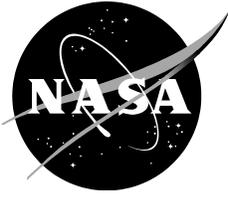


NASA/TP-2017-219579



NASA Downscaling Project

Final Report

Robert Ferraro

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Duane Waliser

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Christa Peters-Lidard

Goddard Space Flight Center, Greenbelt, Maryland

February 2017

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

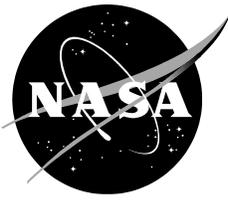
- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.
- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question to help@sti.nasa.gov
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:
NASA STI Information Desk
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

NASA/TP-2017-219579



NASA Downscaling Project

Final Report

Robert Ferraro

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Duane Waliser

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Christa Peters-Lidard

Goddard Space Flight Center, Greenbelt, Maryland

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
Pasadena, CA 91109-8099

February 2017

Acknowledgments

Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

© 2016. All rights reserved.

Level of Review: This material has been technically reviewed by technical management.

Available from

**NASA STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
757-864-9658**

**National Technical Information Service
5301 Shawnee Rd.
Alexandria, VA 22312
webmail@ntis.gov
703-605-6000**

This report is also available in electronic form at

<http://ntrs.nasa.gov>

Table of Contents

| | | |
|-----|---|----|
| 1. | Executive Summary..... | 4 |
| 2. | Project Overview | 5 |
| | 2.1. Motivation | 5 |
| 3. | Experiment Design Considerations..... | 7 |
| | 3.1. Metrics Objectives..... | 7 |
| | 3.3. Available Data for Comparisons | 8 |
| | 3.4. Metrics Employed in the Evaluations..... | 8 |
| | 3.5. Model Considerations | 12 |
| | 3.6. High Performance Computing Resources | 14 |
| 4. | Experiments Description and Model Configuration..... | 14 |
| | 4.1. NU-WRF Model Configuration | 15 |
| | 4.2. Non-NASA Physics Parameterization Schemes | 15 |
| | 4.3. Atmospheric Initial and Boundary Conditions..... | 16 |
| | 4.4. Land Surface Initialization..... | 16 |
| | 4.5. M2R12K/GEOS Replay Runs | 17 |
| | 4.6. Utilization of NASA Computational Resources..... | 17 |
| 5. | Phase I: Pilot Experiment Simulations and Issues Encountered..... | 18 |
| | 5.1. Domain Selection and Nesting..... | 18 |
| | 5.2. Spectral Nudging Considerations | 19 |
| | 5.3. Datasets Produced..... | 19 |
| | 5.4. Code Issues and Fixes..... | 20 |
| | 5.5. Pilot Experiments Metrics..... | 21 |
| 6. | Phase II: Single-Decade Production Experiments..... | 28 |
| | 6.1. Initial Production Simulations..... | 28 |
| | 6.2. Metrics Evaluation of Experiments | 30 |
| 7. | Conclusions | 44 |
| | 7.1. CONUS PDF | 44 |
| | 7.2. ARs..... | 45 |
| | 7.3. MCS..... | 45 |
| | 7.4. NES | 46 |
| 8. | Recommendations for Future Study | 46 |
| 9. | Conference Presentations and Papers in Preparation | 47 |
| | 9.1. Presentations | 47 |
| | 9.2. Papers | 48 |
| 10. | Downscaling Team Members..... | 49 |
| 11. | References..... | 49 |

1. Executive Summary

A team of researchers from NASA Ames Research Center, Goddard Space Flight Center, the Jet Propulsion Laboratory, and Marshall Space Flight Center, along with university partners at UCLA, conducted an investigation to explore whether downscaling coarse resolution global climate model (GCM) predictions might provide valid insights into the regional impacts sought by decision makers. Since the computational cost of running global models at high spatial resolution for any useful climate scale period is prohibitive, the hope for downscaling is that a coarse resolution GCM provides sufficiently accurate synoptic scale information for a regional climate model (RCM) to accurately develop fine scale features that represent the regional impacts of a changing climate. As a proxy for a prognostic climate forecast model, and so that ground truth in the form of satellite and in-situ observations could be used for evaluation, the MERRA and MERRA-2 reanalyses were used to drive the NU-WRF regional climate model and a GEOS-5 replay. This was performed at various resolutions that were at factors of 2 to 10 higher than the reanalysis forcing. A number of experiments were conducted that varied resolution, model parameterizations, and intermediate scale nudging, for simulations over the continental US during the period from 2000–2010. The results of these experiments were compared to observational datasets to evaluate the output.

The team adopted a metrics-based approach to evaluation, rather than examining detailed diagnostics, as the RCM output fields were not expected to exactly match the weather during the historical period. The study focused on the ability to reproduce regionally impactful events such as atmospheric rivers on the West Coast, mesoscale convective systems in the central US, and north-east winter storms (NES) in the Mid-Atlantic and New England states. Simple metrics were devised to assess how well the downscaling experiments matched the historical observations, to observe the impact of varying spatial resolution, two level nesting, domain size, parameterization schemes, and nudging of the simulations to the synoptic scale. A series of 10-month pilot experiment simulations were conducted to test the sensitivity of the RCM to these different conditions, and to decide which experiments to pursue for longer, decadal-scale simulations. Ultimately, a set of 10 decadal scale experiments were conducted at 24 km, 12 km, and 4 km spatial resolution, with and without nudging at 2000 km and 600 km (Because of code problems, computational and time limitations, the 4-km experiment only completed about half a decade). The metrics developed for this project were applied to each of these experiments.

Based on the metrics developed and applied, the results did not show dramatic improvement of the downscaled fields in most cases compared to the reanalysis fields (which were at approximately 0.5 degree resolution). The effects of higher resolution resulted in only marginal improvements in most cases. Nudging showed some benefit in the metrics over un-nudged simulations, but again the effects were not dramatic. Overall, the mixed results of experiments did not demonstrate a compelling value for downscaling, at least over the test conditions employed in this study. However, it is extremely important to recognize the limitations associated with the observations, metrics and evaluation process. To do the evaluations, model results down to grid resolutions on the order of 1-10 km must be examined, at scales where physical processes (e.g. convection, extreme weather) and relevant boundary conditions (e.g. land-sea contrast, topography) are expected to have considerable bearing on climate impacts, decision support, and associated assessments. Satellite products and even reanalysis products have very little useful information at these scales. Gridded in-situ data of precipitation and temperature exist – but in many places are highly under-sampled (e.g. Stage IV, PRISM). Point wise in-situ data (e.g. ARM site, or other amalgams of surface and sounding observations) might provide more local observation constraints, but are themselves sparse. Overall, the lack of high quality, high resolution, high frequency, and long time series data limits what can be assessed and concluded at the finest scales.

The study encountered several difficulties during its execution which, together with resource limitations, prevented the team from doing a more exhaustive (and perhaps conclusive) set of experiments, as initially intended. What the study ultimately explored was the value of

downscaling reanalysis. Still, the lack of demonstrable improvement due to resolution increase was surprising. However, within the limitations of time and resources of the project, the team is not able to explain why such improvement was not observed.

Despite the lack of conclusive proof one way or the other on the study question, the team found significant value in the project. For example, a four-center NASA team came together to productively work on a new area for NASA; improved NU-WRF versions are now able to carry out climate-scale experimentation; and means to perform new evaluations of weather and climate variability from both the NU-WRF and downscaled MERRA-replay simulations are now available. It also developed a number of model fidelity and climate impact metrics relevant to synoptic/regional and continental scales, and thus to the NCA or even CMIP/IPCC assessments.

This study provides NASA a means to contribute to regional-scale climate modeling and impacts assessments, whether through dynamic downscaling of limited area models (e.g. NU-WRF), downscaling via GEOS5 (replay) capabilities, and/or model evaluation and impacts assessments. This capability, as well as the team's productive and collegial effort, was recognized by the two external advisors to the project. An additional outcome of this study should be a broader recognition that as yet our observational, satellite-based resources are still sorely limited for addressing relatively high-resolution climate processes and associated impacts relevant to applications and decision support.

2. Project Overview

2.1. Motivation

In the context of climate modeling, “dynamic downscaling” refers to the practice of driving a limited-area model with boundary conditions derived from a previously executed GCM simulation. Its potential advantages include higher spatial resolution and/or a more comprehensive treatment of physical processes given the savings in computational resources by only focusing on a specific region of the Earth. In this context, the limited-area model is typically referred to as an RCM.

Despite the significant use of dynamic downscaling with RCMs, e.g. North American Regional Climate Change Assessment Program (NARCCAP) (Mearns 2009), Coordinated Regional climate Downscaling Experiment (CORDEX) (Giorgi & Gutowski 2015), ENSEMBLES (Hewitt & Griggs 2004), particularly for climate assessment and impact studies, the credibility of dynamic downscaling has been called into question (e.g. Pielke 2012; Hall 2014). Concerns are raised over: 1) the mismatches in parameterized physical processes between the GCM and RCM that may lead to process discontinuities that influence the fidelity of the downscaled climate, 2) the inconsistencies between coarse resolution topography and associated flow in the GCM as lateral boundary conditions for an RCM domain which represents fine (e.g. taller) features of the topography, 3) missing regional-scale information in the lateral forcing, and 4) possible ramifications associated with the inability for the evolution in the RCM model to feed back onto the large-scale flow/forcing, etc.

Given the important role that dynamically downscaled climate projections are playing in climate change impacts assessment, it is vital to quantify the specific contributions, if any, that such downscaling offers. Figure 1 illustrates the typical framework of a climate downscaling experiment. Step I is used to evaluate the basic fidelity of the RCM in representing regional climate characteristics given observed boundary condition (BC) forcing. Step II is used to evaluate the fidelity of the RCM when used in conjunction with BCs, and thus the strengths and shortcomings, of a given GCM. For Steps I and II, observations can be utilized to characterize model performance, where emphasis is usually on the statistical (e.g. means, variance, skewness, extremes), rather than deterministic, character of the simulation. Step III is used to characterize the statistical character of the future climate of the given domain, with the present-day (modeled) reference state taken from Step II.

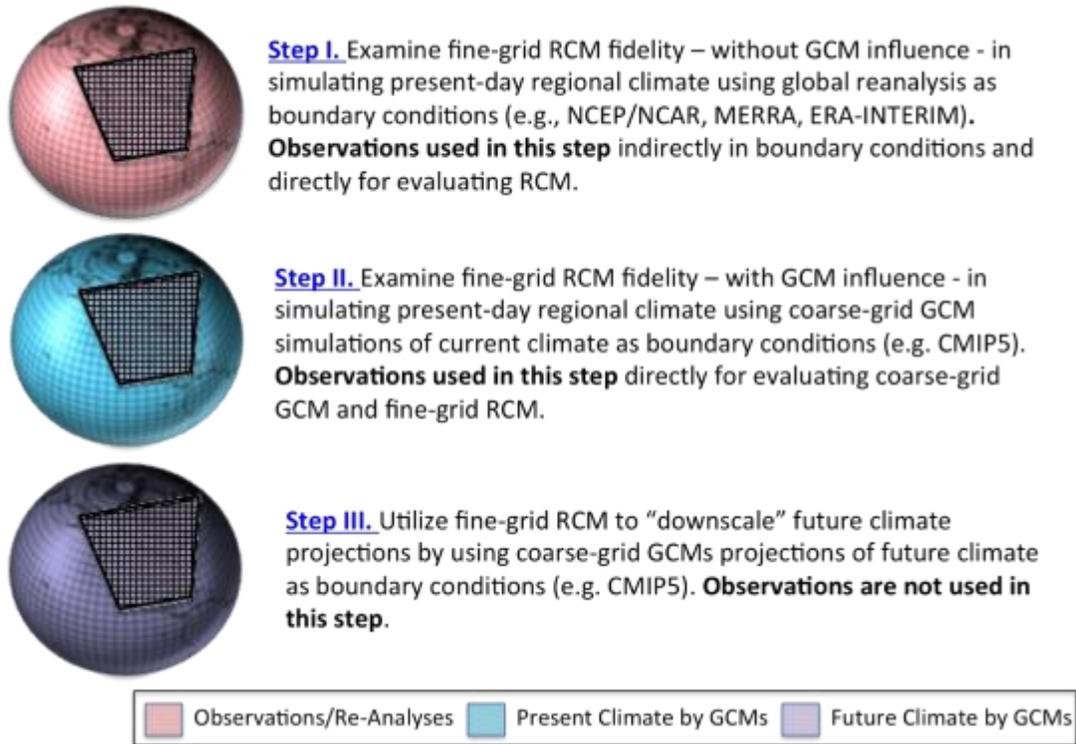


Figure 1. Schematic illustrating typical configurations utilized by downscaling. In Step I, the RCM uses BCs from a reanalysis and evaluation is of standalone RCM. In Step II, BCs are from a GCM simulation of the present climate, and the evaluation is of the statistical character of the climate produced by the combination of the GCM and RCM. In Step III, BCs are from a GCM projection of future climate, with no direct evaluations against observations possible. Rather, Steps III and II are compared to determine how the external forcing has changed the statistical character of regional climate.

Establishing the credibility of climate downscaling is crucial given both the role it plays in climate impacts assessment studies and the growing potential for these studies to influence policy and agencies supporting climate change decisions. In general, consensus community activities, such as systematic modeling experiments, targeted diagnostic analysis, and development and application of performance metrics, as well as programmatic ownership and priority for funding support for regional modeling have significantly lagged results from GCMs alone. This includes the robust and systematic utilization of a wide range of observations that can identify strengths and weaknesses in cloud, radiation, precipitation, mesoscale structure, surface fluxes, aerosol properties and interactions, etc.

In this project, we focus on Step I evaluation of NU-WRF (a NASA RCM) and GEOS5 (a NASA GCM) in downscaled replay mode. Our objective is to determine what value is added by driving an RCM with boundary conditions and forcings from a lower resolution model. In order to have an objective measure of that value, we need ground truth for evaluating the results. We focus on two quantities with reasonably dense observational datasets – temperature and precipitation over the continental US – as these are arguably amongst the most sought after predictions for future climate. We force both the RCM and the GCM with MERRA or MERRA-2 reanalysis over a historical period and evaluate the results with both satellite-based and ground-based observations (which are used to represent the ground truth), with the belief that the reanalysis will provide the best boundary condition forcing of the RCM, given true synoptic scale conditions.

This report summarizes the model configurations used, the evaluation methods employed, the experiments performed, and the results of the evaluations. Additional reports were also generated that document the models and experiments performed, and the metrics employed, in greater detail:

- Downscaling Project Model Configuration and Experiment Details
- AR Metrics and Evaluation Results
- MCS Metrics and Evaluation Results
- NES Metrics and Evaluation Results
- Hourly Precipitation Metrics and Evaluation Results
- CONUS-Wide Metrics and Evaluation Results

This report was drawn from information captured in these documents.

3. Experiment Design Considerations

3.1. Metrics Objectives

Detailed comparisons of simulated fields to observations are expected to show differences, since reproducing the weather at fine scales is not expected from downscaling. However, the perceived value of downscaling is in predicting the changing climatology, and the end use of downscaling is expected to be in understanding the potential changes to the frequency, duration, and local impact of severe weather events. We therefore desire metrics that will indicate if the downscaled results from historically driven boundary conditions reproduce the climatology of these events at regional scales compared to the observations. To that end, we have focused on three regional phenomena with consistent annual impacts in the US: Atmospheric Rivers (AR) on the west coast, Mesoscale Convective Systems (MCS) in the central plains, and NES over the mid-Atlantic and New England states. We also looked at the CONUS-wide precipitation and temperature statistics to complete the evaluation. If downscaled simulations of GCM-resolution boundary conditions do produce statistically reasonable local instances of these events, compared to observations, then downscaling may indeed be of value in interpolating GCM predictions to local scales. If they do not, then the value of downscaling to ascertain regional impacts of a changing climate may be called into question.

It is important to note that we are not examining the prognostic capabilities of GCMs. If the lower resolution predictions are wrong to begin with, it is unlikely that downscaling will fix the problem. This study is only looking at what additional value downscaling might bring if correct synoptic scale forcing is applied to the RCM.

There are many diagnostics that could be applied to the downscaled results, which will certainly point to issues with the details of the simulated results, and would be highly valuable in improving the RCM performance. That is not our aim here. Rather, we want to look at metrics that arguably capture the climatology of the regional phenomena and can be applied to the downscaled results for objective comparison. These metrics must also be based on observations available at high enough resolution to capture the local variability. As this project provides the opportunity to test a variety of downscaled resolutions (down to 4 km), we need observations of similar resolution in order to do a fair comparison. As it turns out, high spatial and temporal resolution datasets for temperature and precipitation over a decade are not plentiful, and satellite based observations are not available at the highest spatial scale of interest. The datasets and metrics we ultimately chose for the downscaling evaluation are discussed in the following sections.

3.3. Available Data for Comparisons

The precipitation and temperature data used as the reference against which the model outputs were evaluated are summarized in Table 1 and Table 2, respectively.

Table 1. Reference Precipitation Data

| Dataset Name | Resolution | Info |
|-----------------|-------------------|---|
| NLDAS forcing A | 0.125 deg, hourly | Gauge + RS, Land only |
| TRMM | 0.25 deg, daily | RS, Land and Ocean |
| TRMM | 0.25 deg, 3 hr | 3B42 V7 precip |
| MERRA-Land | 2/3 deg x 0.5 deg | GEOS + gauge + RS, Land only |
| CPC | 0.25 deg, daily | Gauge |
| Stage IV | 4 km, hourly | Radar-based with gauge data corrections |
| PRISM | 4 km, daily | Gauge, Kriging with topography |

Table 2. Reference Temperature Data

| Dataset Name | Resolution | Info |
|----------------------|-------------------|---|
| NLDAS forcing A | 0.125 deg, hourly | Gauge + RS, Land only |
| MERRA-Land | 2/3 deg x 0.5 deg | GEOS + gauge + RS, Land only |
| PRISM | 4 km, daily | Gauge, Kriging with topography |
| Wang & Zeng (WZ) SAT | 0.5 deg, hourly | http://rda.ucar.edu/datasets/ds193.0/ |

Integrated water vapor transport (IVT) from MERRA (for the pilot experiments) and MERRA-2 (for the decadal experiments) provides the basis for AR detection and evaluating simulated AR frequency and strength.

3.4. Metrics Employed in the Evaluations

CONUS-Wide Precipitation Probability Distribution Function

If we denote each datum from the reference and model simulations as x and y , respectively, then the relationship between the two can be described mathematically as their joint distribution, $p(y, x)$. Model verification essentially is characterizing the joint distribution $p(y, x)$ (Murphy et al. 1989).

In practice, direct characterization of $p(y, x)$ is not straightforward, especially when it is complex. Often it is easier and more illustrative to characterize the various components of $p(y, x)$, especially its two marginal distributions $p(y)$ and $p(x)$ and their lower-order statistical moments.

These moments can be further translated into the following four metrics – bias, coefficient of variation (CV), mean square error (MSE), and linear correlation coefficient (CC):

$$\begin{aligned}
 \text{Bias} &= \bar{y} - \bar{x}, \\
 \text{CV} &= \sigma_y / \bar{y}, \\
 \text{MSE} &= \frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2 = \text{Bias}^2 + \sigma_x^2 + \sigma_y^2 - 2\sigma_{yx},
 \end{aligned}$$

and

$$CC = \frac{\sigma_{yx}}{\sigma_y \sigma_x},$$

where \bar{x} , \bar{y} , σ_x^2 , σ_y^2 , and σ_{xy} , are the mean, variance, and covariance of (between) x and y, respectively. These constitute our basic metrics set for evaluation over CONUS. In addition, we found it instructive to compare the shape of the two marginal distributions, p(y) and p(x), and we refer to such a comparison as “Probability Distribution Function (PDF) comparison.”

Atmospheric Rivers

The AR landfall chronologies for constructing AR-related climatology were developed by applying the AR detection algorithm documented in Guan and Waliser (2015, under review) on the basis of MERRA/ MERRA-2 IVT. To evaluate the overall representation of AR activities in the downscaling experiments, AR frequency and mean AR IVT are considered. To quantify the similarity between simulated and reference AR frequency and mean AR IVT, Taylor metrics and Tian scores (see below, Mesoscale Convective Systems section) are calculated.

Performance of individual simulations with respect to these diagnostics are quantified using various metrics including the mean bias, Taylor diagram, and skill scores based on the mean-square-error and correlations (Tian score hereafter). The PDF skill score, which measures the overlap between two PDFs (Perkins et al. 2008), is also applicable.

Mesoscale Convective Systems

Many previous studies have proposed metrics to evaluate the performance of climate models. Some of them have relied primarily on correlation coefficient R and others have mainly relied on mean square error (MSE). The larger (or closer to 1) R is, the more similar two fields are in terms of structure and phase. However, it is impossible to determine whether these two fields have the same amplitude or not from a large R alone. In addition, it is impossible to determine how much of the error is due to the difference in structure/phase or amplitude from a small R (close to 0) alone. MSE approaches zero as two fields become more alike in magnitude. However, it is impossible to determine whether two fields have the same structure or not from a small MSE alone. Similarly, it is impossible to determine how much of the error is due to the difference in structure/phase or amplitude from a large MSE alone.

From the above discussion, we can see that both R and MSE provide important and complementary statistical information quantifying the correspondence between two fields but neither is complete individually. Accordingly, we propose a new metric T that combines both R and MSE in a way that is similar to the Taylor diagram (Taylor 2001):

$$T = [(1+R)/2] \times [1 - \text{MSE}/[V(f)+V(r)+[G(f)-G(r)]^2]]$$

Here, G(f) and G(r) are the means of two fields, f and r. V(f) and V(r) are the variances of f and r. R and MSE denote the correlation coefficient and mean square error between f and r in both space and time as defined above. The metric provides a skill score that has a maximum possible value of 1 (for R=1 and MSE=0) that indicates a perfect skill and a possible minimum value of zero that indicates no skill at all (negative values are rare in our context). In this report, this new metric T is referred to as the *Tian score*.

This metric T uses both R and MSE, instead of only one of these two fundamental statistics to quantify the correspondence between two spatial-temporal fields and is expected to be a better metric than either used alone. Furthermore, this metric T is a generic one and can be applied to any model simulated field f of any dimension (either spatial only or temporal only or spatial-temporal combined) and any resolution depending on the available reference data r. Because of the use of the ratio of the standard deviations of f and r ($\sigma(f)/\sigma(r)$), the skill score S proposed by Taylor [2001] also implicitly uses MSE information too.

North-East Winter Storms

Metrics for evaluating the ability of the downscaled simulations to reproduce climatologies of observed impact features of NES are based on individual precipitation-based storm events. Each event is defined as one or more consecutive days with daily total accumulated precipitation of 2.5 mm or greater and is referred to as a *storm event*. Storm events are identified for cold season months when NES are most active, spanning November to March. Metrics are then computed on the storm events, which are identified at each grid point.

Four primary storm characteristic metrics are developed and employed. *Storm intensity* is the maximum daily storm total within a storm event. For example, if the storm was three days long with day one receiving 3 mm, day two 20 mm, and day three 10 mm of precipitation, the storm intensity for that storm event would be 20 mm. *Storm duration* is the length, in days, of a storm event. For the previous example the storm duration would be three days. *Storm frequency* is the number of storm events that occur per season. Storm frequency is computed for all storms and for heavy storms separately with heavy storms having a minimum storm intensity of 25 mm. *Storm total* is the total accumulated precipitation across a storm event. For the example above, the storm total would be 33 mm. Metrics are computed for all storm events and then separately for storm events where only days that have average daily temperatures below freezing are included. This is to evaluate the ability of the models to realistically partition frozen and non-frozen precipitation. Additionally, storm frequency is also computed separately for heavy storms where the storm intensity is at least 25 mm. For comparison with observations, metrics are summarized into four climatological variables: mean storm intensity, mean storm duration, storm frequency per season, and median storm total. Median is used for storm total to reduce the impact of rare, localized extremely high storm totals.

Metrics are computed over the US National Climate Assessment Northeast Region, which includes West Virginia northeastward to Maine and is entirely state-defined. This domain was chosen because it provides an objective boundary for analysis and is appropriate for the potential integration of results into the next National Climate Assessment. Reference data is primarily from the PRISM dataset. PRISM provides daily temperature and precipitation data at 4 km horizontal resolution, making it comparable to the highest resolution NU-WRF simulation. This eliminates the possibility of having to re-grid any datasets from higher resolution to lower resolution for intercomparison. At first, NLDAS precipitation was used, which is also a high-quality dataset for CONUS; however at 12 km it is considerably coarser than PRISM. Furthermore, NLDAS temperature is reanalysis derived and therefore subject to substantial biases while PRISM is based on in-situ surface measurements, albeit with a model based interpolation scheme. Therefore PRISM was chosen instead of NLDAS because it provides a consistent and high quality reference for model evaluation at very high resolution.

Hourly Precipitation over the Central US

Despite continuous efforts to improve climate models, accurate simulation of extreme precipitation remains a challenge. Recently, Kendon et al. (2014) showed that an RCM with a very high spatial resolution of 1.5 km better simulates rainfall characteristics than coarse resolution simulations using the same model. In their study, the rainfall characteristics, including extreme precipitation on the sub-daily time scale in observational and model data, were defined as a joint probability distribution function (JPDF) between duration and peak intensity of rainfall events.

We adopted the JPDF to evaluate summertime rainfall characteristics over the Great Plains in the NU-WRF simulations. To evaluate the pilot experiment simulations for the summer of 2005, NCEP's Stage IV product based on radar and gauge observations was used as a reference dataset. The analysis domain for evaluating NU-WRF simulations is shown in Figure 2(a). When evaluating Phase II simulation results, we considered spatial homogeneity of precipitation characteristics. So, the comparison between NU-WRF and the observations was made for the three sub-regions shown in Figure 2(b), Northern, Central and Southern Plains from Bukovsky [2011]. In addition to the Stage IV dataset, we used hourly precipitation data from NASA's Global Precipitation Measurement (GPM) satellite mission to evaluate

climatological JPDFs from the Phase II simulations over a decade. A preliminary sensitivity study shows that even the two years of GPM data (2014–2015) can be used to build climatological JPDFs that show reasonable agreement with those from Stage IV.



Figure 2. (a) The Great Plains domain (35-45N, 110-90W) used in evaluating pilot experiment simulations. (b) Northern (blue), Central (Green), and Southern (Orange) Plains for regional comparison of JPDF between observations and simulations. The three regions are defined in Bukovsky [2011].

Following the definition of Kendon et. al. (2014), a wet spell is a continuous period of rainfall larger than 0.1 mm/hr. At each grid point in a region, hourly time series of precipitation between June 1st and August 31st (JJA) in each year is analyzed to find wet spells. The duration and peak rainfall intensity of each wet spell is used to bin the event within a JPDF representing the region’s summertime precipitation characteristics for a year. (This JPDF is essentially a two-dimensional histogram of wet spells binned by peak rainfall rate – duration bins). To build a climatological JPDF, nine JJA JPDFs between 2002 and 2010 were made with B24, B12, and Stage IV. B4’s climatological JPDF is from five JJA JPDFs between 2000 and 2004. With GPM, only the observations in 2014 and 2015 were used to make a JPDF. Figure 3 displays GPM’s JPDF over the Northern Plains. The JPDF consists of bins scaling probability [%] with green colors for certain ranges of rainfall duration (x-axis) and peak intensity (y-axis). The sum of probability over all the bins is 100 %.

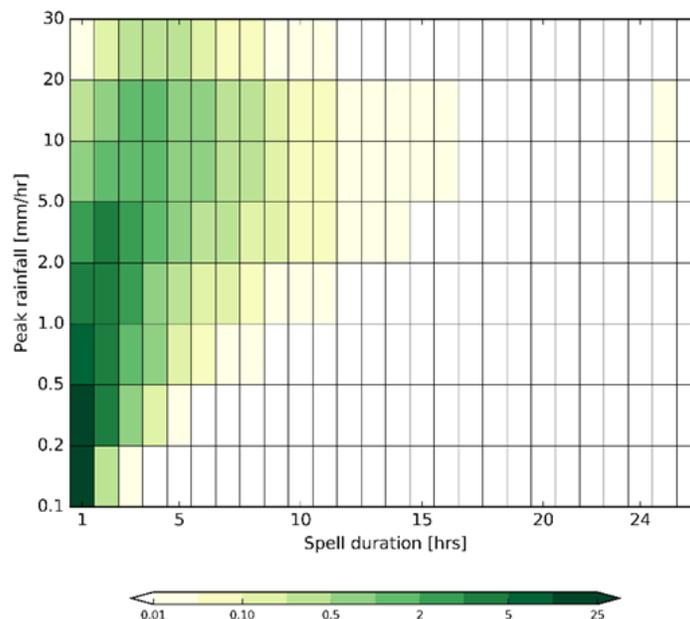


Figure 3. JPDF of wet spell duration and peak rainfall in GPM data over the Northern Plains for summer (June-August) in 2014 and 2015. The probability of each spell duration – peak rainfall bin is expressed as a percentage between 0 and 100%.

3.5. Model Considerations

As noted above, this study included both an RCM as well as a GCM in downscaled “replay” mode. The team focused on a NASA-centric modeling system and configuration that it was already familiar with. This approach enabled us to explore the impact of dynamical downscaling at satellite-resolved scales, with an ultimate goal of approaching convection-allowing scales (e.g., Kain et al. 2008, 2010). Below, we describe the two main atmospheric modeling systems used in this project: NU-WRF and GEOS-5.

RCM: NU-WRF

The RCM chosen for this project is the NASA Unified-Weather Research and Forecasting (NU-WRF) modeling system (Peters-Lidard et al. 2015). NU-WRF is an observation-driven integrated modeling system that represents aerosol, cloud, precipitation and land processes at satellite-resolved scales $O(1-25\text{ km})$, thereby bridging the continuum between local (microscale), regional (mesoscale), and global (synoptic) processes. NU-WRF is a superset of the National Center for Atmospheric Research (NCAR) Advanced Research WRF (ARW; Skamarock et al. 2008) dynamical core model, achieved by fully integrating the GSFC Land Information System (LIS, Kumar et al. 2006; Peters-Lidard et al. 2007), the WRF/Chem enabled version of the Goddard Chemistry Aerosols Radiation Transport (GOCART; Chin et al. 2000a,b) model, the Goddard-Satellite Data Simulation Unit (G-SDSU; Matsui et al. 2014), and custom initial / boundary condition preprocessors into a single software release. While this project does not invoke many of the coupling capabilities of NU-WRF, we do utilize the NASA-developed physical parameterization schemes developed at GSFC that are available within NU-WRF. These include the GSFC single-moment, 3-ice microphysical scheme, and shortwave and longwave radiation schemes, each of which are described in more detail below

Goddard Microphysical Scheme

The Goddard Cumulus Ensemble (GCE; Tao and Simpson, 1993, Tao et al. 2003; Lang et al. 2007) single-moment bulk microphysical scheme has been implemented into both the community ARW model and the NU-WRF. These schemes are mainly based on Lin et al. (1983) with additional processes from Rutledge and Hobbs (1984). However, the Goddard microphysics schemes have several modifications. First, there is an option to choose either graupel or hail as the third class of ice hydrometeors (McCumber et al. 1991). These differences can affect not only the description of the hydrometeor population and formation of the anvil-stratiform region but also the relative importance of the microphysical-dynamical-radiative processes. Second, there are various saturation techniques (Tao et al. 1989, 2003; Lang et al. 2003) available, designed to ensure that super saturation (sub-saturation) cannot exist at a grid point that is clear (cloudy). Third, all microphysical processes that do not involve melting, evaporation, or sublimation (i.e., transfer rates from one type of hydrometeor to another) are calculated based on one thermodynamic state. This ensures that all of these processes are treated equally. Fourth, the sum of all sink processes associated with one species will not exceed its mass, thus ensuring that the water budget is balanced in the microphysical calculations.

Goddard Shortwave and Longwave Radiation Schemes

Goddard long- and short-wave radiative transfer processes (Chou and Suarez 1999 and 2001) have been implemented within NU-WRF and are used within the GCE model.

The use of a fully explicit microphysics scheme (with size distributions of liquid and ice) and a fine horizontal resolution can provide relatively realistic cloud optical properties, which are crucial for determining the radiation budgets and diurnal variation of precipitation processes. With high spatial resolution, each atmospheric layer is considered either completely cloudy (overcast) or clear. No partial cloudiness or overlap assumptions are needed. In addition, the definition of cloud is based on a small cloud optical depth ($1E-4$) to account for thin clouds.

It is known that Goddard SW radiation underestimates molecular absorption compared to schemes calibrated against the more recent HITRAN database (such as RRTMG; Oreopoulos et al. 2012). As a result, Goddard SW radiation tends to overestimate clear-sky SW downwelling surface radiation by up to ~ 30 W/m² relative to the RRTMG. As a remedy, the 2014 scheme arbitrarily increased molecular absorption by about 50%, which reduced biases overall.

Land Information System (LIS)

The NASA LIS is a high-performance land surface modeling and data assimilation system that integrates satellite-derived datasets, ground-based observations, and model reanalyses to force a variety of LSMs (Kumar et al. 2006; Peters-Lidard et al. 2007). LIS can run LSMs offline globally with a grid spacing as fine as 1 km (or better) to characterize land surface states and fluxes. The system also supports an optimization and uncertainty analysis for calibrating land surface model parameters to observations (Santanello et al. 2013). LIS has also been coupled to the ARW dynamical core (Kumar et al. 2007) for NWP applications using the NU-WRF modeling framework (Peters-Lidard et al. 2015). In this study, LIS was used to run multi-year offline spinups of the LSM prior to coupled NU-WRF initialization, thus improving upon coarsely resolved initial soil conditions obtained from the reanalysis data alone.

GCM: GEOS-5

The Global Modeling and Assimilation Office (GMAO) at NASA/GSFC hosts a seasonal forecast system consisting of an atmospheric general circulation model coupled to the Catchment land surface model and a full ocean GCM imported from the NOAA Geophysical Fluid Dynamics Laboratory. The Goddard Earth Observing System Model, Version 5 (GEOS-5) is a system of models integrated using the Earth System Modeling Framework (ESMF). The GEOS-5 Data Assimilation System integrates the GEOS-5 AGCM with the Gridpoint Statistical Interpolation (GSI) atmospheric analysis developed jointly with NOAA/NCEP/EMC. The GEOS-5 systems are being developed in the GMAO to support NASA's earth science research in data analysis, observing system modeling and design, climate and weather prediction, and basic research. See <http://gmao.gsfc.nasa.gov/research/climate/> for more information on the forecast system.

The GEOS-5 system has a replay capability (akin to nudging) that adds a forcing term to the model equations to constrain it to follow a specified trajectory. This capability is adapted here to produce a downscaled version of an existing, relatively low-resolution reanalysis. The replay mechanism is similar to the standard GEOS atmospheric data assimilation system (ADAS) procedure in the sense that it uses an incremental analysis update (IAU, see Bloom et al. 1996 and Rienecker et al. 2011) to constrain the model to the pre-existing analysis. In both the GEOS ADAS and the GEOS replay, analysis increments are applied to the model over a 6-hour period during the IAU corrector step. The simulation produced using replay is thus a blend of the forcing analysis combined with the atmospheric model in use. This mechanism has been verified to reproduce the original assimilation quite well. A singular advantage of this replay mode is that a different model and/or resolution may be used than that used to generate the original sequence of analyses.

MERRA and MERRA-2 Reanalysis Products

The Modern Era Retrospective-Analysis for Research (MERRA) is described in Rienecker et al. 2011. The objective is to provide reanalysis data for the science community but also to make some improvement of the water cycle beyond existing reanalyses. While the system uses a three dimensional variational assimilation method, it also includes the use of incremental analysis updates, which reduce the shock of the analysis initialization on the forecast model, thus reducing spin-down spikes in fields like precipitation. The MERRA data has been reviewed and proved to be competitive with contemporary reanalyses from ECMWF, NCEP and JMA, but weaknesses were identified for future development efforts. Specifically, the changing observing system induced a spurious effect on low frequency variability. Also, precipitation biases from the

model forecast were identified as a significant cause of bias in the near-surface states and fluxes.

MERRA-Land

An offline version of the MERRA-land surface model (i.e., Catchment) was developed using the existing MERRA surface forcing data (radiation, temperature, water vapor, wind, and surface pressure) to reprocess the land quantities. However, the modeled precipitation data were replaced with an observation-corrected forcing designed to ameliorate, if not remove, the MERRA forecast model precipitation bias. This new ancillary data set was provided as MERRA-Land (Reichle et al. 2011), and has been useful in surface hydrology studies and testing land model development.

MERRA-2

From the lessons learned from MERRA and forecasting studies, along with the inability of the MERRA system to assimilate recent satellite instruments, an update to the reanalysis data products was developed. MERRA-2 includes interactive and assimilated aerosols, numerous model upgrades (including the cube sphere grid) and the capability to assimilate many more satellite instruments than MERRA. The MERRA-2 system also addresses some of the noted deficiencies in the MERRA data. Observation corrected precipitation similar to that used for MERRA-Land, was included during the production of MERRA-2. This is expected to provide the benefits of MERRA-Land, but built into the MERRA-2 data. Also, a mass constraint was applied to the assimilation of dry mass and water vapor, which ensures global conservation of mass (Takacs et al 2015). This mass constraint reduces the spurious effects of the changing observing system on the temporal variability of the reanalysis system. While the strengths and weaknesses in the MERRA-2 data are only beginning to be understood, MERRA-2 marks significant progress of the GEOS-5 data assimilation system to produce reanalysis data (Bosilovich et al. 2015).

3.6. High Performance Computing Resources

In support of the downscaling project, two NASA High Performance Computing (HPC) systems were used: Discover at the NASA Center for Climate Simulation (NCCS) at NASA/GSFC and Pleiades at the NASA Advanced Supercomputing (NAS) facility at NASA Ames Research Center. Existing systems in both the NCCS and NAS were used during the first phase of the project while the NCCS Discover cluster was upgraded with additional computing resources. The upgrade was initially installed in late fall of 2014 and provided to the downscaling project by early 2015. Details on the utilization of these computational resources for the regional and global downscaling experiments are presented in Section 4.6.

4. Experiments Description and Model Configuration

The dynamical downscaling project was conducted in two phases. Phase I involved a 10-month pilot study spanning 1 November 2005 to 1 September 2006 in order to implement and test appropriate model configurations, develop meaningful metrics for each phenomena being studied, and to examine preliminary results. The results and lessons learned from the Phase I pilot study were then applied to a decade-long Phase II set of simulations spanning 1 November 1999 to 1 November 2010. A final set of metrics were produced from the 11 years of simulation/replay for each of the three specific phenomena, as well as overall statistics of precipitation and temperature distribution. The following sub-sections describe in greater detail the various model configuration characteristics for physics, dynamics, and initial and boundary condition data.

4.1. NU-WRF Model Configuration

Grids were designed to exactly cover identical areas over the Western Hemisphere, adequately cover the regions where the three evaluation phenomena occur (AR, MCSs, NES), keep the map scale factors close to unity to avoid reducing the computational time step, allow testing of grid resolution sensitivity in regional climate downscaling, and enable nesting of outer “A” domains to inner “B” grids (Note that direct nesting in WRF was only employed in Phase I).

The domains configured for the NU-WRF simulations are illustrated in Figure 4. In the Phase I pilot experiment, the effects of domain size (A vs. B) and nesting (B nested within A simulation) on the three phenomena of interest were examined to determine the best method with which to proceed when conducting the decade-long simulations in Phase II.

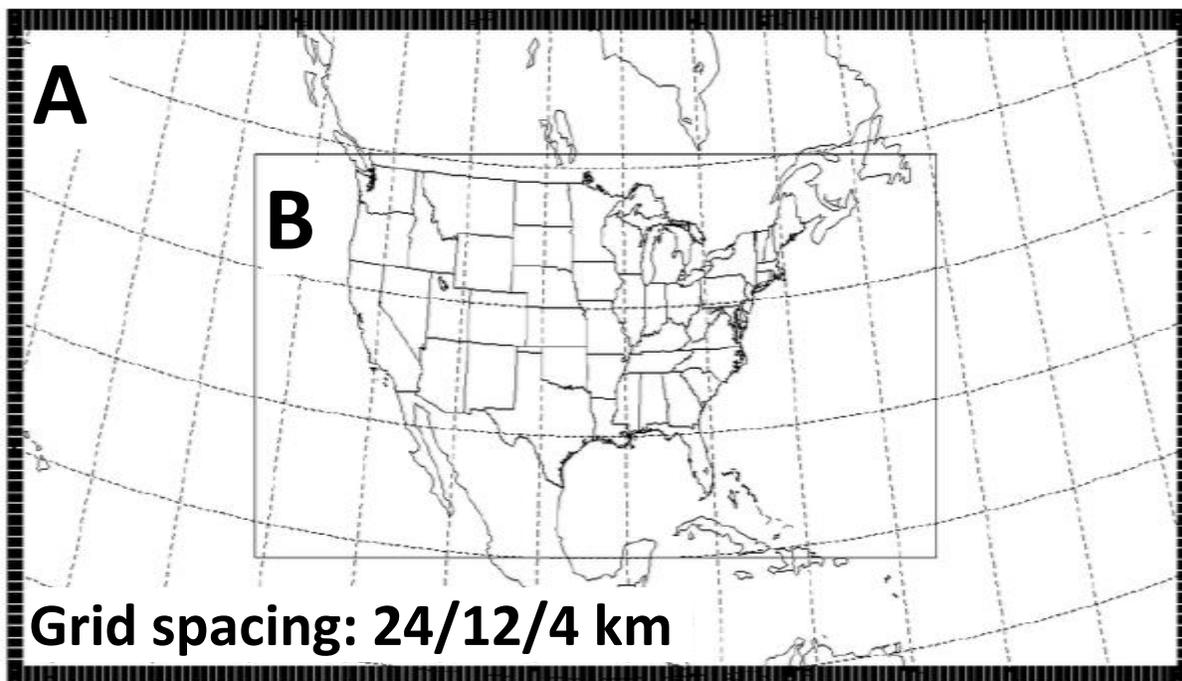


Figure 4. Geographical coverage of the larger “A” domain and smaller “B” domain used for the regional dynamical downscaling simulations with the NU-WRF modeling system. Both domains were used in the Phase I pilot experiments, whereas only domain B was utilized for the decade-long simulations in Phase II.

4.2. Non-NASA Physics Parameterization Schemes

In addition to the NASA/GSFC microphysical and shortwave/longwave radiation schemes described in Section 3.4, appropriate non-NASA physics parameterization schemes were chosen to represent sub-grid scale processes for planetary boundary layer processes and convection, and for land surface/surface energy processes. For the planetary boundary layer and sub-grid scale turbulence, the level 2.5 Mellor-Yamada-Janjic turbulence scheme (MYJ; Janjić 1990, 1996, 2002) was chosen, as this scheme has a long history of reliability in the National Weather Service operational models (e.g. Eta and North American Mesoscale models). The corresponding Monin-Obukhov Janjic Eta surface scheme (Janjić 1996, 2002) was required when running the MYJ turbulence scheme. For convection parameterization, the Grell-3D (G3D) deep cumulus scheme was selected based on the results of Qiao and Liang (2015), who showed that the G3D had relatively good precipitation verification scores over the contiguous United States during June-August 1993 and May-July 2008, compared to other available schemes within NU-WRF. However, this study also showed a pronounced high precipitation bias from the G3D over the Gulf Stream in 1993. Since tropical cyclone development (which could be affected by the bias over the Gulf Stream) was not one of the phenomena to be

evaluated, and since we expected East Coast storms and West Coast ARs to be primarily affected by grid scale (microphysics) effects, we selected the G3D scheme. In later Phase II sensitivity tests, we also invoked the Betts-Miller-Janjic, new Kain-Fritsch, and New Simplified Arakawa-Schubert schemes in place of G3D to examine the impacts of alternative convective parameterization schemes.

The GEOS-5 global replay run employed the relaxed Arakawa-Schubert scheme (RAS; Moorthi and Suarez 1992). The RAS scheme adopts a simple implementation style of the original Arakawa-Schubert scheme with the following two major modification: modifying the entrainment relation and relaxing the state toward equilibrium. The first modification introduced the assumption that the normalized mass flux for each cloud type is a linear function of height instead of the original exponential function. The second modification is relaxing the assumption in the original scheme that the interaction among various cloud types occurs instantaneously. This assumption might result in an ill-posed problem: No mass-flux distribution could produce an exact balance for all clouds with positive buoyancy, or multiple distributions could satisfy an over-adjustment problem (Silva-Dias and Schubert 1977). In RAS, the problem was removed by selecting an equilibrium distribution that is dependent on the time scales specified for the adjustment of individual cloud types.

4.3. Atmospheric Initial and Boundary Conditions

Initial and boundary conditions were provided in six-hourly intervals from MERRA reanalyses for the 10-month pilot phase, followed by MERRA-2 reanalyses for the Phase II simulations, since MERRA-2 reanalyses were still being processed in simultaneous streams during the beginning of the NASA downscaling project. Within the NU-WRF modeling system is a utility called MERRA2WRF that enables the MERRA reanalyses to be interpolated and converted into the WRF Preprocessing System intermediate binary format. All the primary variables required for WRF initial and boundary conditions (wind vector components, geopotential height, temperature, specific humidity, and surface pressure) were processed through the MERRA2WRF utility.

4.4. Land Surface Initialization

Initial soil moisture and temperature states generated from large-scale atmospheric initialization (e.g., via reanalyses) of WRF forecasts are typically limited by coarse spatial resolution and lack of heterogeneous or observation-based land surface conditions. Substantial improvement can therefore be obtained by simply performing an offline land surface model (LSM) spinup (Santanello et al. 2013). The importance of an accurate, high-resolution LSM spinup for coupled prediction has been highlighted in previous studies, ranging from impacts on land surface states and fluxes (Lawston et al. 2014) to ambient weather and precipitation on short-term (Chen et al. 2007; Kumar et al. 2008; Case et al. 2008, 2011; Wen et al. 2012) to seasonal (Hirsch et al. 2014) scales.

The specific release of NU-WRF used in this project includes LIS version 7. The LSM employed is version 3.3 of the Noah LSM (Chen and Dudhia 2001; Ek et al. 2003), and is identical to the version of Noah packaged in the community version of the WRF-ARW. Noah is used in operational and research modes by a number of institutions, and as such is a well-supported, developed, and utilized LSM for both offline and coupled weather prediction applications.

Noah was spun-up offline for 10-years (1990-1999) prior to the kickoff of the Phase I Pilot and Phase II Decade experiments. The atmospheric forcing for these spinups was acquired from the MERRA/MERRA-Land and MERRA-2 for the Pilot and Decade runs, respectively. This ensured consistency between the forcing applied to the LSM spinup and that used for the initial and boundary conditions of the NU-WRF simulations. Forcing variables required for spinup included 2-meter temperature and specific humidity, precipitation, wind speed and direction, shortwave and longwave radiation, and surface pressure. Noah spinups were performed on

each of the case study grids of Figure 4 (A24, A12, A4, B24, B12, B4) following the procedure described above, and then used to initialize each respective coupled NU-WRF experiment (thus overwriting the coarser/default initial soil conditions from MERRA/MERRA-2). Once the coupled runs began, the LSM reverted to the Noah LSM embedded within the WRF-ARW (identical to that in LIS).

4.5. M2R12K/GEOS Replay Runs

The GEOS-5 replay capability was used to produce a 15-year downscaled global reanalysis product at ~12.5-km global resolution using a non-hydrostatic version of the GEOS-5 atmospheric model. This high-resolution version of GEOS-5 was nudged to the recent MERRA-2 reanalysis produced by the GEOS-5 ADAS. This downscaled global simulation is referred to as the MERRA-2 Replay at 12.5 km (M2R12K). While the MERRA-2 product is produced at ~50-km global resolution, a spectral analysis of the kinetic energy in MERRA-2 and M2R12K show that the MERRA-2 forcing stops adding value to the simulation below a spectral truncation of effectively T60 (~666 km). Thus all increments from MERRA-2 added to the M2R12K simulation are filtered below T60, permitting the replay to closely follow the trajectory of the large-scales (effectively T60) in the underlying MERRA-2 reanalysis, while allowing the M2R12K downscaling model to embed its own, internally developed, mesoscale organization. This allows us to directly compare the simulated mesoscale with observations for specific synoptic situations.

For this project, we used the GEOS-5 M2R12K replay run output covering the period from December 1999 through October 2010 over CONUS. In addition to replaying the analyzed meteorology, M2R12K includes reactive gases (CO₂, CO, and SO₂) as well as aerosol assimilation for sulfates and carbon using the Goddard Aerosol Assimilation System (GAAS, Buchard et al. 2015), including dynamically emitted aerosols for dust and sea-salt. Thus, the M2R12K product provides a comprehensive downscaled reanalysis product for weather and climate.

The M2R12K output includes full-resolution output at hourly frequency for 2-dimensional fields and 3-hourly frequency for 3-dimensional fields. This includes native resolution on the c720 (720x4320 cubed-sphere grid) and an interpolated 0.125-degree (2880 longitudes by 1441 latitudes) atmospheric grid. In addition, the entire MERRA-2 0.5-degree (576 longitudes by 361 latitudes) file-specification is reproduced from the M2R12K downscaling to facilitate inter-comparison with the native MERRA-2 reanalysis.

4.6. Utilization of NASA Computational Resources

Table 3 shows the resources used for the NU-WRF runs during Phase I of the project. In this phase, computer resources at both the NCCS and NAS were used along with three different Intel processors (shown in the processor column of the table). The listed wall times represent end-to-end timings, and do not include additional limited reruns made after applying a sea-ice-related patch for the restart files (see Section 5.4). The range in timings for a given domain/system/processor combination is due to the use or non-use of spectral nudging – nudging simulations consistently took longer to run. Due to extremely poor performance with the original spectral nudging code, the B4 simulation on Pleiades was aborted early and the configuration was instead used to test changes to the nudging code (see Section 5.4).

Table 4 shows the resources used for the NU-WRF runs during Phase II of the project. During this phase, only the upgraded resources within the NCCS Discover cluster were utilized.

In addition to the above resources used for NU-WRF, the GEOS M2R12K replay was completed using two different streams: stream 1 ran from 1999 through 2004 and stream 2 ran from 2005 through 2012. Table 5 shows the resources used by the GEOS M2R12K replay for these two streams.

Table 3. Resources used for the WRF runs during Phase I of the project.

| Domain | System | Processor | Cores | Wall Time (Days) |
|------------------|----------|--------------|-------|------------------|
| A24 | Discover | Westmere | 648 | 1.9 – 2.5 |
| | Pleiades | Westmere | 648 | 0.7* |
| A12 | Discover | Westmere | 2,592 | 5.2 – 7.0 |
| B24 | Discover | Westmere | 232 | 1.3 – 1.5 |
| B12 | Discover | Westmere | 892 | 3.4 – 4.5 |
| | Pleiades | Sandy Bridge | 660 | 3.5 |
| B4 | Pleiades | Ivy Bridge | 3,260 | See text |
| A24-B12 NDOWN | Discover | Westmere | 900 | 3.3 – 4.3 |

*Only ran 3-month simulation for comparison with Discover run

Table 4. Resources used for the WRF runs during Phase II of the project.

| Domain | System | Processor | Cores | Wall Time (Days) |
|--------|----------|-----------|-------|------------------|
| B24 | Discover | Haswell | 245 | 10.1 – 10.3 |
| B12 | Discover | Haswell | 865 | 24.1 – 30.3 |
| B4 | Discover | Haswell | 6,083 | 80 |

Table 5. Resources used for the two GEOS replay (i.e. M2R12K) streams: stream 1 covered 1999 through 2004 and stream 2 covered 2005 through 2012.

| Run | System | Processor | Cores |
|----------|----------|-----------|-------|
| Stream 1 | Discover | Haswell | 7,168 |
| Stream 2 | Discover | Haswell | 7,168 |

5. Phase I: Pilot Experiment Simulations and Issues Encountered

5.1. Domain Selection and Nesting

One of the original objectives of the pilot study was to investigate the role of domain size on long-term regional dynamical downscaling results. To support this investigation, 10-month simulations spanning 1 November 2005 to 1 September 2006 were conducted on both domains A and B. Using the HPC resources denoted in Table 3, single-domain runs were made at 24-km and 12-km grid spacing on both domains, along with an additional 4-km simulation on the B domain only (hereafter denoted as A24, B24, A12, B12, and B4, respectively). Due to computational resource constraints prior to the installation of the new Haswell hardware on Discover, an A4 run was not conducted for the pilot phase.

Besides the resolution comparison, another objective was to examine the influence of nesting on subsequent results of the inner-nested domain. Due to the complexity of optimizing computational resources for nested domains running concurrently in a single simulation, the

team opted to run the outer A24 domain to its 10-month completion and then use the utility “NDOWN” (part of the community WRF model distribution) to interpolate at fixed 3-hourly time intervals; the A24 solution served as boundary conditions for an inner-nested B12 simulation during the same 10-month period. As the metrics were developed for each phenomena, the metric evaluation team subsequently examined the results of each of these domains and nested experiments to determine the best method for proceeding into the decade-long Phase II experiments (e.g., A12-direct vs. B12-direct vs. A24-B12 NDOWN nest).

5.2. Spectral Nudging Considerations

Another consideration for dynamical downscaling is the use of nudging at specified wavelengths to help constrain the model during long-term integration. Since the downscaling community’s perspective on applying nudging is mixed (e.g., Lo et al. 2008), we opted to examine two configurations¹: (1) a “Control” run that did not use any nudging, and (2) a nudging configuration that employed spectral nudging at preferentially long wavelengths to constrain the larger-scale features while enabling mesoscale circulations to develop at finer scales within the model. The MERRA (Phase I) and MERRA-2 reanalyses (Phase II) provided six-hourly analyses of the u- and v-wind components, geopotential height, and temperature to which the NU-WRF simulations were spectrally nudged.

Following the results and methodology in Miguez-Macho et al. (2004, 2005), Gula and Peltier (2012), and Glisan et al. (2013), the spectral nudging in the NU-WRF simulations were configured to be constrained by wavelengths on the order of ~2000 km and greater in both the zonal and meridional directions for Phase I and both O(2000 km) and O(600 km) in Phase II, since the GEOS-5 replay (M2R12K) utilize a tighter T60 (~666 km) wavelength constraint in their spectral nudging-like procedure. Therefore, inter-comparisons could be made between the global M2R12K and NU-WRF spectrally nudged runs using constraints at more similar wavelengths, as well as the effect of nudging to different wavelengths within NU-WRF [O(2000 km) vs. O(600 km)]. We used nudging coefficients of 1E-4 for the u- and v-wind components, temperature, and geopotential height. Specific humidity was not used in the spectral nudging; rather, NU-WRF was allowed to generate its own moisture features at the native model resolutions based on the constraints provided by the other MERRA/MERRA-2 variables and the lateral boundary conditions.

5.3. Datasets Produced

In both phases of this project, five different output streams were produced for analysis by the metrics team: (1) standard 3D WRF output files at 3-hourly intervals, (2) 2D WRF output fields at hourly intervals, (3) WRF diagnostics/statistical output of 2D fields at hourly intervals, (4) isobaric output at 850 hPa at hourly intervals, and (5) 3D WRF restart files every 7 days. These five output streams required the use of at least five dedicated output groups, so as to prevent a back-log in I/O processing time in the event that all five output streams were being written simultaneously.

Table 6 provides an overall summary of the Phase I Pilot study simulations that were made for the experiments described in the sub-sections above. Note that the B4 spectral nudging simulation in Phase I could not be completed due to the spectral nudging bug that was corrected for the Phase II simulations, as discussed in the next sub-section.

¹ To further examine the impacts of nudging on dynamical downscaling results, additional experiments were conducted to use grid analysis nudging (to include specific humidity nudging).

Table 6. Experiment configurations of 10-month pilot simulations. In all grid domains, 41 vertical levels were used in a stretched configuration from the surface to 10 hPa.

| Simulation name | Horizontal resolution / grid points / CPU cores | Spectral nudging wavelength | HPC system / hardware platform |
|-----------------|---|------------------------------|--------------------------------|
| A24_0 | 24 km / 575×262 / 648 | OFF | Discover / Westmere |
| A24_2000 | 24 km / 575×262 / 648 | 2000 km | Discover / Westmere |
| A12_0 | 12 km / 1149×523 / 2592 | OFF | Discover / Westmere |
| A12_2000 | 12 km / 1149×523 / 2592 | 2000 km | Discover / Westmere |
| B24_0 | 24 km / 332×157 / 232 | OFF | Discover / Westmere |
| B24_2000 | 24 km / 332×157 / 232 | 2000 km | Discover / Westmere |
| B12_0 | 12 km / 663×313 / 892 | OFF | Discover / Westmere |
| B12_2000 | 12 km / 663×313 / 892 | 2000 km | Discover / Westmere |
| B12_0_NEST | 12 km / 663×313 / 900 | OFF | Discover / Westmere |
| B12_2000_NEST | 12 km / 663×313 / 900 | 2000 km (A24 Parent only) | Discover / Westmere |
| B4_0 | 4 km / 1987×937 / 3260 | OFF | Pleiades / Ivy |
| B4_2000* | 4 km / 1987×937 / 3260 | 2000 km | Pleiades / Ivy |

5.4. Code Issues and Fixes

During the process of conducting NU-WRF simulations, the team encountered three problems related to the community ARW model that were identified and corrected. The code fixes were passed along to the ARW code maintainers at the NCAR Developmental Testbed Center. These code issues are described in the three sub-sections below.

Divergence Problem Associated with Restart Files

A restart inconsistency was found during the pilot project period when two B12 runs were compared on Pleiades. Both simulations ran from 1 November 2005 to 1 January 2006; one simulation ran continuously over the two-month period while the other stopped on 20 December 2005 and then continued on from a restart file to 1 January 2006. However, small differences between the two runs were noticed shortly after the restart point (20 December 2005), which were then accumulated over the course of the simulation and became visually apparent by 1 January 2006. Because the settings of the two runs were otherwise exactly the same, the “warm-start” of the second simulation from a restart file became the only reason that the two simulations diverged. This issue was confirmed by subsequent experiments on both Pleiades and Discover, further indicating that the emergence of differences between two runs may depend on the particular restart point. Furthermore, the same issue was identified in simulations with the community WRF model, indicating that the problem was not induced by the NU-WRF modification of the code.

After some experimentation, it was determined that the divergence issue appeared to be associated with the appearance of sea ice in the above simulations, which started to show on 13 December 2005. A fix was made to address the identified problem in the WRF registry file and the source code. After applying the code fix, NU-WRF simulations running with or without

restart points always produced the same results (at least on the same computer platform). We reported the code fix back to the NCAR WRF team, and it is now incorporated into the latest release of the community WRF.

Revision of FFT Algorithms to Improve B4 Spectral Nudging Efficiency

During the course of the Phase I Pilot project, turning on the spectral nudging options on the B24 and the B12 domains resulted in a substantial compute performance penalty. For example, it took about 2 hours of wall time to complete one day of simulation on ~3,000 CPU cores without nudging. Meanwhile, with nudging, the corresponding wall time increased to 6 to 8 hours for a one-day simulation on the same CPU configuration. Increasing the number of CPUs didn't help solve the problem, either. Therefore, the very slow performance of the spectral nudging on the B4 domain became a significant obstacle for conducting the experiments.

The performance issue was traced to the FFT of the nudged 3-dimensional boundary condition fields at each time step. The efficiency of the FFT algorithm highly depends on the prime factorization of the grid dimensions. Unfortunately, the dimensions of B4 (1987 × 937) are both large prime numbers. One solution to the problem would have been to resize the grid dimensions so that they factored into a set of small prime numbers. This approach would have required that we re-configure the B4 (and correspondingly the B24 and the B12) domains and restart the experiments from the very beginning. Therefore, we decided to use an alternative approach previously developed in other engineering applications that pads zero values to the original data grid so that the new dimensions are factors of 2 and 5. Because padding zeroes in the space/time domain is equivalent to interpolating in the spectral domain, we do not expect this technique to induce any substantial effects on the simulation results.

We revised the WRF code to pad the nudged fields with zeroes. The original dimensions of these fields are recovered from the outputs of the inverse FFT algorithm by truncating the highest order wave numbers. Comparing the results from the revised algorithm (padding zeroes) to the original algorithm (no padding) indicated high agreement. At the same time, the model execution efficiency with the revised code was significantly improved for the B4 domain, with the run-time overhead reduced to 20-30%, comparable to the overhead seen in B24 and B12.

5.5. Pilot Experiments Metrics

CONUS

For the pilot study, we evaluated B12 and B4 runs, with the STIV data as the reference. All the performance metrics listed above and the PDFs were computed. Figure 5 shows the temporal correlation between each model output and the reference data, for winter and summer, respectively. It can be seen that the temporal correlations have strong seasonal dependence: higher in winter and lower in summer. Similar seasonal dependencies are also evident with other metrics.

In addition, the difference in performance between B12 and B4 (e.g. Figure 5a and Figure 5b) is not very dramatic. The DJF performance is largely dominated by large-scale forcing and frontal systems, so these small differences make sense. The small differences in JJA are more surprising; one would expect the B4 run to better represent the spatial patterns and diurnal cycle of convection.

Atmospheric Rivers

Overall, the results from the evaluation of AR frequency (Figure 6) and IVT (see full report) suggest that the smaller domain (domain B) outperforms the larger domain (domain A), and the one-step simulation configuration outperforms the two-step configuration. Spectral nudging converges, but not necessarily improves, the simulations. The effect of spatial model resolution is not obvious.

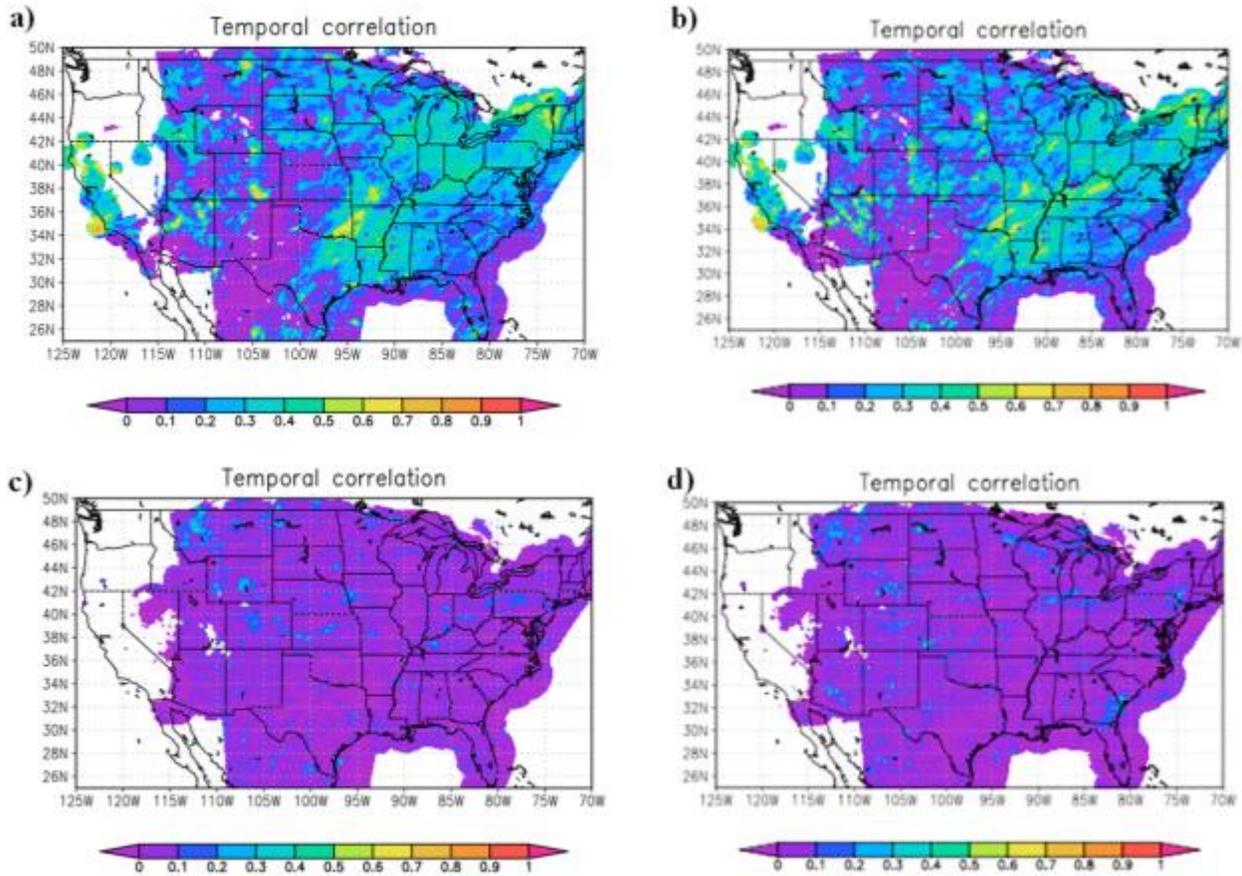


Figure 5. Temporal correlation for the pilot experiments between hourly model precipitation and STIV, for a) B12, winter (DJF); b) B4, winter; c) B12, summer (JJA), and d) B4, summer.

The effects of the size and resolution of a model domain on the model performance in simulating the AR-related precipitation and temperature anomalies in WUS for the 2005–2006 winter (November 2005–March 2006) are covered in the 1-year pilot simulation. A total of eight NU-WRF simulations were evaluated in the pilot study. The runs are stratified according to the domain size (A vs. B), spatial resolutions (24 km vs. 12 km), and either the presence or absence of spectral nudging. The AR chronology from the SSMI-based identification (Neiman et al. 2008) was used in the pilot experiment.

Comparison of the observed winter mean for the 10-year period (1998–2007) and the one winter of the pilot study (2005–2006) shows that AR landfalls leave qualitatively similar effects on the long-term and 1-year means, but the associated details can be quite different (not shown). An objective collective evaluation of the pilot run was performed using the Taylor diagrams and Tian scores. Figure 7 suggests that performance of the NU-WRF model in simulating the spatial structure of the precipitation fraction associated with all AR landfalls increases for smaller domain, finer resolution, and with spectral nudging: all NU-WRF runs generate similar standard deviations but the pattern correlation with the reference data, the NLDAS in this evaluation, varies systematically according to the size and resolution of the model domain and the presence of nudging. Anomalies of T2 and FZ associated with AR landfalls also support the same conclusions (not shown).

Conclusion from the AR pilot study:

- The NU-WRF model shows reasonable performance in simulating AR frequency, AR IVT, and the effects of AR landfalls on precipitation, surface air temperature, and freezing-level altitude over the WUS region in the 2005-06 winter.

- Fidelity of the NU-WRF simulations vary according to the domain size, resolution, and the use of spectral nudging.
- Smaller domain, finer spatial resolution and spectral nudging appear to improve the simulated AR precipitation fraction.

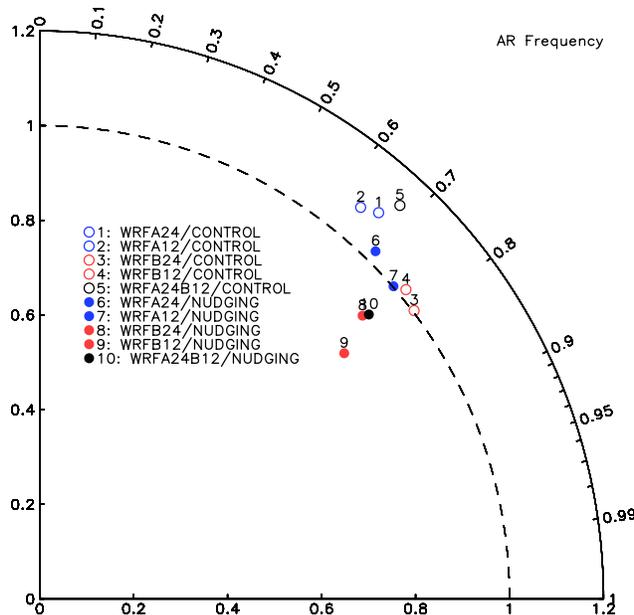


Figure 6. Taylor diagram based on comparing the spatial patterns of AR frequency between MERRA (the reference) and different downscaling experiments.

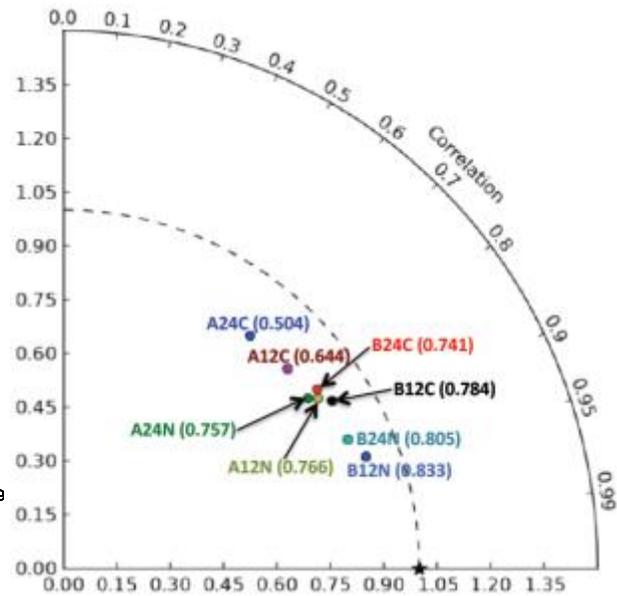


Figure 1. Taylor diagram evaluation of the spatial variations in the precipitation fraction corresponding to all AR landfalls from the NU-WRF pilot runs. Numbers in parenthesis are Tian scores.

Mesoscale Convective Systems

For the MCS metric, we apply the previously described Tian score to several physical phenomena related to MCSs over the US Great Plains. The area considered for these metrics is 110W – 90W longitude by 35N – 45N Latitude. Metrics are computed for the summer (JJA) mean precipitation pattern, summer mean precipitation diurnal cycle pattern, summer mean precipitation eastward propagation, summer mean Surface Air Temperature (SAT) pattern, and summer mean SAT diurnal cycle pattern. For example, Figure 8 shows the 2006 summer mean precipitation fields from TRMM and MERRA over the US Great Plains. Here we treat the summer mean precipitation field from TRMM as the reference field r and the summer mean precipitation field from MERRA as the model simulated field f . The Tian score of the summer mean precipitation field from MERRA (based on that from TRMM) is calculated using the two-dimensional (longitude and latitude) precipitation maps. The final Tian score is 0.83, indicating a high agreement between TRMM and MERRA, which is supported by the visual inspection of the maps. In another example, Figure 9 shows the 2006 summer mean precipitation Hovmöller diagrams from TRMM and MERRA over the US Great Plains.

Here we treat the summer mean precipitation field from TRMM as the reference field r and the summer mean precipitation field from MERRA as the model simulated field f . The Tian score of the summer mean precipitation Hovmöller diagram from MERRA (based on that from TRMM) is calculated using the 2-dimensional (longitude and hour) precipitation Hovmöller diagrams. The final Tian score of 0.08 indicates a very poor agreement between TRMM and MERRA, which is supported by the visual inspection of the Hovmöller diagrams. TRMM shows the eastward propagating precipitation systems in the diurnal time scale, while the MERRA

shows mainly the standing precipitation systems with little eastward propagation features. This indicates that MERRA represents the eastward propagation of MCSs very poorly. The Tian score for other physical phenomena can be also calculated similarly.

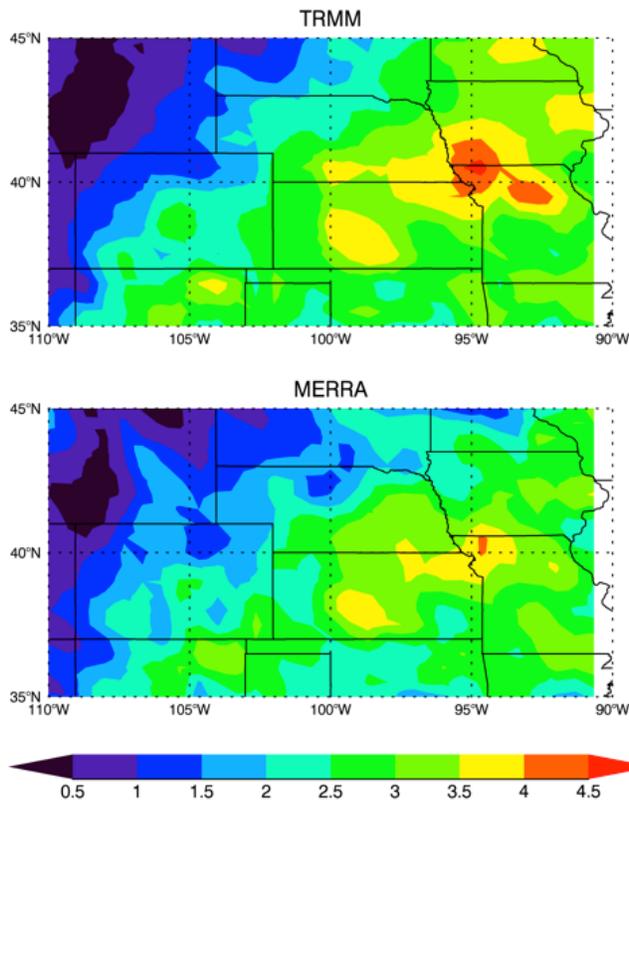


Figure 8. The 2006 summer (JJA) mean precipitation fields from TRMM and MERRA over the US Great Plains to show the application of Tian score.

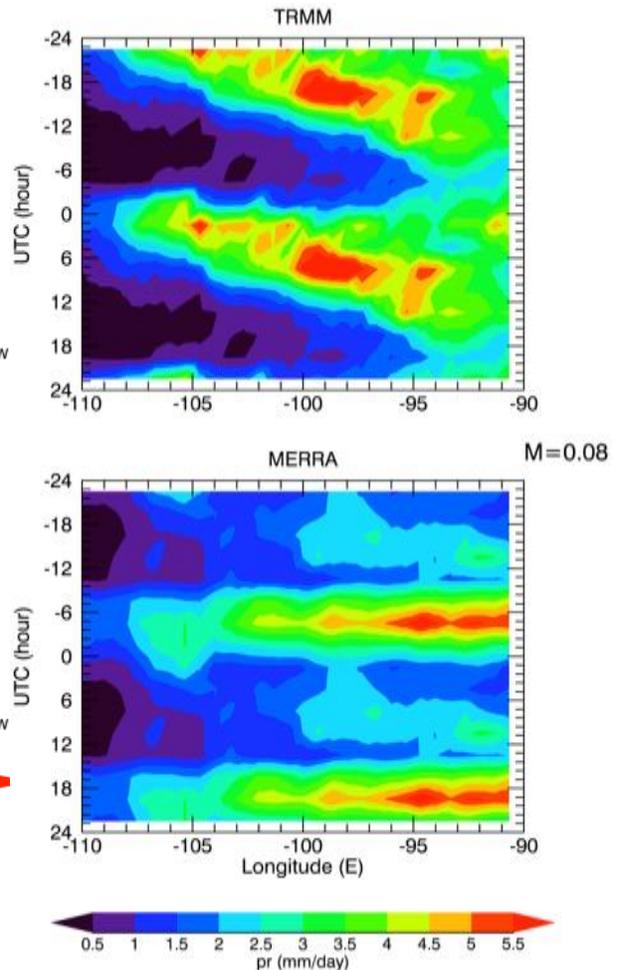


Figure 9. The 2006 summer (JJA) mean precipitation Hovmöller diagrams from TRMM and MERRA over the US Great Plains to show the application of Tian score.

Table 7 summarizes the MCS metric results for MERRA, M2R12K and various NU-WRF runs from the pilot experiments. The first column indicates the physical phenomena that each metric is based on, such as summer mean precipitation, summer mean precipitation diurnal cycle, summer mean precipitation eastward propagation, summer mean SAT, and summer mean SAT diurnal cycle. The second column indicates the reference datasets for the metrics, that is, TRMM and Stage IV (ST4 in tables) for precipitation and WZ and NLDAS2 for SAT. The metric of one reference dataset based on another, such as Stage IV based on TRMM, represents the uncertainties of reference datasets and the calculated metrics based on them. For example, the metric for summer mean precipitation of Stage IV based on TRMM is 0.70. This means that the uncertainties of precipitation reference datasets, TRMM and Stage IV, are around 30%, so a 30% variance in the metrics calculated for the model runs should be considered to be within the observational uncertainty. The third column indicates the metrics for various model runs.

The main conclusions drawn from the one-year pilot experiments are:

- The calculated metrics values depend on the reference datasets chosen.

- All NU-WRF model runs have poor performance in summer mean precipitation pattern and summer mean precipitation diurnal cycle, and it is unclear whether nudging is better or not.
- All NU-WRF model runs have very good performance in summer mean SAT pattern and summer mean SAT diurnal cycle pattern and nudging is better than no nudging. The model performance is clearly better in SAT simulations than precipitation simulations.
- The poor performance of MERRA on the diurnal cycle and eastward propagation indicates that the poor scores for these NU-WRF runs are due more to forcing/nudging than to a lack of downscaling validity.

Table 1. MCS metric results for the pilot experiments.
Scores are shown with only the significant digit (0.1) and are colored according to their values.

| | | Green (> 0.6) | | | | Yellow (0.4 - 0.6) | | | | Red (< 0.4) | | | | |
|---|------------------|-------------------|--------|---------|---------|--------------------|---------|---------|---------|-------------|---------|-------------|-------------|--------|
| Physical phenomena | Reference data r | Simulated field f | | | | | | | | | | | | |
| | | MERRA | M2R12K | A24 Con | A24 Nud | A12 Con | A12 Nud | B24 Con | B24 Nud | B12 Con | B12 Nud | A24 B12 Con | A24 B12 Nud | B4 Con |
| Summer mean precip | TRMM | 0.9 | 0.5 | 0.4 | 0.4 | 0.6 | 0.4 | 0.4 | 0.5 | 0.2 | 0.4 | 0.5 | 0.4 | 0.1 |
| | ST4, 0.70 | 0.7 | 0.4 | 0.3 | 0.3 | 0.4 | 0.3 | 0.3 | 0.3 | 0.2 | 0.3 | 0.4 | 0.3 | 0.1 |
| Summer mean precip diurnal cycle | TRMM | 0.1 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.2 | 0.4 | 0.3 | 0.3 | 0.2 |
| | ST4, 0.58 | 0.1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.1 | 0.3 | 0.2 | 0.3 | 0.1 |
| Summer mean precip eastward propagation | TRMM | 0 | 0.6 | 0.5 | 0.6 | 0.6 | 0.5 | 0.5 | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 | 0.3 |
| | ST4, 0.84 | 0 | 0.5 | 0.4 | 0.5 | 0.6 | 0.5 | 0.4 | 0.4 | 0.3 | 0.5 | 0.5 | 0.4 | 0.2 |
| Summer mean precip frequency distribution | TRMM | 0.2 | 0.1 | 0.5 | 0.5 | 0.4 | 0.5 | 0.4 | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 | 0.5 |
| | ST4, 0.76 | 0.4 | 0 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| Summer mean SAT | WZ | | | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 |
| Summer mean SAT diurnal cycle | WZ | | | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 | 0.9 | 0.8 |

North-East Winter Storms

Analysis of the pilot experiments largely predated the precipitation-based storm event metrics. It was overall less specific to NES and more specific to characteristics of the precipitation probability distribution function (PDF), and to liquid-to-frozen precipitation partitioning. This was because the storm-based metrics arose out of discussions within the downscaling team and over time through experimenting with ways to evaluate NES in a meaningful way. Additionally, the very short period of the pilot runs did not provide a robust climatology of storm-based metrics to evaluate. Figure 10 shows a summary of the biases in precipitation across the precipitation probability density function from the 50th percentile to the 99th percentile, computed as a ratio of simulated/PRISM precipitation values at the indicated percentile of the PDF. Overall, spatially aggregated biases in precipitation are larger for the median of the distribution compared with the more extreme values. Also, the NU-WRF simulations under-simulated heavy precipitation (not shown) and over-simulated the amount of heavy precipitation that fell concurrently with below freezing temperatures. This coincided with a cold bias across the entire NES domain (not shown).

Conclusions from the NES pilot study:

- The NU-WRF simulations under-simulated heavy precipitation (as defined as the 90th percentile) over most of the NES domain.
- Resolution didn't play a major role in model performance, however the B domain showed some improvement over the A domain.
- Nudging had a small but noticeable positive effect on the simulation of temperature and precipitation across the NES domain.
- Longer runs are needed to gain meaningful statistics on the climatology of NES precipitation-based impact metrics.

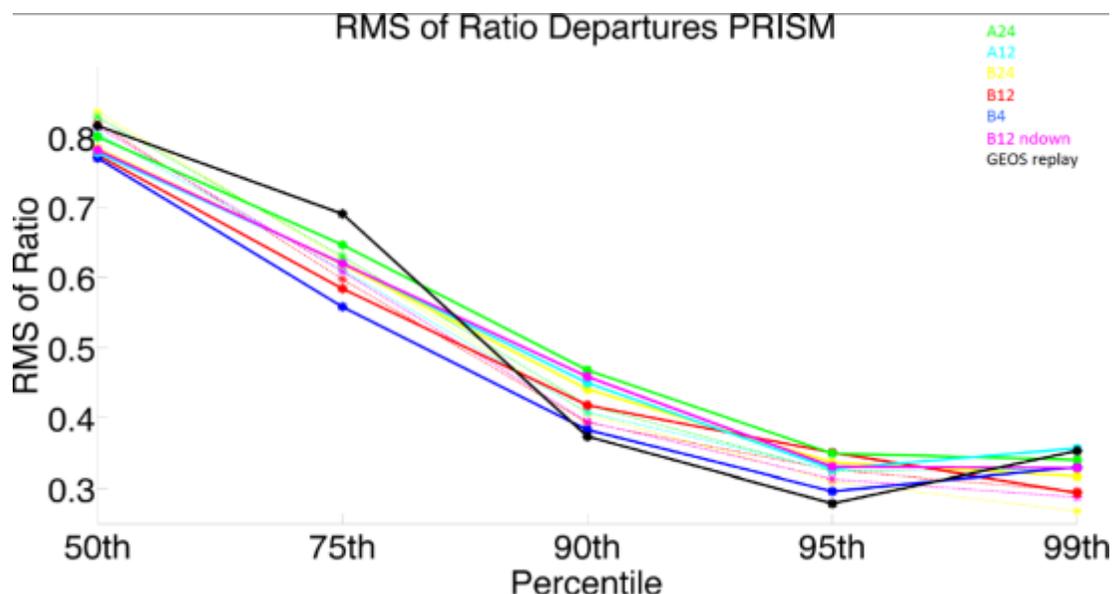


Figure 10: The root mean square of the spatial field of precipitation ratios calculated at each grid point for the pilot experiments. Precipitation ratios are the ratio of the model-simulated precipitation at the indicated percentile to the PRISM precipitation value for the same percentile. To assess the overall bias in the spatial distribution of the precipitation ratios, all ratios are computed as a deviation from 1 (a perfect agreement between model and PRISM); so that negative and positive biases are equally contributing to the RMS value. An RMS of 0 would indicate perfect agreement between model and observations. Control runs are plotted with solid lines and nudged runs are the same color but with dotted lines. Note that GEOS replay is used incorrectly as M2R12K was erroneously not being employed at this step in the analysis process. Nudging improves simulation to some extent in most cases, but differences between control and nudged are difficult to discern overall.

Hourly Precipitation over the US Great Plains

In this section, we analyzed impacts of 1) spatial resolution, 2) domain size, and 3) spectral nudging on the NU-WRF's performance in simulating the JPDF of wet spell duration and peak intensity. Before detecting wet spells and calculating JPDF over the Great Plains, the original hourly output from NU-WRF and Stage IV was temporally averaged into three-hourly mean precipitation. In summer 2006, there were 3,482,532 wet spells in Stage IV. So 1% in Figure 11(a) accounts for about 30,000 rainfall events. The JPDF from Stage IV data in Figure 11(a) indicates that most of summertime wet spells in 2006 showed peak rainfall of less than 0.5 mm/hr and lasted less than 9 hours.

Figure 11(b) and (c) exhibit the difference in JPDF between the control simulations in the A domain and Stage IV. The negative biases indicate that short-duration downpouring is less frequent in NU-WRF than Stage IV. In NU-WRF, rainfall duration is longer than Stage IV. With the higher spatial resolution, A12 shows more frequent short-duration precipitation than A24. This indicates the improved performance of A12 over A24.

Figure 12 shows the biases in JPDFs from two simulations, B12 and A12 control. In B12 control, the peak intensity of rain events is slightly smaller than A12 control. However, the difference between A12 and B12 is smaller than their biases from Stage IV JPDF.

Unlike the domain size, spectral nudging (Figure 13) significantly improves the performance of NU-WRF by making short-duration downpouring more frequent than in the control simulations. In both domains, the nudged run better simulates short-duration rainfall, whose peak intensity is between 0.2 and 5 mm/hr.

In summary, NU-WRF simulates less frequent events of intense short-duration rainfall than the Stage IV observations. With higher resolution (12 km) and spectral nudging, NU-WRF's performance regarding the duration-intensity characteristics of rainfall is improved.

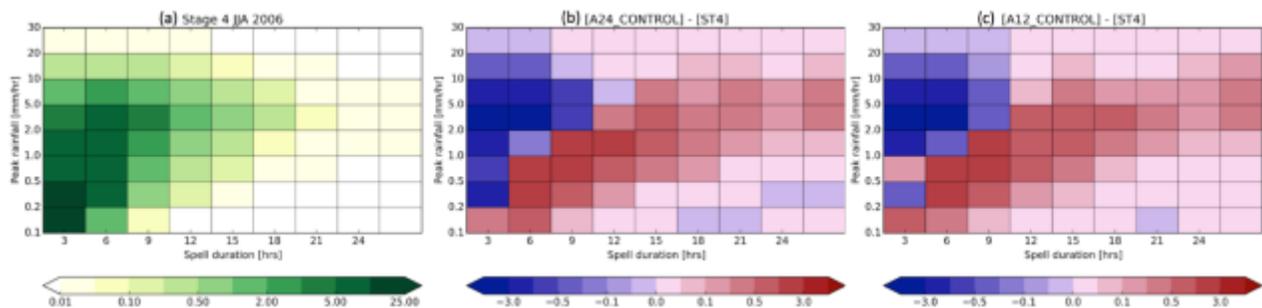


Figure 11. (a) JPDF from three-hourly averaged Stage IV data over the Great Plains in summer (June-August) 2006. The JPDF biases of (b) A24 and (c) A12 control simulation compared to Stage IV. A number of rainfall events in each bin is divided by the total wet spells (3482532 in Stage IV).

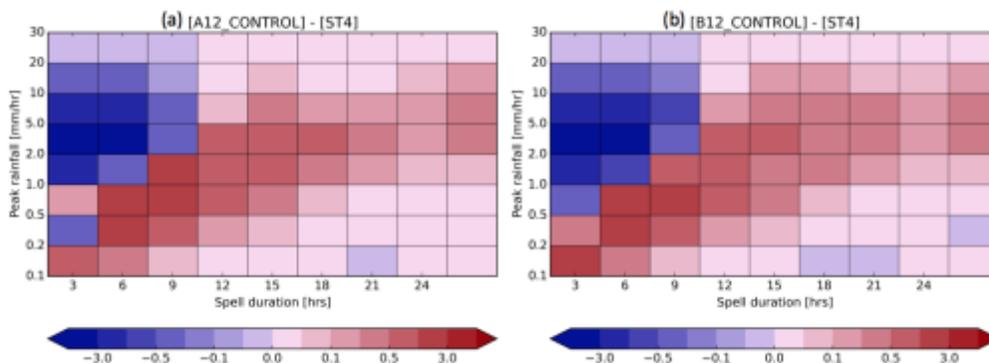


Figure 12. JPDF biases of (a) A12 and (b) B12 control simulation compared to Stage IV.

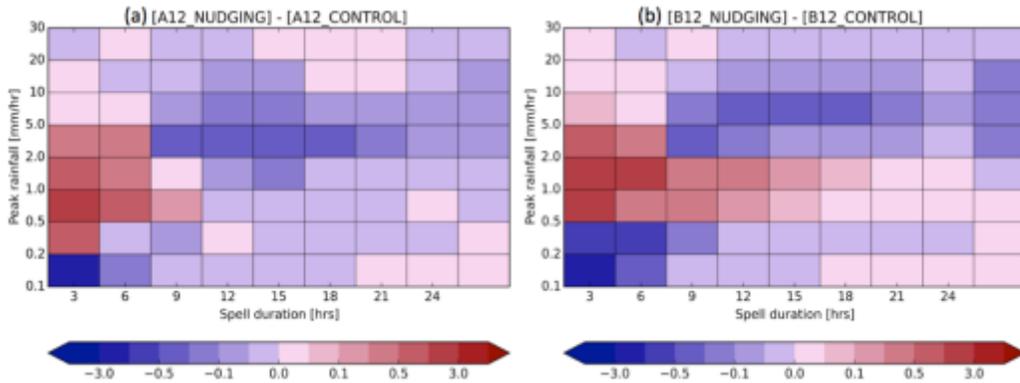


Figure 13. JPDF difference plot between nudged and control simulations with 12-km resolutions in domains (a) A and (b) B.

6. Phase II: Single-Decade Production Experiments

6.1. Initial Production Simulations

We designed ten experiment configurations for the single-decade production runs. Table 8 summarizes the abbreviated name, horizontal grid spacing, number grid dimensions, number of CPU cores used, as well as strength of spectral nudging and employed cumulus parameterization scheme in each configuration. Six runs with horizontal grid spacing of 24 km adopted different strengths of spectral nudging and different cumulus parameterization schemes. Three runs with horizontal grid spacing of 12 km were configured using different strengths of spectral nudging.

Two different strengths of spectral nudging were tested in our single-decade production experiments. The strengths were dependent on wavenumber parameters determining the shortest wavelengths to nudge specified variables, set to approximately 600 km and 2000 km by configuring the corresponding wavenumbers. The nudging strength featured with the shorter wavelength of 600 km was nearly consistent with that assumed in the M2R12K GEOS-5 replay runs. The 2000-km nudging strength had been employed commonly in typical regional climate simulations based on the previous studies noted in Section 5.2.

Table 8. Experiment configurations of single-decadal production runs. In all grid domains, 41 vertical levels were used in a stretched configuration from the surface to 10 hPa

| Abbreviation | Horizontal resolution / grid points / CPU cores | Spectral nudging wavelength | Cumulus parameterization |
|--------------|---|-----------------------------|--------------------------|
| B24_0 | 24 km / 332×157 / 245 | OFF | G3D* |
| B24_600 | 24 km / 332×157 / 245 | 600 km | G3D |
| B24_2000 | 24 km / 332×157 / 245 | 2000 km | G3D |
| B12_0 | 12 km / 663×313 / 865 | OFF | G3D |
| B12_600 | 12 km / 663×313 / 865 | 600 km | G3D |
| B12_2000 | 12 km / 663×313 / 865 | 2000 km | G3D |
| B4_600 | 4 km / 1987×937 / 6083 | 600 km | G3D |

*G3D: The Grell 3D ensemble scheme (Grell 1993, Grell and Devenyi 2002).

Each run with horizontal grid spacing of 24 km employed 245 Intel Haswell CPU cores; 220 cores were assigned to the main computation and 25 cores were used for the quilting I/O system. The total CPU wall time for completing the 11-year integration from November 1999 to October 2010 was approximately two weeks. The control run without spectral nudging was slightly faster than other runs with spectral nudging. The total size of the output files in the 11-year integration was approximately 4.2 terabytes.

In contrast to the 24-km runs, WRF boundary files (i.e., wrfinput, wrfbdy, wrflowinp, wrffdda) were built every two-year cycle in the pre-processing for the 12-km grid runs. Reducing the file sizes improves the computational performance by decreasing the time required to scan the boundary files for the correct boundary condition time in a restart run, particularly when the domain size is very large. Each 12-km run employed 865 Intel Haswell CPU cores; 840 cores were assigned to the main computation and 25 cores were used for the quilting I/O management. The total CPU wall time for the 11-year integration was approximately one month, with the total file size of the output files at approximately 17 terabytes.

The pre-processing for the 4-km runs used boundary files with a one-year interval. Each 4-km run employed 6083 Intel Haswell CPU cores initially, of which 200 cores were assigned to the quilting I/O management. The use of more CPU cores for I/O was necessary to assign enough memory to the quilting I/O for the B4 domain. The total number of employed CPU cores was increased to 7575 for approximately a 10% speedup as extra cores became available toward the end of the project. The total CPU wall time for a one-year integration was approximately two weeks, with the estimated total time for the 11-year integration of approximately five months. Due to the source code bugs encountered late in the project, the B4 production integration was only completed through the end of 2004 within the computational allocation on the Discover cluster. The total size of the output files for the one-year integration was approximately 13 terabytes.

Lateral Boundary Condition Error Build-Up Issue

This problem was (unfortunately) only encountered when the long-term B4 run went into the seventh year of the planned 10-plus-year simulation (1 November 1999 to 1 January 2010). The B4 run crashed on 2 Apr 2007 with an instability (CFL) error. The model crash and CFL error were caused by very large values ($\sim 50 \text{ m s}^{-1}$) of the vertical wind component (w) along the lateral boundaries of the domain from the mid to upper atmosphere levels. At the same, some unrealistic “wave” patterns occurred in the vertical wind, propagating across the south-north boundaries at these atmospheric levels. Additional diagnosis indicated that the high value of w had accumulated over the course of the simulation.

In reviewing notes from the latest WRF workshop (June 2015) we found a reported bug that was likely related to the problems encountered with the B4 run. New code has been added in the latest version (3.7) of WRF to address this problem, which updates the LBCs at every time step with interpolated tendency values. This corrected source code from version 3.7 was subsequently back-ported to the version of NU-WRF used in the DSCALE project. Test runs confirmed that the code revisions successfully suppressed the error accumulation in the w -wind field along the boundaries. It was unfortunate that this issue only became detectable in the long-term simulations; therefore, we did not detect it in the 10-month Pilot simulation. As a result, we had to abandon the decadal runs already completed, and start over. We were able to complete full decades for the B24 and B12 domains, but were only able to complete four years on the B4 domain.

Deep Soil Temperature Lower Boundary Condition

When conducting long-term climate simulations, WRF provides an option to use a dynamic deep soil temperature parameterization (Salathe et al. 2008). This parameterization treats deep soil temperature as a function of lagged surface skin temperature, and is activated by setting `tmn_update = 1` in the WRF name list. The scheme is further tunable by modifying the value of

'lagday', which represents the number of days used in calculation of a lagged average value of skin temperature (Tn). The deep soil temperature is then calculated as

$$T_{\text{deep}} = 0.6 \cdot T_{365} + 0.4 \cdot T_n$$

where T365 is the average skin temperature from the previous year (365 days). The ARW User Guide states that lagday will default to a physically reasonable 150 days if not specified in the namelist (as was the case with our simulations). However, the WRF Registry file defaults this value to 1 day instead of 150. This implies that Tdeep was tied to the daily skin temperature, and as a result likely varied too much, and too quickly. The WRF developers were notified of their incorrect setting in the Registry.

Since this problem was discovered after all of the decadal runs had been completed, additional experiments on the B24 domain were conducted with lagday = 150 to investigate the potential impact of the default setting. Although there are noticeable effects in the hourly surface temperatures, results from the climate metrics analysis of the lagday = 150 vs. 1 day setting show that overall the setting has very minimal impacts on either the results and statistics of ARs or on temperature and precipitation in the Great Plains. A detailed analysis of the AR metrics (not shown) indicates that the lagday setting shows no measurable impacts on simulating winter-mean/AR-mean precipitation or T2 in the three Bukovsky regions analyzed. Table 9 shows the impact on the precipitation analysis over the Great Plains region as compared to Stage IV observations from the B24, B24 re-run, and B12 and B4 runs. There is some improvement (2 percent) when using lagday = 150, but it is much smaller than the impact of resolution alone (12/4 km vs. 24 km). That the impacts are positive supports the realism of the 150-day setting. However, the small effect seen in the B24 simulations leads us to believe that the results for the other experiments with lagday = 1 remain valid.

Table 9. Skill scores of precipitation from the various NU-WRF simulations vs. Stage IV observations over the Northern, Central, and Southern Great Plains regions.

**Effect of the delayed soil temperature
(observation data: Stage IV)**

| [%] | Northern Plains | Central Plains | Southern Plains |
|-------------------|--------------------|-------------------|--------------------|
| B24 | 70.0 | 67.4 | 61.6 |
| B24 (lag 150days) | 72.3 | 68.3 | 63.4 |
| B12 | 79.3 | 78.1 | 74.4 |
| B4 | 83.9 | 81.2 | 82.3 |

6.2. Metrics Evaluation of Experiments

CONUS-Wide Precipitation PDF

To examine the temporally continuous variation of the performance metrics, we produced both bias and spatial correlation (Figure 14) calculated from monthly mean precipitation. Both metrics showed strong seasonal dependence. For bias, the models tend to produce strong overestimates in summer. Similarly the spatial correlation also shows a strong seasonal cycle – higher correlation in winter and lower in summer. No substantial differences can be seen in these metrics among the three NU-WRF model resolutions (B24, B12, and B4), although there is a noticeable increase in bias in both the NU-WRF and M2R12K. There is also a notable degradation in correlation in the M2R12K while the NU-WRF runs seems to maintain a correlation similar to MERRA-2.

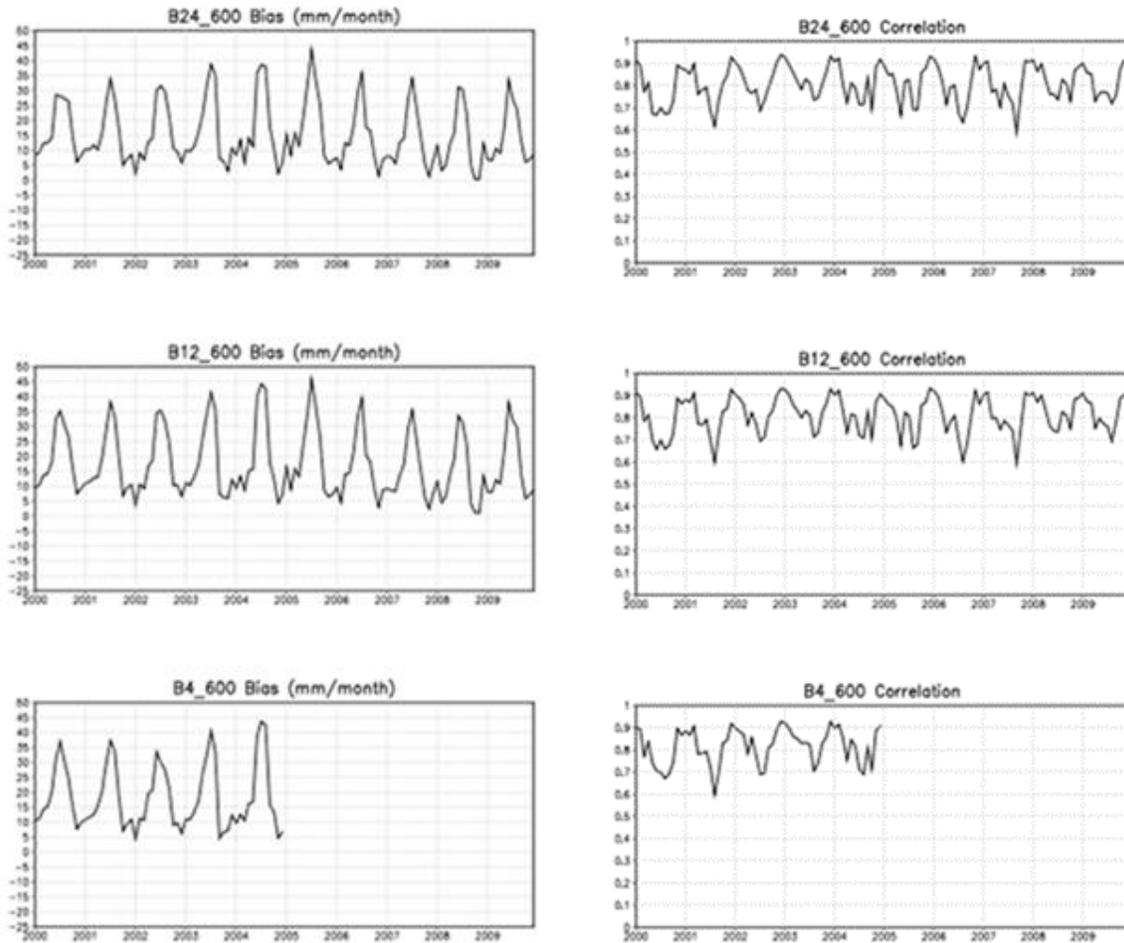


Figure 14. Area-mean bias (left) and spatial correlation (right) of monthly precipitation between models and PRISM, for the three model resolutions: B24_600 (top row), B12_600 (middle row), and B4_600 (bottom row).

The PDFs of the precipitation rate (Figures 15 and 16) revealed more interesting features. The original MERRA-2 rainfall PDF for CONUS proved to be quite similar to the observations, with a slight tendency to overestimate lighter rainfall. In contrast, the MERRA-2 replay seems to significantly overestimate the heavy precipitation rates over CONUS. This overestimation is further revealed as a systematic overestimate in winter (DJF) combined with a mix of low rain rate underestimation and high rain rate overestimation in the summer (JJA). The SE JJA appears to be particularly problematic for both MERRA-2 and M2R12K.

Over CONUS, all the NU-WRF model outputs consistently have more precipitation, especially in the intermediate rainrate range. The excess precipitation at the intermediate range can be traced to summer, and further to the northeastern and southeastern CONUS. The highest resolution (B4_600) exacerbates the excessive precipitation at the intermediate range, still mostly over the two eastern regions (NE and SE). Curiously, the B4_600 run produced less intensive rainfall (> 30 mm/day) than the coarser resolution counterparts (B24_600 and B12_600).

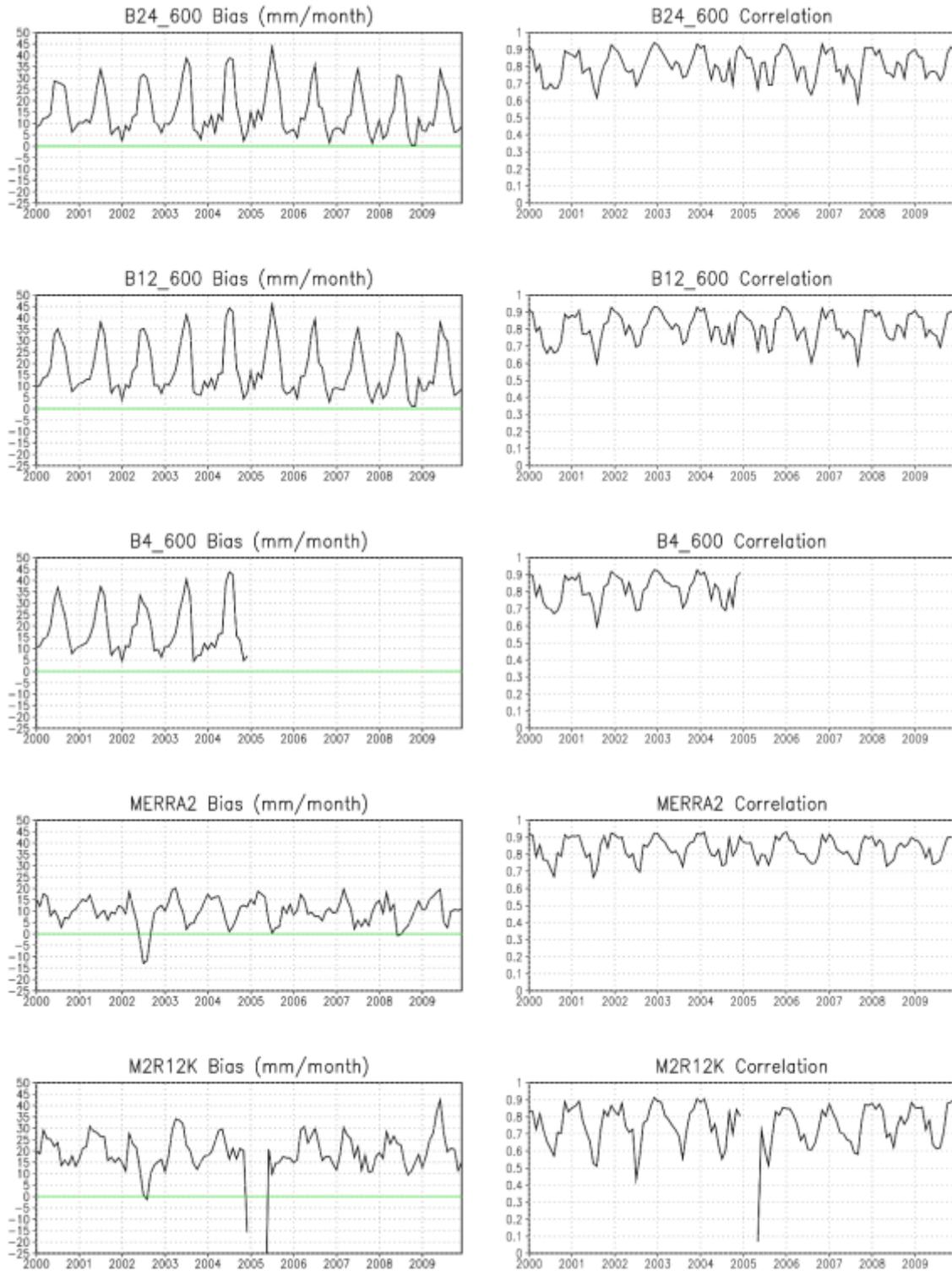


Figure 15. Area-mean bias (left) and spatial correlation (right) of monthly precipitation between models and PRISM for the three NU-WRF model resolutions: B24_600 (top row), B12_600 (second row), and B4_600 (third row), in addition to MERRA-2 (fourth row) and the MERRA-2 12-km replay (bottom row).

Rain amount distribution vs. rainrate, 2000-2009

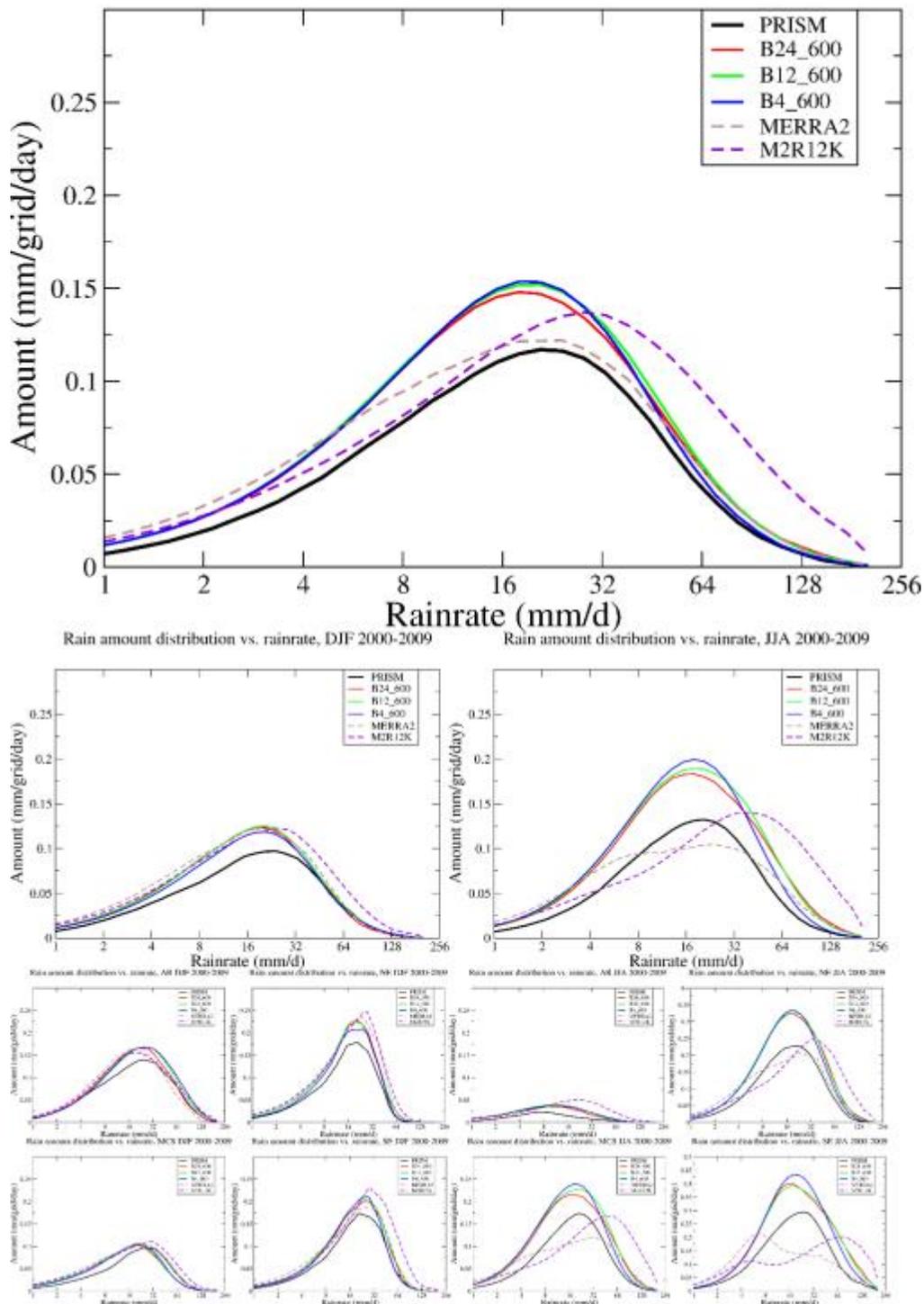


Figure 16. Area-mean daily precipitation amount as a function of precipitation rate for three model resolutions: B24_600, B12_600, and B4_600 (red, green, and blue curves). There are three levels of comparisons between the model output and PRISM for each model: the top level is for the entire spatial domain (CONUS) and study period (2000-2009); the middle level is for the entire spatial domain (CONUS) but for separate winter (DJF; left) and summer (JJA, right); and the bottom level is for four regions associated with the respective winter and summer season.

Atmospheric Rivers

AR frequency patterns (north-south and east-west gradients) are well represented in B simulations, although in terms of correlation B4_600 is not as high as others (0.9 compared to 0.95-0.98). Negative biases exist in simulated AR frequency, around 0.3 (in units of standard deviation of the reference map) in most cases. B12_600 has the weakest bias (-0.16), and B4_600 is among the better ones (-0.25) (Figure 17, upper panel). AR zonal IVT (integrated water vapor transport) patterns (land-sea gradient) are well represented in B simulations. All B simulations have similar performance in terms of pattern correlations and standard deviation ratios. There are positive biases in simulated AR zonal IVT, around 0.45 standard deviation in most cases. B4_600 has the weakest bias (0.33) (Figure 18, upper panel). AR meridional IVT patterns (land-sea gradient) are also well represented in B simulations. Performance is similar among all B simulations in terms of pattern correlations, with B12_600 being the best by small margins. Negative biases of ~ 0.1 standard deviation exist in most cases, much weaker than in the case of zonal IVT. B6_600 has the weakest bias (-0.01 standard deviation) (Figure 19, upper panel).

Based on Tian scores (lower panels in Figures 17-19), B12_600 is one of the best (by small margins) in terms of AR frequency; B4_600 has the lowest score although still above 0.8. B4_600 is one of the best (by small margins) in terms of AR zonal IVT. There are no remarkable differences between the various experiments in the case of AR meridional IVT.

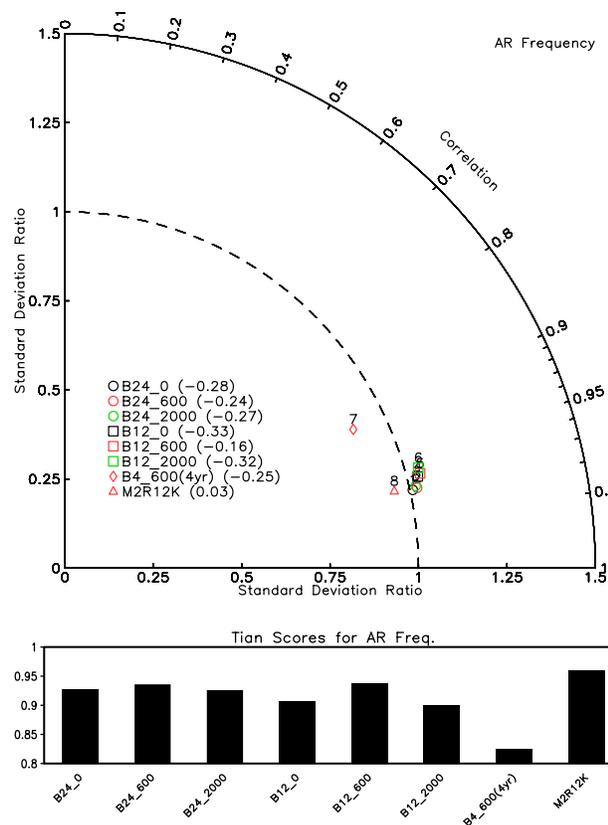


Figure 17. (upper) Taylor diagram based on comparing the spatial patterns of AR frequency between MERRA-2 (the reference) and different downscaling experiments. Numbers in brackets indicate the mean bias between the two maps being compared normalized by the standard deviation of the reference map. (lower) Tian scores. Note that the vertical scale starts from 0.8 to better visualize the differences.

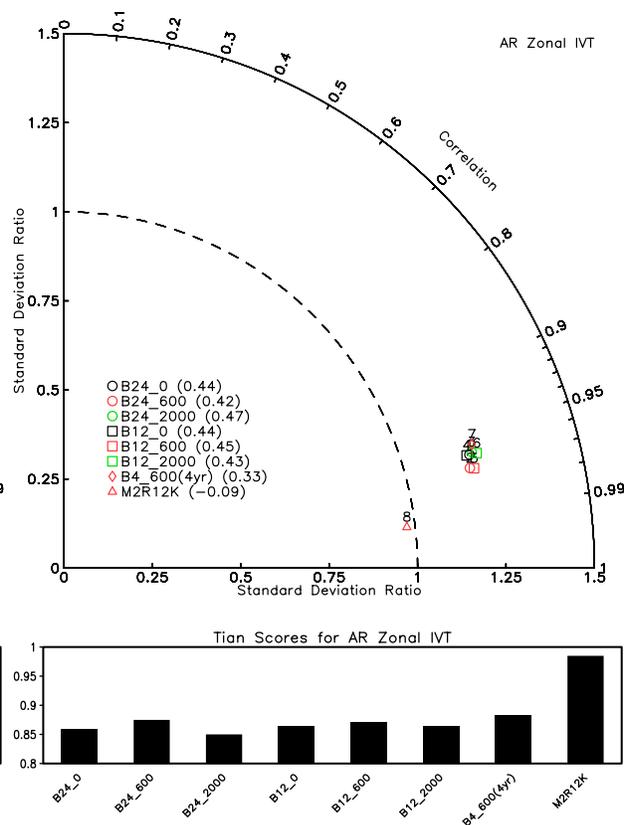


Figure 18. Same as Figure 17 but for AR zonal IVT.

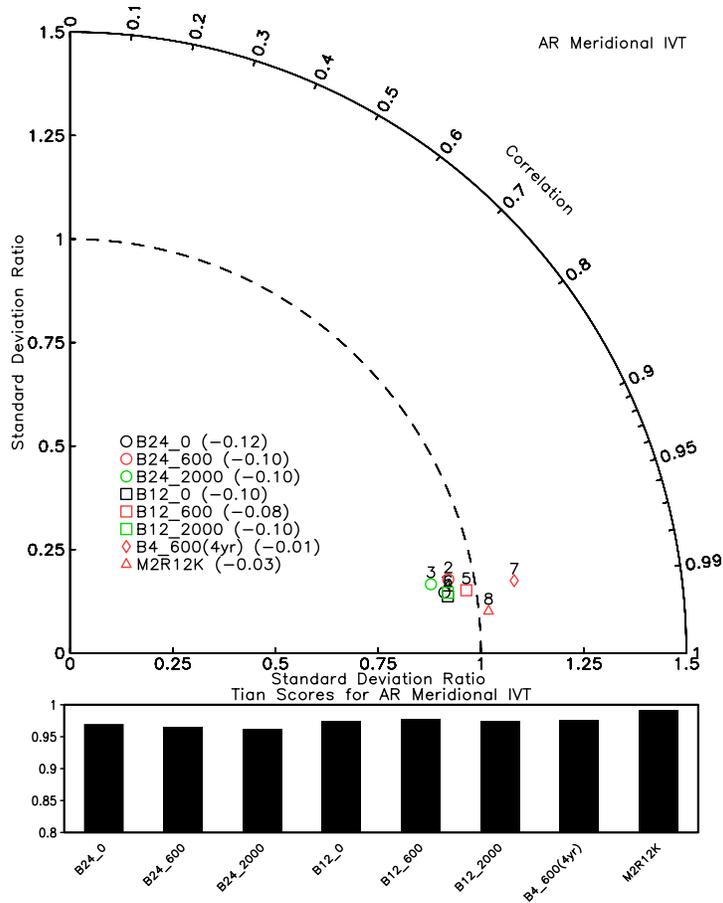


Figure 19. Same as Figure 17 but for AR meridional IVT

In order to examine the regional variations in model performance, the evaluation has been performed separately for three Bukovsky regions in WUS including PNW, PSW, and GB. Three metrics—Taylor diagram, Tian score and the PDF skill score—are employed to quantify model performance.

As shown in Figure 20, the higher resolution nudging (600 km vs. 2000 km) generally improves both correlation and standard deviation ratio in all three regions, with the highest correlations in the PSW for B12.

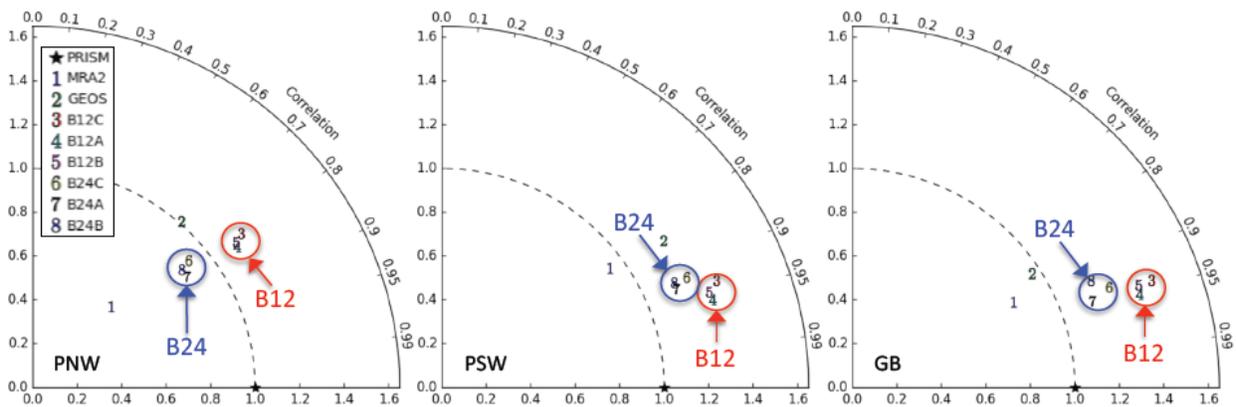


Figure 20. Taylor diagrams for winter precipitation climatology against PRISM. B12A and B24A (B12B and B24B) represent the NU-WRF runs with 600 km (2000 km) spectral nudging.

Evaluation of the precipitation and temperature for the entire winter shows that higher model resolutions generally correspond to larger spatial variability in the simulated data. In addition, among NU-WRF simulations, higher resolution runs show higher correlation with the PRISM data (Figure 21). Such effects of spatial resolution on simulating the winter-mean climatology is also clear for T2 (not shown). The effects of nudging scales are not clear for either the simulated precipitation or temperature.

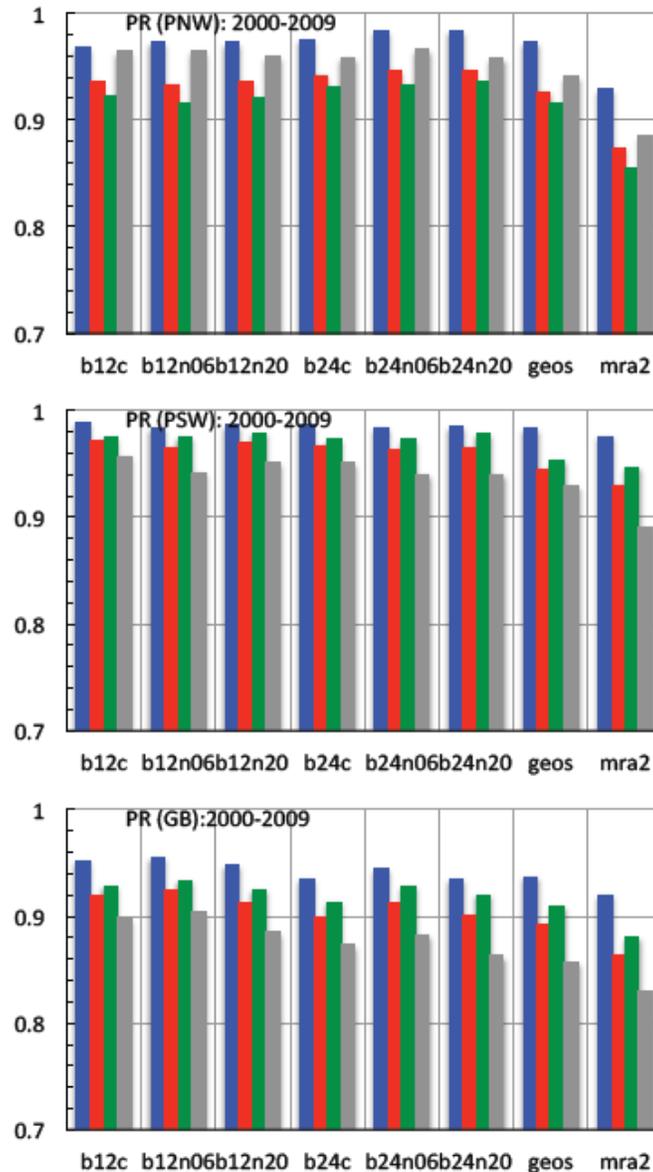


Figure 21. The PDF skill score of the model data against PRISM in the three regions for all landfall categories

All model simulations and MERRA-2 simulate AR-related precipitation fraction in WUS very well; however, they also show significant biases in simulating the locations and magnitudes of the AR-related temperature anomalies. Unlike for the winter-mean climatology, the effects of model resolution in simulating the AR-related precipitation and temperature anomalies are not clear. These effects vary widely according to regions and AR landfall locations without any systematic behavior.

All model data and MERRA-2 accurately represent the observed variations in the precipitation intensity associated with AR landfalls for all regions. The effects of spatial resolution on the daily precipitation intensity can be identified in the PDFs for all regions.

MERRA-2 data underestimates the frequency of heavy precipitation in all three regions. All 12-km NU-WRF runs and GEOS5 overestimate heavy precipitation events (>160 mm/day range) for all regions; 24-km runs show similar frequency as in PRISM in the heavy precipitation range. Thus, higher spatial resolution tends to generate higher frequency of heavy precipitation events. Despite the systematic differences in representing the heavy-tail side of the PDF, all model data but MERRA-2 show a similar PDF skill score that measures the overlapping area between two PDFs (Figure 21). The PDF skill score varies according to regions, with the highest scores in PSW and the lowest scores in GB.

Mesoscale Convective Systems

Figure 22 shows the 10-year (2000-2009) summer mean precipitation pattern over the US Great Plains from the two reference datasets (TRMM and Stage IV), MERRA2, M2R12K, and various other NU_WRF model runs. The metric results for MERRA2, M2R12K, and the NU_WRF runs are listed in red at the top right of each panel. The first result (left) is based on TRMM as the reference and the second (right) is based on Stage IV as the reference. Clearly, the metrics differ and depend on the reference datasets used but their difference is typically within ~15%. This is consistent with the metric of Stage IV based on TRMM, 0.86, which represents the uncertainties of reference datasets.

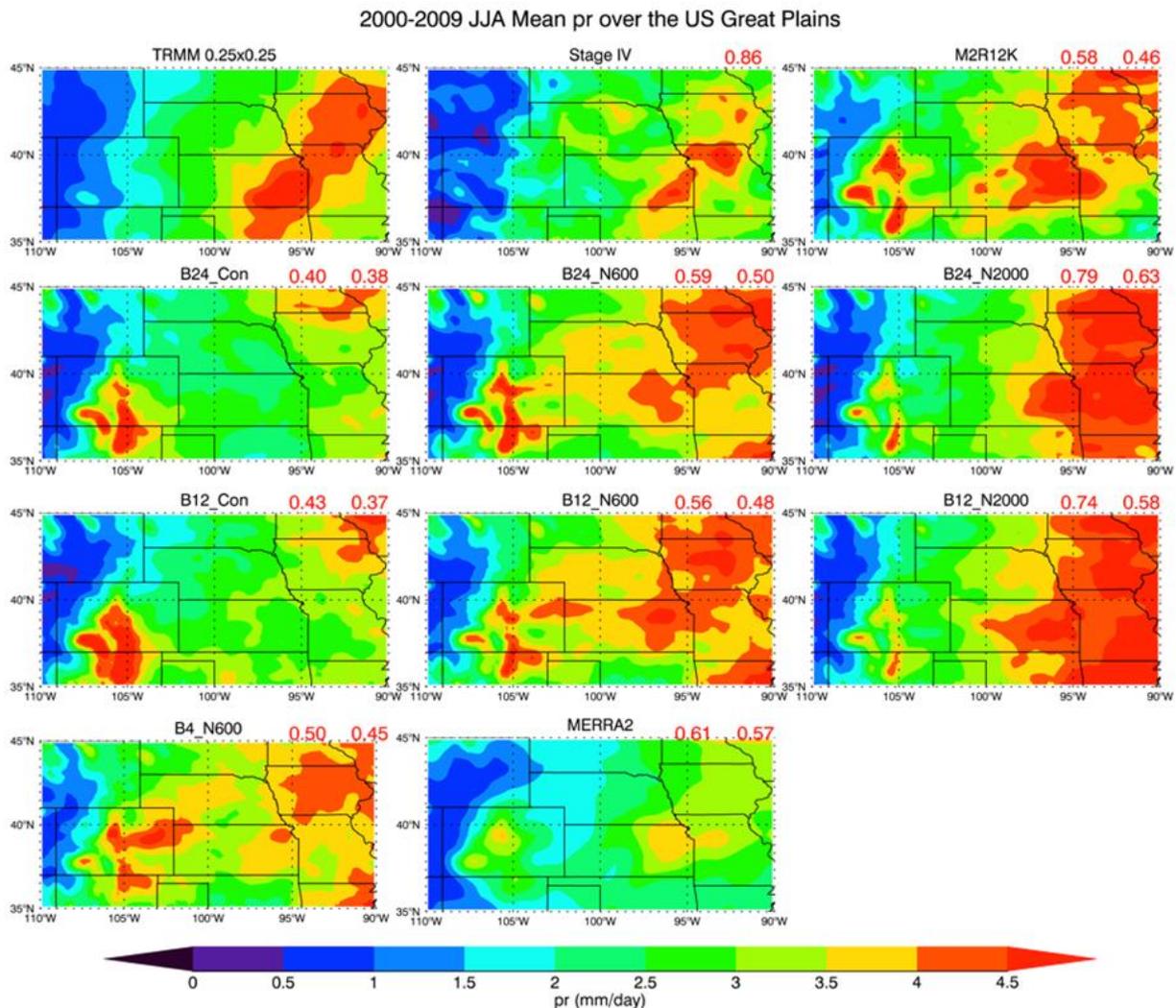


Figure 22. JJA mean precipitation over the US Great Plains, 2000-2009

Figure 23 shows the 10-year (2000-2009) summer mean precipitation eastward propagation pattern over the US Great Plains from the two reference datasets (TRMM and Stage IV), MERRA, M2R12K, and various other NU_WRF model runs. The metric results for MERRA2, M2R12K, and the NU_WRF runs are listed in red at the top right of each panel. The first result (left) is based on TRMM as the reference and the second (right) is based on Stage IV as the reference. Precipitation diurnal cycle and mean SAT were also examined, but plots are not included here.

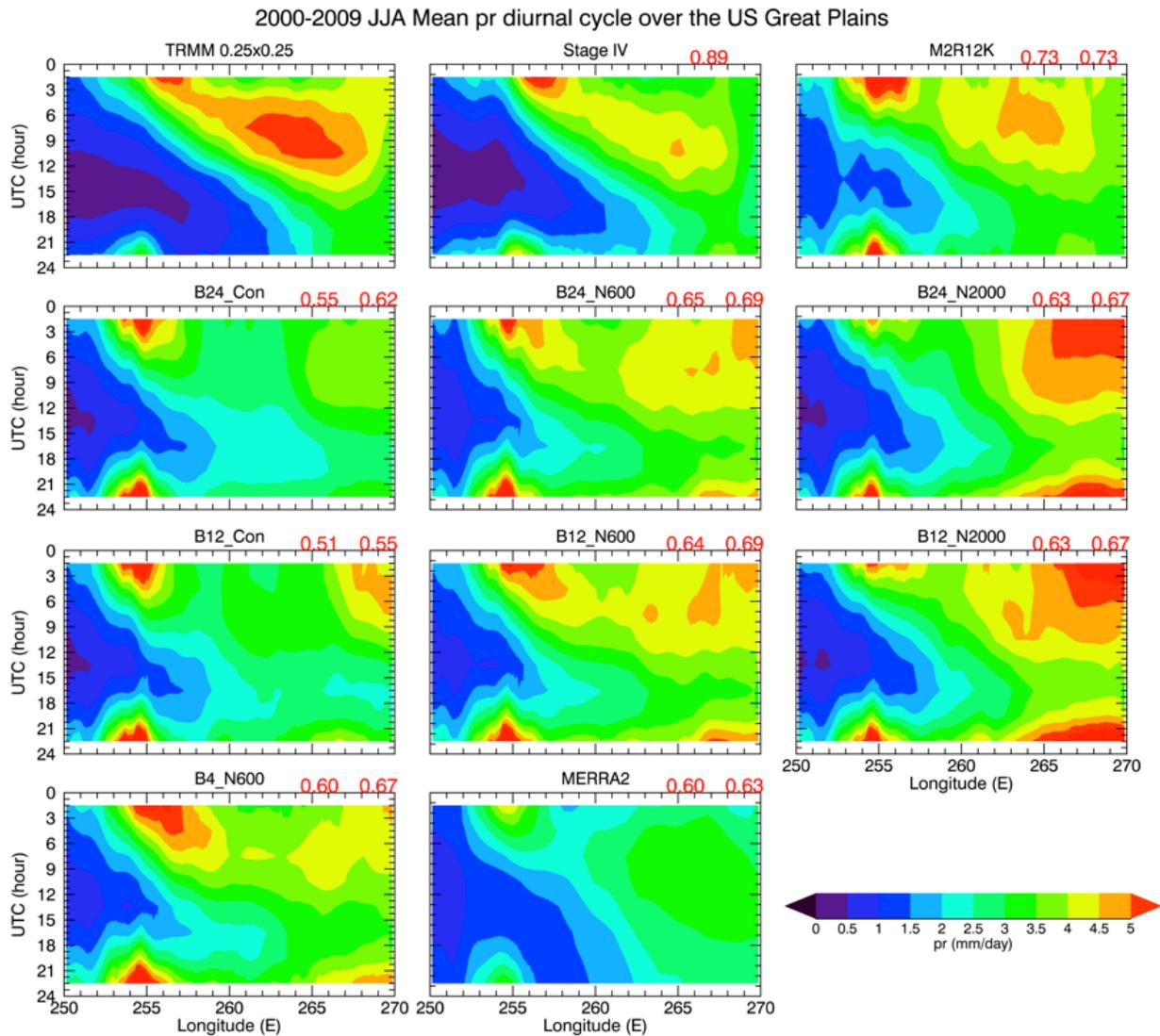


Figure 23. JJA mean precipitation eastward propagation over the US Great Plains, 2000-2009.

Table 10 summarizes the MCS metric results for MERRA2, M2R12K and various NU_WRF runs from the single-decade experiments. The first column indicates the physical phenomena that each metric is based on, such as summer mean precipitation pattern, summer mean precipitation diurnal cycle, summer mean precipitation eastward propagation, summer mean SAT pattern, and summer mean SAT diurnal cycle. The second column indicates the reference datasets for the metrics, that is, TRMM and Stage IV for precipitation and WZ and NLDAS2 for SAT. The metric of one reference dataset based on another, such as Stage IV based on TRMM, represents the uncertainties of reference datasets and the calculated metrics based on them. For example, the summer mean precipitation metric of Stage IV based on TRMM is 0.84. This means that the uncertainties of precipitation reference datasets, TRMM and Stage IV, are

around 16%, so a 16% variance in the metrics calculated for the model runs should be considered to be within the observational uncertainty. The third column indicates the metrics for various model runs.

The main conclusions drawn from the single-decade experiments are:

- The metric values depend on the reference datasets used but the overall results and conclusions do not seem to change.
- No significant differences are found among model runs at different resolutions. For example, the metrics for B24_N600, B12_N600, and B4_N600 are very similar and within their uncertainties.
- The nudging runs seem to be better than the control runs.
- The model performance is clearly better in SAT simulations than precipitation simulations.

Table 10. MCS metric results for the single-decade experiments. Scores are shown with only the significant digit (0.1) and are colored according to their values

| | | Green (> 0.6) | Yellow (0.4 - 0.6) | Red (< 0.4) | | | | | | |
|-------------------------------------|------------------|-------------------------|--------------------|-------------|----------|-----------|---------|----------|-----------|---------|
| Physical phenomena | Reference data r | Model simulated field f | | | | | | | | |
| | | MERRA2 | M2R12K | B24 Con | B24 N600 | B24 N2000 | B12 Con | B12 N600 | B12 N2000 | B4 N600 |
| Summer mean pr | TRMM | 0.6 | 0.6 | 0.4 | 0.6 | 0.8 | 0.4 | 0.6 | 0.7 | 0.5 |
| | ST4, 0.84 | 0.6 | 0.5 | 0.4 | 0.5 | 0.6 | 0.4 | 0.5 | 0.6 | 0.5 |
| Summer mean pr diurnal cycle | TRMM | 0.5 | 0.6 | 0.4 | 0.5 | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 |
| | ST4, 0.79 | 0.5 | 0.5 | 0.4 | 0.5 | 0.6 | 0.4 | 0.5 | 0.5 | 0.5 |
| Summer mean pr eastward propagation | TRMM | 0.6 | 0.7 | 0.6 | 0.7 | 0.6 | 0.5 | 0.6 | 0.6 | 0.6 |
| | ST4, 0.89 | 0.6 | 0.7 | 0.6 | 0.7 | 0.7 | 0.6 | 0.7 | 0.7 | 0.7 |
| Summer mean SAT | WZ, 0.80 | 0.9 | 0.9 | 0.8 | 0.9 | 0.9 | 0.8 | 0.9 | 0.9 | 0.9 |
| | NLDAS2 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| Summer mean SAT diurnal cycle | WZ, 0.86 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |
| | NLDAS2 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 | 0.9 |

North-East Winter Storms

The four primary storm-based metrics were computed for the decadal experiments; Figure 24 shows three of them: storm frequency, storm intensity, and storm total. A map of the PRISM values for each metric is presented to show what the field used as a reference looks like. A Taylor diagram summarizing the evaluation results of each metric is provided below each map.

In addition to these panels, the same comparison was applied to the results for frequency of heavy storms (storms with intensity greater than 25 mm) and storm duration (not shown). Also not shown are the results repeated on storms that were concurrent with temperatures below freezing (i.e. “frozen storms”). Figure 25 summarizes the model performance as referenced against PRISM with the root mean squared difference/error (RMSE) between the spatial field of each metric and PRISM. All RMSE values are normalized by the spatial standard deviation of the PRISM field for easier interpretation and inter-metric comparison.

Overall it is difficult to discern a major difference between nudging and non-nudging as well as between the three different resolutions. It is somewhat apparent in the Taylor diagrams that the non-nudged runs show slightly inferior performance, particularly with a systematic over simulation of the spatial standard deviation of the each field. It is also apparent that while different nudging schemes don't make a large difference in model skill, simulations using the same schemes do tend to cluster together in the Taylor diagram. For example, the control simulations (in red) are adjacent in all three Taylor diagrams while the N06 runs (blue) not only align together but generally exhibit better accuracy than the other configurations. One feature common amongst all simulations, regardless of resolution or nudging scheme, is an over simulation of the metric values for all storm metrics (not apparent in figures provided here). This effect is reduced in the MERRA-2 and M2R12K datasets.

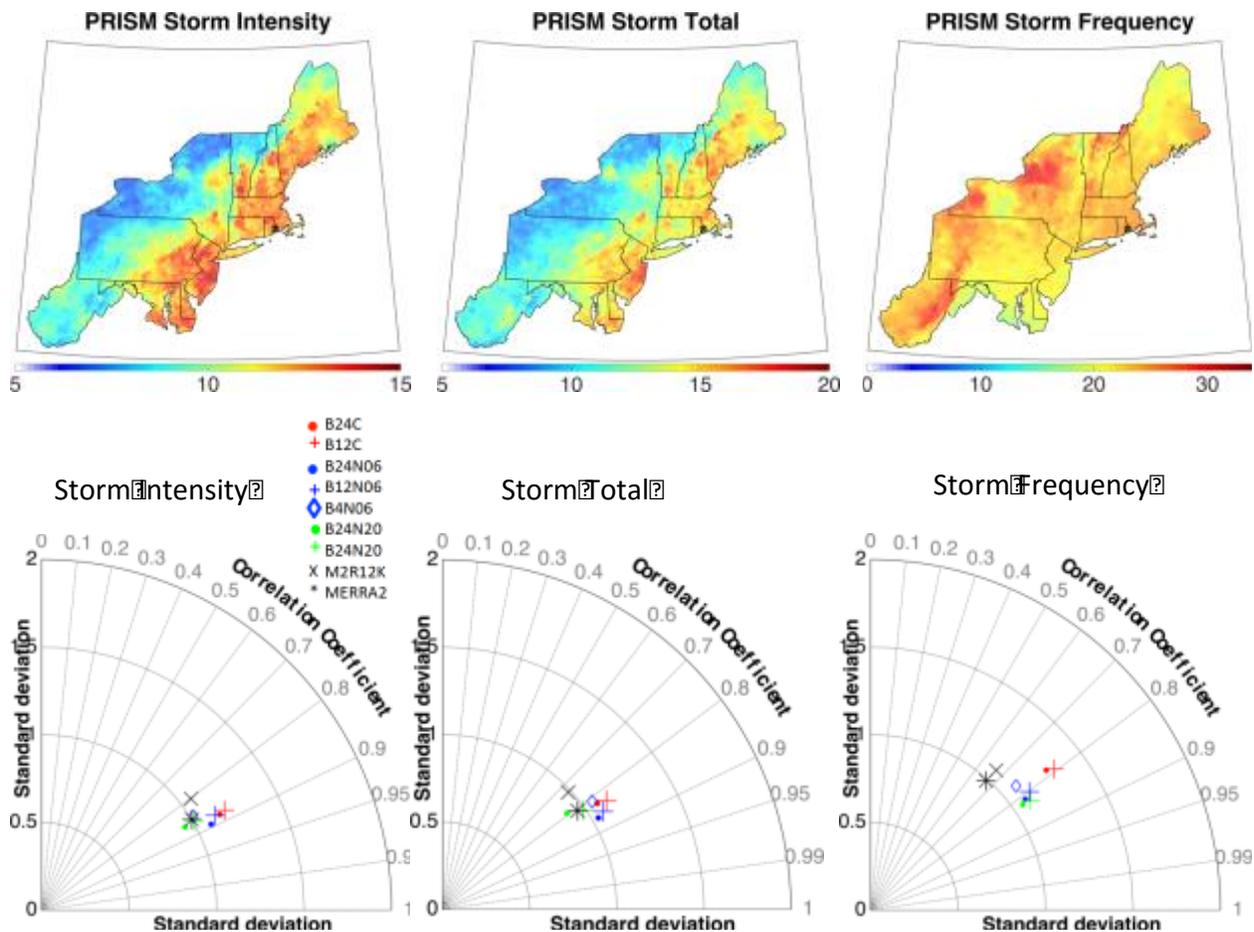


Figure 24. Three of the five metrics for all storms (frozen and non-frozen): (left) Storm Intensity, (center) Storm Total, and (right) Storm Frequency. Maps show the reference field for each metric and Taylor diagrams summarize the model performance for each field.

| All Storms | B24C | B24N20 | B2406 | B12C | B12N20 | B12N06 | B4N06 | M2R12K | MERRA2 | Mean RMSE |
|------------|------|--------|-------|------|--------|--------|-------|--------|--------|-----------|
| Total | 0.78 | 0.91 | 0.83 | 0.85 | 0.86 | 0.82 | 0.67 | 0.8 | 0.72 | 0.80 |
| Duration | 1.8 | 1.6 | 1.4 | 1.8 | 1.7 | 1.6 | 1.5 | 1.2 | 1.2 | 1.5 |
| Intensity | 0.64 | 0.7 | 0.65 | 0.71 | 0.66 | 0.68 | 0.55 | 0.67 | 0.55 | 0.65 |
| Frequency | 1.1 | 1.5 | 1.4 | 1.2 | 1.6 | 1.5 | 0.8 | 2.0 | 1.9 | 1.4 |
| Freq Heavy | 1.2 | 1.1 | 0.9 | 1.3 | 1.1 | 0.94 | 0.64 | 1.2 | 0.95 | 1.0 |
| Mean RMSE | 1.1 | 1.2 | 1.0 | 1.2 | 1.2 | 1.1 | 0.8 | 1.2 | 1.1 | 1.1 |

| Froz Storms | B24C | B24N20 | B2406 | B12C | B12N20 | B12N06 | B4N06 | M2R12K | MERRA2 | Mean RMSE |
|-------------|------|--------|-------|------|--------|--------|-------|--------|--------|-----------|
| Total | 0.83 | 0.93 | 0.87 | 0.89 | 0.89 | 0.92 | 0.94 | 0.77 | 0.65 | 0.85 |
| Duration | 1.9 | 1.8 | 1.7 | 2.0 | 1.8 | 1.8 | 1.9 | 1.5 | 1.2 | 1.7 |
| Intensity | 0.67 | 0.78 | 0.73 | 0.71 | 0.72 | 0.75 | 0.76 | 0.66 | 0.61 | 0.71 |
| Frequency | 1.0 | 0.99 | 0.91 | 1.1 | 1.0 | 0.98 | 0.89 | 1.1 | 0.95 | 1.0 |
| Freq Heavy | 1.7 | 1.6 | 1.4 | 1.8 | 1.5 | 1.4 | 1.1 | 1.3 | 0.98 | 1.4 |
| Mean RMSE | 1.2 | 1.2 | 1.1 | 1.3 | 1.2 | 1.2 | 1.1 | 1.1 | 0.88 | 1.1 |

Figure 25. A portrait diagram of the RMS difference (normalized by the spatial standard deviation of the reference field) for each dataset in reference to PRISM. The top table is for all storms and the bottom panel is for frozen storms. The bottom row for both tables is the mean RMS difference for each dataset and the rightmost column is the mean RMSE for each metric. Color shading corresponds to the RMSE magnitude, with warmer colors indicating larger errors and greener colors indicating lower errors.

The summary of all metric calculations in Figure 25 reiterates the small differences between different simulation configurations. The mean RMSE values only vary a few tenths across the suite; however, there is some indication that N06 runs have fewer errors than the other configurations, especially for the all storms metrics. Also, B4 stands out as superior among the NU-WRF runs, especially for the all storms metrics. For frozen storms, MERRA-2 outperforms the other simulations with a mean RMSE of 0.88, when the reanalysis is treated as a model. More apparent are the differences between the ability of the datasets to capture different storm metrics. For example, storm total and storm intensity are relatively well simulated with mean RMSE values of 0.8 and 0.65 for all storms, and 0.85 and 0.71 for frozen storms, respectively. Conversely, all storms duration and storm frequency show the highest RMSE values of 1.5 each. Frozen storms duration and heavy storm frequency show relatively large RMSE values, at 1.7 and 1.4, respectively.

Many key features are not apparent in the summary figures provided here. For example, all datasets are able to reasonably capture the regional variations in each storm metric (such as high storm intensity) along the coast in comparison with the interior or the orographic effect on storm frequency. Overall, however, this reasonable reproduction is associated with a systematic high bias for all metrics domain-wide, with bias most apparent over higher elevations. Also only apparent upon visual inspection is the commonly superior performance of the M2R12K simulation compared with NU-WRF. This does not always stand out in the summary metrics because there is some spatial noise in the metric fields for both PRISM and M2R12K. While this noise may increase the RMSE and decrease the pattern correlation, the general spatial distribution of metric values matches up fairly well. Noise would likely be reduced if a longer climatology were employed than the single decade available here. Lastly, visual inspection of maps of each metric suggests that the higher resolution simulations capture key regional to local scale features, especially related to topography, more realistically than the lower resolution simulations. This follows also from the fact that some features that can be resolved at 4 km simply cannot be resolved at 24 km.

Hourly Precipitation

When comparing JPDFs made out of three-hourly precipitation datasets with Stage IV, B4 did not show noticeable improvement over B12 and B24. It is well known that the advantage of high-resolution simulations in modeling precipitation is their ability to reproduce variability in

high-frequency. Therefore, all JPDFs from the Phase II simulation results were built using the original hourly output (assuming JPDF bin width for wet spell duration is one hour).

Figure 26 shows the JPDF from Stage IV, and the biases of three WRF simulations for the Northern Plains. Similar to the pilot experiment results, NU-WRF simulations commonly show negative biases for the short-duration downpouring events. Nevertheless, B4 shows the overall improvement in reproducing high-frequency rain events, especially those that last between one and two hours.

In Figure 27, the reference data is the JPDF from GPM precipitation for the Central Plains. Not surprisingly, B4 better represents wet spells that last less than an hour than B12 and B24. This improvement of B4 is consistent with the previous comparison with Stage IV JPDF.

To measure the similarity of a simulated JPDF to an observed JPDF, our evaluation metric is calculated as

$$overlap = \sum_x \sum_y \text{minimum} (F_0(x, y), F_1(x, y))$$

where x is wet spell duration and y is rainfall intensity [Lee et al. 2015]. F_0 and F_1 are JPDFs from observational and model datasets respectively. The overlap ranges from 0 to 100%, and performance of a simulation is better with larger overlap values.

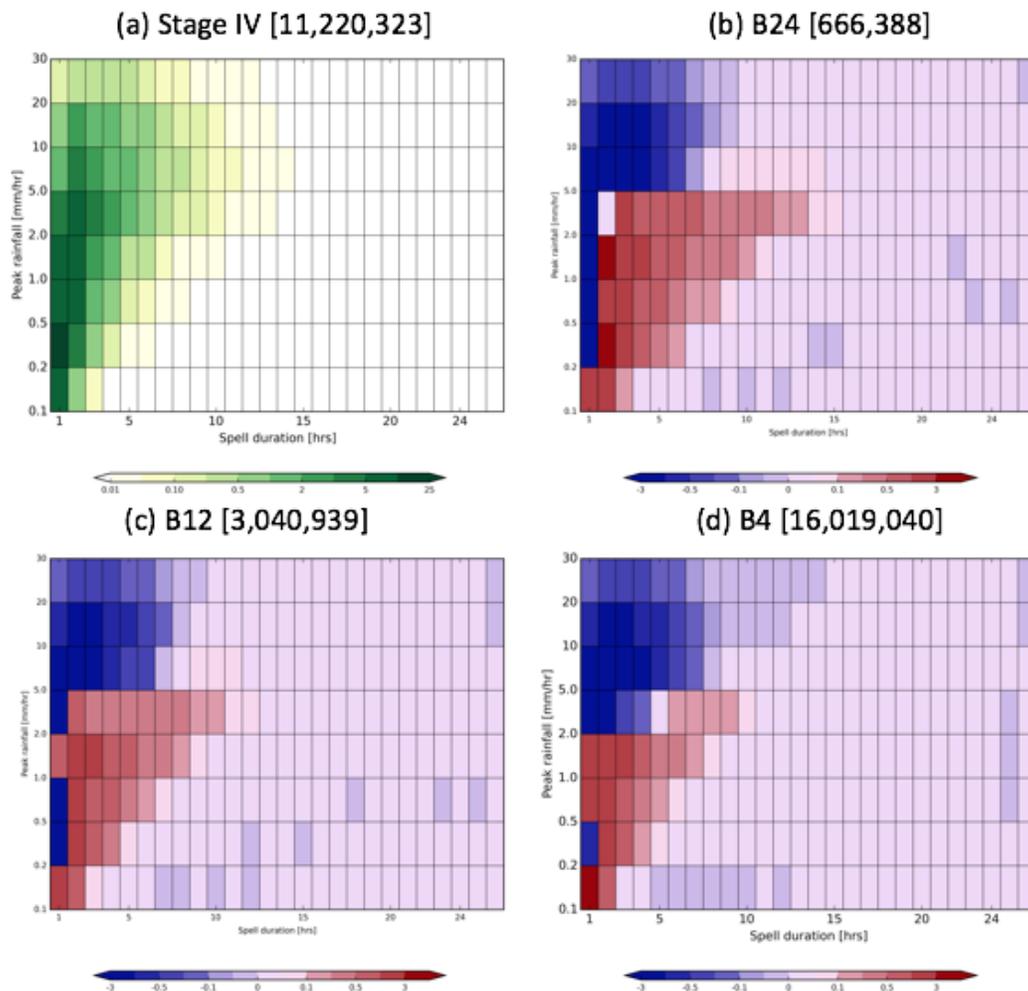


Figure 26. (a) Climatological summertime (June-August) JPDF from hourly Stage IV data over the central Great Plains between 2002 and 2010. JPDF differences of (b) B24, (c) B12 and B4 simulations with spectral nudging at 600-km scale from the Stage IV JPDF in (a). The numbers in the square bracket represent the total wet spell events to build the JPDF in each dataset.

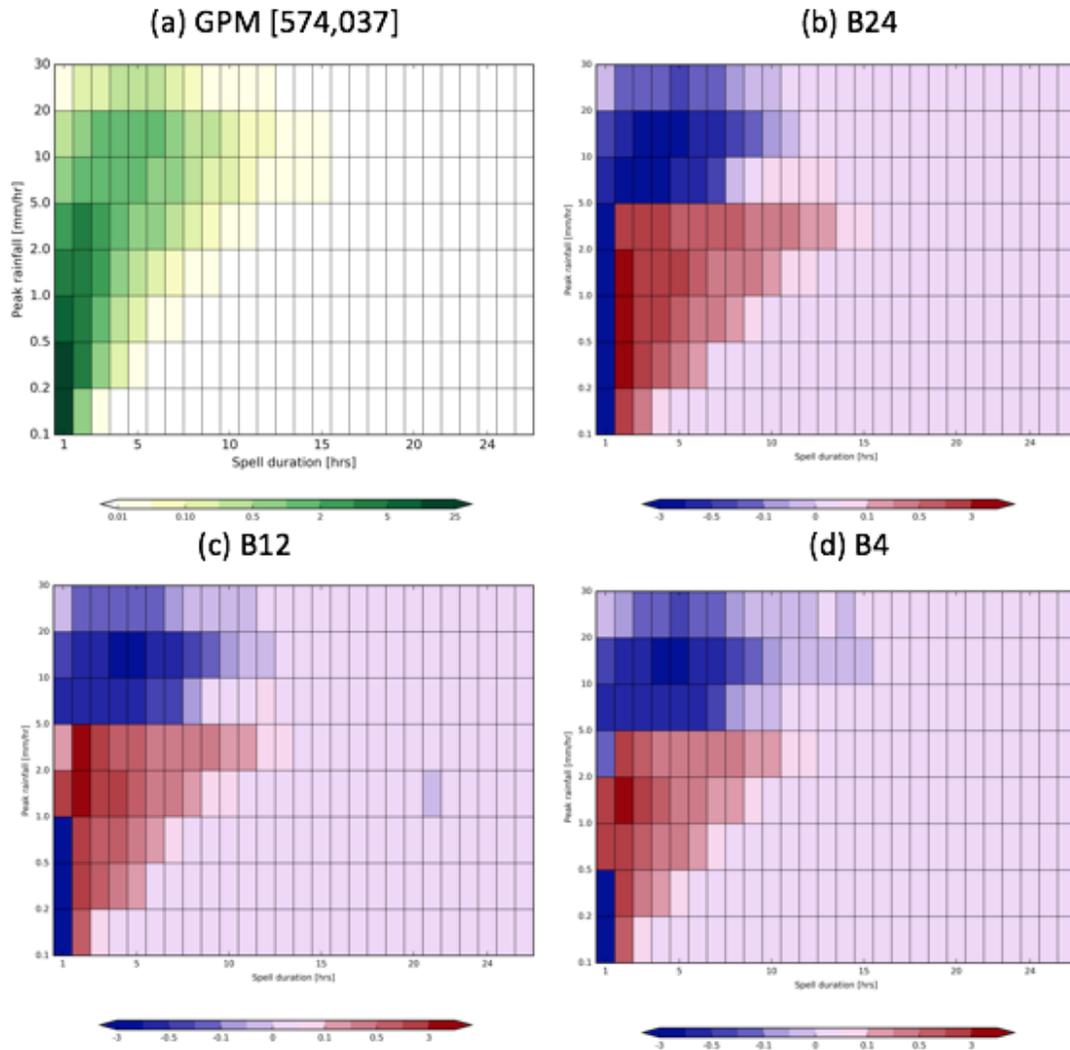


Figure 27. Same as 26, but the reference JPDF (a) is from GPM data from two summers, 2014 and 2015.

In Table 11, the overlap ratios of JPDF between simulation and the two observational datasets, Stage IV and GPM, are given. It is obvious that the NU-WRF simulation with 4 km resolution shows the best agreement with both observations.

For 24 km and 12 km runs, there are runs with a different spectral nudging scale and runs without spectral nudging. Table 12 summarizes the similarity in JPDF between Stage IV and the runs with various nudging options. In the Northern Plains, nudged runs show better agreement with the observation. However, it is hard to say that nudging significantly improves the performance of models in simulating rainfall characteristics in the Central and Southern Plains.

In summary, in the Great Plains, B4 can simulate the most similar rainfall characteristics to Stage IV and GPM at its original grid points. The spectral nudging effect is not significant in the JPDF of precipitation peak intensity and duration over the Great Plains. The most significant improvement in the results is due to higher horizontal resolution.

Table 11. Overlap [%] of the JPDFs between simulations with 600-km spectral nudging and StageIV/GPM in the Northern, Central, and Southern Plains.

| Simulations | Northern Plains | Central Plains | Southern Plains |
|--------------------|------------------------|-----------------------|------------------------|
| B24 | 70/67 | 67/62 | 62/59 |
| B12 | 79/76 | 78/73 | 74/71 |
| B4 | 84/82 | 81/77 | 82/79 |

Table 12. Overlap [%] of the JPDFs between simulations with different spectral nudging scales and Stage IV in the Northern, Central, and Southern Plains.

| Simulations, nudging scale | Northern Plains | Central Plains | Southern Plains |
|-----------------------------------|------------------------|-----------------------|------------------------|
| B24, 600 km | 70 | 67 | 62 |
| B24, 2000 km | 70 | 67 | 62 |
| B24, no nudging | 66 | 68 | 64 |
| B12, 600 km | 79 | 78 | 74 |
| B12, 2000 km | 80 | 79 | 75 |
| B12, no nudging | 74 | 77 | 75 |

7. Conclusions

7.1. CONUS PDF

We examined the performance of 12 models with several conventional performance metrics: bias, coefficient of variation (CV), and correlation coefficient. In addition, we calculated the PDFs of precipitation from each model and compared them with PRISM and NLDAS as references. We paid special attention to investigate and document how the models' performances was dependent on the following factors:

- Model resolution
- Spatial and temporal locations, especially climate regions and seasons
- Model nudging strength

Our findings are summarized as follows:

- Overall, all the models reproduced winter climatology better than summer. Winter enjoys lower biases and higher spatial and temporal correlations than summer. The PDFs are also closer to the reference for winter. Summer tends to witness strong positive or negative biases, lower correlation coefficients, and too much heavy precipitation over the eastern CONUS.
- Nudging resulted in fundamental differences in the models' performances. In fact, there was little skill in the model without nudging. The two nudging strengths used in our experiment (2000-km and 600-km) produced no perceptible differences in the metrics.
- Finally, across the different model resolutions, from 24 km, to 12 km, to 4 km, changes to the metrics, if any, were not evident. No significant degradation in bias or correlation was observed from low- to high-resolution model runs, except that the PDF at 4-km has more intermediate-rate precipitation than its lower-resolution counterparts during summer. No particularly higher temporal variability was detected in the high-resolution output.

7.2. ARs

Key findings in the evaluation of the 10-year regional climate simulation are:

- All NU-WRF runs and MERRA-2 simulate the winter precipitation and temperature climatology very well.
- Spatial patterns of AR frequency, and zonal and meridional IVT are well represented in all WRF runs. However, typically there are negative biases of ~ 0.3 standard deviation in simulated AR frequency, positive biases of ~ 0.45 standard deviation in zonal IVT, and negative biases of ~ 0.1 standard deviation in meridional IVT. With respect to AR frequency and IVT, B12_600 performs the best among B24 and B12 simulations, although by small margins. In that regard, the result for B4_600 is mixed: it has the weakest bias of AR IVT, the lowest pattern correlation and Tian score for AR frequency, and the lowest agreement rate/strongest bias in landfall dates, which could be all or partly related to the 4-year analysis period.
- The effects of nudging and resolution clearly appears in simulating winter climatology.
- Higher spatial resolution yields higher spatial σ in the simulated spatial distribution of the winter-mean precipitation and temperature climatology.
- Higher spatial resolution and spectral nudging shows positive effects in simulating the winter temperatures and precipitation distributions.
- All models and MERRA-2 perform reasonably in simulating the precipitation fraction in the three Bukovsky regions.
- Models and MERRA-2 data show lower performance for the PNW region compared to the PSW/GB region.
- Observations show that the PDF of daily precipitation and temperature in the three regions vary according to AR landfall locations.
- AR landfalls in the PNW coast increase the frequency of heavy precipitation in the PNW region while reducing it in the PSW region.
- AR landfalls in the California coast increase the heavy precipitation frequency in the PSW and GB regions while reducing it in the PNW region.
- The most noticeable effects of AR landfalls on the temperature PDF are the reduction in the variability mostly by reducing cold events.
- All model simulations and MERRA-2 data depict the variations in the daily precipitation and temperature PDFs found from the fine-resolution PRISM analysis data.

Overall, unlike for simulating the winter-mean climatology, no systematic effects of resolution and/or spectral nudging have been identified in simulating the AR-related precipitation and temperature anomalies.

7.3. MCS

Evaluation of all model runs from the one-year pilot experiment and a single-decade experiment provides very similar results regarding the MCS simulations in NU-WRF model, MERRA, MERRA-2, and GOES5-M2R12K replay runs. It seems that MERRA does well in simulating the summer mean precipitation and SAT patterns but does poorly in simulating the diurnal cycle of the precipitation. The NU-WRF model, MERRA-2, and GOES5-M2R12K replay runs can simulate most major features of the MCSs well (a metric value > 0.5). These include the summer mean precipitation pattern, summer mean precipitation diurnal cycle, summer mean precipitation eastward propagation, summer mean SAT pattern, and the summer mean SAT diurnal cycle. It seems that there are some credibility and value added of downscaled climate model simulations of the summertime MCSs over the US Great Plains ($110\text{--}90^\circ\text{W}$; $35\text{--}45^\circ\text{N}$) in comparison to the coarser-resolution GCMs. In addition, the downscaled climate model performance is clearly better in SAT simulations than precipitation simulations. Furthermore, the downscaled climate model performance of the NU-WRF model depends on the domain size and

nudging strength. The nudging runs seem to be better than the control runs. However, no significant differences were found among model runs at different spatial resolutions.

7.4. NES

- NU-WRF simulations are able to capture the main geospatial features of the storm event metrics.
- There is subtle improvement in N06 compared with N20 and control simulations.
- The high-resolution B4 run shows lower error overall and visually matches PRISM better than the higher resolution NU-WRF runs.
- In general, M2R12K visually reproduces the climatology of most storm metrics better than NU-WRF, as compared with PRISM.
- All models over simulate the storm metrics.
- It is difficult to confidently conclude whether downscaling to higher resolutions strongly improves the simulation of precipitation-based impacts of NES.

8. Recommendations for Future Study

While we strived to identify the impact of model resolutions on model performance, this impact seems to be quite weak from our current analysis. One reason is that the impact of other factors, including nudging, model setup and structure differences, turned out to be much larger than that of the model resolution. Thus the benefit of the high-resolution modeling, if any, was overwhelmed by the differences caused by these factors. In addition, for a given configuration, any of the models already have fairly considerable errors at the coarser resolution, and subsequently, at higher resolution, it is difficult to separate the errors from the contribution of the coarser model forcing, and the errors from the high-resolution internal dynamics. Therefore, for future studies, we need to start with a well-calibrated, trustworthy coarser resolution model.

Currently there is no clear definition of “model performance improvement with resolution.” For example, high-resolution models will naturally produce a lower correlation coefficient due to increased variability at finer scales. But it’s not clear how much of the change can be deemed “improvement” vs. “degradation”. We propose to develop metrics-resolution relationships based on the argument that if a downscaling experiment is to exhibit value, it has to produce more information than what can be simply inferred from information sources already available. Information sources could include both initial and boundary conditions, the coarse resolution model in which the higher-resolution models are embedded, and the set of physical parameterizations. These sources define an “information threshold” as a function of the spatial and temporal resolution. This threshold then serves as a benchmark to quantify the information gain from the downscaling experiments, or any other approaches. For a downscaling experiment to show any value, the information has to be above this threshold.

The notable biases in AR frequency (~0.3 standard deviation in most experiments) and AR zonal IVT (~0.45 standard deviation in most experiments) need to be understood in future studies. The weakest bias in the case of B4_600 suggests higher spatial resolution could be helpful for reducing biases in simulated AR frequency and IVT, but it remains to be investigated using longer term simulations.

Evaluation of climate model performance suggests that simulating the relationship of AR landfalls with the corresponding anomalies in winter precipitation and temperatures is relatively insensitive to resolution. Thus the impact of downscaling is not so evident for ARs – at least based on the level and type of analyses undertaken by us. These findings may be of limited applicability for understanding and quantifying the effects of downscaling for a number of key climate and regional impacts (e.g., flooding, water resources), due to the fact that the performance of model simulations have been evaluated mostly for relatively large scales (e.g. Bukovsky regions) and with only very limited observational resources.

A number of hydrometeorology phenomena directly related to practical applications (i.e., rainfall-snowfall partitioning, runoff, snowpack, soil moisture) are associated with much smaller-scale processes such as the regional moisture transport by the Sierra barrier jet, the portion of terrain above the freezing level, and the shape of individual watersheds. Such fine-scale processes – key to understanding both impacts and associated decision support – are not evaluated in this study because of the fine-scale observations that are needed, but not yet unavailable. Fine-scale observations needed to address these concerns are difficult to acquire from satellite and conventional climate data. Thus, it will be necessary to search for/develop methodologies of applying existing data (e.g., current/new satellite missions, station observations, special observation campaigns) that can more readily examine the fine-scale features afforded by high-resolution (O[1 km]) downscaling. In addition, future experiments need to perform simulations for longer periods. Decade-long simulations cannot capture the effects of key large-scale variability closely related with the climate in WUS, such as ENSO. These shortcomings, as well as the lack of fine-scale and physically comprehensive observations, and the appropriate length of simulations need to be addressed in future experiments and analyses of the type explored in this project.

Specific to Northeast winter storms, future study would benefit from the analysis of in-situ observations to provide finer detail and as an uncertainty constraint on PRISM. Taking a closer view of the highest resolution data, including in situ observations, will determine if high-resolution downscaling is able to provide information on features that are not producible at lower resolutions. This question of whether useful detailed information is gained from high-resolution downscaling was not adequately answered in any of the analyses and would warrant future study. Implicit in this is the careful examination of B4, which was not simulated for a complete decade as of the writing of this report.

It is crucial that future efforts are focused on climate-scale analysis of existing data. This would most beneficially target phenomena, quantities, and processes where it could be hypothesized that high resolution will provide added value over coarser resolution. Potential data would include extremes, mesoscale features and processes, orographically influenced processes, and features with sharp horizontal gradients. If new data were to be created through additional simulations, the best use of time would be to run 30-year hindcasts, which will help obtain more statistically robust climatologies. This could be done with any experiment configuration (nudging, non-nudging, any convective parameterization) as long as these choices are consistent across the simulation suite. The purpose of longer simulations is to focus strongly on whether or not downscaling is beneficial for producing climatologies of high-impact features that require high resolution, such as the examples presented above.

9. Conference Presentations and Papers in Preparation

9.1. Presentations

- Kim, J., B. Guan, D.E. Waliser, R.D. Ferraro, J. Case, T. Iguchi, E. Kemp, W. Putnam, W. Wang, and D. Wu, 2015: Effects of resolution and spectral nudging in simulating the effects of wintertime atmospheric river landfalls in the Western US. 2015 AGU Fall meeting, San Francisco, A11F-0106.
- Kim, J., B. Guan, D.E. Waliser, and R.D. Ferraro, 2016: Effects of atmospheric river landfalls on surface air temperatures in the western US. 2016 AGU Fall meeting, San Francisco, A54D-05.
- Kim, J., B. Guan, D. Waliser, R. Ferraro, 2016: Effects of AR landfalls in the US Pacific coast on the winter precipitation and temperatures in western US. ICRC-CORDEX 2016, Stockholm, Sweden.

- Loikith, P. C., D. E. Waliser, J. Kim, and R. Ferraro, 2015: Evaluation of Northeast United States Winter Storm Impacts in a Suite of High Resolution Downscaled Climate Model Hindcasts. AGU Fall meeting, December 2015, San Francisco, CA.
- Loikith, P. C., D. E. Waliser, J. Kim, and R. Ferraro, 2016: Winter storms in the Northeast United States in a suite of high resolution NASA Unified-WRF regional climate model hindcasts: Evaluation of precipitation-based impact metrics. ICRC-CORDEX 2016, Stockholm, Sweden.
- Matsui, T., T. Iguchi, and W.-K. Tao, 2016: Precipitation Variability and Diurnal Cycle of Convection-Permitting Deterministic Simulations versus Mesoscale Multi-Physics Ensemble Simulations: A Preliminary Result from the NASA Downscaling Project, GEWEX Convection-Permitting Climate Modeling workshop, Boulder, CO, 6-8 September 2016.
- Tian, Y., C. Peters-Lidard, et al., 2015: Quantifying information gain from dynamic downscaling experiments. AGU Fall meeting, December 2015, San Francisco.

9.2. Papers

- Iguchi, T., Y. Tian, C. Peters-Lidard, D. Wu, J. Case, R. Ferraro, B. Guan, E. Kemp, J. Kim, H. Lee, P. Loikith, W. Putman, J. Santanello, W.-K. Tao, B. Tian, D. Waliser, W. Wang, and B. Zavadsky, 2017: Evaluating Information Gain from Convection-Permitting Seasonal Simulations over the Continental United States, in preparation (to be submitted to *Climate Dynamics*).
- Iguchi, T., W.-K. Tao, D. Wu, C. Peters-Lidard, J. Santanello, E. Kemp, Y. Tian, J. Case, W. Wang, R. Ferraro, D. Waliser, J. Kim, H. Lee, B. Guan, B. Tian, P. Loikith, 2016: Sensitivity of CONUS summer rainfall to the selection of cumulus parameterization scheme in NU-WRF climate simulations, *Journal of Hydrometeorology*, in revision.
- Kim, J., B. Guan, D. E. Waliser, R. D. Ferraro, J. Case, T. Iguchi, E. Kemp, W. Putman, W. Wang, D. Wu, and B. Tian, 2016: Assessments of dynamic downscaling for the effects of atmospheric river landfalls on the winter precipitation characteristics in WUS, *Climate Dynamics*, in revision.
- Kim, J., B. guan, D.E. Waliser, R.D. Ferraro, J. Case, T. Iguchi, E. Kemp, W. Putnam, W. Wang, D. Wu, and B. Tian, 2017: Winter precipitation characteristics in Western US related to atmospheric river landfalls: Observations and model evaluations. *Climate Dynamics*, submitted.
- Kim, J., B. Guan, D.E. Waliser, and R.D. Ferraro, 2017: AR landfall effects on the winter-season surface air temperatures in WUS, in preparation.
- Lee, H., D. E. Waliser, T. Iguchi; B. Tian, P. C. Loikith, 2017, Evaluating hourly rainfall characteristics over the US Great Plains in dynamically downscaled climate model simulations using NASA Unified WRF (NU-WRF) , in preparation.
- Loikith, P. C., D. E. Waliser, J. Kim, and R. Ferraro: Evaluation of Cool Season Precipitation Event Characteristics over the Northeast US in a Suite of Downscaled Climate Model Hindcasts, submitted to *Clim. Dyn.*
- Tian, B., H. Lee, D. Waliser, R. Ferraro, J. Kim, J. Case, T. Iguchi, E. Kemp, D. Wu, W. Putman, and W. Wang, 2016: Development and application of performance metrics to assess the downscaled climate simulations of summer mesoscale convective systems over the US Great Plains, *Journal of Hydrometeorology*, submitted.

10. Downscaling Team Members

Jonathan Case ENSCO, Inc/MSFC
Daniel Duffy GSFC
Robert Ferraro JPL
Takamichi Iguchi UMD/GSFC
Eric Kemp SSAI/GSFC
Jinwon Kim UCLA/JPL
Bin Guan UCLA/JPL
William Gutowski* Iowa State Univ.
Huikyo Lee JPL
Tsengdar Lee NASA HQ
Paul Loikith Portland State/JPL
Linda Mearns* NCAR
Christa Peters-Lidard GSFC
Steven Pawson GSFC
William Putman GSFC
Brent Roberts MSFC
Joseph Santanello GSFC
Max Suarez GSFC
Wei-Kuo Tao GSFC
Baijun Tian JPL
Yudong Tian UMD/GSFC
Duane Waliser JPL
Weile Wang CSU/ARC
Di Wu UMD/GSFC
Brad Zavodsky MSFC

* *External Advisors*

11. References

- Bloom, S., L. Takacs, A. DaSilva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, 124, 1256-1271.
- Buchard, V., A. da Silva, P. Colarco, A. Darmenov, R. Govindaraju, and R. Spurr, 2015: Using OMI Aerosol Index and Aerosol Absorption Optical Depth to evaluate the NASA MERRA Aerosol Reanalysis. *Atmos. Chem. Phys.*, 15, 5743–5760, doi:10.5194/acp-15-5743-2015.
- Bukovsky, M.S., 2011: Masks for the Bukovsky regionalization of North America, Regional Integrated Sciences Collective, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO. Downloaded 2016-01-19. [<http://www.narccap.ucar.edu/contrib/bukovsky/>]
- Bosilovich, M. G., and Coauthors, 2015: MERRA-2: Initial Evaluation of the Climate, Technical Report Series on Global Modeling and Data Assimilation, Volume 43, NASA/TM–2015-104606/Vol. 43, 139 pp. [Available online at <http://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/docs/>]
- Case, J. L., W. L. Crosson, S. V. Kumar, W. M. Lapenta, C. D. Peters-Lidard, 2008: Impacts of High-Resolution Land Surface Initialization on Regional Sensible Weather Forecasts from the WRF Model. *J. Hydrometeor.*, 9, 1249–1266.
- Case, J. L., S. V. Kumar, J. Srikishen, G. J. Jedlovec, 2011: Improving Numerical Weather Predictions of Summertime Precipitation over the Southeastern United States through a High-Resolution Initialization of the Surface State. *Wea. Forecasting*, 26, 785–807.

- Chen, F., and J. Dudhia, 2001: Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, 129, 569-585.
- Chen, F. and Coauthors, 2007: Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. *J. Appl. Meteor.*, 46, 694–713.
- Chin, M., R. B. Rood, S.-J. Lin, J.-F. Müller, and A. M. Thompson, 2000a: Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties. *J. Geophys. Res.*, 105 (D20), 24671-24687. <http://dx.doi.org/10.1029/2000JD900384>.
- Chin, M., D. L. Savoie, B. J. Huebert, A. R. Bandy, D. C. Thornton, T. S. Bates, P. K. Quinn, E. S. Saltzman, and W. J. De Bruyn, 2000b: Atmospheric sulfur cycle simulated in the global model GOCART: Comparison with field observations and regional budgets. *J. Geophys. Res.*, 105 (D20), 24689-24712. <http://dx.doi.org/10.1029/2000JD900385>.
- Chou M.-D., and M. J. Suarez, 1999: A solar radiation parameterization for atmospheric studies. NASA Tech. Rep. NASA/TM-1999-10460, vol. 15, 38 pp.
- Chou M.-D., and M. J. Suarez, 2001: A thermal infrared radiation parameterization for atmospheric studies. NASA/TM-2001-104606, vol. 19, 55pp.
- Ek M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, 108, 8851, doi:10.1029/2002JD003296.
- Giorgi, F. & W.J. Gutowski Jr, Regional Dynamical Downscaling and the CORDEX Initiative, *Annual Review of Environment and Resources*, 2015 Vol. 40: 467-490. DOI: 10.1146/annurev-environ-102014-021217,
- Glisan, J. M., W. T. Gutowski Jr., J. J. Cassano, and M. E. Higgins, 2013: Effects of spectral nudging in WRF on Arctic temperature and precipitation simulations. *J. Climate*, 26, 3985-3999.
- Guan, B. and D. E. Waliser (2015), Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies, in review at *J. Geophys. Res. Atmos.*
- Gula, J., and W. R. Peltier, 2012: Dynamical downscaling over the Great Lakes basin of North America using the WRF regional climate model: The impact of the Great Lakes system on regional greenhouse warming. *J. Climate*, 25, 7723-7742.
- Hall A, 2014: Projecting regional change. *Science*, 346 (6216): 1461–1462. DOI: 10.1126/science.aaa0629
- Hewitt, C.D and Griggs, D.J., 2004: Ensembles-based predictions of climate changes and their impacts. Published article appears in *Eos*, 85, 566.
- Hirsch, A., Jatin Kala, Andy J. Pitman, Claire Carouge, Jason P. Evans, Vanessa Haverd, and David Mocko, 2014: Impact of Land Surface Initialization Approach on Subseasonal Forecast Skill: A Regional Analysis in the Southern Hemisphere. *J. Hydrometeorol*, 15, 300–319.
- Janjić, Z. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.*, 118, 1429–1443.
- Janjić, Z. I., 1996: The surface layer in the NCEP Eta Model. Preprints, Eleventh Conference on Numerical Weather Prediction, Norfolk, VA, Amer. Meteor. Soc., 354–355.
- Janjić, 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., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levitt, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, 25, 1536-1542.
- Kain, J. S., et al., 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, 23, 931-952.

- Kendon EJ, Roberts NM, Fowler HJ, Roberts MJ, Chan SC, and Senior CA, 2014: Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nat Clim Change* 4: 570-576. doi: 10.1038/nclimate2258.
- Kumar, S. V., and Coauthors, 2006: Land Information System - An Interoperable Framework for High Resolution Land Surface Modeling. *Environmental Modelling & Software*, 21, 1402-1415.
- _____, C. D. Peters-Lidard, J. L. Eastman, and W.-K. Tao, 2007: An integrated high-resolution hydrometeorological modeling testbed using LIS and WRF. *Environmental Modeling & Software*, 23 (2), 169-181, doi: 10.1016/j.envsoft.2007.05.012.
- _____, R.H. Reichle, C.D. Peters-Lidard, R.D. Koster, X. Zhan, W.T. Crow, J.B. Eylander, and P. R. Houser, 2008: A Land Surface Data Assimilation Framework using the Land Information System: Description and Applications, *Advances in Water Resources*, 31, 1419-1432, DOI:10.1016/j.advwatres.2008.01.013
- Lang, S., W. K. Tao, J. Simpson, and B. Ferrier, 2003: Modeling of convective-stratiform precipitation processes: Sensitivity to partitioning methods. *Journal of Applied Meteorology*, 42, 505-527.
- Lang, S., W. K. Tao, J. Simpson, R. Cifelli, S. Rutledge, W. Olson, and J. Halverson, 2007: Improving simulations of convective systems from TRMM LBA: Easterly and westerly regimes. *Journal of the atmospheric sciences*, 64, 1141-1164.
- Lawston, P. M., J. A. Santanello, B. Zaitchik, and M. Rodell, 2015: Impact of irrigation methods on land surface model spinup and initialization of WRF forecasts. *J. Hydrometeor.*, 16, 1135–1154.
- Lee H, Kim J, Waliser DE, Loikith PC, Mattmann CA, and McGinnis S, 2015: Using joint probability distribution functions to evaluate simulations of precipitation, cloud fraction and insolation in the North America Regional Climate Change Assessment Program (NARCCAP). *Clim Dynam* 45: 309-323. doi: DOI 10.1007/s00382-014-2253-y.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteorol.*, 22, 1065-1092.
- Lo, J. C.-F., Z.-L. Yang, and R. A. Pielke Sr. (2008), Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model. *J. Geophys. Res.*, 113, D09112, doi:10.1029/2007JD009216.
- Matsui, T., and Coauthors, 2014: Introducing multisensor satellite radiance-based evaluation for regional Earth System modeling. *J. Geophys. Res.: Atmos.*, 119, 8450-8475.
- McCumber, M., W.-K. Tao, J. Simpson, R. Penc, and S.-T. Soong, 1991: Comparison of ice-phase microphysical parameterization schemes using numerical simulations of tropical convection. *Journal of Applied Meteorology*, 30, 985-1004.
- Mearns, L. O., W. J. Gutowski, R. Jones, L.-Y. Leung, S. McGinnis, A. M. B. Nunes, and Y. Qian: A regional climate change assessment program for North America. *EOS*, Vol. 90, No. 36, 8 September 2009, pp. 311-312.
- Miguez-Macho, G., G. L. Stenchikov, and A. Robock, 2004: Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *J. Geophys. Res.*, 109, D13104, doi:10.1029/2003JD004495.
- Miguez-Macho, G., G. L. Stenchikov, and A. Robock, 2005: Regional climate simulations over North America: Interaction of local processes with improved large-scale flow. *J. Climate*, 18, 1227–1246.
- Moorthi, S. and M. J. Suarez, 1992: Relaxed Arakawa-Schubert. A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, 120, 978–1002.
- Murphy, A.H., B.G. Brown and Y.-S. Chen, 1989: Diagnostic verification of temperature forecasts. *Wea. Forecat.*, 4, 485-501.

- Oreopoulos, L., and Coauthors, 2012: The continual intercomparison of radiation codes: Results from Phase I. *J. Geophys. Res.: Atmos.*, 117, D06118, doi:10.1029/2011JD016821.
- Perkins, S., A. Pitman, N. Holbrook, and J. McAneney, 2008: Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J. Climate*, 20, 4356-4376.
- Peters-Lidard, C.D., and Coauthors, 2007: High-performance earth system modeling with NASA/GSFC's Land Information System. *Innovations in Systems and Software Engineering*, 3(3), 157-165.
- Peters-Lidard, C.D., and Coauthors, 2015: Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales. *Environmental Modelling & Software*, 67, 149-159.
- PielkeSr., R. A. and R. L. Wilby, (2012), Regional climate downscaling: What's the point?, *Eos Trans. AGU*, 93(5), 52.
- Qiao, F., and X.-H. Liang, 2015: Effects of cumulus parameterizations on predictions of summer flood in the Central United States. *Climate Dynamics*, 45, 727-744, doi:10.1007/s00382-014-2301-7.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, 24, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.
- Rienecker and Coauthors, 2011: MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, 24, 3624-3648, doi:10.1175/JCLI-D-11-00015.1.
- Rutledge, S. A., and P. V. Hobbs, 1984: The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cold-frontal rainbands. *J. Atmos. Sci.*, 41, 2949-2972.
- Santanello, J. A., S. V. Kumar, C. D. Peters-Lidard, K. Harrison, and S. Zhou, 2013: Impact of land model calibration on coupled land-atmosphere prediction. *J. Hydrometeor.*, 14, 1373-1400.
- Silva-Dias, P. L., and W. H. Schubert, 1977: Experiments with a spectral cumulus parameterization theory. *Atmos. Sci. Pap.*, No. 275, Colorado State University, 132 pp.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3, NCAR Technical Note, NCAR/TN-475+STR, 123 pp. [Available on-line at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf]
- Takacs, L. L., M. Suarez, and R. Todling, 2015. Maintaining Atmospheric Mass and Water Balance Within Reanalysis. NASA/TM-2014-104606, Vol. 37. [Available online at <http://gmao.gsfc.nasa.gov/pubs/docs/Takacs737.pdf>]
- Tao, W.-K., and J. Simpson, 1993: The Goddard cumulus ensemble model. Part I: Model description. *Terr. Atmos. Oceanic Sci.*, 4, 35-72.
- Tao, W.-K., J. Simpson, and M. McCumber, 1989: An ice-water saturation adjustment. *Mon. Wea. Rev.* 117, 231-235.
- Tao, W. K., and Coauthors, 2003: Microphysics, radiation and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteorol. Atmos. Physics*, 82, 97-137.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106(D7), 7183-7192, doi:10.1029/2000jd900719.
- Wen, X., S. Lu, J. Jin, 2012: Integrating remote sensing data with WRF for improved simulations of oasis effects on local weather processes over an arid region in northwestern China. *J. Hydrometeor.*, 13, 573-587.

| REPORT DOCUMENTATION PAGE | | | Form Approved OMB No. 0704-0188 | | |
|---|-------------|---|---|------------------------------|---|
| <p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p> | | | | | |
| 1. REPORT DATE (DD-MM-YYYY) 01/02/2017 | | 2. REPORT TYPE Technical Publication | | 3. DATES COVERED (From - To) | |
| 4. TITLE AND SUBTITLE NASA Downscaling Project Final Report | | | 5a. CONTRACT NUMBER NNN12AA01C | | |
| | | | 5b. GRANT NUMBER | | |
| | | | 5c. PROGRAM ELEMENT NUMBER | | |
| 6. AUTHOR(S) Ferraro, Robert D Waliser, Duane E Peters-Lidard, Christa | | | 5d. PROJECT NUMBER | | |
| | | | 5e. TASK NUMBER | | |
| | | | 5f. WORK UNIT NUMBER 929099.03.03.01.23 | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Jet Propulsion Laboratory California Institute of Technology 4800 Oak Grove Drive Pasadena, CA 91009 | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001 | | | 10. SPONSORING/MONITOR'S ACRONYM(S) NASA | | |
| | | | 11. SPONSORING/MONITORING REPORT NUMBER NASA/TP-2017-219579 | | |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified—Unlimited | | | | | |
| Subject Category 47 Meteorology and Climatology | | | | | |
| Availability: NASA CASI (757) 864-9658 Distribution: Nonstandard | | | | | |
| 13. SUPPLEMENTARY NOTES | | | | | |
| 14. ABSTRACT A team of researchers from NASA Ames Research Center, Goddard Space Flight Center, the Jet Propulsion Laboratory, and Marshall Space Flight Center, along with university partners at UCLA, conducted an investigation to explore whether downscaling coarse resolution global climate model (GCM) predictions might provide valid insights into the regional impacts sought by decision makers. Since the computational cost of running global models at high spatial resolution for any useful climate scale period is prohibitive, the hope for downscaling is that a coarse resolution GCM provides sufficiently accurate synoptic scale information for a regional climate model (RCM) to accurately develop fine scale features that represent the regional impacts of a changing climate. As a proxy for a prognostic climate forecast model, and so that ground truth in the form of satellite and in-situ observations could be used for evaluation, the MERRA and MERRA-2 reanalyses were used to drive the NU-WRF regional climate model and a GEOS-5 replay. This was performed at various resolutions that were at factors of 2 to 10 higher than the reanalysis forcing. A number of experiments were conducted that varied resolution, model parameterizations, and intermediate scale nudging, for simulations over the continental US during the period from 2000–2010. The results of these experiments were compared to observational datasets to evaluate the output. | | | | | |
| 15. SUBJECT TERMS Dynamical Downscaling, Model Evaluation, NU-WRF, GEOS5 | | | | | |
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT UU | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON HQ-STI-INFODESK at hq-sti- infodesk@mail.nasa.gov |
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | |

| | | | | | |
|---|---|---|--|--|---|
| U | U | U | | | 19b. TELEPHONE NUMBER (Include area code) 757-864-9658 |
|---|---|---|--|--|---|

JPL 2659 R 8 / 13

Standard Form 298 (Rev. 8-98)

Prescribed by ANSI Std. Z39-18

NASA Supplementary Instructions To Complete SF 298 (Rev. 8-98 version)

NASA uses this inter-governmental form that does not allow customization. Look for special notes (NOTE) if NASA's procedures differ slightly from other agencies.

- Block 1 NOTE: NASA uses month and year (February 2013) on the covers and title pages of its documents. However, this OMB form is coded for block 1 to accept data in the following format: day, month, and year (ex.: day (23), month (02), year (2013) or 23-02-2013, which means February 23, 2013. For this block, use the actual date of publication (on the cover and title page) and add 01 for the day. Example is March 2013 on the cover and title page, and 01-03-13 for block 1.
- Block 2: Technical Paper, Technical Memorandum, etc.
- Block 3: Optional for NASA
- Block 4: Insert title and subtitle (if applicable)
- Block 5a: Complete if have the information
- b: Complete if have the information
- c: Optional for NASA
- d: Optional for NASA; if have a cooperative agreement number, insert it here
- e: Optional for NASA
- f: Required. Use funding number (WU, RTOP, or UPN)
- Block 6: Complete (ex.: Smith, John J. and Brown, William R.)
- Block 7: NASA Center (ex.: NASA Langley Research Center)
City, State, Zip code (ex.: Hampton, Virginia 23681-2199)
You can also enter contractor's or grantee's organization name here, below your NASA center, if they are the performing organization for your center
- Block 8: Center tracking number (ex.: L-17689)
- Block 9: National Aeronautics and Space Administration
Washington, DC 20546-0001
- Block 10: NASA
- Block 11: ex.: NASA/TM-2013-123456
- Block 12: ex.:
Unclassified – Unlimited
Subject Category <http://www.sti.nasa.gov/sscg/subcat.html>
Availability: NASA STI (757) 864-9658
Distribution: (Standard or Nonstandard)
If restricted/limited, also put restriction/limitation on cover and title page
- Block 13: (ex.: Smith and Brown, Langley Research Center. An electronic version can be found at [http:// _____](http://_____), etc.)
- Block 14: Self-explanatory
- Block 15: Use terms from the NASA Thesaurus <http://www.sti.nasa.gov/sti-tools/#thesaurus>,
Subject Division and Categories Fact Sheet <http://www.sti.nasa.gov/sscg/subcat.html>,
or Machine-Aided Indexing tool <http://mai.larc.nasa.gov/>
- Block 16a,b,c: Complete all three
- Block 17: UU (unclassified/unlimited) or SAR (same as report)
- Block 18: Self-explanatory
- Block 19a: STI Information Desk at email: HQ-STI-INFODESK at hq-sti-infodesk@mail.nasa.gov