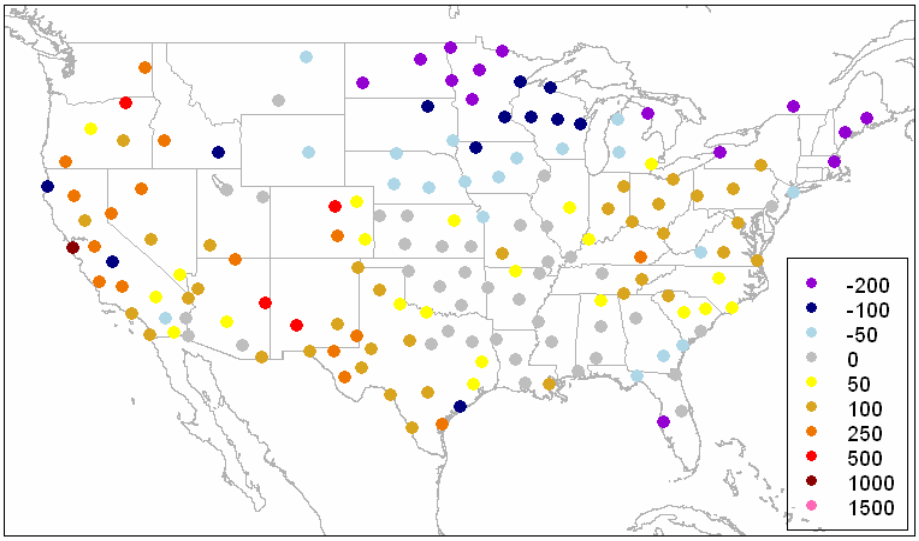


Figure 53. Difference in future versus current number of hourly-modeled temperature values above 100°F (311 K).

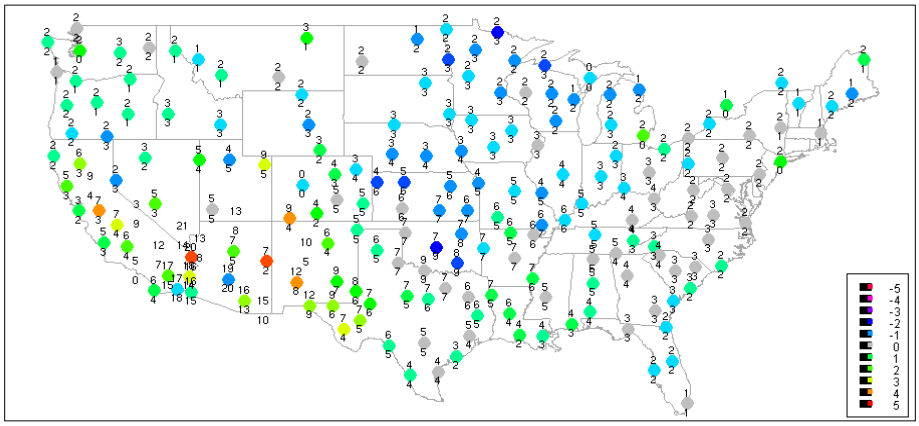


Figure 54. The mean length (days) of heat episodes: Future – Current Scenarios. The numbers on above (below) each marker represent the mean number of days in a heat wave episode in the future (current) scenario. Heat episodes are defined as the number of consecutive days with at least one simulated hourly temperature greater than 90°F (305 K).

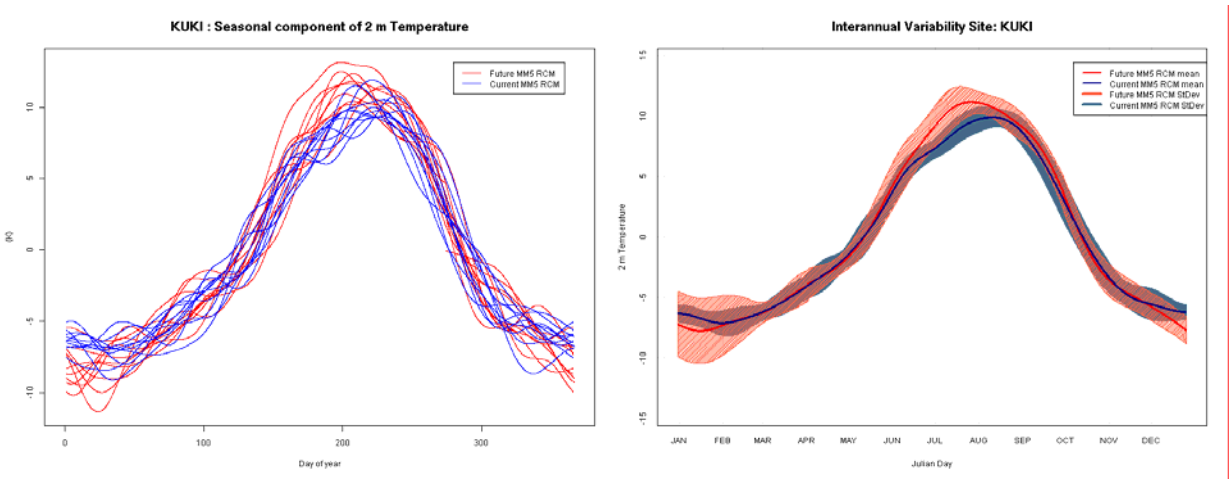


Figure 55. The seasonal component of the current and future modeled temperatures at the NWS site in Ukiah, CA. (a) shows the seasonal temperature component for each year separately to assess interannual variability, and (b) shows the average seasonal temperature component for all ten years and the ±1 σ standard deviation around that mean for all ten years of current and future simulations.

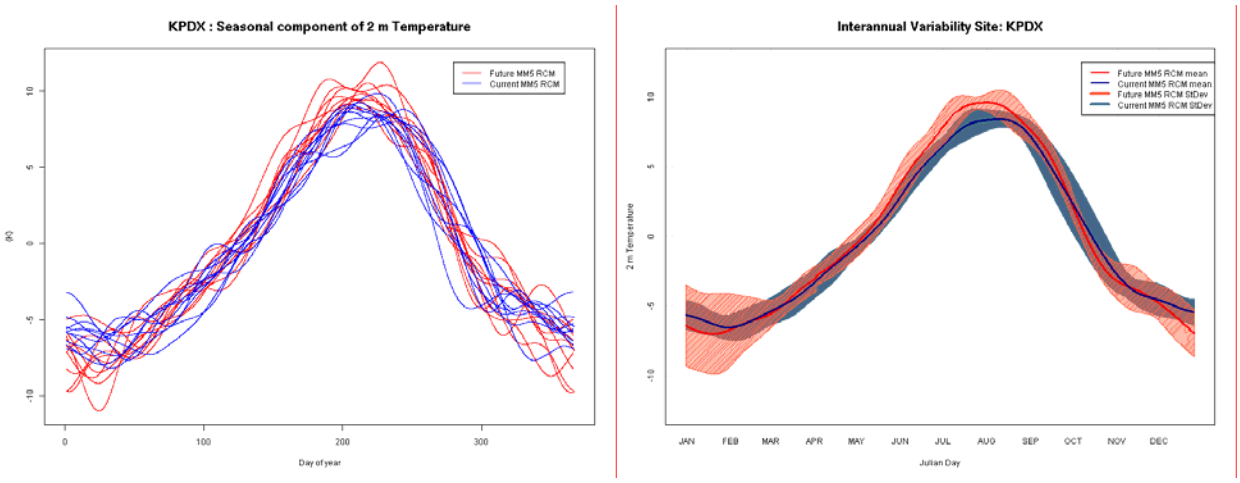


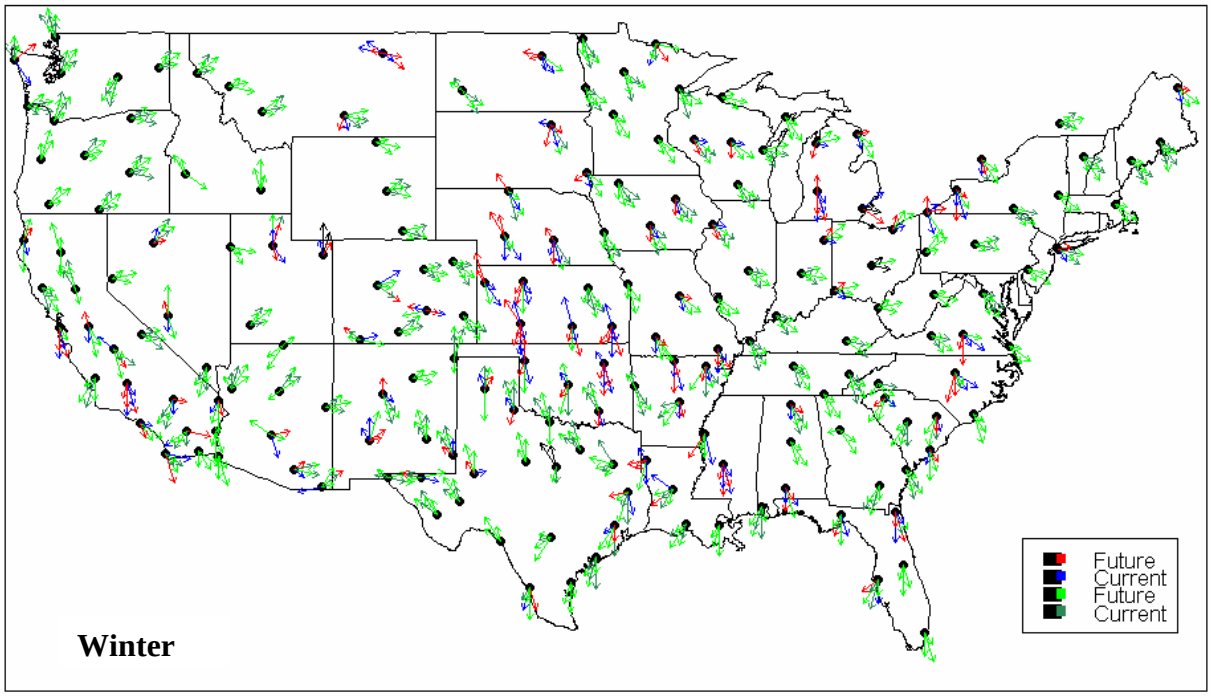
Figure 56. The seasonal component of the current and future modeled temperatures at the NWS site in Portland, OR. (a) shows the seasonal temperature component for each year separately to assess interannual variability, and (b) shows the average seasonal temperature component for all ten years and the ±1 σ standard deviation around that mean for all ten years of current and future simulations.

4.3.2. Wind patterns

A comparison of the current and future global climate simulations that are used as boundary conditions to these RCM simulations is presented elsewhere [*Mickley et al.*, 2004]. Their results suggested an increase in stagnation events in the Eastern U.S. in the future simulation based on global scale modeling. Another preliminary comparison of the current and future RCM simulations [*Leung and Gustafson*, 2005] includes a comparison of stagnation days and unvented hours to consider stagnation event changes. Their comparisons show a significant increase in stagnation days in the future simulation in the Southwestern U.S. during summer and autumn seasons. A decrease in the number of stagnation days was evident in the Southeastern U.S. during summer. Results do suggest that the forthcoming air quality simulations may see some impacts from the predicted changes in stagnation; however, until the air quality simulations are completed, the degree of impact from these changes is unknown.

To consider the impact of surface wind direction and flow patterns on the surface temperature changes shown in the previous section, dominant wind direction comparisons and sea-level pressure patterns are shown in Figure 57 and Figure 58 for the current and future simulations for winter and summer.

In the winter, differences in the future and base case dominate wind direction patterns directly relate to differences in the dominant mean sea-level pressure patterns. Figure 57 indicates that the main difference in the dominant wind flow patterns is over the central portion of the United States, specifically Nebraska, Oklahoma, north Texas and Arkansas. The flow vectors are similar in direction (generally northerly and southerly winds), but the order of dominance of particular modes (vector length) is different at most of the sites. This is a signal that the mean sea-level pressure patterns are generally similar, but the dominance of particular patterns that constitute the mean are reversed. This finding agrees well with results presented in Chapter 3. Figure 58 provides the mean sea-level pressure of both the current and future simulations for the same period as the wind flow. The sea-level pressure patterns indicate that the current simulation has stronger high-pressure axis from the northern Plains into Arkansas. A stronger high in the mean sea-level pressure pattern can indicate the winter high-pressure (Arctic High) is either stronger on average in the current simulation, or that the Arctic High is more persistent/frequent over the center of the country. The latter is likely the main reason, as a more persistent Arctic High will induce differences in the dominant wind directions near the surface. Evidence of this is also present in the mean winter temperature difference between the base case and future scenarios, which indicate that the future simulation has 1-2 K warmer temperatures during the



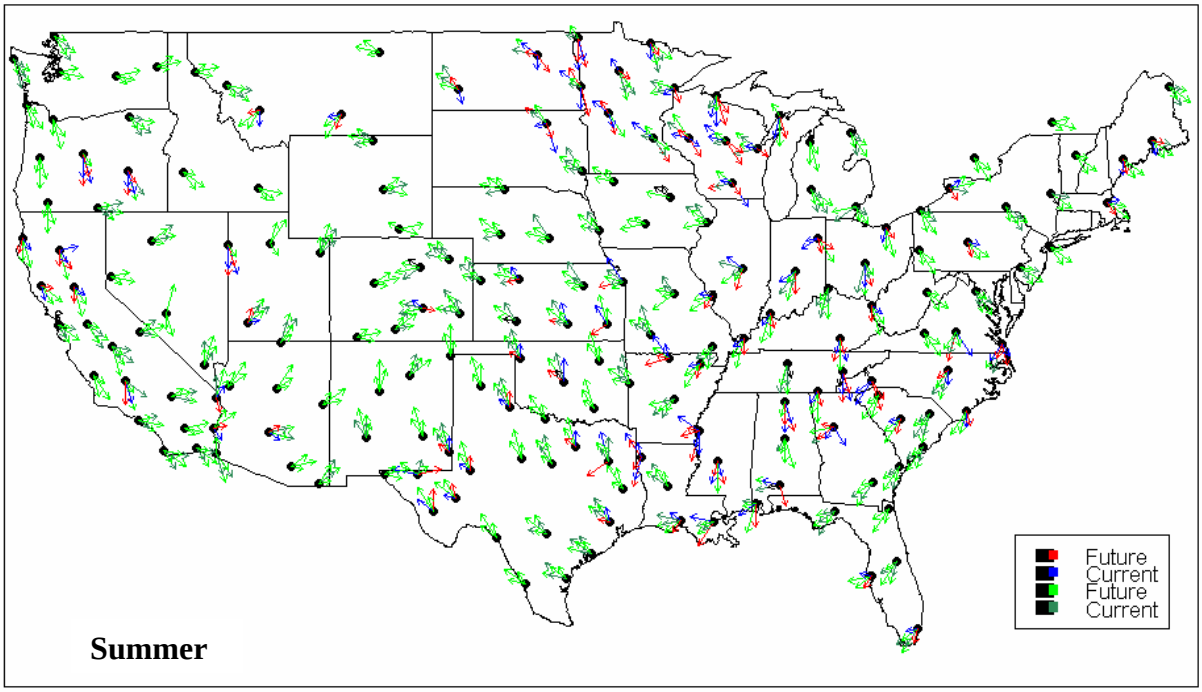


Figure 57. Current and future RCM predicted dominant 10-m wind directions at the climate reference sites for the winter and summer. The length of the vectors represents the frequency of occurrence. Green colored vectors indicate the future (dark green) wind direction is within 40° of the current (light green). If the wind directions are not within 40° of each other, the wind vectors follow a different color patter of blue (current) and future (red).

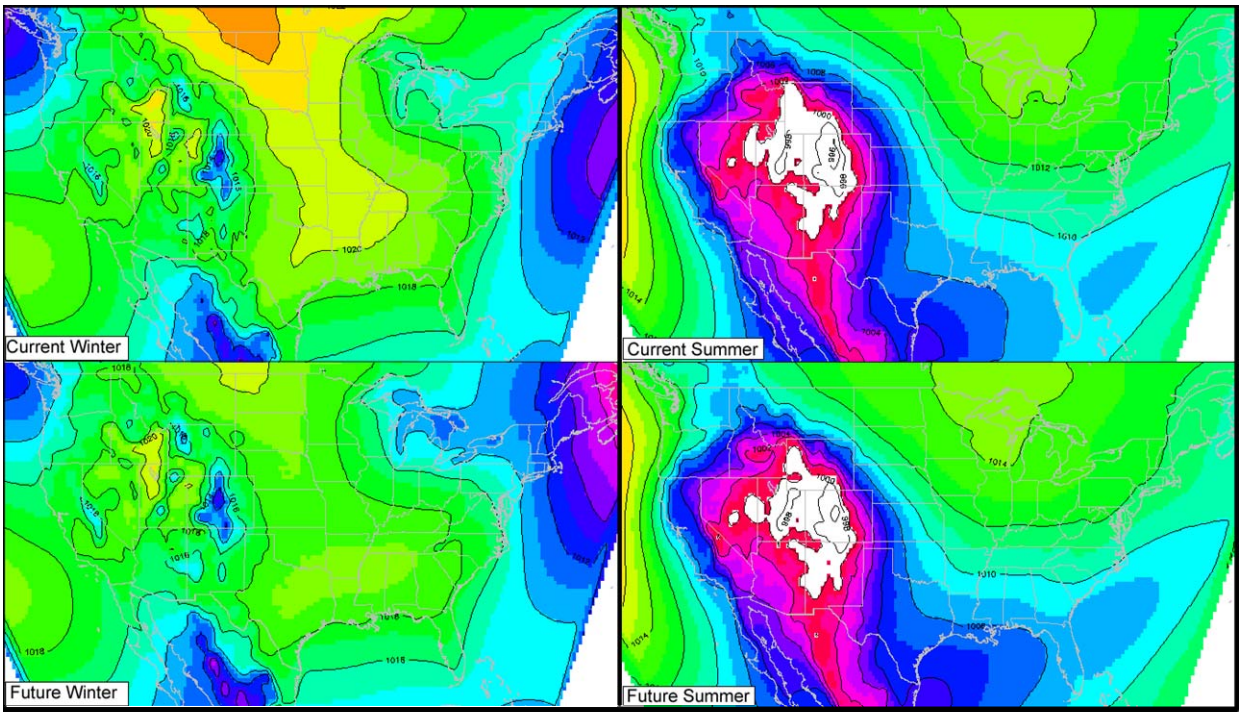


Figure 58. Current (top) and future (bottom) MM5-RCM mean sea-level pressure patterns for the winter (left) and summer (right).

winter over the center of the United States (Figure 50) consistent will a less frequent Arctic high. Weaker overall wind speeds in the future simulation over this region also point to few Arctic high-pressure systems propagating from Canada, which are typically accompanied by strong winds. Few differences exist in the dominant wind modes during the winter over the western U.S. The sea-level pressure patterns support this as they are about the same. Temperature differences in the winter are consistently small over the western United States.

In the summer, Figure 57 indicates few overall differences in the wind flow patterns relative to the wintertime. As with the winter wind directions, more differences exist over the Great Lakes region. Most other differences are a result of slight differences in the rank of the wind direction modes. The mean sea level pressure patterns in Figure 58 for the summer indicate there

is very little difference in the overall summer patterns. A slight difference in the high-pressure placement near the Great Lakes exists. In the future simulation the high-pressure axis is over Wisconsin rather than Michigan, which is likely linked to the differences in the dominant summer wind modes over this area. Sites in Wisconsin and Michigan have more northerly (future, red) than westerly (current, blue) winds, which is consistent with more westward placement of the summer high-pressure. Mean temperatures over the Great Lakes are similar between the two, so no clear reaction of temperature to small changes in this high-pressure placement is seen. Differences in the dominant wind direction mode exist over the Pacific Northwest, but the difference is in the rank of the modes. Sea-level pressure patterns are similar over the entire western United States, yet temperatures in the western United States are much warmer in the future simulation compared to the current. Warmer sea temperature could influence the coastal region, but not inland. Summer rainfall is about the same over most of this area where the temperatures are much warmer. The temperature difference may be a long-term warming trend related to radiative changes that are not reflected in these other weather parameters.

4.3.3 Precipitation

In Figure 59, the average precipitation for the current and future 10 years of simulation is presented along with the difference between the base case and future predictions. Differences show only small changes in average precipitation across most of the United States. An average increase between 10 and 20 cm is evident in the model predictions for the Southeast. This is an increase of approximately 20-30%. Some larger increases of about 50% are also seen at a few Southeastern coastal sites such as Wilmington, NC. Differences seen on the western coast are less than a 10% since the precipitation totals are much higher in that region. Differences in total

precipitation amounts are particularly relevant to atmospheric deposition issues such as nutrient deposition to coastal bay areas. Increased (decreased) precipitation in the future predictions could also lead to decreases (increases) in summertime pollution events. Figure 60 shows similar average precipitation plots and the different between the future and base case simulations for summer seasons only. Many of the differences in Figure 59 range between ±10 cm. These are relatively small changes in precipitation volume as compared to some seasons; however, the change represents an approximate 50% increase in precipitation amounts because total volumes are low. This could suggest that there will be an increase in predicted removal and deposition of air pollutants, but the relative impact on air pollution events will depend on the frequency and duration of the precipitation events.

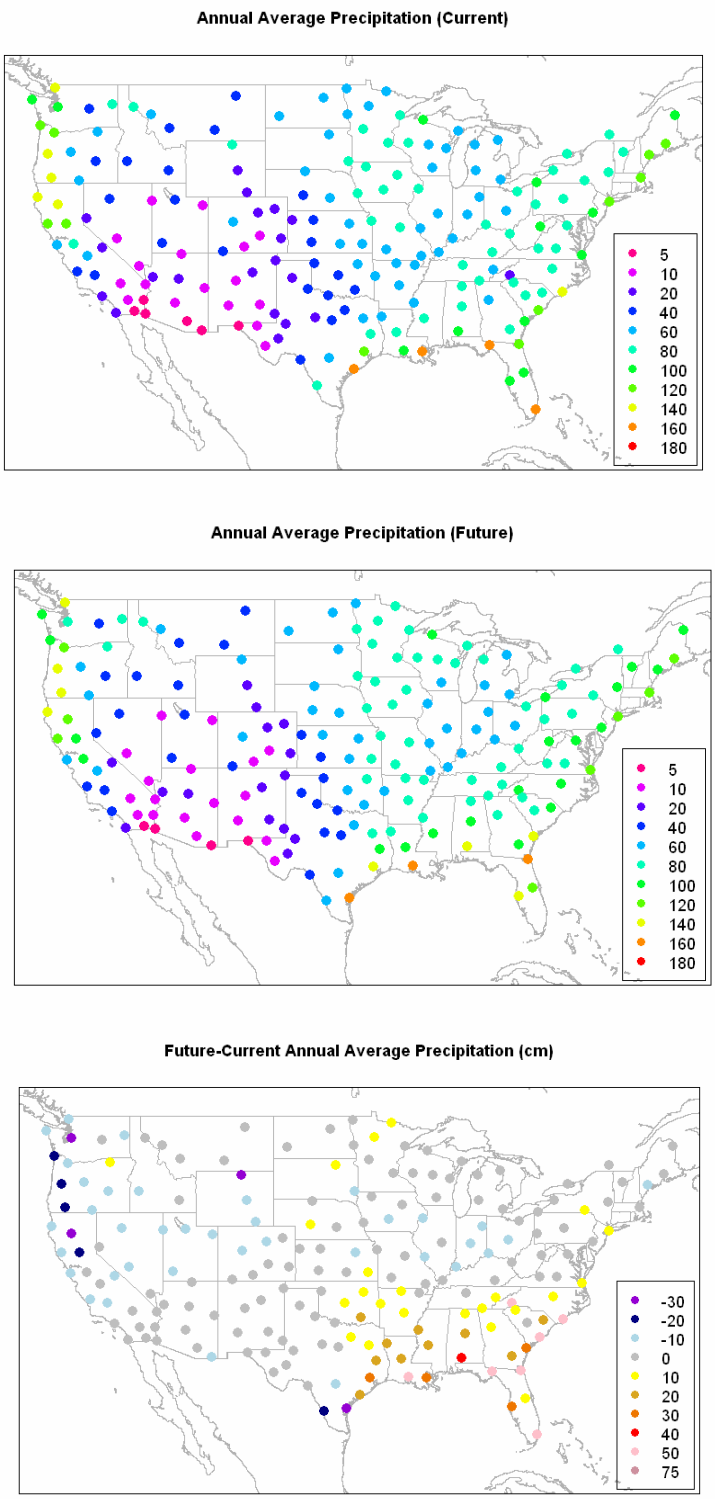


Figure 59. Average rainfall (cm) for the future and current scenarios and the different between Future-Current Scenario.

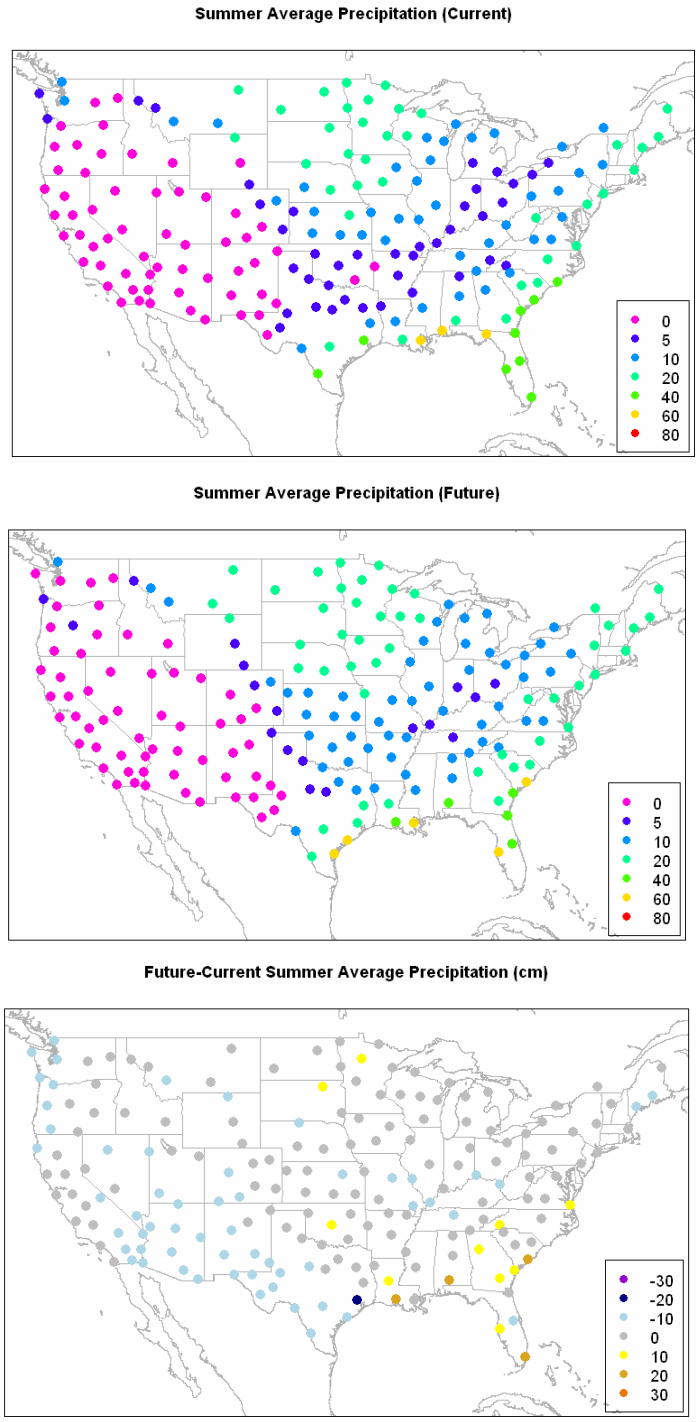


Figure 60. The summer average rainfall (cm) for the future and base case scenarios and the different between Future-Base Case scenario.

5.0 Summary

Two decades of hourly base case and future regional climate model scenarios have been analyzed with regard to 1) their temporal and spatial characterization of climate processes relative to present-day observations and 2) their description of projected climate changes in preparation for air quality simulations that are currently under development. Detailed discussions of current base case and future surface pressure, wind, temperature and precipitation and steering level wind scenarios have been provided in Sections 3 and 4 of this report. Key findings are summarized below. Some of these findings suggest clear opportunities for meaningful air quality assessment applications. Other findings suggest geographic regions or time periods for which interpretation of air quality model projections based on these scenarios should be approached with caution.

5.1 Base Case Climate Simulation

- The GISS/RCM performs well relative to observed climate for some seasons and geographic regions of the United States. On the whole, spring season comparisons show the best surface and steering level climate agreement for the U.S. Winter season dominant near-surface wind flow regimes, temperature and precipitation are also simulated well over the Southern Great Plains. Areas in the northeastern United States and New England are also reasonably well represented by the climate model during the winter.

- The general representation of the summer climate over the eastern one-half of the United

States is poor, including the Northeast, Mid-Atlantic, Southeast, Deep South, Ohio Valley, Mississippi Valley and much of the Midwest. There are major differences in the mean sea-level pressure, surface and steering level wind flow patterns, weaker and more variable steering level wind speeds, a noticeable warm (but rarely significant) bias in temperature, more variability/amplitude in the diurnal temperature signal, and significantly drier summer climate in terms of rainfall. The source of these large-scale summer differences appears to be poor climate model simulation of the location of warm Gulf Stream water, resulting in poor characterization of anti-cyclonic circulation (Bermuda High) over the Southern U.S. The significance of these differences for pollutant concentrations and transport are yet to be determined pending completion of the air quality model simulations.

- With the exception of the Northeast and New England, the simulated current climate over the Eastern United States during the wintertime is poor. The evidence suggests that the simulated climate has a suppressed polar Jet Stream, which frequently allows unusually cold air masses to anchor over most of the Eastern United States. This also holds the active storm track much further south and east of its climatological position, preventing winter storms from reaching into the central United States and East Coast. Sea-level pressure and wind flow patterns indicate that the GISS/RCM does not reproduce the observed southerly wind component frequency over the southeastern quadrant of the United States. The GISS/RCM temperatures exhibit a statistically significant 3-5 K cold bias in these areas and precipitation that is 50% or less than the observed climate.

- Comparisons for the western one-half of the United States show mixed results during all seasons. Mean sea-level pressure patterns in both winter and summer were similar along and off the West Coast, but significant steering level differences were noted. In particular, winter season steering level features suggest differences that could influence the frequency and direction of on-shore transport of globally transported pollutants into the continental U.S. as well as the geographic distribution of winter monsoon rainfall. Winter season air quality scenario results for the West Coast of the U.S. should be interpreted with caution until these meteorological differences can be more thoroughly investigated.

- There is substantial uncertainty in estimates of the frequency with which dominant steering level transport patterns occur during summer, autumn and winter seasons in both the reanalysis and base case GISS/RCM datasets. Given the relatively high level of visual similarity noted previously, the most likely source of this uncertainty lies in the temporal correlation structure of the datasets, i.e., pattern persistence and transitions from one dominant pattern to another. Caution should be exercised when considering statements regarding the significance of transport frequency differences until the underlying driver for these differences is better understood.

5.2 Future Climate Conditions

- Initial results show that indeed a temperature increase is evident during many seasons in the regional climate simulations. Preliminary statistical analysis indicates significant summertime temperatures changes along the West Coast, the Gulf of Mexico, Southeastern Atlantic Coast and Southern Texas only, with some increases in the frequency and

maximum temperatures of extreme temperature events. Winter and autumn season temperature changes appear to be significant throughout the continental domain. Mean spring season base case temperatures are generally not statistically different than future scenario means.

- Precipitation changes in the model predictions suggest that increased precipitation in the Eastern U.S. that may lead to an increase in atmospheric wet deposition of pollutants during the summer season, but the overall impact on air quality episodes is to be determined based on the air quality model simulations.

- Dominant steering level transport patterns suggest little visual change from base case to future time periods, but there may be noticeable changes in the frequency and/or persistence of the base case patterns that could significantly impact future air quality conditions.

6.0 References

Anderberg, M.R., *Cluster Analysis for Applications*, 359 pp., Academic Press, New York, 1973.

Byun, D.W., and J.K.S. Ching, Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, United Stated Environmental Protection Agency, Office of Research and Development, Washington D.C., 1999.

Climate Change Science Program (CCSP), Strategic Plan for the U.S. Climate Change Science Program, 202, 2003.

Cohn, R.D., B.K. Eder, S.K. LeDuc, and R.L. Dennis, Development of an Aggregation and Episode Selection Scheme to Support the Models-3 Community Mulitscale Air Quality Model, *Journal of Applied Meteorology*, *40*, 210-228, 2001.

Dirks, R.A., J.P. Kuettner, and J.A. Moore, Genesis of Atlantic Lows Experiment (GALE): An overview, *Bulletin of the American Meteorological Society*, *69*, 148-160, 1988.

Dudhia, J., Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, *Journal of Atmospheric Science*, *46*, 3077-3107, 1989.

Eder, B.K., J.M. Davis, and P. Bloomfield, An Automated Classification Scheme Designed to Better Elucidate the Dependence of Ozone on Meteorology, *Journal of Applied Meteorology*, *33* (10), 1182-1199, 1994.

Eskridge, R.E., J.Y. Ku, S.T. Rao, P.S. Porter, and I.G. Zurbenko, Separating Different Scales of Motion in Time Series of Meteorological Variables, *Bulletin of the American Meteorological Society*, *78* (7), 1473-1483, 1997.

Fernau, M.E., and P.J. Samson, Use of Cluster Analysis to Define Periods of Similar Meteorology and Precipitation Chemistry in Eastern North America. Part I: Transport Patterns, *Journal of Applied Meteorology*, *29*, 735-750, 1990.

Grell, G., Prognostic evaluation of assumptions used by cumulus parameterizations, *Monthly Weather Review*, *119*, 764-787, 1993.

Grell, G.A., J. Dudhia, and D. Stauffer, A description of the fifth-generation Penn State/NCAR

Mesoscale Model (MM5), in *NCAR Technical Note*, National Center for Atmospheric Research, Boulder, 1994.

Hirsch, M.E., A.T. DeGaetano, and S.J. Colucci, An East Coast Winter Strom Climatology, *Journal of Climate*, *14*, 882-899, 2001.

Hogrefe, C., J. Biswas, B. Lynn, K. Civerolo, J.-Y. Ku, J. Rosenthal, C. Rosenzweig, R. Goldberg, and P.L. Kinney, Simulating regional-scale ozone climatology over the eastern United States: model evaluation results, *Atmospheric Environment*, *38*, 2627-2638, 2004a.

Hogrefe, C., B. Lynn, K. Civerolo, J.-Y. Ku, J. Rosenthal, C. Rosenzweig, R. Goldberg, S. Gaffin, K. Knowlton, and P.L. Kinney, Simulating changes in regional air pollution over the eastern United States due to changes in global and regional climate and emissions, *Journal of Geophysical Research*, *109* (D22301), 22301-223021, 2004b.

Hogrefe, C., S.T. Rao, I.G. Zurbenko, and P.S. Porter, Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the eastern United States, *Bulletin of the American Meteorological Society*, *81*, 2083-2106, 2000.

Hurrell, J.W., and K.E. Trenberth, Global Sea Surface Temperature Analyses: Multiple Problems and Their Implications for Climate Analysis Modeling and Reanalysis, *Bulletin of the American Meteorological Society*, *80* (12), 2661-2678, 1999.

Kalkstein, L.S., G. Tan, and J.A. Skindlov, An evaluation of three clustering procedures for use in synoptic climatological classfiication, *Journal of Climate and Applied Meteorology*, *26*, 717-730, 1987.

Kalnay, E., and Coauthors, The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, *77*, 437-471, 1996.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J.J. Hnilo, M. Fiorino, and G.L. Potter, NCEP-DOE AMIP-II Reanalysis (R-2), *Bulletin of the American Meteorological Society*, *83*, 1631-1643, 2002.

Kasahara, A., A comparison between geostrophic and nongeostrophic numerical forecasts of hurricane movement with the barotropic steering model, *Journal of Meteorology*, *16*, 371-384, 1959.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H.v.d. Dool, R. Jenne, and M. Fiorino, The NCEP-NCAR 50-Year reanalysis: Monthly means CD-ROM and documentation, *Bulletin of the American Meteorological Society*, *82* (2), 247-267, 2001.

Leung, L.R., and W.I. Gustafson, Potential Influence of Regional Climate Change on U.S. Air Quality, *Geophysical Research Letters*, *32*, L16711, 2005.

Leung, L.R., Y. Qian, and X. Bian, Hydroclimate of the Western United States Based on Observations and Regional Climate Simulation of 1981-2000. Part I: Seasonal Statistics, *Journal of Climate*, *16* (12), 1892-1911, 2003.

Lynn, B.H., L. Druyan, C. Hogrefe, J. Judhia, C. Rosenzweig, R. Goldberg, D. Rind, R. Healy, J. Rosenthal, and P. Kinney, Sensitivity of Present and Future Surface Temperatures to Precipitation Characteristics, *Climate Research*, *28* (1), 53-65, 2004.

Lynn, B.H., C. Rosenzweig, R. Goldberg, D. Rind, C. Hogrefe, R. Healy, J. Dudhia, J. Rosenthal, and P. Kinny, Sensitivity of Simulated Current and Future Climate to Model Physics Configuration and Grid Resolution, *Journal of Applied Meteorology*, submitted.

Mass, C.F., M. Albright, D. Ovens, R. Steed, M. MacIver, E. Grimit, T. Eckel, B. Lamb, J. Vaughan, K. Westrick, P. Storck, B. Colman, C. Hill, N. Maykut, M. Gilroy, S.A. Ferguson, J. Yetter, J.M. Sierchio, C. Bowman, R. Stender, R. Wilson, and W. Brown, Regional environmental prediction over the Pacific Northwest, *Bulletin of the American Meteorological Society*, *84*, 1353-1366, 2003.

Mickley, L.J., D.J. Jacob, B.D. Field, and D. Rind, Effects of future climate change on regional air pollution episodes in the United States, *Geophysical Research Letters*, *31*, L24103, 2004.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough, Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave, *Journal of Geophysical Research*, *102* (D14), 16663-16682, 1997.

Mo, D.C., and M. Ghil, Cluster analysis of multiple planetary flow regimes, *Journal of Geophysical Research*, *D93*, 10927-10952, 1988.

National Envrionmental Scientific Computing Center, User Guide Version 9.2,

OEI/OTOP/NTSD/NESCC, 2004.

NRC, *Global Air Quality: An Imperative for Long-Term Observational Strategies*, 56 pp., Committee on Atmospheric Chemistry, Board on Atmospheric Sciences and Climate, National Research Council, National Academy Press, Washington, D.C., 2001.

Rao, S.T., I.G. Zurbenko, R. Neagu, P.S. Proter, J.Y. Ku, and R.F. Henry, Space and Time Scales in Ambient Ozone Data, *Bulletin of the American Meteorological Society*, *78* (10), 2153-2166, 1997.

Reisner, J., J. Rasmussen, and R. Bruintjes, Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model, *Quarterly Journal of the Royal Meteorological Society*, *124B*, 1071-1107, 1998.

Rind, D., J. Lerner, K. Shah, and R. Suozzo, Use of on-line trancers as a diagnostic tool in general circulation model development2. Transport between the troposhere and stratosphere, *Journal of Geophysical Research*, *104(D8)*, 9151-9168, 1999.

Rolph, G.D., Eta Data Assimilation System (EDAS) Archive Information (TD-6141), pp. 9, NOAA-Air Resources Laboratory, 2002.

Russell, G.L., J.R. Miller, and D. Rind, A Coupled Atmosphere-Ocean Model for Transient Climate Change Studies, *Atmosphere-Ocean*, *33* (4), 683-730, 1995.

Ward, J.H., Hierarchical Groupings to Optimize an Objective Function, *Journal of the American Statistical Association*, *58*, 236-244, 1963.

Wilks, D.S., *Statistical Methods in the Atmospheric Sciences*, 467 pp., Academic Press, San Diego, 1995.

Wise, E.K., and A.C. Comrie, Meteorologically adjusted urban air quality trends in the Southwestern United States, *Atmospheric Environment*, (in press).

Blank Page

Appendix A. List of National Weather Service (NWS) sites used in climate analysis

Site ID	Latitude	Longitude	City State	
CYTR	44.12	-77.53	Trenton, Ont.	
CYUL	45.47	-73.75	Montreal / Dorval International, Que	
KABI	32.42	-99.68	Abilene, Abilene Regional Airport	TX
KABQ	35.05	-106.62	Albuquerque, Albuquerque International Airport	NM
KABR	45.45	-98.43	Aberdeen, Aberdeen Regional Airport	SD
KACV	40.98	-124.1	Arcata / Eureka, Arcata Airport	CA
KAEX	31.33	-92.55	Alexandria, Alexandria International Airport	LA
KAKO	40.17	-103.22	Akron, Akron-Washington County Airport	CO
KALB	42.75	-73.8	Albany, Albany County Airport	NY
KALO	42.55	-92.4	Waterloo, Waterloo Municipal Airport	IA
KALS	37.45	-105.87	Alamosa, San Luis Valley Regional Airport	CO
KAMA	35.23	-101.7	Amarillo, Amarillo International Airport	TX
KAMG	31.53	-82.5	Alma, Bacon County Airport	GA
KAPN	45.07	-83.57	Alpena, Alpena County Regional Airport	MI
KAQR	34.4	-96.15	Atoka, Atoka Municipal Airport	FL
KARA	30.03	-91.88	New Iberia, Acadiana Regional Airport	LA
KARG	36.13	-90.92	Walnut Ridge Automatic Weather Observing / Reporting System	AR
KASE	39.22	-106.87	Aspen, Aspen-Pitkin County Airport	CO
KAST	46.15	-123.88	Astoria, Astoria Regional Airport	OR
KATL	33.65	-84.42	Atlanta, Hartsfield Atlanta International Airport	GA
KAVL	35.43	-82.54	Asheville, Asheville Regional Airport	NC
KAXN	45.87	-95.4	Alexandria, Chandler Field	MN
KBFL	35.43	-119.05	Bakersfield, Meadows Field Airport	CA
KBGM	42.22	-75.98	Binghamton, Binghamton Regional Airport	NY
KBGR	44.8	-68.83	Bangor, Bangor International Airport	ME
KBIH	37.37	-118.37	Bishop, Bishop Airport	CA
KBIL	45.8	-108.53	Billings, Billings Logan International Airport	MT
KBJI	47.5	-94.93	Bemidji	MN
KBLH	33.62	-114.72	Blythe, Blythe Airport	CA
KBLI	48.8	-122.53	Bellingham, Bellingham International Airport	WA
KBNA	36.13	-86.68	Nashville, Nashville International Airport	TN
KBNO	43.58	-118.95	Burns, Burns Municipal Airport	OR
KBOI	43.57	-116.22	Boise, Boise Air Terminal	ID
KBOS	42.37	-71.02	Boston, Logan International Airport	MA
KBTM	45.95	-112.5	Butte, Bert Mooney Airport	MT
KBUF	42.93	-78.73	Buffalo, Greater Buffalo International Airport	NY
KCAE	33.95	-81.12	Columbia, Columbia Metropolitan Airport	SC
KCAO	36.45	-103.15	Clayton, Clayton Municipal Airpark	NM
KCAR	46.87	-68.02	Caribou, Caribou Municipal Airport	ME
KCDC	37.7	-113.1	Cedar City, Cedar City Municipal Airport	UT
KCDS	34.43	-100.28	Childress, Childress Municipal Airport	TX
KCHA	35.03	-85.2	Chattanooga, Lovell Field	TN
KCHS	32.9	-80.03	Charleston, Charleston Air Force Base	SC
KCLE	41.42	-81.85	Cleveland, Cleveland-Hopkins International Airport	OH

Site ID	Latitude	Longitude	City State	
KCMH	40	-82.88	Columbus, Port Columbus International Airport	OH
KCNU	37.67	-95.48	Chanute, Chanute Martin Johnson Airport	KS
KCOU	38.82	-92.22	Columbia, Columbia Regional Airport	MO
KCPR	42.9	-106.47	Casper, Natrona County International Airport	WY
KCRP	27.77	-97.5	Corpus Christi, Corpus Christi International Airport	TX
KCVG	39.05	-84.67	Covington / Cincinnati, Cincinnati / Northern Kentucky International Airport	KY
KCWA	44.78	-89.67	Mosinee / Central Wisconsin	WI
KDAG	34.85	-116.78	Daggett, Barstow-Daggett Airport	CA
KDDC	37.77	-99.97	Dodge City, Dodge City Regional Airport	KS
KDEC	39.83	-88.87	Decatur, Decatur Airport	IL
KDEN	39.87	-104.67	Denver, Denver International Airport	CO
KDFW	32.9	-97.03	Dallas / Fort Worth, Dallas / Fort Worth International Airport	TX
KDIK	46.8	-102.8	Dickinson, Dickinson Municipal Airport	ND
KDLH	46.85	-92.18	Duluth, Duluth International Airport	MN
KDRO	37.15	-107.75	Durango, Durango-La Plata County Airport	CO
KDRT	29.37	-100.92	Del Rio, Del Rio International Airport	TX
KDSM	41.53	-93.65	Des Moines, Des Moines International Airport	IA
KDTW	42.23	-83.33	Detroit, Detroit Metropolitan Wayne County Airport	MI
KDUG	31.47	-109.6	Douglas Bisbee, Bisbee Douglas International Airport	AZ
KDVL	48.12	-98.92	Devils Lake Automatic Weather Observing / Reporting System	ND
KDVN	41.62	-90.58	Davenport, Davenport Municipal Airport	IA
KEAU	44.87	-91.48	Eau Claire, Chippewa Valley Regional Airport	WI
KEED	34.77	-114.62	Needles, Needles Airport	CA
KEET	33.18	-86.78	Alabaster, Shelby County Airport	AL
KEKN	38.88	-79.85	Elkins, Elkins-Randolph County-Jennings Randolph Field	WV
KELP	31.8	-106.4	El Paso, El Paso International Airport	TX
KENV	40.73	-114.03	Wendover / Air Force Auxillary Field	UT
KEPH	47.3	-119.52	Ephrata, Ephrata Municipal Airport	WA
KERI	42.08	-80.18	Erie, Erie International Airport	PA
KESC	45.75	-87.03	Escanaba	MI
KEUG	44.12	-123.22	Eugene, Mahlon Sweet Field	OR
KEVV	38.05	-87.53	Evansville, Evansville Regional Airport	IN
KFAR	46.9	-96.8	Fargo, Hector International Airport	ND
KFAT	36.77	-119.72	Fresno, Fresno Air Terminal	CA
KFLG	35.13	-111.67	Flagstaff, Flagstaff Pulliam Airport	AZ
KFLO	34.18	-79.72	Florence, Florence Regional Airport	SC
KFLP	36.3	-92.47	Flippin Automatic Weather Observing / Reporting System	AR
KFSD	43.58	-96.75	Sioux Falls, Foss Field	SD
KFSM	35.33	-94.37	Fort Smith, Fort Smith Regional Airport	AR
KFST	30.91	-102.91	Fort Stockton, Fort Stockton-Pecos County Airport	TX
KFVX	37.35	-78.43	Farmville	VA
KFWA	41	-85.2	Fort Wayne, Fort Wayne International Airport	IN
KGAG	36.3	-99.77	Gage, Gage Airport	OK
KGDP	31.83	-104.8	Pine Springs, Guadalupe Mountains National Park	TX

Site ID	Latitude	Longitude	City State	
KGEG	47.63	-117.53	Spokane, Spokane International Airport	WA
KGGW	48.22	-106.62	Glasgow, Glasgow International Airport	MT
KGLD	39.37	-101.7	Goodland, Renner Field	KS
KGLH	33.48	-90.98	Greenville, Mid Delta Regional Airport	MS
KGRB	44.48	-88.13	Green Bay, Austin Straubel International Airport	WI
KGRI	40.97	-98.32	Grand Island, Central Nebraska Regional Airport	NE
KGRR	42.88	-85.52	Grand Rapids, Gerald R. Ford International Airport	MI
KGSP	34.88	-82.22	Greer, Greenville-Spartanburg Airport	SC
KGZH	31.42	-87.03	Evergreen, Middleton Field	AL
KHCO	48.76	-96.93	Hallock, Hallock Municipal Airport	MN
KHLC	39.38	-99.83	Hill City, Hill City Municipal Airport	KS
KHOB	32.68	-103.22	Hobbs / Lea County	NM
KHSV	34.65	-86.77	Huntsville, Huntsville International / Jones Field	AL
KHTS	38.37	-82.55	Huntington, Tri-State Airport	WV
KIAD	38.95	-77.45	Washington DC, Washington-Dulles International Airport	VA
KIAH	29.97	-95.35	Houston, Houston Intercontinental Airport	TX
KICT	37.65	-97.43	Wichita, Wichita Mid-Continent Airport	KS
KIGM	35.27	-113.95	Kingman, Kingman Airport	AZ
KILM	34.27	-77.9	Wilmington, Wilmington International Airport	NC
KIND	39.73	-86.27	Indianapolis, Indianapolis International Airport	IN
KINL	48.57	-93.38	International Falls, Falls International Airport	MN
KIPL	32.83	-115.58	Imperial, Imperial County Airport	CA
KIWD	46.53	-90.13	Ironwood Automatic Weather Observing / Reporting System	MI
KJAN	32.32	-90.08	Jackson, Jackson International Airport	MS
KJAX	30.5	-81.7	Jacksonville, Jacksonville International Airport	FL
KJFK	40.65	-73.78	New York, Kennedy International Airport	NY
KLAA	38.07	-102.68	Lamar, Lamar Municipal Airport	CO
KLAR	41.32	-105.67	Laramie, Laramie Regional Airport	WY
KLAS	36.08	-115.17	Las Vegas, McCarran International Airport	NV
KLAX	33.93	-118.4	Los Angeles, Los Angeles International Airport	CA
KLBF	41.13	-100.68	North Platte, North Platte Regional Airport	NE
KLEW	44.05	-70.28	Auburn-Lewiston	ME
KLFK	31.23	-94.75	Lufkin, Angelina County Airport	TX
KLIT	34.73	-92.23	Little Rock, Adams Field	AR
KLKV	42.17	-120.4	Lakeview Automatic Weather Observing / Reporting System	OR
KLOZ	37.08	-84.08	London, London-Corbin Airport-Magee Field	KY
KLRD	27.55	-99.47	Laredo International Airport	TX
KLVS	35.65	-105.15	Las Vegas, Las Vegas Municipal Airport	NM
KMAF	31.95	-102.22	Midland, Midland International Airport	TX
KMCF	27.85	-82.52	Macdill Air Force Base, Fl.	FL
KMCI	39.32	-94.72	Kansas City, Kansas City International Airport	MO
KMCO	28.43	-81.32	Orlando, Orlando International Airport	FL
KMFR	42.37	-122.87	Medford, Rogue Valley International Airport	OR
KMHK	39.13	-96.67	Manhattan, Manhattan Municipal Airport	KS
KMIA	25.82	-80.28	Miami, Miami International Airport	FL
KMLP	47.45	-115.68	Mullan Pass, Mullan Pass Vor	ID

Site ID	Latitude	Longitude	City State	
KMOB	30.68	-88.25	Mobile, Mobile Regional Airport	AL
KMOD	37.63	-120.95	Modesto, Modesto City-County-Sham Field	CA
KMPV	44.2	-72.57	Barre / Montpelier, Knapp State Airport	VT
KMRF	30.37	-104.02	Marfa	TX
KMSN	43.13	-89.35	Madison, Dane County Regional-Truax Field	WI
KMSO	46.92	-114.08	Missoula, Missoula International Airport	MT
KMSP	44.88	-93.22	Minneapolis, Minneapolis-St. Paul International Airport	MN
KMSY	29.98	-90.25	New Orleans, New Orleans International Airport	LA
KMYV	39.1	-121.57	Marysville, Yuba County Airport	CA
KOKC	35.4	-97.6	Oklahoma City, Will Rogers World Airport	OK
KOMA	41.3	-95.9	Omaha, Eppley Airfield	NE
KORF	36.9	-76.2	Norfolk, Norfolk International Airport	VA
KPAH	37.07	-88.77	Paducah, Barkley Regional Airport	KY
KPDT	45.68	-118.85	Pendleton, Eastern Oregon Regional At Pendleton Airport	OR
KPDX	45.6	-122.6	Portland, Portland International Airport	OR
KPGA	36.93	-111.45	Page, Page Municipal Airport	AZ
KPHL	39.87	-75.25	Philadelphia, Philadelphia International Airport	PA
KPHX	33.43	-112.02	Phoenix, Phoenix Sky Harbor International Airport	AZ
KPIH	42.92	-112.57	Pocatello, Pocatello Regional Airport	ID
KPIT	40.5	-80.22	Pittsburgh, Pittsburgh International Airport	PA
KPOF	36.77	-90.32	Poplar Bluff, Poplar Bluff Municipal Airport	MO
KPRB	35.67	-120.63	Paso Robles, Paso Robles Municipal Airport	CA
KPSX	28.73	-96.25	Palacios, Palacios Municipal Airport	TX
KPUB	38.28	-104.52	Pueblo, Pueblo Memorial Airport	CO
KRDD	40.5	-122.3	Redding, Redding Municipal Airport	CA
KRDM	44.25	-121.15	Redmond, Roberts Field Airport	OR
KRDU	35.88	-78.79	Raleigh / Durham, Raleigh-Durham International Airport	NC
KRNO	39.5	-119.78	Reno, Reno Tahoe International Airport	NV
KROA	37.32	-79.97	Roanoke, Roanoke Regional Airport	VA
KROW	33.3	-104.53	Roswell, Roswell Industrial Air Center Airport	NM
KSAN	32.73	-117.17	San Diego, San Diego International-Lindbergh Field	CA
KSAT	29.53	-98.47	San Antonio, San Antonio International Airport	TX
KSAV	32.12	-81.2	Savannah, Savannah International Airport	GA
KSEA	47.45	-122.3	Seattle, Seattle-Tacoma International Airport	WA
KSEP	32.22	-98.18	Stephenville, Clark Field Municipal Airport	TX
KSFO	37.62	-122.38	San Francisco, San Francisco International Airport	CA
KSGF	37.23	-93.38	Springfield, Springfield Regional Airport	MO
KSHR	44.77	-106.97	Sheridan, Sheridan County Airport	WY
KSHV	32.47	-93.82	Shreveport, Shreveport Regional Airport	LA
KSJN	34.52	-109.38	St. Johns, St. Johns Industrial Airpark	AZ
KSLC	40.78	-111.97	Salt Lake City, Salt Lake City International Airport	UT
KSPS	33.98	-98.5	Wichita Falls, Sheppard Air Force Base	TX
KSPW	43.17	-95.19	Spencer, Spencer Municipal Airport	IA
KSTL	38.75	-90.37	St. Louis, Lambert-St. Louis International Airport	MO
KTCS	33.23	-107.27	Truth Or Consequences, Truth Or Consequences Municipal Airport	NM

Site ID	Latitude	Longitude	City State	
KTLH	30.4	-84.35	Tallahassee, Tallahassee Regional Airport	FL
KTPH	38.05	-117.08	Tonopah, Tonopah Airport	NV
KTRM	33.63	-116.17	Thermal, Desert Resorts Thermal Regional Airport	CA
KTUL	36.2	-95.9	Tulsa, Tulsa International Airport	OK
KTUS	32.12	-110.93	Tucson, Tucson International Airport	AZ
KTVC	44.73	-85.58	Traverse City, Cherry Capital Airport	MI
KTYR	32.35	-95.4	Tyler, Tyler Pounds Field	TX
KTYS	35.82	-83.98	Knoxville, McGhee Tyson Airport	TN
KUIL	47.95	-124.55	Quillayute, Quillayute State Airport	WA
KUKI	39.13	-123.2	Ukiah, Ukiah Municipal Airport	CA
KUNV	40.85	-77.85	State College, University Park Airport	PA
KVEL	40.43	-109.52	Vernal, Vernal Airport	UT
KVTN	42.87	-100.55	Valentine, Miller Field	NE
KWMC	40.9	-117.8	Winnemucca, Winnemucca Municipal Airport	NV
KYUM	32.65	-114.6	Yuma / Yuma International	AZ

APPENDIX B. Data Management and Quality Assurance

This Appendix documents the quality control measures that were implemented for CIRAQ Phase I, generation of "CMAQ ready" meteorological inputs. Additionally, quality assurance issues that arose during data processing are highlighted, as well as the specific steps of the data management process. This section begins by providing a brief overview of the meteorological data. Next, a complete description of the data processing process, from receiving the raw meteorological model output to the generation of air quality "model-ready" input, is provided,. This Appendix closes with a description of the quality control process and a graphical summary of the quality control log. The CIRAQ Quality Assurance Project Plan (QAPP), approved 12/04 follow these discussions.

B.1 Meteorological Data Processing

The MM5 based RCM (MM5-RCM) downscaling process described in Chapter 2 produces hourly meteorological output. This fact, when combined with the continental coverage at a grid spacing of 36 km (versus ~275 km in the GISS model), results in a relatively large volume of raw model output (~3 terabytes). A number of steps are required to procure the model output from the DOE Pacific Northwester National Laboratory (PNNL) and to generate Communitity Multiscale Air Quality (CMAQ) ready meteorological fields. This exposes the data to potential for file corruption that could result in inaccurate data inputs to subsequent models and analyzes. Hence, it is necessary to document both the data processing/quality control procedure and to highlight

any significant issues that occurred. This summary meets the data quality documentation and reporting requirements described in the Quality Assurance Project Plan for the Climate Impacts on Regional Air Quality (CIRAQ) Project approved 12/14/04.

MM5-RCM model output is transferred from PNNL to the EPA facility at Research Triangle Park (RTP), North Carolina by way of 300-gigabyte portable hard drives. This was the only practical method of data delivery because of the amount of data, computer security restraints, and bandwidth constraints. Each drive holds approximately 2 years of MM5-RCM output. Over the period from late 2003 to early 2005, all 20 years of model output was incrementally received by the U.S. EPA. As model output was received, it was loaded onto a temporary drive on the NOAA/EPA High Performance Computing Cluster (HPCC) network. After initial data processing, the raw model output was transferred to the National Environmental Scientific Computing Center's (NESC) tape archival system for long-term storage [*National Envrionmental Scientific Computing Center*, 2004].

The data processing step is controlled by a semi-automated data management script, which is a multi-stage and multi-function procedure that was performed on the model output in yearly batches. The main function of the data management process is to convert the raw MM5 output to a form that the air quality model can utilize. The model that will be used in subsequent air quality assessments is the Community Multiscale Air Quality (CMAQ) Modeling System version [*Byun and Ching*, 1999] version 4.5. This is accomplished by running the Meteorology-Chemistry Interface Processor (MCIP) to produce daily meteorological files. A customized version of MCIP Version 2.0 is used here to eliminate several output variables that were not needed in any planned analysis. Additionally, slight modifications were made at PNNL to include additional output variables from the MM5-RCM simulations needed to calculate aerodynamic

resistance and stomatal resistance land-surface parameters required by the CMAQ deposition module [*Byun and Ching*, 1999].

Once the MCIP output is generated for a particular day, a series of data quality checks was conducted. After MCIP data files for a specified year are generated and quality checks performed, a meteorological variable extraction program extracts hourly meteorological fields that correspond to specific locations throughout the model domain where meteorological observations are routinely gathered and archived. These data are used in the analysis portion of the climate simulation presented in this report. In addition, daily 700 mb wind fields at 1800 UTC were extracted for the cluster analysis in a separate step (Chapter 3). Once these processing steps for a particular year were completed, the MCIP files are transferred to the NESC tape archive. These yearly datasets will be transferred back to the local computer cluster when the CMAQ simulations are ready to be executed.

B.2 Specific Data Processing Issues

B.2.1 MCIP Settings

The MCIP is configured to transform raw MM5-RCM output into CMAQ-ready meteorological input. MCIP files were generated with 25-hourly grids starting at 00 UTC and ending 00 UTC the following day. This required us to use share grids across original 24-hour MM5-RCM files. In addition, standard practice when modeling for climate applications ignores the designation of leap years. As a result, February 29th is missing from the data archive during 1996, 2000, 2004, 2048, 2052. This omission produces CMAQ run-time errors and so a new data record is

created for February 29th and the meteorology for the 28th of February copied to the new record during these five data years.

3.2.2 Data Extraction for Various Climate Analyses

In order to compare the current climate simulation with the observed climate over the last 10-years, data corresponding with observation sites was extracted from the first layer of the gridded MCIP files. For comparisons of model output to data from meteorological observation sites around the country, model output from the nearest grid cell to the observation site were used (no interpolation) to represent the climate of the site. 850mb pressure level wind and temperature data were extracted for each observation site. Again, data from the nearest grid point to the observation site was used for the 850 mb meteorological fields. However, the vertical interpolation to 850mb was performed by extracting the pressure profile from the nearest grid cell along with the temperature and wind profile. Then using linear interpolation the value was calculated. 700 mb wind was also extracted for use in the cluster analysis described in Chapter 3. In this case, data were extracted once a day at 1800 UTC and then horizontally interpolated (bilinear) to an alternative latitude-longitude reanalysis grid. Once again, linear interpolation between the grid cell points bounding the 700 mb level was used for vertical interpolation. Interpolation errors will result, but are generally limited to areas where large gradients in topography or near land-water boundaries.

B.3 Quality Assurance Procedures

B.3.1 File existence and integrity – Layer I Quality Control

The most elementary of the quality control steps performed on the meteorological files was a file existence check for each day of the simulations. A second step was to ensure that the file size was consistent with the other files. File size for the same model configuration covering the same period should be identical. Thirdly, an integrity check, inherent to other processes was performed to ensure that the format of the model output was readable. This is a necessary check because data handling and archiving raises the potential that files could become corrupted.

A program script was developed and run on files located in the NESC archive to confirm the existence of a file for each model day in each model year. Yearly log files were produced for both the MM5-RCM files and MCIP files. These yearly log files were then combined into one file that was scanned for missing file flags. The log files are available on the CIRAQ website (http://snow.rtpnc.epa.gov/asmd/gilliam/ciraq/qaqc.php). The initial scan identified several missing days of MM5 model (April 21-23, 2005, May 21-23, 2005, and Sep. 30, 2005). Subsequently, these files were retrieved from PNNL, and are now stored in the long-term archive. The scan of file sizes reveal that all files (3-day files are 1697072084 bytes, 2-day files are 1131420596 bytes, and 1-day files are 565769108 bytes), were complete. Therefore, this quality control script confirmed that all MM5 files were present in the CIRAQ archive.

The MCIP logfile (http://snow.rtpnc.epa.gov/asmd/gilliam/ciraq/qaqc.php) was scanned and it was confirmed that each of the MCIP output file types (GRIDCRO2D, GRIDCRO3D, GRIDDOT2D, METBDY3D, METCRO2D, METCRO3D and METDOT3D) are present in the long-term archive for each day of the simulation. In addition, the file sizes are consistently the same size for each type. It should be stated that the files for Feb 29 on leap years 1996, 2000, 2004, 2048, 2052 are not present because of the standard practice not to consider leap years in

climate simulations. Additionally, since the daily MCIP files cover a period from 00 UTC one day to 00 UTC the following day (25 hours), the Feb 28 files of all leap years were missing the last hour, which represents Feb 29, 00 UTC. Prior to CMAQ execution, Feb 28 files will need to be copied to represent the meteorology on Feb 29. In addition, the 00 UTC data for March 1 will need to be appended to the Feb 28 files as well as to the modified Feb 29 file. The headers for that time (March 1, 00 UTC) will need to be changed to represent February 29, 00 UTC.

B.3.2 Data Scan – Spatial and Temporal Range Checks – Layer II Quality Control

A scan of the hourly values from a set of variables (see Table B1) that are critical to the air-quality model results was performed on the MCIP output files. This quality control scan is necessary because any time a large volume of data is handled, corruption can occur. We wanted to document that no data corruption was detected prior to the delivery of the raw data to the EPA or during internal transfers. Two distinct scans were performed on the hourly grids. The first scan was a range check, and the second scan was of hourly change in these pre-determined variables.

The range check is an inspection of the grid-wide minimum and maximum value of each variable for each hour of the simulation. The acceptable range of each variable generally follows the quality assurance procedures for meteorological data that was developed for NOAA's Forecast System Laboratory's Meteorological Assimilation Data Ingest System (MADIS) (http://www-sdd.fsl.noaa.gov/MADIS/madis_sfc_qc.html) and National Weather Service Technique Specification Package 88-21-R2. A few variables were not presented in these quality assurance plans, so limits were subjectively determined using experience. Table B1 provides the range of acceptable values that were specified in the QC range check. Variables assigned a subjective limit are marked by an asterisk (*).

Table B1. Range and hourly change limits used in the quality control check of CIRAQ meteorological model output.

Variable	Range		Hourly Change
	Min	Max	Max
2-m Temperature (K)	224	328	20
Boundary Layer Height (m)	0	5000	3000
Surface Pressure (mb)	568	1100	15
Surface Shortwave Rad. (w/m^2) *	0	1200	500
Cloud Faction (%)*	0	100	100
Convective Precip. (inches)	0	44	44
Non-Convective Precip. (inches)	0	44	44
850 mb Temperature (K) *	180	330	10
850 mb U-Wind Comp. (m/s) *	-100	100	50
850 mb V-Wind Comp. (m/s) *	-100	100	50

Similarly, the temporal (hourly) change check determines the grid-wide minimum and maximum hourly change at each grid point for each variable. The acceptable range of the hourly-change was also adopted from the NOAA FSL's MADIS quality control plan (Table B1). For the variables not in the FSL list (indicated in Table B1 by asterisk), the acceptable range was subjectively determined using experience of working with model output. For all of the maximum and minimum values the grid-point location (latitude and longitude) was calculated and all variables at this grid point were placed in a CIRAQ database. Also calculated and logged was the mean, median, and root-mean-square (RMSE) of the grid for each variable and each hour. This information can collectively provide confidence that the meteorological data is reasonable and consistent in time. The following subsections highlight results from an analysis of the hourly quality control data.

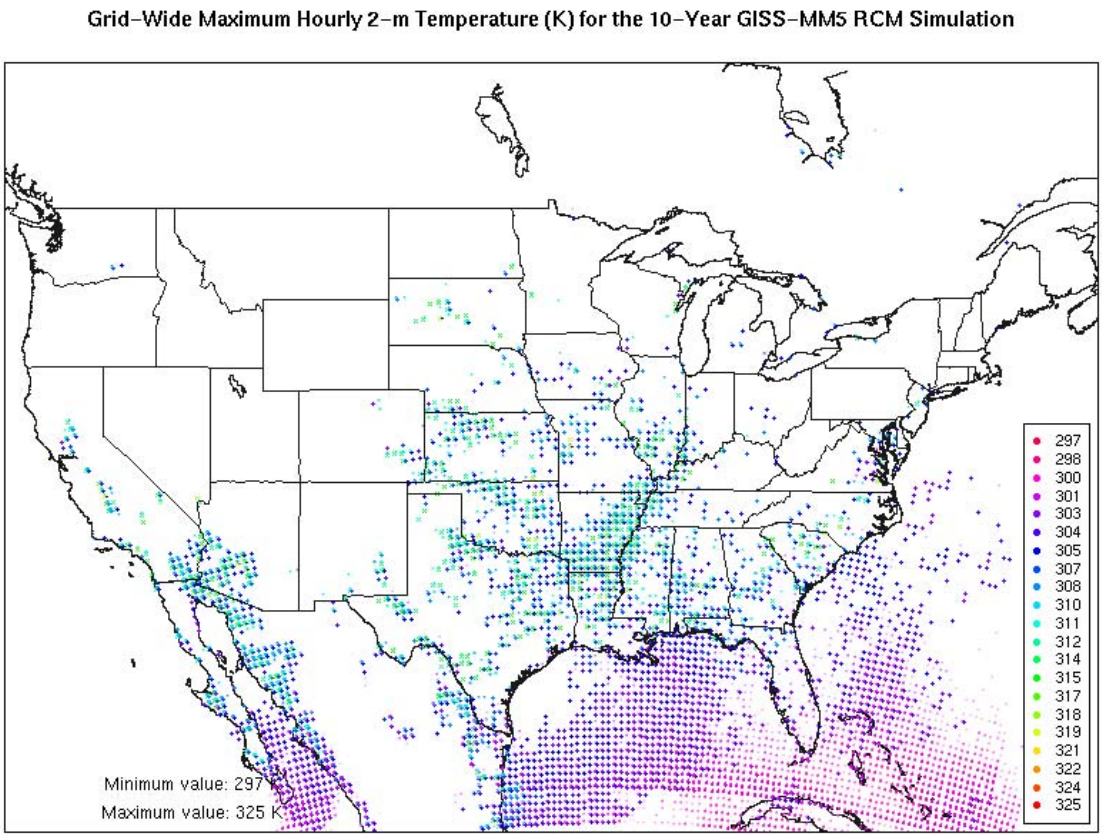


Figure B1. Hourly maximum grid-wide 2-m temperatures from the 10-year current scenario regional climate model simulation.

Range Check

Although many variables, important to air quality applications, were scanned during the quality control process (variables indicated in Table B1), and many types of checks were logged, results of the scan will be abbreviated. It should be stated upfront that no corrupted or unreasonable values were found during the scan; however, a few values were flagged as questionable. Temperature near the surface was used as one of the main variables in the QA checks. Figure B1

indicates the location and value of the maximum 2-m temperature in the domain each hour of the 10-year simulation. The minimum and maximum value of 2-m maximum temperature is indicated near the lower-left corner; the maximum simulated temperature over the 10-year simulation (325 K, in the Desert Southwest) does not exceed the range (224-328 K) guideline in Table B1, and the values are reasonable in magnitude and spatial distribution. Note that when multiple maximum grid values occur for the same grid point, only the last plotted value is visible since the previous dot is over-written. Figure B2 is a slightly different view of the hourly minimum grid-wide temperature. The time series clearly shows reasonable values with the exception of the winter when a few instances exist where temperatures do fall slightly below the recommended minimum of 224 K (very end of time series in Figure B2). A spatial plot (not shown) of these sub-224 K temperature reveal the location is isolated to a few grid points that represent the climate of the Rocky Mountains in southwest Canada, so the values are very low, but are only a few degrees below the recommendation and only exist for a limited number of instances. Table B2 gives the minimum and maximum values that were summarized from similar plots for the other variables. In all cases, the minimum and maximum hourly values are within range standards. Similar information for the other variables is documented in Table B2. All minimum and maximum values for the other variables are within the range specified by Table B2.

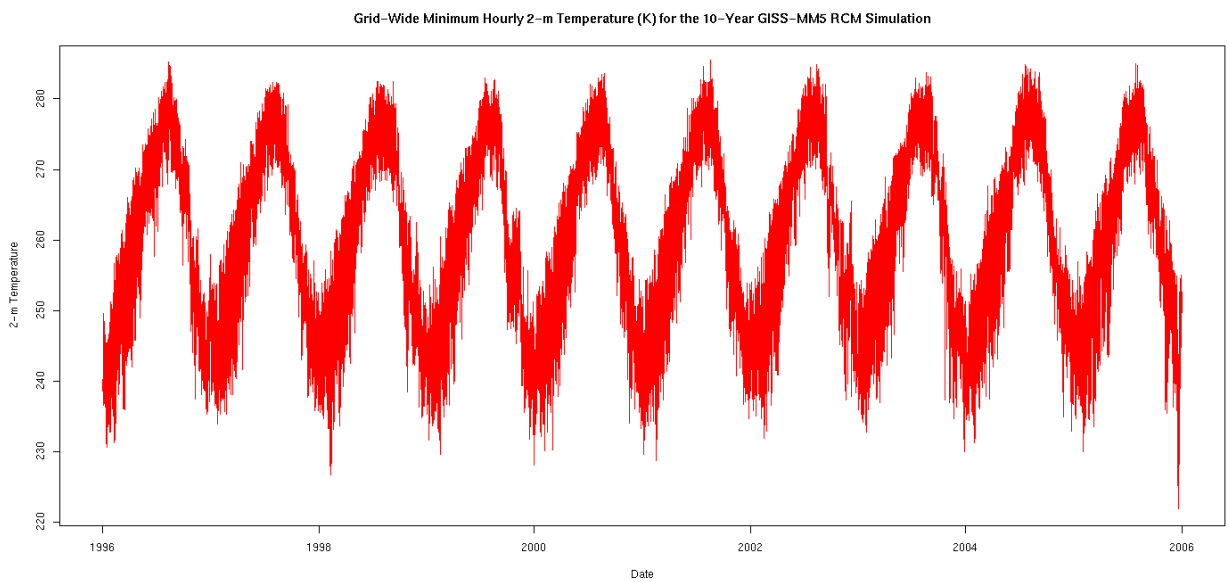


Figure B2. A time series of hourly minimum grid-wide 2-m temperatures from the 10-year current scenario regional climate model simulation.

Table B2. Maximum and minimum values and hourly change results for CIRAQ meteorological model output variables on which quality control checks were performed.

Variable	Value		Hourly Change
	Min	Max	Max
TEMP1P5 (K)	222	325	26
PBL (m)	33	4478	4437
PRSFC (mb)	652	1050	15
RGRND (w/m^2)	0	1191	1019
CFRAC (%)	0	1	1
RC (inches)	0	N/A	N/A
RN (inches)	0	N/A	N/A
TA (K)	239	314	26
UWIND (m/s)	-62	64	39
VWIND (m/s)	-62	72	35

Temporal Check

The same variables scanned in the range check were also scanned and plotted as the

hourly change or temporal consistency check. The minimum and maximum hourly change for each variable is provided in Table B2. For temperature, the maximum one-hour change at one grid-point over the 10 years was 26 K, which is greater than the NOAA FSL guideline of 20 K. All the variables are within the guidelines except for the PBL height and insolation. Table B2 indicates the PBL at a particular grid point changed ~4000 m in one hour. In retrospect, this QA range limit may be too rigid as the model is capable of simulating deep tropospheric boundary layers in excess of 5000 m in parts of the country within deep convective cells, then dropping to less than 1000 m as the storm cell advects to another grid cell. An hourly change in insolation of 1100 $W \cdot m^{-2}$ occurred at some of the grid cells along the tropical southern boundary, which is likely a result of deep convection that is frequent along limited-area model boundaries. Other than these instances, the overall temporal consistency check verifies the range check results that no persistent inconsistencies exist in the CIRAQ meteorological data that will be used for the CMAQ simulations.

B.3.3 Incidental Data Integrity Checks –Layer III Quality Control

Many of the various analyses of the climate data as well as the MM5-RCM data processing procedures implicitly scan the data for quality as well as integrity. This final section of the quality control highlights these various programs and how they further add to the quality control process.

B.4 MCIP Processing

The first and most rigorous of these incidental or implicit data checks was the processing of the raw MM5-RCM data by the MCIP program. MCIP will not execute nor will daily files be generated if the dataset is corrupted. Although abnormally large values of a variable like PBL and temperature are not flagged during this process, complete execution of MCIP ensures that the input model data is in the proper format. The quality control log documents that each daily file was present in the base case and future scenarios and verifies that MCIP was properly executed for each day.

B.4.1 Climate Data Extraction and Analysis

The second program that was executed on the 20-year set of model output was the extraction of the climate data that could be associated with selected climate sites across the country. As with MCIP, the program inherently would not execute if any of the daily files were corrupt. Additionally, the data that was extracted has been analyzed, and if grossly unreasonable data were present, these abnormal data would have been found in the climate analysis presented in the report. A few examples of these analyzes are mean sea-level pressure and precipitation analyses. Unreasonable data was not found which further provides support that the meteorological data, both the MM5-RCM raw output and MCIP output are of high quality and ready for use by CMAQ.

B.4.2 700 mb Wind Extraction for Cluster Analysis

The final analysis that was performed on the dataset was the extraction of wind at 700 mb.

These data were extracted from the MCIP files for an independent analysis. This independent cluster analysis indicated no abnormal data was present. In addition, daily 700 mb wind vector plots were simultaneously generated. These plots have been visually scanned in yearly animations, and show reasonable flow patterns.

B.5 Quality Assurance Project Plan for the Climate Impact on Air Quality(CIRAQ) Project (Phase I, 2005-2007)

1.0 Project Management

1.1 Problem Definition

Past research suggests that future climate may differ substantially from current conditions because of greenhouse gas affects on the radiation budget. As greenhouse gases increase, longwave radiation can be trapped and cause higher temperatures within the troposphere. The Intergovernmental Panel on Climate Change (IPCC) has conducted a series of global climate model simulations testing this hypothesis using a range of future emission scenarios. Climate model predictions varied, but all illustrated a positive trend in temperature with increasing greenhouse gases.

Future scenarios of changing climate suggest it is plausible that air quality may be affected. The exploration of possible responses to future climate conditions is addressed by ORD/NERL through the Climate Impact on Air Quality (CIRAQ) project. A principle goal of CIRAQ is to explore the uncertain nature of future air quality. This is done by adopting a phased sensitivity analysis approach. Phase 1 of CIRAQ develops a series of tests to consider the sensitivity of air quality to potential climate change conditions while holding most emissions at present day levels (Task 12913). Phase 2 will expand that analysis to include future emissions scenarios. A phased sensitivity modeling approach to uncertainty estimation is necessary because of the predictive nature of this study where future conditions cannot be observed. Air quality simulations will be performed using the USEPA Community Multiscale Air Quality (CMAQ) model under current and future climate scenarios to test the sensitivity of ozone and particulate matter predictions to changing climate scenarios. Findings from these studies will provide background information to U.S. EPA Global Change Research Program (GCRP) scientists housed within the National Center for Environmental Assessment (NCEA). This Quality Assurance Project Plan (QAPP) deals with data, modeling and evaluation activities associated with Phase 1 of the CIRAQ project.

1.2 Project Organization

The CIRAQ project team consists of six principle members (see table below). The project Team Lead, Ellen Cooter, is responsible for the project management. CIRAQ is based in the USEPA Atmospheric Modeling Division (AMD), which is located within the Office of Research and Development's National Exposure Research Laboratory (NERL). Since the project is based solely within AMD, the Division has responsibility for developing the model simula-

tions, analyses, and documentation required for the project. All emissions processing to model-ready format, meteorology processing to model-ready format and air quality simulations will be performed on the NERL/AMD Linux cluster GLOBAL (see description in Section 2.3).

Contractor support for the CIRAQ project is provided through the FAIR II contract 68-W-01-032, Task Order Number 2045 with Computational Sciences Corporation (CSC). A FAIR II Task Order Number 2045 Quality Assurance Project Plan (QAPP) has been developed by CSC

CIRAQ Project Team

Ellen Cooter , Air-Surface Processes Modeling Branch (APMB)
Team Lead (project management, regional climate assessment, coordination of external research in vegetation change

Alice Gilliland, Model Evaluation and Applications Research Branch (MEARB) Coordination with global models, air quality modeling

Robert Gilliam, Atmospheric Model Development Branch (AMDB)
Evaluation of regional climate simulations, coordination of simulations to produce model-ready meteorology for air quality model

Jenise Swall, MEARB
Statistical guidance in spatial and temporal analyses of regional climate and air quality simulations

William Benjey, APMB
Guidance on emission inventory development for control base simulations, participates in cross-laboratory research on future air quality emission scenarios

Gary Walter, MEARB
Computational and data management infrastructure design for CIRAQ

(approved 03/03/04) and describes the procedures that CSC follows for emissions processing and model simulations. When providing technical support to the CIRAQ project in these areas, CSC will follow their approved QAPP.

1.3 Project/Task Description and Schedule (FY05-FY07)

The CIRAQ project is responsible for the following tasks (and schedule for each task):

1. Production, evaluation and analysis of current and future model-ready regional climate scenarios.
   a. Schedule: two year activity, FY04-FY05
   b. Product: Report on the performance of regional climate simulations and analysis of temporal or spatial trends in predictions **(FY05 APM-551)**.
   c. Persons Responsible: Robert Gilliam, Ellen Cooter, Alice Gilliland, Jenise Swall

2. Development of model-ready emissions for CMAQ simulations
   a. Schedule: two year activity, FY04-FY05
   b. Product: Model-ready emission files (**FY05 APM-551**)
   c. Persons Responsible: William Benjey, Alice Gilliland, Ellen Cooter

3. Performance of CMAQ simulations under current and future conditions (1999 ± 5 years, and 2050 ± 5 years
   a. Schedule: two year activity, FY05 - FY06
   b. Product: CMAQ model output of concentrations for 5 current and 5 future years
   c. Persons Responsible: Alice Gilliland, Ellen Cooter

4. Evaluation and analysis of CMAQ simulations
   a. Schedule: two year activity, FY06 - FY07
      Product: Report on the performance of CMAQ and analysis of temporal and spatial variations or trends in the predictions (**FY07 APM-87**).
   c. Persons Responsible: Ellen Cooter, Alice Gilliland

1.4 Input Data Quality Objectives

Model simulations will be performed to test the sensitivity of air quality to potential climate change scenarios. To insure that model results represent a sensitivity response to modeled climate change scenarios rather than model input errors, several steps will be taken to test and maintain input data quality. The climate and emission input data are described in Section 2.1.

1.4.1 Regional Climate Data (RCM) Quality Control

Quality Control (QC) of the Regional Climate Model (RCM) data is performed when it is in model-ready Meteorology-Chemistry Interface Processor (MCIP) format. Variables at all the model grid points and each hour are compared against set tolerance limits. Additionally, the hourly changes of each variable at each gridpoint are compared to a set temporal change limits. Tolerance and temporal change limits are outlined in a quality control procedure devised by

NOAA's Forecast System Lab (FSL) and found at "http://www.sdd.fsl.noaa.gov/MADIS/madis_sfc_qc.html". Values that fail the quality control test and the locations where these values occur are stored in a database along with the hourly grid statistics. Additionally, a back check of the meteorological data from MCIP is compared with the raw RCM output for consistency. Statistics are generated from this back check comparison and archived in the CIRAQ database. The CIRAQ quality control database can be found at EPA intranet address "http://snow.rtpnc.epa.gov/asmd/gilliam/ciraq_log.php". Significant inconsistencies noted during this process are referred to the RCM database creator, Ruby Leung of the U.S. Department of Energy, Pacific Northwest National Laboratory (DOE/PNNL) for further investigation and rectification. If , after further investigation, a systematic error is detected, data correction *via* regeneration will be negotiated jointly between the contractor (DOE/PNNL), the AMD scientists performing the QC checks (Gilliam) and the NCEA Project Officer (Anne Grambsch). See Section 1.5 for specific reporting and documentation requirements.

1.4.2 Emission Data Quality Control

QC checks will be used to make sure that the processed emissions still maintain the same mass budget as in the official USEPA 2001 National Emission Inventory (referred to as *2001ad*) prepared for the Office of Air Quality Planning and Standards. CSC contractors will perform the emissions processing following QC procedures for emissions processing outlined on pages 7-10 in their approved QAPP. In addition, CIRAQ scientists will review the contractor QA reports and will conduct visual checks throughout processing using tools such as PAVE. The *2001ad* database is widely used in current applications and so it is not anticipated that significant QC issues will arise. However, upon detection, the AMD scientist providing QC oversight (Benjey) will document and refer the issue to the CSC Work Assignment manager. If there does indeed prove to be an unresolved QC issue, a plan for rectification will be negotiated between the EPA Project Officer (Benjey) and CSC management. CIRAQ Phase 1 does not include the development of future anthropogenic or technology-driven emission scenarios. The 2001ad inventory will be used for both present and future time periods.

1.5 Documentation and Records

Documentation and records for the CIRAQ project include 1) documentation describing task activities and planning and 2) model evaluation and analyses records.

Documentation describing CIRAQ Phase 1 plans and activities is maintained in the USEPA NERL Task Information System (TIS) under Task 12913. The task describes the goals of the CIRAQ project, the schedule and product milestones, the staff involved, and the extramural resources. Tasks are updated every year and are submitted for management approval. Through

TIS, the approval process is tracked, and the current approval status shown. Backups of the TIS system are regularly maintained by NERL. Laboratory notebooks will be used to document significant project activities.

Model (RCM and CMAQ) evaluation and analysis summaries will be maintained by the CIRAQ Team Lead on his/her local computer (with backup). Individual CIRAQ investigators will document sections for which they bear major responsibility more thoroughly in Laboratory notebooks, but the Team Lead will be responsible for maintaining an official summary. A QC summary (e.g., no errors detected, error description, etc.) will be generated at the completion of the 10 year base case processing to model-ready format and the 10 year future model-ready processing. Final summaries of the RCM and CMAQ model evaluations will be provided to the NERL Assistant Lab Director who is assigned to Global Climate research, currently Dr.Michele Aston, and the US EPA Global Change Research Program (GCRP) air focus group Lead, currently Dr. Anne Grambsch. Research abstracts will be produced for APM outputs (see section 1.3).

Other documentation records for model evaluation are the log files generated during processing of model results. Log files will be archived along with the model output files. In addition, if non-physical results are identified as part of the screening processes and are noted in the log file, additional notes will then be added to the record describing the steps that were taken to correct or replace the model output before proceeding.

All model output and log files will be located on the CIRAQ Linux GLOBAL cluster containing nodes named "global, global2, global3, global4, and global5." Record of the machine and directory location of all model simulation results, log files, and analysis results will be maintained by the CIRAQ team. All of these data files will be backed up along with the model output files as part of a regular tape backup and archival process.

2.0 Data Acquisition

The CIRAQ project requires no direct data measurement.

2.1 Climate and Emissions Data Acquisition

The CIRAQ project requires downscaled Regional Climate Model (RCM) scenarios and emissions input data. RCM scenarios have been produced by Ruby Lueng (DOE/PNNL) under an EPA/DOE Inter-Agency Agreement (PI is Anne Grambsch, ORD/NCEA). These model-generated data are required for the development of present and future climatological conditions to drive emissions and atmospheric chemistry and transport models. RCM acceptance is based on model-ready RCM (i.e., MCIP) quality acceptance criteria illustrated in Figure 1 and described in

QAPP section 1.4.1.

Emissions inventory data are the most recently available 2001 modeling inventory (referred to as *2001ad*) prepared for the Office of Air Quality Planning and Standards (OAQPS). Model-ready emission data for CMAQ are prepared from the inventory using the Sparse Matrix Operator Kernel Emission (SMOKE) model. Plume rise is computed for all possible point sources in SMOKE. Major point sources, such as Electric Generating Unit (EGU) stacks, are not treated differently from other point sources. Instead, annual EGU values from the inventory are disaggregated with temporal profiles. Data acceptance criteria are described in section 1.4.2. Although OAQPS uses annual aggregated Continuous Emission Monitoring (CEM) data to estimate inventory EGU emissions, hourly CEM data are not appropriate for this application because of the climatological (long-term) nature of the analysis. Biogenic emissions will be estimated using current landuse/landcover databases and the Biogenic Emission Inventory System, Version 3.12 (Pierce, et al., 1998; Pierce et al., 2002). On-road mobile source emissions will be modeled by the U.S. EPA MOBILE6 mobile source emission model (Office of Transportation and Air Quality, 2002). Both BEIS and MOBILE6 are standard components of the SMOKE processing system. All emissions input data and SMOKE processing is performed by CSC under their approved FAIR II QAPP and SOPs.

2.2 Data Management

Downscaled RCM scenarios are transferred to the EPA from the PNNL on portable hard drives. A log entry is made on receipt of the drive, its contents are loaded onto the GLOBAL LINUX cluster and the drive is returned to PNNL for re-use. Subsequent management of these data through input to the Chemical Transport Module is illustrated in Figure 1 below.

A copy of the RCM data are stored locally on the GLOBAL cluster with a backup copy on the National Environmental Scientific Computing (NESC) Center archive. NESC Center backup procedures are described in the NESC Center User Guide (2004). Sparse Matrix Operator Kernal Emission (SMOKE) model processed and merged emission files will be saved locally with a mirror backup on a separate cluster disk. Surface CMAQ output conditions along with selected "re-start" files will be saved locally and maintained on the mirror backup disk.

2.3 Hardware/Software Configuration

The current publically available version of the CMAQ modeling system (e.g., CMAQ version 4.4 or more recent) will be used for this application. This release was produced following the criteria and procedures outlined under the Quality Assurance Project Plan for the Development of the Community Multiscale Air Quality (CMAQ) Model (in review).

Global Linux cluster hardware configuration

- 5 nodes comprised of global, global2, global3, global4 and global5
- Each node has....
  - Dell Poweredge 2650
  - 2 x 2.4 Ghz,
  - 412K Cache
  - 533 Mhz
  - Front side bus
  - 2G memory

3.0 Monitoring Progress and Reports to Management

3.1 Project Assessment and Plans for Science and Product Peer Review

CIRAQ project progress is continuously monitored by the Team Lead and periodically reviewed by the Team Lead Branch Chief (Tom Pierce) and AMD Global Change Program coordinator (Alice Gilliland). The Project task description and progress are reviewed and approved, if needed, on an annual basis by the AMD Division Director, the NERL Assistant Lab Director for Global Change and the Associate Lab Director. Review of CIRAQ science and products will take place at the Agency level (standard EPA and NOAA review process) as well as at the Program Level (e.g., BOSC review, Fall 2004; PART review scheduled for summer, 2005).

Major AMD products expected under the CIRAQ project include CMAQ model-ready regional climate (i.e., MCIP) and emission (i.e., SMOKE) input datasets for current and future conditions, temporal and spatial evaluation of the MCIP database, and evaluation CMAQ estimates of 5 years base case and 5 years future atmospheric concentrations of ozone and particulates. Products designated as APMs will undergo standard APM clearance and acceptance procedures. New evaluation methodologies developed for CIRAQ and results will be presented at scientific workshops and meetings as well as appearing in peer reviewed Journal publications. Teleconference calls between the ALD, EPA GCRP Air Lead and CIRAQ team members to report progress and to discuss project plan adjustments are scheduled as needed, but will occur at least once per quarter.

3.2 Model Performance Evaluation

3.2.1 Climate Model Performance Evaluation

Evaluation tools will be used to measure the performance of the regional climate model against observational data. Meteorological observations or gridded observation-driven reanalysis databases will be used. Observational data resources include NCDC surface and upper-air

observations (see http://lwf.ncdc.noaa.gov/oa/ncdc.html), the NOAA Forecast Systems Lab observational database (see http://www.fsl.noaa.gov/data), Daily Weather Maps (see http://docs.lib.noaa.gov/resur/dwm/data_rescue_daily_weather_maps.html) and seasonal weather summaries reported in the open literature. Reanalysis datasets that may be used for evaluation include the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis (Kalnay, et al., 1996), the NCEP-Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP-II) reanalysis (Kanamitsu, et al., 2002) and the European Center for Medium Weather Forecast (ECMWF) ERA-40 reanalysis data (Kållbert et al., 2004).

Since the RCM simulations are driven by Global Climate Model (GCM) information rather than assimilated observation or reanalysis, they do not necessarily reproduce day-to-day or exact year-to-year observed variations but rather, represent time periods under representative climatological conditions. Keeping these special characteristics in mind, an evaluation of the simulated meteorology will be performed for the current RCM scenario (1995-2004). Several methods will be used to examine how well the RCM depicts the present climate including comparison of model and observed distributions of hourly meteorological variables (e.g., temperature, wind, moisture and solar radiation) by way of a cumulative distribution function (CDF), statistical cluster analysis comparing RCM seasonal patterns of mean and extreme 700mb wind, and time series analysis to characterize the diurnal, synoptic, seasonal and interannual components of selected variables such as surface temperature. The 700mb wind analysis will compare RCM output to gridded reanalysis datasets covering the entire RCM domain. The CDF and time series evaluations will be performed for selected grid-cells or relatively small geographic regions. Cells or regions will be selected to represent known seasonal and annual geographic signatures in the meteorological data. For instance, variations in climate over a region (e.g., mountains, Piedmont, coastal plain of North Carolina) are evident in the distribution of meteorological variables such as temperature. Grid cells representing each of these climatological regimes will be selected and paired with available observations. If the regional climate model is indeed simulating the regional climate properly, the distributions should be similar. Time series analysis will be performed on a sub-set of CDF locations. Combined, these three analyses will provide an assessment of how well the RCM, driven by a global climate model, represents present climate. Further discussion of these techniques, including references to peer reviewed method descriptions are provided in the TIS entry for Task 12913.

3.2.2 CMAQ Model Performance Evaluation

Plans for CMAQ model performance evaluations are discussed in the Task 12913 description on the TIS system. As soon as the reference emission fields become available and are quality assured, the CMAQ simulations will begin. Initial and boundary conditions will be estimated based on published values from global chemical model studies or from GCTM output. Analysis

approaches will be developed to assess variability and trends in the ozone and particulate matter levels throughout the continental model domain. Where possible, base case CMAQ results will be compared to distributions of ambient air quality data from the following networks: IMPROVE, STN, CASTNet, NADP, and AIRS. In contrast to the meteorological reanalysis datasets listed in Section 3.2.1, the relatively sparse geographic coverage and short periods of complete record for observed ambient atmospheric concentration limits the number of such comparison that can be performed. As CMAQ results become available, efforts will accelerate to establish more formal evaluation procedures. The CIRAQ Team will coordinate development of appropriate air quality model evaluation techniques with procedures developed by others for other CMAQ applications, e.g., Task 18339-Air Quality Forecast Database and Analysis, Task 20463-Evaluation Techniques and Tool Development for FY08 CMAQ Release. Air quality model evaluation is not expected to begin until FY06. CMAQ evaluation techniques and results will be presented at scientific workshops and meetings as well as appearing in peer reviewed Journal publications.

3.3 CMAQ Model Sensitivity Analysis

Air Quality model output uncertainty regarding potential responses to future climate will be addressed through CMAQ model sensitivity analyses. Principle sensitivity analyses to be performed under CMAQ focus on the isolation of potential climate and emission impacts on present and future air quality. This is accomplished through a two-stage approach described in TIS tasks 12913 (Phase 1) and 20478 (Phase 2). In Phase 1, with the exception of changes that are a response to climate factors, future emissions will remain constant at present levels. CIRAQ Phase 2 will revisit these air quality scenarios using an alternative emission inventory that includes some estimate of future economic, population and technological change in the continental U.S. If time and resources permit, Phase 2 may also include an evaluation of the sensitivity of alternative air quality futures to future climate scenario (i.e., repeat the previous analysis for a suite of possible future climate conditions).

4.0 REFERENCES

Atmospheric Model Development Branch: Quality Assurance Project Plan for the Development of the Community Multiscale Air Quality (CMAQ) Model. ORD/NERL/AMD/AMDB (In review, 2004).

Computer Sciences Corporation: Quality Assurance Project Plan in support of National Exposure Research Laboratory, U.S. Environmental Protection Agency, Task Order Number 2045, Contract 68-W-01-032. (Approved March 3, 2004).

Kållberg, P., A. Simmons, S. Uppala and M. Fuentes. 2004: The ERA-40 archive. ERA Project Report Series No. 17. Available in .pdf format from http://www.ecmwf.int/publications/library/do/references/list/192

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, 77, 437-471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S-K Yang, J.J. Hnilo, M. Fiorino and G.L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, 83, 1631-1643.

National Environmental Scientific Computing Center, 2004: User Guide Version 9.2. OEI/OTOP/NTSD/NESCC. Available at http://www.epa.gov/nesc/10_publications/Customer_Document/UserGuide.pdf

Office of Transportation and Air Quality, 2002: User's Guide to Mobile6.1 and Mobile6.2: Mobile Source Emission Factor Model. EPA420-R-02-028, U.S. Environmental Protection Agency, Ann Arbor, MI, 264 pp.

Pierce, T., C. Geron, L. Bender, R. Dennis, G. Tonnesen, and A. Guenther, 1998: Influence of Increased Isoprene Emissions on Regional Ozone Modeling. *J. Geophys. Res.*, 103, 25611-25629.

Pierce, T., C. Geron, G. Pouliot, E. Kinnee , and J. Vukovich, 2002: Integration of the Biogenic Emissions Inventory System (BEIS3) into the Community Multiscale Air Quality Modeling System. In Proceedings of the Annual Meeting of the Air and Waste Management Association, Air and Waste Management Association: Pittsburgh, PA, 2pp.