Performance assessment of the Pleim-Xiu LSM, Pleim surface-layer and ACM PBL physics in version 3.0 of WRF-ARW

Robert C. Gilliam and Jonathan E. Pleim

Atmospheric Modeling Division, National Exposure Research Laboratory,

United States Environmental Protection Agency, Research Triangle Park, NC

27711

Submitted to Journal of Applied Meteorology and Climate

(September 18, 2008)

Corresponding author address: Robert C. Gilliam, US EPA, 109 TW Alexander Dr, Research Triangle Park, NC 27709
E-mail: gilliam.robert@epa.gov

Abstract

The Pleim-Xiu land-surface model, Pleim surface-layer and Asymmetric Convective Model (Version 2) are now options in version 3.0 of the Weather Research and Forecasting (WRF) model. These physics options were developed for the MM5 mesoscale model and used extensively by the air quality modeling community, so there was a need to extend these physics to WRF. The new physics in WRF are compared to the MM5 counterpart and several other WRF configurations in this study with a focus on the replication of 2-m temperature, 2-m moisture and precipitation.

The new physics implementation is recommended for retrospective air quality simulations, in particular, those used so simulate warm season meteorology. In the summer, the performance was similar to the MM5 counterpart and has a smaller 2-m temperature and 2-m moisture error than other WRF simulations. For cold season cases, the model simulation was not as accurate as the other simulations as a whole, but did comparatively well in terms of less 2-m temperature error in the western part of the model domain (Plains and Rockies) and parts of the northeast United States. Also, 2-m mixing ratio was well simulated by the new physics suite in WRF. Monthly precipitation was well simulated, especially in the winter. Summer precipitation was not as well simulated as winter because of the scattered and random nature of warm season convection, but difference among models was not large. It is expected that once more precise analyses are available for WRF, like the RAWINDS analyses used to drive MM5, the performance of the new physics will further improve.

1. Introduction

Mesoscale models require land-surface, surface-layer and planetary boundary layer (PBL) parameterizations to represent the transfer of heat, moisture and momentum between the surface and atmosphere. A new suite of land-surface and PBL physical parameterizations have been implemented in version 3.0 of the Weather Research and Forecasting (WRF) model, the Advanced Research WRF (ARW) core (Skamarock et al., 2005). The Pleim-Xiu land-surface model (PX LSM; Xiu and Pleim, 2001), Pleim surface-layer scheme (Pleim, 2006), and Asymmetric Convective Model version 2 (ACM2) (Pleim, 2007a,b) for the planetary boundary layer (PBL) have been used extensively as physics options in the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5; Grell et al., 1995). Many users of the Community Multiscale Air Quality (CMAQ) modeling system (Byun and Schere, 2006) have employed MM5 with this physics configuration for various air quality models. The ACM2 PBL is preferred because it results in consistent turbulent mixing in the meteorological (WRF) and air quality (CMAQ) model. Additionally, the soil moisture nudging capability (Pleim and Xiu, 2003) of the PX LSM model, along with four dimensional data assimilation (FDDA) using model analyses (Stauffer and Seaman 1990, 1994; Stauffer et al. 1991), results in high-quality meteorological fields (Gilliam et al., 2006). This implementation of additional physics also serves a need expressed at the WRF Users' Workshop in June 2006 to develop additional PBL and LSM options.

The PX LSM, surface layer, and ACM2 PBL options in WRF-ARW Version 3.0 (referred to as WRF PXACM herein) have been evaluated. This paper describes the

performance of the WRF PXACM relative to observations, a similar MM5 simulation, and three commonly-used WRF configurations. The goal is not necessarily to rank the models or physics options according to accuracy since some model configurations are suited for certain applications or regions, and may require special configuration methods. The main objective of the assessment is to ensure the WRF PXACM is of similar skill as the MM5 version (MM5 PXACM), and in the range of common WRF solutions. The evaluation will mainly focus on the model's ability to simulate 2-m temperature with a secondary focus on 2-m mixing ratio and precipitation. Also examined are the distribution characteristics of the diurnal variations of selected meteorological variables of the WRF and MM5 PXACM simulations. Together, these comparisons will provide an assessment of the WRF PXACM's ability to represent key characteristics of the atmosphere that have an impact on air quality, while considering the performance of other models and physics options.

2. Methodology

a. Overview of PX LSM, surface layer scheme and ACM2 PBL Model

The PX LSM coupled with the ACM2 PBL and surface-layer schemes, historically available only in MM5, has been found to be well-suited for extended (weeks, months or even years) retrospective simulations where the data assimilation scheme provides realistic tracking of soil moisture and vegetation trends. The PX LSM simulates the evolution of soil moisture and temperature in two layers (surface – 0-1 cm and root zone – 1-100 cm) as well as canopy moisture. There are three pathways for evaporation in the PX LSM: soil surface, canopy, and evapotranspiration. Grid-cell-aggregate surface

parameters to describe roughness, thermal, and moisture retention properties of the surface are derived from fractional land-use and soil texture data unlike other land-surface models that use a single value corresponding to the dominant land-use. The ACM for the PBL (Pleim and Chang, 1992), which was a derivative of the Blackadar convective model, has been recently updated to the hybrid non-local scheme of the original ACM, and an eddy diffusion scheme (ACM2; Pleim, 2007a,b). These physics options from the MM5 model have been disconnected and implemented as separate LSM, surface-layer and PBL physics options in WRF Version 3.0. Each of these new physics options has been successfully tested with the other ARW physics for compatibility or what is commonly referred as “plug-and-play” capability.

The PX LSM currently does not contain a process to account for the accumulation, sublimation or melting of snow. Rather, it uses 3-hourly gridded snow-water equivalent from the driving analysis to compute snow coverage. It has been noted in past evaluations (i.e., Gilliam et al., 2006; Gilliam et al., 2004) that the PX LSM does not perform as well as some other land-surface models in areas of frequent snowfall. Several improvements for snow cover were made in the PX LSM including an updated volumetric heat capacity for snow and a fractional snow coverage that is a function of land-use and snow depth, following the method used in the NOAH land-surface model (Ek et al. 2003). Fractional snow coverage is used by the PX LSM to compute a surface heat capacity that is weighted according to the fraction of the surface that is covered by snow.

b. Model Configuration

The simulations were executed on a 12 km eastern U.S. grid with 34 vertical levels, extending from the surface to the 100 hPA level. All WRF and MM5 simulations were configured on a similar grid in the horizontal and same sigma levels in the vertical (note that the heights of the vertical layers varies in time for WRF but not for MM5). Simulations were done for cold and warm season cases (January and August of 2006). All simulations utilized the FDDA using the National Centers for Environmental Prediction (NCEP) North American Mesoscale model (NAM) analysis for the 00, 06, 12 and 18 UTC times and the 3-hour NAM (WRF-NMM) forecast for the 03, 09, 15 and 21 UTC times. As configured by Stauffer et al. (1991) and Otte (2008), no temperature or moisture nudging was done within the ACM2 diagnosed PBL, but the nudging of wind was applied at all levels of the model. Also following these conventions, the nudging strength was greater for temperature and wind (3.0×10^{-4} s^{-1}) than moisture (1.0×10^{-5} s^{-1}). The FDDA option not only activates the three dimensional analysis nudging of wind, temperature and moisture, but also initiates the soil moisture and temperature nudging described by Pleim and Gilliam (2008) and Pleim and Xiu (2003).

The month-long simulations that used the PX LSM were started 10 days prior to the first of the month. The first 10 days were used to fully spin-up the PX soil model. The initial deep soil temperature at the initiation of the spin-up of the simulation (first run segment) was set to the average 2-m temperature of the 5.5 day run segment using the IPXWRF utility (available for download at http://www.wrf-model.org). Next, 5.5 day run segments were executed starting at the first of the month; the first 12 hours were a spin-up and not included in the analysis. The 2-layer soil moisture and temperature were

preserved from one run segment to the next using the same IPXWRF utility referenced above.

Throughout the following assessment of model performance, the simulations will be referred to by the run identification in Table 1, which also provides the physics configuration of each simulation where the physics options other than PBL, LSM and surface-layer schemes are the same. There are minor differences in the microphysics schemes in MM5 and WRF. The simulations used Rapid Radiative Transfer Model (RRTM) for longwave (Mlawer et al., 1997), Dudhia (1989) shortwave radiation, Kain-Fritsch 2 (Kain 2004) convective scheme and the WRF Single Moment (WSM-6) (Hong et al., 2004) microphysics in WRF and Reisner 2 (Reisner et al., 1998) in MM5. The other WRF simulations use combinations of the NOAH LSM (Ek et al., 2003), Rapid-Update Cycle (RUC) LSM (Smirnova et al., 1997; Smirnova et al., 2000), Yonsei University (YSU) PBL (Noh et al., 2003), Mellor-Yamada-Janjic (YSU) PBL (Janjic, 1994) and two different Monin-Obukhov (M-O) surface layer schemes (Janjic, 1996; Janjic, 2002; Dyer and Hicks, 1970). These represent routinely used physics combinations by the WRF user community. It should be stressed that other land-surface models do not have an internal mechanism like the soil moisture and temperature nudging, so a run segment of 5.5 days may or may not be the best run strategy for non-PX simulations. The accuracy of these simulations may have been improved if the simulation length was shortened, but such sensitivity tests are outside the scope of this paper.

c. Model Assessment Techniques

The evaluation of the five simulations was primarily done using the Atmospheric Model Evaluation Tool (AMET, Gilliam et al., 2005) that pairs surface observations from the Meteorological Assimilation Data Ingest System (MADIS) database with the corresponding model simulations in space and time. For this study, the focus is on model-observation comparisons of 2-m temperature and mixing ratio. Since the wind is nudged at all levels in the PBL, and all model simulations have very similar performance statistics, wind will not be examined.

The model performance statistics will be assessed collectively on a monthly basis and parsed by time of day. Also, observation site specific, model performance statistics are presented spatially as are model-analysis comparisons. In particular, model estimations of 2-m temperature are compared to the 2-m temperature analysis from the NAM. This is not a standard evaluation method, but it will be shown that a model-analysis comparison is a valuable technique to assess model performance, because in many instances the analysis is a best-estimate of the state of the atmosphere, which is derived from short-term model forecast and many types of observations to minimize error. We argue that the model simulation should not be expected to outperform the analysis in terms of how they compare against point measurements. In other words, if the analysis has a large error relative to observations over a particular region like the Rocky Mountains, the model should be expected to have a similar (at best) or larger error and uncertainty.

The final analysis utilizes the Kolmogorov-Zurbenko (KZ) filtering technique, described in Rao et al. (1997), Eskridge et al. (1997), and Hogrefe et al. (2000), to examine the diurnal behavior of temperature and moisture. The monthly time series (model and observed) of 2-m temperature and mixing ratio were extracted for each

observation site. The K-Z filter was applied to derive the intra-day, diurnal and synoptic forcing embedded the time series. Then, the standard deviation was computed for each of these signals to measure the variability of strength of the forcing. Aspects of the variability, mainly the diurnal component, are examined spatially for both the WRF PXACM and MM5 PXACM. The main purpose is not only to compare how these models reproduce the features in the observations, but also, to see if the behavior of the WRF PXACM is similar to the MM5 version.

3. Assessment of Model Performance

a. Domain-Wide Model Performance

A convenient and useful method to assess the general performance of a model is to compute model error and bias relative to surface observations in the domain. The bulk or collective monthly model performance statistics are provided in Table 2 for both 2-m temperature and mixing ratio. In this particular comparison, we focus on Root Mean Square Error (RMSE) and mean bias (BIAS) (Wilkes, 1995). As a prelude, the paired model-observation data used to compute the statistics in Table 2 were compared using a paired (model-observation) or two sample t-test with a 95% confidence interval. For all model simulations, the 2-m temperature and mixing ratio was significantly different from the observations. For January 2006, the WRF PXACM had larger temperature error than the MM5 PXACM and two of the other WRF configurations (WRF NOAHYSU and RUCYSU). On the other hand, for the warm season period (August), the new WRF PXACM configuration is very similar to the MM5 PXACM in terms of both error and

bias, and actually simulates less error and bias than each of the other WRF configurations.

The mixing ratio errors of all simulations are low in January (Table 2), but the small error is primarily because the moisture content of the atmosphere in winter, than it is an improvement of model performance. The WRF PXACM does have a slightly higher mixing ratio error in January than most of the other simulations except the WRF RUCYSU. The mixing ratio error of the WRF PXACM in August is similar to the MM5 PXACM, but the WRF PXACM has a lower overall bias. As with temperature, the mixing ratio in August is very well simulated by the WRF PXACM when compared to the other WRF simulations. The WRF RUCYSU has a suspiciously large temperature and moisture error in August. A plot of domain-wide RMSE as a function of day of the month (not shown) indicates the error in the WRF RUCYSU increased more than the other simulations throughout each of the 5 day run segments, which relates to the issue that each model might have some specific configuration practices not explored here that would lower model error.

Mean and RMSE of 2-m temperature is presented as a function of time of day for both January and August of 2006 in Figure 1. Included in this figure is not only the model performance in term of error, but also the amount of error in the various analyses that drive the soil moisture and temperature nudging of the PX LSM. It is clear from Figure 1 (bottom-left) that the MM5 surface analysis (RAWINDS-derived, pink line) has a much lower error relative to the surface observations than the NAM analysis (dark blue) used by WRF. The WRF PXACM (red) has a similar error as the NAM analysis at night, but a much larger error during the day. The MM5 PXACM has a much lower error at all times

than the WRF PXACM and the mean temperature of the time series indicates the WRF PXACM too cold. It is thought that at least part of the difference between the WRF PXACM and MM5 performance is because of the difference in error of the two analyses that drive these models. This hypothesis will be thoroughly explored once a new RAWINDS-like tool is fully tested for WRF, but preliminary tests using an early pre-release version of a tool named *OBSGRID,* indicate the potential for dramatic improvements in winter (RMSE decrease from 2.85 K to 2.36 K). It is necessary to note that all WRF simulations have the largest error and cold bias (1.0-1.5 K) during the late afternoon (January).

The WRF PXACM has a larger error than the other WRF simulations at night (Figure 1, lower-left) and early part of the day in January. All WRF configurations have a large error during the mid-day hours. The mean 2-m temperature is Figure 1 (upper-left) indicate all WRF configurations simulate much colder mid-day temperatures than the MM5. This raises a question that is not within the scope of this project, but should be explored in the future. Is this large middle-of-the-day-error a common bias of the WRF ARW physics options such as the radiation scheme or microphysics? Another noteworthy point is that all WRF ARW simulations have an error at night that is near, in some cases less than, the error of the NAM analysis.

For August, the diurnal error distribution and mean temperature (Figure 1) of the WRF PXACM implementation is similar to the MM5 PXACM version. The MM5 PXACM has a little less error and bias at night than the WRF PXACM, but the WRF PXACM has slightly less error and bias during the day. For August, the MM5 RAWINDS surface analysis has significantly less error than the NAM analysis used by

WRF. Again, as with January, preliminary results suggest that the WRF PXACM will can be closer to, or even surpass, the performance of the MM5 PXACM once the more precise surface analyses that drives the soil nudging are available from the OBSGRIB tool. The WRF PXACM has in most cases significantly less error than the other WRF configurations at night, but the amount of error converges during the day, with the exception of the WRF RUCYSU simulation. The mean bias of the MM5 and WRF PXACM simulations and the WRF NOAHYSU are minimal (+/- 0.5 K) throughout the day, but the WRF RUCYSU has a large warm bias, especially during the day, and the WRF MYJ has a substantial cold bias at night.

Mixing ratio error, partitioned by time of day, for both January and August of 2006 is presented in Figure 2. As with the diurnal temperature error, the mixing ratio error of the WRF PXACM is slightly larger than the other model configuration in January. The WRF PXACM is very close and even lower than the MM5 PXACM during August. Overall, considering the bulk and diurnal statistics of 2-m temperature and moisture, the new WRF PXACM performs at the level of MM5 and has less error than the other WRF configurations in August. However, the WRF PXACM does have more error in the winter.

b. Model-Observation and Model-Analysis Comparisons (Spatial)

The statistics presented in the previous section do not provide detail of the spatial distribution of model error and bias. When one evaluates a model it is essential to look at the model performance using a spatial reference not only the domain-wide statistics. For example, the model can have a no temperature bias when all data is considered, but this

may be a result of a cold bias in one region and warm bias in another. To examine the spatial model error, the 2-m temperature RMSE was computed for each observation site in the domain. The RMSE of each simulation and the two analyses at each observation site was subtracted from the RMSE of the WRF PXACM. These error differences for January 2006 are plotted in Figure 3 to highlight the WRF PXACM model performance relative to the other simulations. The negative (positive) values are cold (warm) colors and represent locations where the WRF PXACM has lower (greater) error relative to the simulation being compared. The gray shading indicates there is little difference of error.

Not all of these comparisons in Figure 3 will be extensively discussed, only the significant features and general points as they relate to our goal to benchmark the WRF PXACM. Panel A in Figure 3 is the WRF PXACM RMSE minus the NAM analysis RMSE. It is significant that the model is actually an improvement over the analysis in areas of the extreme western-central part of the domain (i.e., Colorado). However, the WRF PXACM simulation had substantially more error than the analysis over the northeast, southeast and central parts of the US. Compared to the MM5 RAWINDS analysis (Panel B) the WRF PXACM error is much greater everywhere because these same observations used to derive RMSE are used directly in RAWINDS. Compared to the MM5 PXACM (Panel C), the WRF PXACM has a lower error at only few locations in the Rockies, upper Great Lakes and New England. The MM5 PXACM has lower error particularly in a region from Missouri to the Midwest to Pennsylvania, and only a slightly lower temperature error over other parts of the domain. The other WRF simulations like the NOAHYSU (Panel D) and RUCYSU (Panel F) have less temperature error in many of the same areas as the MM5 PXACM. However, the WRF PXACM has lower error

than most of the other simulations over a large part of the central Plains. In particular, WRF NOAHMYJ (Panel E) has much more error than the WRF PXACM over much of the western and central parts of the model domain. Overall, the evaluation of January indicates some research needs to focus on improving the simulated winter temperatures, especially during the day as the diurnal statistics indicated. Nonetheless, one could argue that the WRF PXACM falls within the bounds of the three other WRF outputs.

An alternative or complementary method to evaluate the 2-m temperature error of the WRF PXACM implementation, and how it compares to the other model configurations, is to use the gridded 2-m temperature analysis. In the case of the simulations using the PX LSM, it should be reiterated that the analysis (2-m temperature and 2-m moisture) drives the indirect soil moisture and temperature nudging, so the model-based analysis may be a more accurate measure of model performance. Figure 4 presents the mean absolute error (MAE) of the WRF PXACM using the gridded NAM 2-m temperature analysis (Panel A) rather than scattered point surface observations. Also presented in Figure 4, similar to Figure 3, is the difference of the WRF PXACM error and the error of the other simulations (panels B-E). This information provides an explicit measure of grid cells where the WRF PXACM has more or less 2-m temperature error relative to the other simulations.

It is clear from Figure 4 (Panel A) that WRF PXACM has larger error over parts of Canada, which are areas with persistent deep snow cover. The error of the WRF PXACM in these areas is 0.5-1.0 K greater than the other simulations. However, the model does have less temperature error in a narrow region from central Canada to the upper Midwest, Great Lakes to the northeast US, especially compared to MM5. The main difference

between WRF and MM5 PXACM implementation (Panel B) is the addition of the fractional snow cover algorithm that is frequently activated over this area, which considers snow depth and landuse to weight the thermal properties of the surface according to snow fraction. Nonetheless, snow processes in the WRF PXACM will be improved in the near future with more advanced snow model that is under development.

The new WRF implementation does have some skill versus the MM5 PXACM over water, much of the intermountain western US, and areas of New England during January. This skill over the MM5 PXACM is supported by the RMSE in Figure 3. The WRF PXACM does well compared to the other WRF simulations (WRF NOAHYSU and RUCYSU, Panels C and E) in areas of the central Plains, but as with the observations (Figure 3), the error is greater over much of the eastern US, particularly the southeast US.

The bulk surface-based model performance statistics in Table 2 and diurnal statistics in Figure 1 indicate the WRF PXACM implementation performs more similar to the MM5 PXACM and other WRF simulations during August in terms of the 2-m temperature error. This general assessment is verified spatially in more detail in Figure 5 and Figure 6. Figure 5 shows the observation-based error of the WRF PXACM relative to other simulations and Figure 6 the error of WRF PXACM relative to the NAM 2-m temperature analysis. First, the RMSE of the WRF PXACM's and the NAM analysis computed using surface observations is virtually the same over much of the eastern US (Figure 5, Panel A). The NAM analysis has less error along much of the US coastline and areas of the central Plains, but the WRF PXACM has less error in parts of the Rocky Mountains and parts of the southeastern and northeastern US. This implies that the WRF

PXACM has generally the same order of uncertainty as the NAM 2-m temperature analysis.

The WRF PXACM, in terms of the 2-m temperature, is comparable to the MM5 over much of the domain as indicated by the gray in Figure 5, panel C. The WRF PXACM has lower error over the Rocky Mountains and at scattered sites in the central part of the domain. The MM5 has lower errors at scattered sites in the northwestern and northeastern parts of the domain. The WRF PXACM implementation has lower 2-m temperature error at most sites when compared to the other WRF simulations (Figure 5, panels D-F), in particular, those observation sites outside of the intermountain west.

Figure 6 further supports the other estimates of model error that generally agree that temperature is well simulated in August 2006 by the WRF PXACM. When the model analysis is used to compute 2-m temperature error, the WRF PXACM has significant improvement over the MM5 PXACM in the western half domain and parts of the southeastern US, while there is slight degradation over parts of the northeast US. Consistent with the observation-based error in Figure 5, the WRF PXACM has lower error relative to the other WRF model configurations (Figure 6, panels C-E) in the eastern US and more error over the western part of the model domain, with virtually no difference over water. These spatial evaluations provide further evidence that the WRF implementation provides sound temperature and moisture (not shown) with similar uncertainty as an analysis during the warmer parts of the year.

c. Monthly Precipitation

The total precipitation for both January and August was computed as well as the estimated observed precipitation using the National Precipitation Analysis (NPA), which is a blend of WSR-88D radar estimated precipitation and gauge data (Seo 1998a; Seo 1998b; Fulton et al. 1998). Figure 7 shows the NPA (observed) and total precipitation amounts from the various simulations for January 2006. The general features and overall pattern of the observed precipitation are well represented in the model simulations. In particular, the 150 mm+ area of rainfall on a southwest to northeast axis across Mississippi and central Tennessee is well represented. Also, the amount and placement of precipitation across the northeast US and areas surrounding the Great Lakes is reasonably replicated. The NPA precipitation is radar and gauge derived, thus, it does not reach far offshore. However, the NPA and WRF PXACM precipitation over the area just off the Gulf of Mexico states and southeast US coasts are reasonably similar, in particular, the enhanced amount off North Carolina.

Differences are noticeable between the WRF PXACM and MM5 PXACM. The areas with the largest differences are over the Gulf of Mexico and off the North Carolina coast. The WRF PXACM generally has more precipitation over the ocean. The MM5 PXACM has more precipitation over the western half of Tennessee, northern Mississippi and parts of Arkansas. Otherwise, the precipitation amounts between the WRF PXACM implementation and the MM5 are not very different and in agreement with the NPA. The differences between WRF simulations are smaller than the difference between the WRF PXACM and MM5 PXACM. The 200 mm + area of precipitation over Tennessee, Alabama and Mississippi are similar among WRF simulations. It is difficult explicitly

state which simulation matches the observed precipitation best as there are inherent errors and biases in the NPA.

The total precipitation for August 2006 is presented in Figure 8. Because of the scattered nature of warm season precipitation, there is less spatial cohesion of the precipitation than seen in January, but several features are well simulated by all models. All models represent the enhanced precipitation over southern Florida, although, the WRF PXACM is more widespread than the NPA analysis and all the other simulations. It could be argued that the area of precipitation in New Mexico, western Texas and Oklahoma is represented best by the WRF PXACM because the MM5 PXACM slightly over predicts the monthly total. All simulations under predict the areas of heavier precipitation in the northern Plains and upper Midwest.

A main point is that although the WRF PXACM has a larger error in the winter relative to the other simulations, when 2-m temperature and mixing ratio are considered, the monthly precipitation totals are not much different. This similarity is likely a result of the three-dimensional analysis nudging of moisture, temperature and wind (FDDA) that is common among simulations.

d. Spectral Analysis

Many meteorological variables have distinct behavior on specific timescales such as diurnal variation of temperature due to the 24 hour cycle of the sun, or sharp increases or decreases in temperature and moisture that accompany cold and warm fronts (Eskridge et al., 1997). Models have little problem resolving synoptic-scale variations, especially those models that utilize FDDA (Gilliam et al, 2006; Hogrefe et al., 2006). The diurnal variations in terms of phase or correlation with observations are also well simulated

(Gilliam et al, 2006; Hogrefe et al., 2006), in particular, the near-surface temperature that is strongly influenced by the accurately represented solar cycle. However, evaluations that consider model performance as a function of time of day find that many models do not replicate the amplitude of the diurnal temperature signal well (Betts et al., 1997; Zehnder, 2002; Liu et al., 2003; Chen and Steenburgh, 2005). For example, it is common for models to overestimate the low temperature at night and underestimate the maximum high temperature during the day (Mass et al., 2002). This can be seen in Figure 1 where several models have a warm bias at night and cold bias during the day. In these instances the amplitude will be lower on average than the observations, thus the variability less. To examine the diurnal signal, the variability of the filtered diurnal component of 2-m temperature and mixing ratio for the observations, MM5 PXACM and WRF PXACM simulations are computed. The variability of each model is divided by the observed diurnal variability at each observations site to examine which areas the models diurnal signal is weaker, similar or stronger than the observations. Figure 9 and Figure 10 present the ratio of simulated to observed diurnal variability of 2-m temperature (Figure 9) and mixing ratio (Figure 10) for both January and August.

Two issues regarding the filtered model signals are important to examine. The first assessment is how well the models replicate the diurnal variability of the observations. The second issue is whether or not the implementation of PX LSM and ACM2 PBL in WRF is consistent with MM5. For January 2006 (Figure 9) the ratio of simulated to observed variability of the diurnal 2-m temperature forcing indicates both the WRF (Panel A) and MM5 (Panel C) underestimate, by a factor of two, the diurnal variability at many sites along the Gulf and Atlantic coast. The specific time series at several of these

sites (not shown) reveal the diurnal amplitude of the model is indeed much dampened. A likely explanation is that the observation sites are in a model grid cell that is partly water, thus, the model grid cell average temperature variation will be lessened.

The models replicate the diurnal forcing of temperature at many other sites in the domain fairly well, although slightly underestimated, as the ratios of variability are generally between 0.70 and 1.10 (blue to teal colors). Also, the spatial pattern of the variability ratio is consistent between the MM5 PXACM and WRF PXACM implementations. The main difference is in the northeast part of the domain and inland parts of the southeast US where the WRF PXACM has a much lower diurnal forcing than the MM5 PXACM. In many of these same locations, in particular the inland parts of the southeast and mid-Atlantic, the WRF PXACM had higher temperature error relative to the temperature analysis than the MM5 PXACM (Figure 3, panel C and Figure 4, panel B). A time series comparison of the two simulations with the 2-m temperature observations at a site in Massachusetts (not shown) reveals that the MM5 PXACM captured the daily high temperatures during several periods of warmer than normal weather in January, while the WRF PXACM was much colder. The WRF PXACM has the same problem capturing the afternoon high temperatures during warm periods in January at sites in central South Carolina, where the WRF model has a small diurnal temperature variability compared to observations.

The WRF ACMPX for January 2006, does show some slight improvement (Figure 9, panels A and C) over the MM5 implementation over parts of the central Plains, Midwest and Great Lake states where the diurnal variability ratio with observations is closer to 1.0. Several sites in Iowa, Kansas, Wisconsin and Michigan were examined and the time

series indicate that the WRF PXACM matched the observations much better on the colder nights in January, which explains the more comparable variability relative to the observations when contrasted with the MM5 PXACM. In fact, these cold nights are on days where the fractional snow cover is greater than zero, which along with some of the other comparisons, indicate that the new fractional snow cover algorithm in the WRF PXACM does indeed improve temperature predictions over surfaces covered by shallow snow.

As with other comparisons of the WRF PXACM and MM5 PXACM, the diurnal 2-m temperature forcing in August 2006 (Figure 9, Panels B and D) is much closer to the observations (dark blues to light greens) than in January, except the coastal sites. Again, these coastal sites are not resolved well by the model because a fraction of the cell is simulated as water, where diurnal temperature forcing is less than land. Spatial patterns, in terms of which areas each model simulation replicates the diurnal amplitude of temperature well, are loosely related to the performance in Figure 5. The WRF PXACM has a variability ratio with the observations closer to 1.0 (light greens) in the states surrounding the Ohio River Valley. The specific time series at several sites in this area were examined (not shown) and the WRF PXACM resolved the daily low temperature better than the MM5 PXACM. Along the front range of the Rocky Mountains, the MM5 PXACM has a diurnal variability ratio with the observations closer to 1.0 than the WRF PXACM. Time series were examined at a few of the sites and the MM5 PXACM resolved the low temperatures much better while both models were similar with the high temperature. Otherwise, many of the other parts of the country were similarly simulated in terms of diurnal temperature forcing.

Figure 10 presents the differences between the MM5 and WRF implementations of PXACM are even more alike, both in August than January 2006, in simulating the diurnal variations of 2-m mixing ratio. The diurnal variability compared to observations is larger in both models by a factor of 1.5 to 2.0 over much of the southern half of the model domain and closer to 1.0 over the northern part of the domain. This again improves the confidence that the new physics in WRF are comparable to the existing MM5 version during the warm season. However, it does raise an issue that diurnal 2-m mixing ratio varies more than observations throughout the day. This may be tied to excessive latent heat flux from the ground or vegetation.

4. Summary

The Pleim-Xiu LSM, Pleim surface-layer scheme, and Asymmetric Convective Model version 2 were implemented in Version 3 of the Advance Research WRF model. This new combination of physics is best used for retrospective simulations, in particular, those used to drive air quality models. Several upgrades were made to the PX LSM in WRF including a new snow heat capacity that accounts for fractional land-use dependent snow cover and deep soil temperature nudging (Pleim and Gilliam, 2008).

Simulations using the new WRF implementation were conducted for a January and August period in 2006. For benchmark purposes, a similar PX LSM and ACM2 MM5 simulation and three other WRF simulations with commonly used physics configurations were also executed. A comparison of model performance was conducted that mostly focused on 2-m temperature and mixing ratio using point observations and model analyses. Cumulative or bulk statistics indicate the new physics implementation in WRF,

considering temperature and mixing ratio, does not perform as well as the MM5 counterpart during winter, but is comparable for the summer month. It was discovered that the analyses that drive the soil moisture and temperature nudging of the PX LSM have different precision when compared to point observations. The analysis used by MM5 was from the RAWINDS utility that re-introduces the point observations to a base model analysis (NAM) that is interpolated to the MM5 grid, while the WRF simulations used the NAM analysis directly interpolated to the WRF grid. It is expected that once a RAWINDS-like utility is developed for WRF (utility called *OBSGRID* is in progress) the WRF simulations using the PX LSM and ACM2 will further improve and merge or even outperform the MM5 implementation, as initial tests support.

Model analyses were used to complete a gridded spatial evaluation of the model simulations. This type of evaluation is beneficial for several reasons. First, a model analysis is designed to be the best guess of the state of the atmosphere that considers a short-term (typically 3-hour) model forecast and multiple types of observations using some method of error minimization. It would be challenging for a model to consistently outperform the analysis in terms of how both compare to point surface observations. Second, the model analysis is directly used for soil nudging in the PX LSM, thus, the most appropriate dataset for evaluating the 2-m temperature from that model.

For January 2006 the WRF PXACM had similar 2-m temperature error as the analysis at night, but much greater error during the day, where not only the WRF PX ACM simulations had a large cold bias, but the three other WRF simulations had large error and cold bias. The new fractional snow cover algorithm improved the simulated 2-

m temperature when compared to the MM5 in areas that are south of the permanent snowpack (Plains, Mid-west and Great Lakes regions).

During the summer month of August 2006, the 2-m temperature simulation of the WRF PXACM was very close to the MM5 implementation in terms of error and bias, with notable improvement over the Rocky Mountains. The uncertainty of the 2-m temperature predictions was in some cases less in the WRF PXACM than the model analysis. Also, the WRF PXACM and MM5 PXACM temperature and moisture simulation contained less error than the other WRF simulations.

An analysis of the variability of the filtered diurnal signal of observations and all model simulations, including both temperature and moisture, indicates the WRF and MM5 implementation of the PXACM are generally similar in terms of diurnal amplitude. However, in the winter, the WRF PXACM had a much depressed diurnal amplitude of temperature over parts of the southeast US and mid-Atlantic regions. There is some evidence of lower shortwave radiation in these same areas, so a difference in cloud cover is suspected. Improvements in the simulated low temperature by the WRF PXACM in the Plains, Mid-west and Great Lakes are attributed to the fractional snow cover algorithm.

Considering the various evaluation approaches presented in this study, and the benchmark of the new implementation in WRF with the existing MM5 version and other WRF configurations, we feel the WRF PXACM is reasonable, especially during the warmer seasons. However, several ongoing investigations and corresponding improvements are necessary. A snow model to improve the temperature and associated PBL properties over snow cover is recommended. Also, we see evidence that during the winter the WRF implementation of the PXACM generates excessive cloud cover at the

top of the PBL, which impacts the daytime high temperature in some areas. The cloud issue needs to be examined in more detail. Finally, it is recommended that model analyses be used more regularly in model evaluation studies, as presented here, since they represent a best guess of the atmosphere considering multiple observations.

Disclaimer – *The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency review and approved for publication.*

Acknowledgments.

It is necessary to acknowledge NCAR and collaborators who have invested much time and effort in the WRF model development and helped to fit our physics parameterizations in the new release. NOAA ESRL/GSD should be recognized for the development of the MADIS observational archive that was critical for our evaluation. Finally, the National Climate Data Center should be thanked for the analyses used to drive the WRF simulations.

References

Betts, A.K., F. Chen, K.E. Mitchell, and Z.I. Janjić, 1997: Assessment of the Land-surface and Boundary Layer Models in Two Operational Versions of the NCEP Eta Model Using FIFE Data. *Mon. Wea. Rev.*, **125**, 2896–2916.

Byun, D., and K.L. Schere, 2006: Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Applied Mechanics Reviews*, **59**, 51-77.

Cheng, W.Y.Y., and W.J. Steenburgh, 2005: Evaluation of Surface Sensible Weather Forecasts by the WRF and the Eta Models over the Western United States. *Wea. Forecasting*, **20**, 812–821.

Dudia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077-3107.

Dyer, A. J., and B. B. Hicks, 1970: Flux-gradient relationships in the constant flux layer. *Quart. J. Roy. Meteor. Soc.*, **96**, 715-721.

Ek, M.B., K.E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, and V. Koren, 2003: Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.

Eskridge, R.E., J.Y. Ku, S.T. Rao, P.S. Porter, and I.G. Zurbenko, 1997: Separating Different Scales of Motion in Time Series of Meteorological Variables. *Bull. Amer. Meteor. Soc.*, **78**, 1473–1483.

Fulton, R.A., J.P. Breidenbach, D.J. Seo, D.A. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377-395.

Gilliam R.C., Bhave, P.V., Pleim, J.E., and T.L. Otte, 2004: A year-long MM5 evaluation using a model evaluation toolkit. 3rd Annual CMAS Conference, October 18-20, 2004, http://www.cmascenter.org/conference/2004/abstracts/poster/gilliam_abstract.pdfGilliam

Gilliam, R.C., W. Appel, and S. Philips, The Atmospheric Evaluation Tool (AMET): Meteorology Module, presented at the 4th Annual CMAS Conference, September 26-28, 2005, http://www.cmascenter.org/conference/2005/abstracts/6_1.pdf.

Gilliam, R.C., Hogrefe, C., and S.T. Rao, 2006: New methods for evaluating meteorological models used in air quality applications. *Atmos. Environ.*, **40**, 5073-5086.

Grell, G. A., J. Dudhia, and D.R. Stauffer: 1995, A Description of the Fifth-Generation PennState/NCAR Mesoscale Model (MM5), NCAR Technical Note NCAR/TN-398+STR. Available at: http://www.mmm.ucar.edu/mm5/doc1.html.

Hogrefe, C., S. T. Rao, I. G. Zurbenko, and P. S. Porter, 2000: Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the eastern United States; *Bull. Amer. Meteor. Soc.*, **81**, 2083 - 2106.

Hogrefe, C., P.S. porter, E. Gego, A. Gilliland, R. Gilliam, J. Swall, J. Irwin, and S.T. Rao, 2006: Temporal features in observed and predicted $pm_{2.5}$ concentrations over the eastern United States. *Atmos. Environ.*, **40**, 5041-5055.

Hong, S.Y., J. Dudhia, and S.H. Chen, 2004: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation. *Mon. Wea. Rev.*, **132**, 103–120.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.

Janjic, Z. I., 1996: The surface layer in the NCEP Eta Model. Eleventh Conference on Numerical Weather Prediction, Norfolk, VA, 19-23 August 1996; Amer. Meteor. Soc., Boston, MA, 354-355.

Janjic, Z. I., 2002: Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model. NCEP Office Note No. 437, 61 pp.

Kain, J.S., 2004: The Kain–Fritsch convective parameterization: An update. *J. of App. Meteor.*, **43**, 170–181.

Liu, G., C. Hogrefe and S.T. Rao, 2003: Evaluating the performance of regional-scale meteorological models: effect of clouds simulation on temperature prediction. *Atmos. Environ.* **37**, 1425-1433.

Mass, C.F., D. Ovens, K. Westrick, and B.A. Colle, 2002: Does Increasing Horizontal Resolution Produce More Skillful Forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative trans-fer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. *J. Geophys. Res.*, **102**(D14), 16663-16682.

Noh, Y., W. G. Cheon, S.-Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.*, **107**, 401–427.

Otte, T.L., 2008: The impact of nudging in the meteorological model for retrospective air quality simulations. Part I: Evaluation against national observations networks. *J. Appl. Meteor*. Clim., In Press.

Pleim, J.E., and J. S. Chang, 1992: A non-local closure model for vertical mixing in the convective boundary layer. *Atmos. Environ.*, **26A,** 965–981.

Pleim, J.E., and A. Xiu, 2003: Development of a Land-surface Model. Part II: Data Assimilation. *J. Appl. Meteor.*, **42**, 1811–1822.

Pleim, J.E., 2006: A Simple, Efficient Solution of Flux–Profile Relationships in the Atmospheric Surface Layer. *J. Appl. Meteor. Clim.*, **45**, 341–347.

Pleim, J.E., 2007a: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395.

Pleim, J.E., 2007b: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part II: Application and Evaluation in a Mesoscale Meteorological Model. *J. Appl. Meteor. Climatol.*, **46**, 1396–1409.

Pleim, J.E., and Gilliam R.C., 2008: An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land-surface model. Submitted to *J. Appl. Meteor. Climatol.*

Rao, S.T., I.G. Zurbenko, R. Neagu, P.S. Porter, J.-Y. Ku, and R.F. Henry, 1997: Space and time scales in ambient ozone data. *Bull. Amer. Meteor. Soc.*, **78**, 2153-2166.

Reisner, J., R. M. Rasmussen, and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **124**, 1071–1107.

Seo, D.J., 1998a: Real-time estimation of rainfall fields using rain gauge data under fractional coverage conditions. *J. of Hydrol.*, **208**, 25-36.

Seo, D.J., 1998b: Real-time estimation of rainfall fields using radar rainfall and rain gauge data. *J. of Hydrol.*, **208**, 37-52.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M.Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF version 2. NCAR Tech Note NCAR/TN 468_STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307.]

Smirnova, T.G., J.M. Brown, and S.G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. *Mon. Wea. Rev.*, **125**, 1870-1884.

Smirnova, T.G., J.M. Brown, S.G. Benjamin, and D. Kim, 2000: Parameterization of cold seasons processes in the MAPS land-surface scheme. *J. Geoph. Res.*, **105** (D3), 4077-4086.

Stauffer, D. R., and N. L. Seaman, 1990: Use of four-dimensional data assimilation in a limited area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Wea. Rev.*, **118**, 1250–1277.

Stauffer, D. R., N. L. Seaman, and F. S. Binkowski, 1991: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation within the planetary boundary layer. *Mon. Wea. Rev.*, **119**, 734–754.

Stauffer, D. R., and N. L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor*., **33**, 416–434.

Xiu, A., and J.E. Pleim, 2001: Development of a Land-surface Model. Part I: Application in a Mesoscale Meteorological Model. *J. Appl. Meteor.*, **40**, 192-209.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences, pp 277-281. Academic Press, New York.

Zehnder, J.A., 2002: Simple Modifications to Improve Fifth-Generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model Performance for the Phoenix, Arizona, Metropolitan Area. *J. Appl. Meteor.*, **41**, 971–979.

List of Figures

Figure 10 The filtered (diurnal) 2-m moisture variability. The plotted values are the ratio of model to observed (model/observation) variability for the MM5 (panel A and B) and WRF (panel C and D) PXACM simulations for January and August 2006.

Table 1 Configuration of various model simulations. Numbers indicate physics option numbers in WRF V3.0 and MM5.

RUN ID	LSM	PBL	Surface-layer
WRF PXACM	PX (7)	ACM2 (7)	Pleim (7)
MM5 PXACM	PX (7)	ACM2 (7)	Pleim
WRF NOAHYSU	NOAH (2)	YSU (1)	M-O (1)
WRF NOAHMYJ	NOAH (2)	MYJ (2)	M-O Janic (2)
WRF RUCYSU	RUC (3)	YSU (1)	M-O (1)
Other Common Configurations	**All WRF**	**MM5**	
Microphysics	WSM 6-class (6)	Reisner 2 (7)	
Convective	Kain-Fritsch 2 (1)	Kain-Fritsch 2 (8)	
Shortwave	Dudhia (1)	Dudhia	
Longwave	RRTM (1)	RRTM (4)	
Initial conditions	NAM-218	NAM-218	
Boundary conditions	NAM-218	NAM-218	
FDDA driver	NAM-218	NAM-218	
Soil moisture nudging	NAM-218	RAWINDS	

Table 2 Summary of surface-based model performance statistics for each simulation

	WRF PXACM	MM5 PXACM	WRF NOAHYSU	WRF NOAHMYJ	WRF RUCYSU
2-m Temperature (K)					
January RMSE	2.80	2.41	2.31	2.92	2.36
January BIAS	-0.90	-0.34	0.04	-1.41	-0.33
August RMSE	2.03	1.97	2.23	2.28	2.84
August BIAS	0.13	-0.07	0.18	-0.47	1.47
2-m Mixing Ratio ($g\ kg^{-1}$)					
January RMSE	0.99	0.90	0.86	0.88	1.04
January BIAS	0.18	0.31	0.30	0.20	0.59
August RMSE	1.90	1.88	2.16	1.85	2.44
August BIAS	0.20	0.78	-0.98	-0.40	-0.85

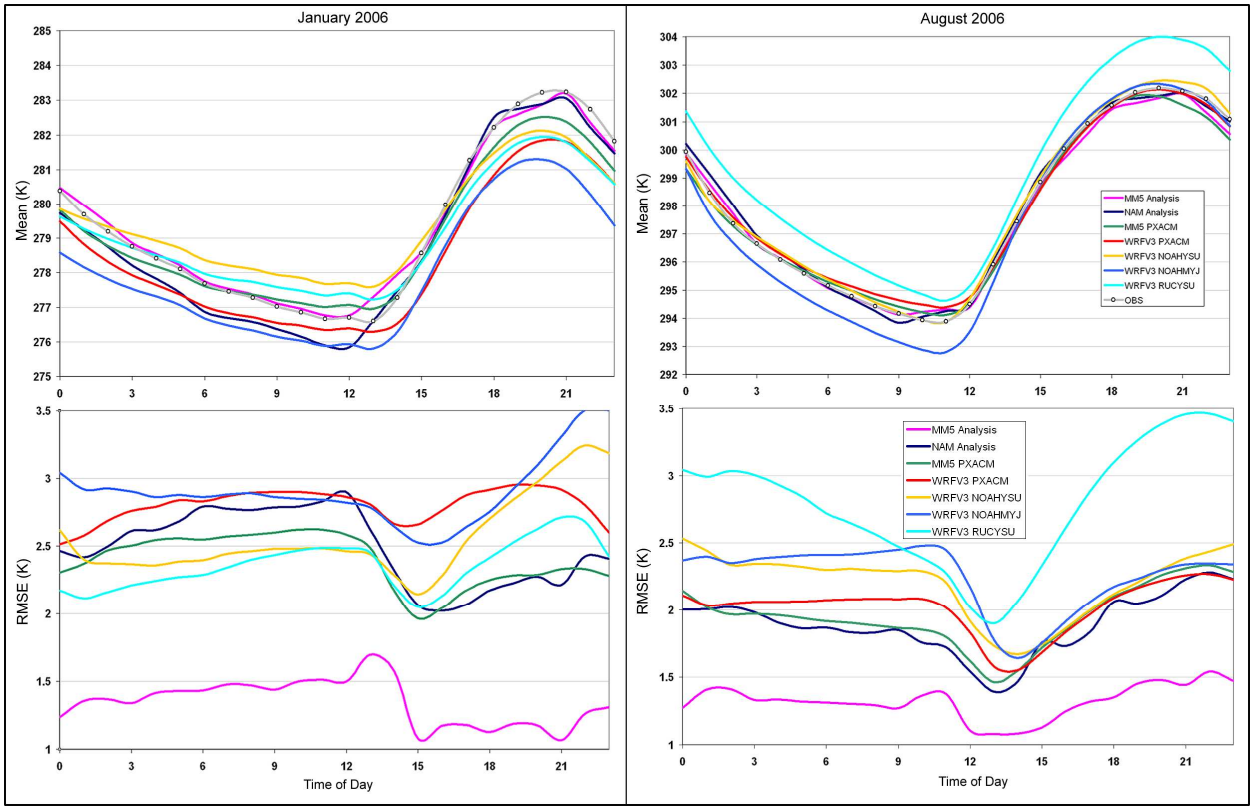

Figure 1 Diurnal (UTC) segregated mean (top) and RMSE (bottom) of model simulated 2-m temperature for January (left) and August (right) of 2006. The statistics include observations combined from all sites in the model domain.

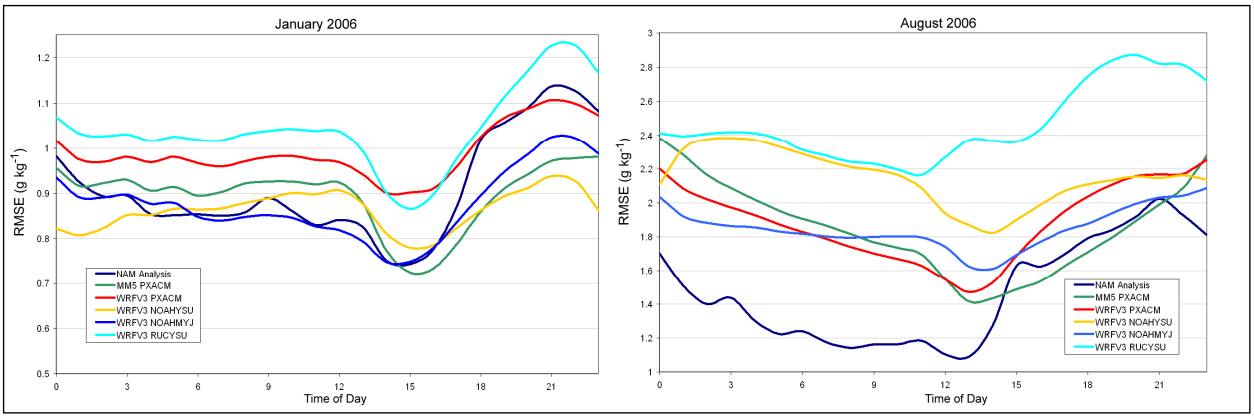

Figure 2 Diurnal (UTC) segregated RMSE of simulated 2-m mixing ratio for January and August of 2006. The statistics include observations from all sites in the model domain.

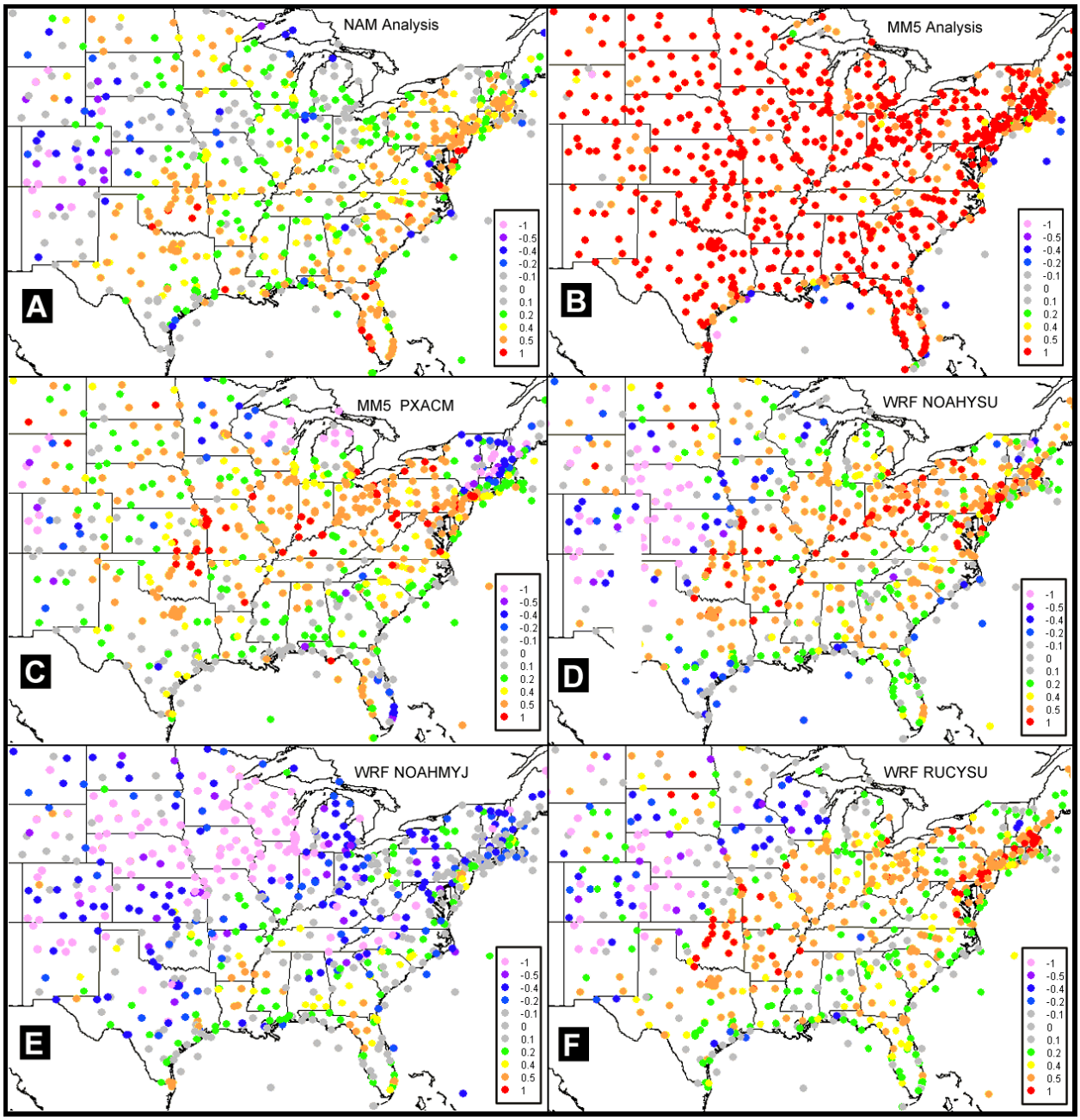

Figure 3 Difference of RMSE of the WRF PXACM simulated 2-m temperature and the other simulations or analyses (noted in upper right corner of each panel) for January 2006. Cold (Warm) colors or negative (positive) values indicate WRF PXACM has less (more) error than the specified simulation/analysis.

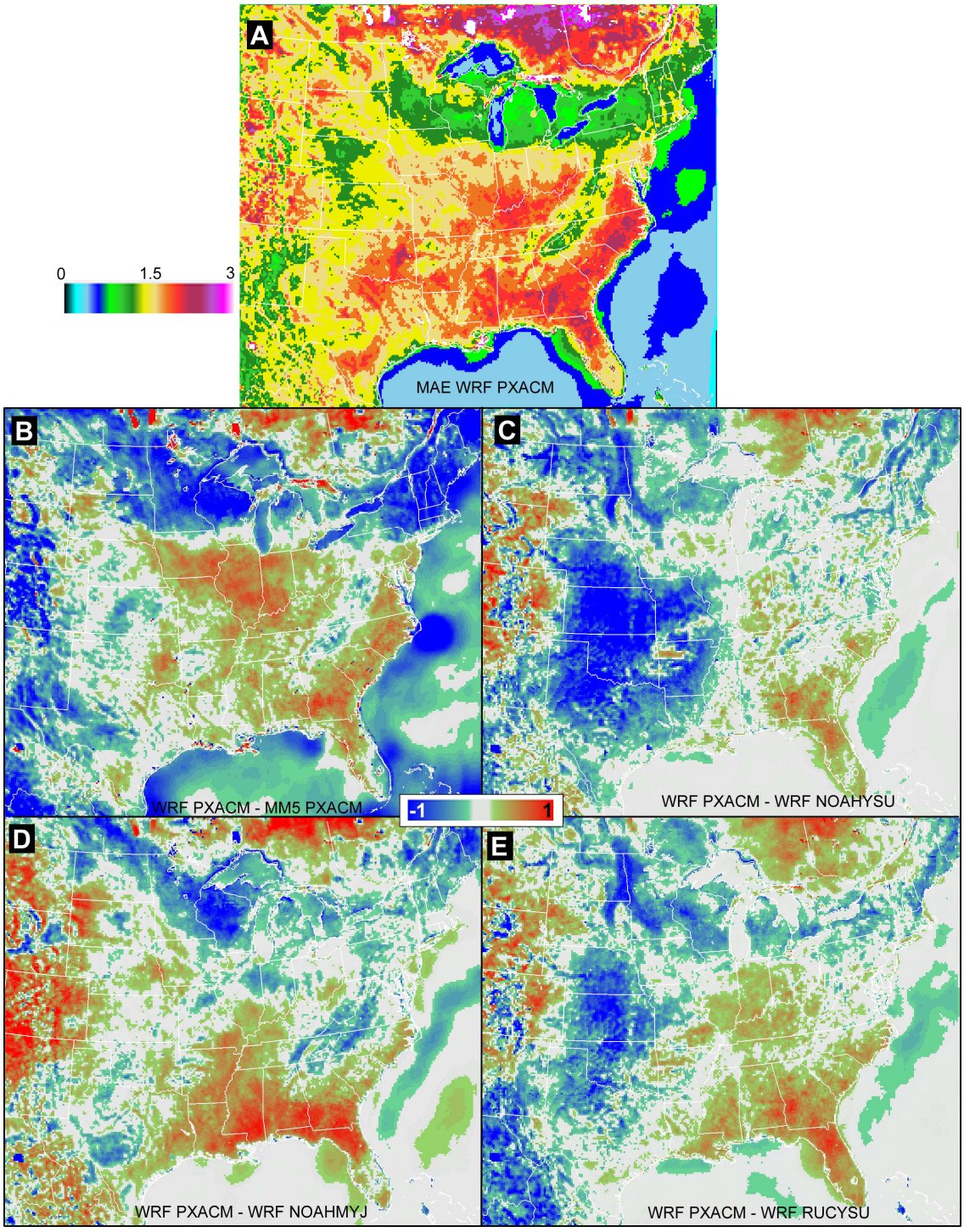

Figure 4 Panel (A) Mean absolute error of simulated 2-m temperature when compared to NAM-12km analysis for the month of January 2006. Panels (B-E) are the difference of MAE of 2-m temperature of the WRF PXACM and each of the other simulations. Note: cool (warm) colors indicate areas were the WRF PXACM has lower (higher) error and the difference scale ranges from -1.0 to 1.0 K

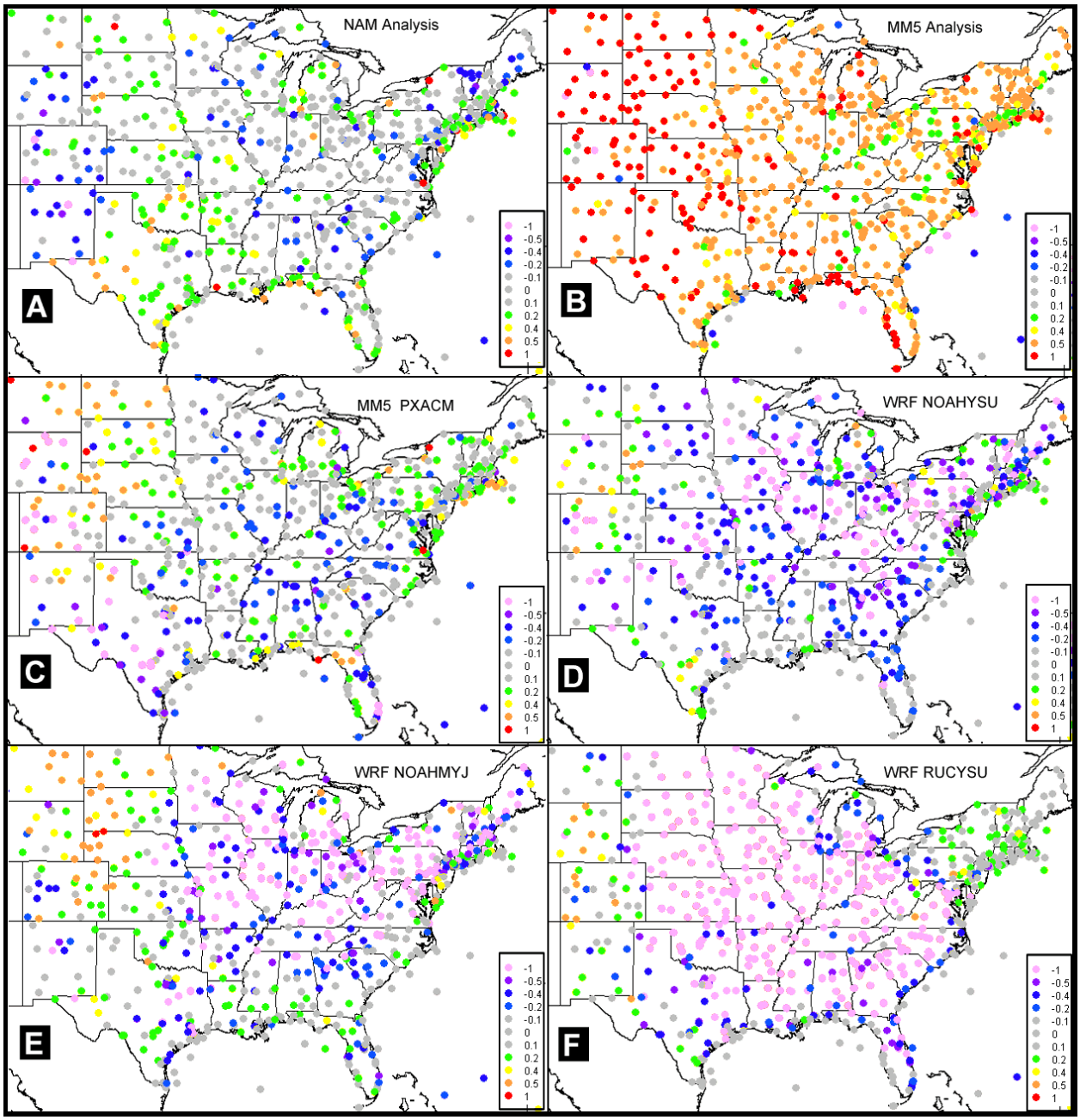

Figure 5 Difference of RMSE of the WRF PXACM simulated 2-m temperature and the other simulations or analyses (noted in upper right corner of each panel) for August 2006. Cold (Warm) colors or negative (positive) values indicate WRF PXACM has less (more) error than the specified simulation/analysis.

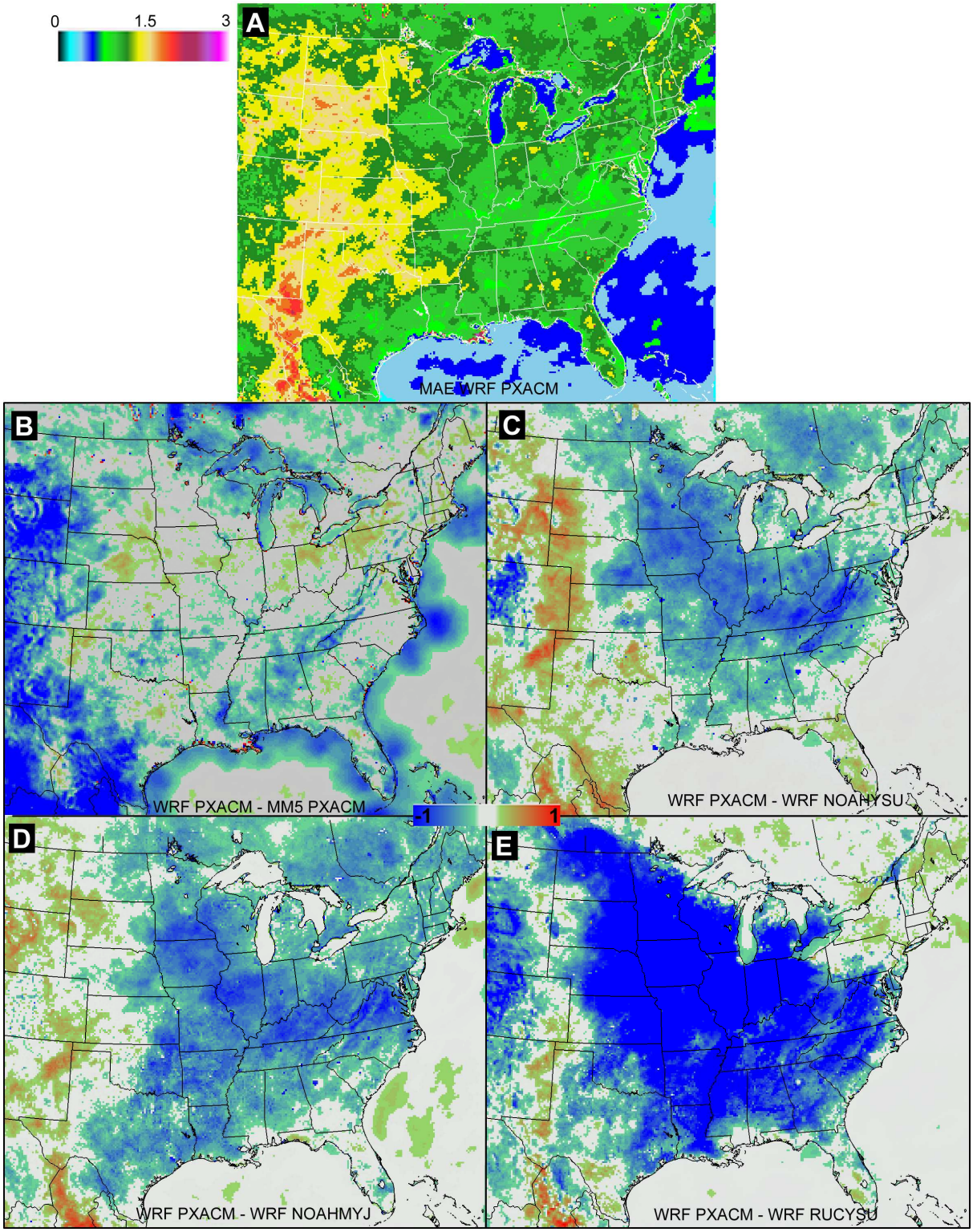

Figure 6 Panel (A) Mean absolute error of simulated 2-m temperature when compared to NAM-12km analysis for the month of January 2006. Panels (B-E) are the difference of MAE of 2-m temperature of the WRF PXACM and each of the other simulations. Note: cool (warm) colors indicate areas were the WRF PXACM has lower (higher) error and the difference scale ranges from -1.0 to 1.0 K

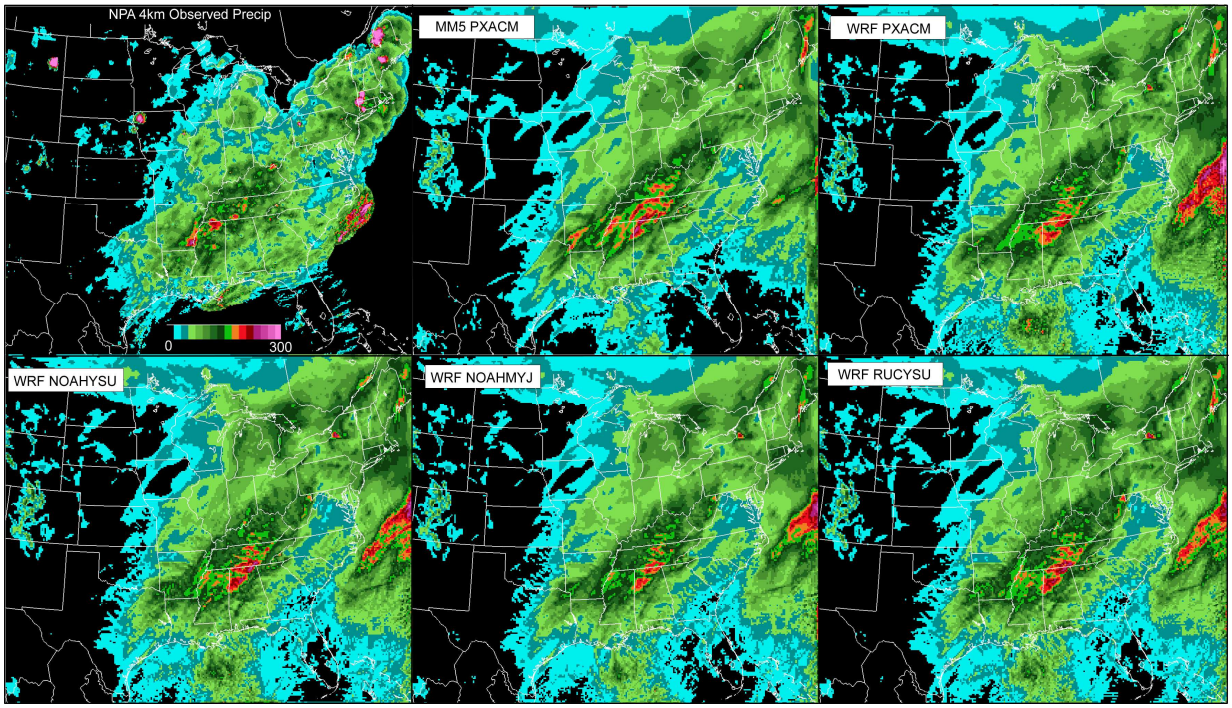

Figure 7 Observed and simulated monthly precipitation (mm) for January 2006.

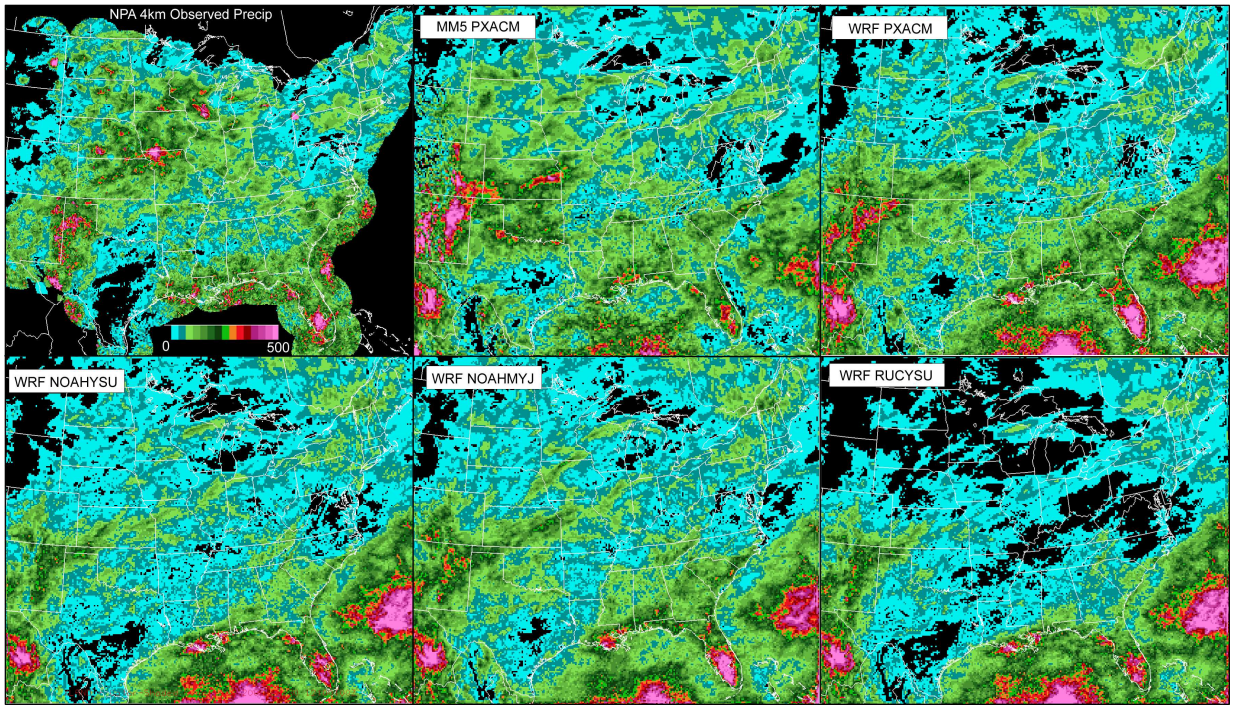

Figure 8 Observed and simulated monthly precipitation (mm) for August 2006.

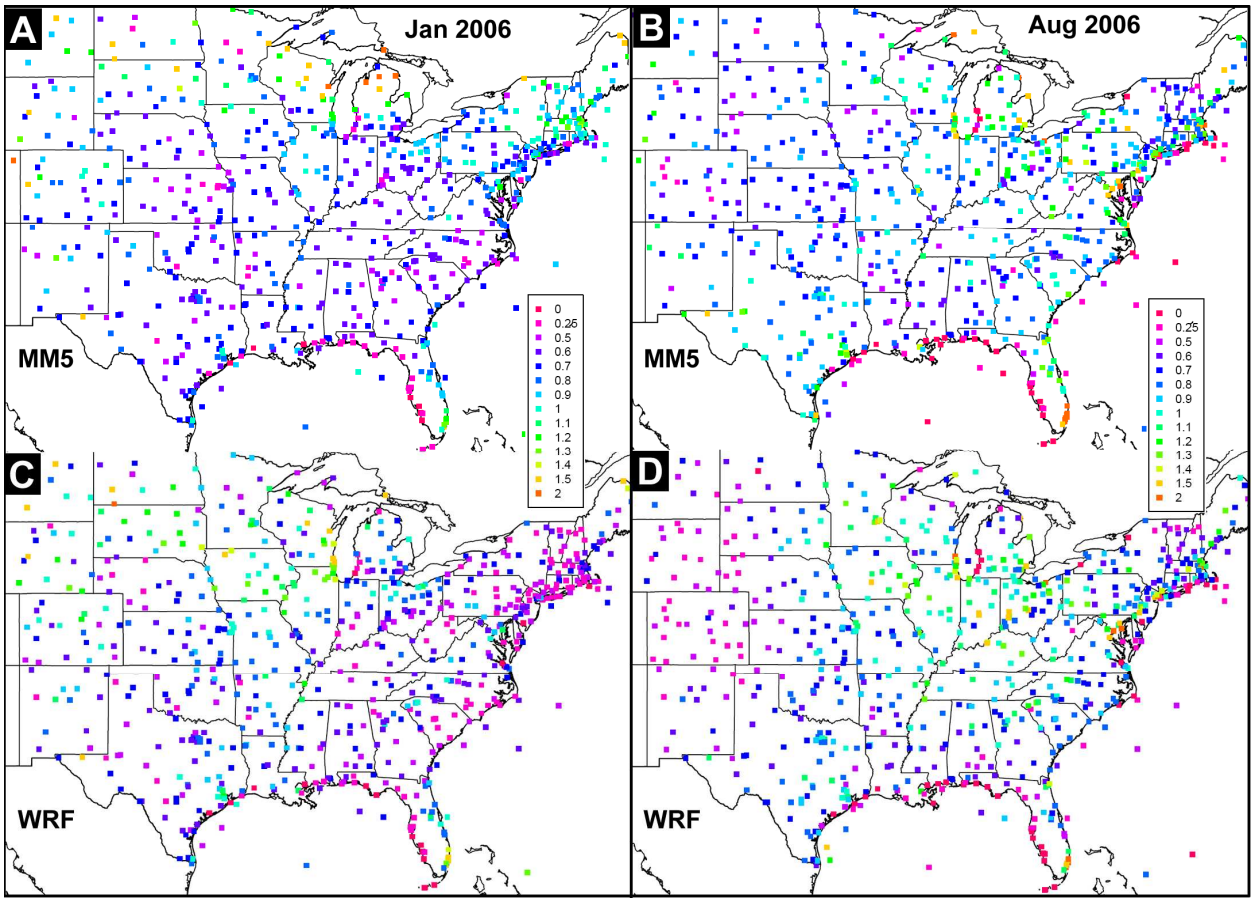

Figure 9 The filtered (diurnal) 2-m temperature variability. The plotted values are the ratio of model to observed (model/observation) variability for the MM5 (panel A and B) and WRF (panel C and D) PXACM simulations for January and August 2006.

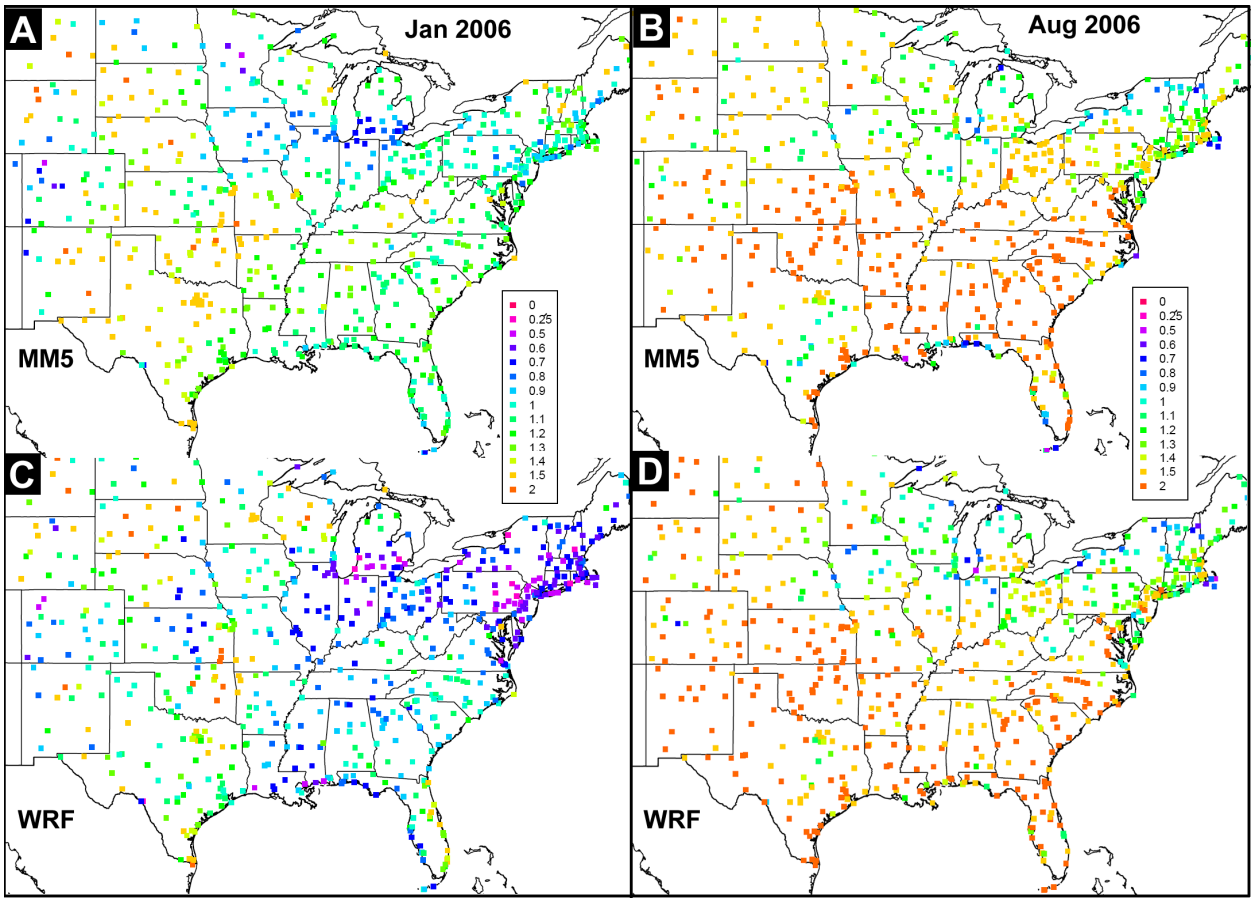

Figure 10 The filtered (diurnal) 2-m moisture variability. The plotted values are the ratio of model to observed (model/observation) variability for the MM5 (panel A and B) and WRF (panel C and D) PXACM simulations for January and August 2006.