Establishing a Numerical Modeling Framework for Hydrologic Engineering Analyses of Extreme Storm Events

Xiaodong Chen, S.M.ASCE[1]; Faisal Hossain, A.M.ASCE[2]; and L. Ruby Leung[3]

Abstract: In this study, a numerical modeling framework for simulating extreme storm events was established using the weather research and forecasting (WRF) model. Such a framework is necessary for the derivation of engineering parameters such as probable maximum precipitation that are the cornerstone of large water-management infrastructure design. Here, this framework was built based on a heavy storm that occurred in Nashville, Tennessee (USA), in 2010, and verified using two other extreme storms. To achieve the optimal setup, several combinations of model resolutions, initial/boundary conditions (IC/BC), cloud microphysics, and cumulus parameterization schemes were evaluated using multiple metrics of precipitation characteristics. The evaluation suggests that WRF is most sensitive to the IC/BC option. Simulation generally benefits from finer resolutions up to 5 km. At the 15 km level, NCEP2 IC/BC produces better results, whereas NAM IC/BC performs best at the 5 km level. The recommended model configuration from this study is: NAM or NCEP2 IC/BC (depending on data availability), 15 km or 15 km–5 km nested grids, Morrison microphysics, and Kain-Fritsch cumulus schemes. Validation of the optimal framework suggests that these options are good starting choices for modeling extreme events similar to the test cases. This optimal framework is proposed in response to emerging engineering demands of extreme storm event forecasting and analyses for design, operations, and risk assessment of large water infrastructures. **DOI: 10.1061/(ASCE)HE.1943-5584.0001523.**

Author keywords: Precipitation; Heavy storms; Nashville; Atmospheric model; Parameterizations.

Introduction

Intense storms, or extreme rainfall events, as they will be called hereafter, pose challenges to infrastructure management and design and trigger other catastrophic events such as floods, landslides, and dam failures (Casagli et al. 2006; Cong et al. 2006; Evans et al. 2000). They are also the cornerstone of engineering design and risk assessment of large infrastructures such as dams, levees, and power plants (Stratz and Hossain 2014). Therefore, it is of great societal interest to physically predict and understand the occurrence and magnitude of such extreme events for both design and operation of engineering infrastructures.

In current engineering practice, the safety of hazardous infrastructures (where lives are at stake with infrastructure failure) is achieved through designs based on probable maximum precipitation (PMP). PMP is defined as the theoretical greatest depth of precipitation for a given duration that is physically possible over a particular drainage area (Huschke 1959). It depicts the precipitation potential of an already intense storm that is maximized to an upper bound using some basic engineering assumptions (Kunkel et al. 2013; Stratz and Hossain 2014). The National Oceanic and Atmospheric Administration (NOAA) has created a database of such intense storms in the United States from approximately 1900–1990 that were maximized to PMPs and publicly released as hydrometeorological reports (HMRs) for the engineering infrastructure community (U.S. Department of Commerce 1999). For engineering practices outside the United States, the World Meteorological Organization (WMO) has outlined several approaches that can be used (WMO 1986). In general, these are local methods (maximization of local storms), transposition methods (storm transposition from same climatological regions), generalized methods (based on some provided PMP distribution maps), and statistical methods such as the one proposed in Hershfield (1965).

PMP is generally expressed mathematically as: $P \times w_{p(\mathrm{maximum})}/w_{p(\mathrm{storm})}$, where P = the observed rainfall accumulation; $w_{p(\mathrm{maximum})}$ = the highest observed precipitable water from historical records; and $w_{p(\mathrm{storm})}$ = the storm precipitable water. This approach is often criticized as being insufficiently physical because it assumes a linear relationship between precipitation and the water-holding capacity of the atmosphere (Abbs 1999; Kunkel et al. 2013). Also, it heavily relies on historical observation data. For very early extreme events used in PMP analysis (such as Storm Elba of 1929), it is difficult to obtain a physically consistent picture because of limitations of record keeping and the linearity assumption (Abbs 1999). In this context, numerical simulation of extreme storms and their consequent physical maximization to a PMP is gaining much more traction among science and engineering communities than before (Kunkel et al. 2013; Stratz and Hossain 2014).

The numerical modeling approach has several key advantages over the traditional approaches. It is able to produce finer details on the spatial-temporal structure of the storms using fewer assumptions and experience-based estimation. It is more tailored to a region that has little or no long-term rainfall record or is rapidly

[1]Graduate Student, Dept. of Civil and Environmental Engineering, Univ. of Washington, More Hall 201, Seattle, WA 98195.

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Washington, More Hall 201, Seattle, WA 98195 (corresponding author). E-mail: fhossain@uw.edu

[3]Laboratory Fellow, Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, 902 Battelle Boulevard, Richland, WA 99352.

Note. This manuscript was submitted on June 4, 2016; approved on January 19, 2017No Epub Date. Discussion period open until 0, 0; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

undergoing changes in weather patterns due to land cover change or global warming. More importantly, a well-established numerical modeling framework is often able to handle various extreme events within the model domain spanning decades (Chen and Hossain 2016). In the study by Tan (2010), the WRF model was calibrated and set up over the American River basin. It was found capable of simulating various PMP-class storms in the basin during 1970–2000. The model also provided better space-time pictures of the historical events that were used in HMRs for PMP estimation in this basin. This is another benefit of the numerical modeling approach.

There have been numerous studies on extreme events or PMPs using numerical atmospheric models. Some conclusions have been reached on the optimal setup of numerical models. For example, optimal grid size ratios of 1:7, 1:5, and 1:3 were validated over eight storms in southwest England (Liu et al. 2012). The study by Pennelly et al. (2014) concluded that for storms in Alberta, Canada, 6-km grids in the WRF model are a balance between simulation quality and time expense. There have also been efforts to optimize the simulated rainfall results by operating models with more information. For example, Giannaros et al. (2016) assimilated lightning data into the atmospheric numerical simulation, and it helped improve precipitation forecast. However, a consistent framework informing the users from the hydrologic engineering community how to systematically set up and analyze numerical models for engineering analyses is still absent from the literature.

Previous studies suggest that the performance of storm simulation heavily depends on the parameterization schemes, which is the mathematical identification of physical processes in the numerical models (Stensrud 2009). Though a wise choice of parameterization schemes results in improved simulations of big storms, it often has to be achieved by trial and error. For example, several numerical studies for the Mumbai July 2005 storm (Chang et al. 2009; Kumar et al. 2008; Rao et al. 2007; Vaidya and Kulkarni 2007) show steady progress in reconstructing the high precipitation values in the various modeling platforms with different parameterization schemes. Rajeevan et al. (2010) revealed that the optimal combinations of parameterization schemes and IC/BC in the model can be quite different for southeast Indian thunderstorms. These high heterogeneities within optimal model configurations make it difficult for engineering communities to set up and operate these models.

Given that the engineering community is relatively new to the setup and operation of numerical models, as well as the use of models for maximization of extreme storms in PMP estimation, a framework to explore the role of various parametrizations and IC/BC on extreme storm simulation accuracy can provide a baseline for optimal criteria for PMP simulation. Such a comprehensive study will also illustrate ways to identify optimal model configurations for extreme storm simulations and help the engineering infrastructure community that engages in hydrologic analyses for design and operations embrace numerical models for PMP estimation and further advance the methodology.

In this study, ways to establish a generic numerical modeling framework over a given area are investigated. Taking the Nashville, Tennessee, USA, 2010 storm as a test case, the procedures required to achieve a good storm reconstruction using the WRF model are illustrated. Various combinations of parameterization schemes, IC/BCs, and grid sizes are evaluated. Using this framework, three questions are addressed:

1. What combinations of model options in WRF are most skillful for extreme storm event simulation?
2. What are the strengths and weaknesses of each model option in reference to simulation accuracy of extreme precipitation?
3. What are the optimal model configurations for engineering operations and infrastructure implications?

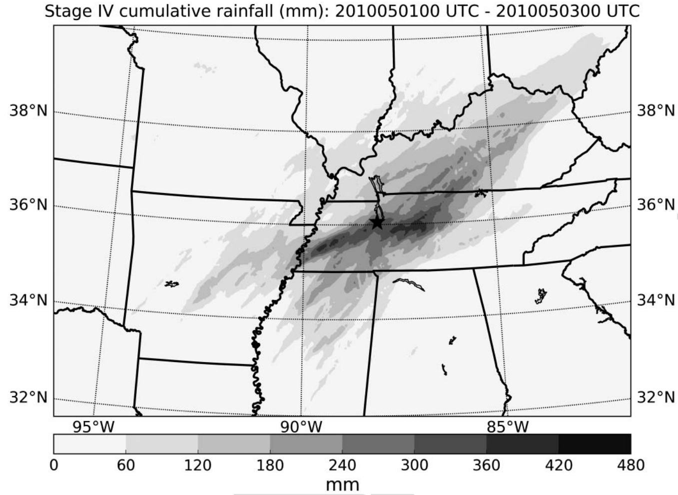


Fig. 1. 48-h (0000 UTC 1 May–0000 UTC 3 May, 2010) total rainfall from Stage IV data

Nashville, USA 2010 Extreme Storm

During May 1 and May 2, 2010, the west and middle Tennessee region of the USA experienced a record-breaking storm. This 2-day rainfall event brought huge amounts of water to western Tennessee, with 48-h cumulative rainfall exceeding historical records at several gauge stations (such as the Nashville and Camden station in Tennessee). Fig. 1 shows the 48-h cumulative rainfall from this storm as observed from the NEXRAD network, which shows a southwest–northeast pattern.

This storm, hereafter referred to as the Nashville 2010 storm, led to a flood in the following days that NOAA categorized as a 1000-year return period flood event (NOAA National Weather Service and Weather Forecast Office, NWSWF 2010). The maximum 48-h total precipitation observed was 493 mm (19.41 in.) at the Camden COOP station (36.05°N, 88.08°W, the star in Fig. 1). This value is quite close to the 5,000 mi^2 48-h design PMP (495 mm, or 19.5 in.) for west Tennessee (an area in the HMR 1951 report, hereafter referred to as the HMR51 region). Nashville's international airport recorded its first and third highest 24-h total rainfall in history on 1 and 2 May, respectively (NWSWF 2010). These statistics qualify this rainfall event as a reference extreme storm for PMP design for the HMR 51 region. During the ensuing flood event, 21 deaths were reported, and over 30 counties were declared major disaster areas by the government. This unique rainfall record and infrastructure-damaging impact make this event worth revisiting with numerical simulation (Durkee et al. 2012). There have not been many numerical simulation efforts on this storm. Thus, a successful model reconstruction of this event would provide an important baseline for studying other local events or events in similar environmental conditions for engineering infrastructure applications.

The Nashville 2010 storm was among a series of big storms (tornados 41, 43, and 57) hitting the mid-southern United States in the same period. Analysis of reanalysis products suggests that the event was associated with a synoptic system with significant atmospheric moisture. The Atlantic ridging associated with the negative phase of the North Atlantic Oscillation (NAO) helped amplify and slow the eastward propagating synoptic wave pattern that generated heavy precipitation from mesoscale organized convective systems (Durkee et al. 2012). An atmospheric river originating from the Intertropical Convergence Zone (ITCZ) in Central America provided the moisture source for this record-breaking event (Durkee et al. 2012). Surface topography in the Appalachians provided orographic forcing for moisture convergence, and land surface heating helped maintain atmospheric instability, so precipitation continued until 2 May, 2010. Previous studies have identified several key atmospheric factors such as the superposition of the polar and subtropical jet (Winters and Martin 2014) and the atmospheric river (Durkee et al. 2012; Moore et al. 2012). Because some elements present in the Nashville 2010 event are common ingredients in other extreme storms, reconstructing this extreme event may serve as an important test case for evaluating the ability of the WRF model for simulating other storms.

The Numerical Atmospheric Model

The WRF model is employed for big storm reconstruction. WRF is an atmospheric modeling system (Skamarock et al. 2008) that features two nonhydrostatic solvers, the advanced research WRF (ARW) core for atmospheric research, and the nonhydrostatic nesoscale model (NMM) core for the operational forecast. This study adopted WRF-ARW v3.6.1 for the storm simulation. WRF-ARW has been employed in various big storm studies and demonstrated to be capable of simulating several big storms across the world (Chen and Hossain 2016; Kumar et al. 2008; Rajeevan et al. 2010; Tan 2010).

WRF-ARW is designed for mesoscale meteorological simulation with spatial resolution ranging from 1 to 100 km. Accordingly, the time step used in the model varies from seconds to minutes. It simulates atmospheric motion using compressible, nonhydrostatic Euler equations with consideration of mass, energy, and momentum conservation. These equations are formulated and solved using the Arakawa-C grid with terrain-following mass vertical coordinates (Laprise 1992). WRF-ARW uses various parameterization schemes to estimate the atmospheric processes at the subgrid scale, and atmospheric moisture is considered in various phases in the cloud microphysics parameterization schemes. For example, in the Morrison microphysics scheme, water is considered in vapor, cloud droplets, cloud ice, rain, snow, and graupel or hail phases (Morrison et al. 2009). This ensures an accurate description of moisture in the air. By default, the WRF-ARW model uses the USGS or MODIS land use dataset to depict the surface feedback. As a platform, the WRF-ARW model provides multiple choices for major physics processes that affect the atmospheric state: cloud microphysics, cumulus processes, radiation processes, planetary boundary layer processes, and land surface processes. With this modular design, it exhibits great flexibility for mesoscale atmospheric activities across a wide range of temporal and spatial scales while maintaining the capability of incorporating recent advances in atmospheric sciences.

Experimental Design

Previous studies suggest that the performance of numerical atmospheric models is mostly affected by cloud physics parameterization, model resolution, and initial and boundary conditions in the model, as well as the simulation period. The subsequent steps illustrate the workflow needed by engineers to establish the optimal modeling framework based on WRF. A schematic is shown in Fig. 2, and the details of each step are explained subsequently with an example of the Nashville 2010 storm simulation.

1. Study previous modeling efforts to understand the background of the study domain;
2. Determine the atmospheric numerical model(s) of interest;
3. Determine the study domain and simulation period. Prioritize the main physical factors in the model that affect the simulation quality. This can be gained from step 1. Outline the model options (i.e., combination of parameterizations) to be tested;
4. Collect the input data, set up the model, and make model runs;
5. Determine the main purpose of the modeling framework and the evaluation criteria. As shown subsequently in the Nashville

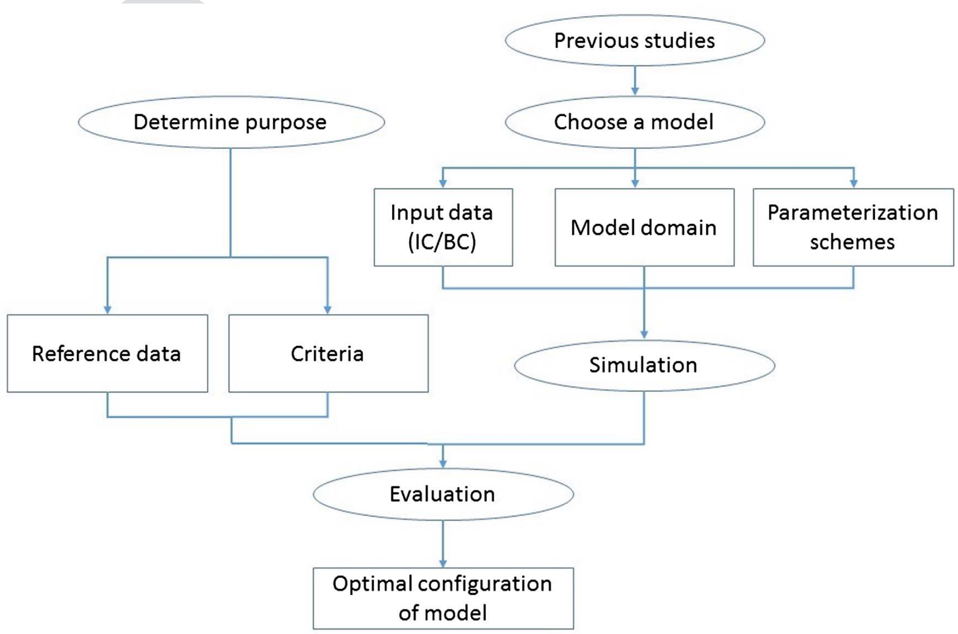
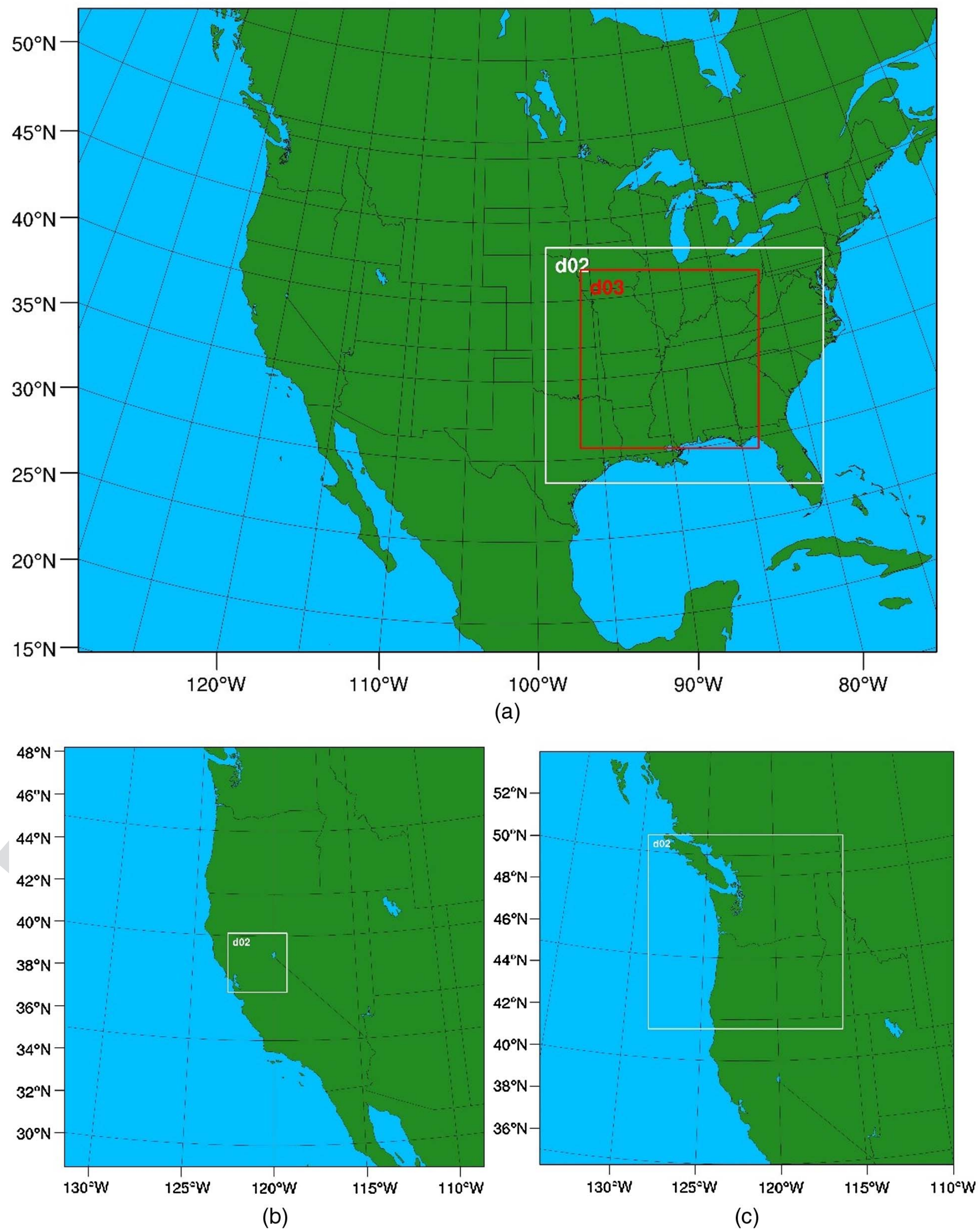


Fig. 2. Generic framework for exploring optimal model configuration for extreme storms

2010 case, different purposes of the modeling framework require different criteria and lead to different configurations in the optimal atmospheric model. Collect the reference data;
6. Evaluate the simulation results using the metric(s) that best serve the purpose.

The Nashville 2010 storm period is May 1–2, 2010, and previous studies (Mahoney 2013) concluded that a long spin-up would result in less rainfall during the event. Thus, the simulation period is chosen as 0000 UTC 1–3 May 2010. Here, three configurations of nested domains were tested to evaluate model performance at 15-, 5-, and 1.6-km (the latter is referred to as "2 km" for convenience) grid sizes. Fig. 3 shows the domains in the simulation of the Nashville 2010 storm and two verification events (which will be discussed later). The three nested domains in Fig. 3(a) are all centered over western Tennessee. In the first configuration [g15, the outmost domain in Fig. 3(a)], the domain covers the contiguous US at 15-km grid spacing. In the second configuration [g5, the whole domain plus the white box in Fig. 3(a)], a d02 domain at 5-km resolution (white box) is nested inside the larger 15-km domain. The third configuration [g2, Fig. 3(a)] further includes a d03 domain of 1.6-km spatial resolution (red box) to better resolve convection at 1.6-km grid spacing. When there is more than one domain involved in the simulation, WRF runs in a two-way nesting mode, which means the coarse grid results are updated using results in finer grids where available. This experiment design allows evaluation of the impacts of higher resolution achieved through nesting with the same placement of the outermost lateral boundaries for all simulations. Nominal time steps of 60, 20, and 6.7 s were used for



F3:1 **Fig. 3.** Spatial domain in modeling framework of Nashville 2010 storm, 1997CA event and 1980PNW event the 15-, 5-, and 1.6-km grids, respectively. Model outputs are archived hourly between 0000 UTC 1 May 2010 and 0000 UTC 3 May 2010, similar to Moore et al. (2012).

Three sources of data were used to generate IC/BCs: (1) NCEP/DOE reanalysis product (NCEP2) at 2.5-degree resolution; (2) NCEP/NCAR reanalysis product (NNRP) at T62 (209-km) resolution; and (3) North America nesoscale (NAM) forecast output at T221 (32-km) resolution. For this study, the NAM forecast initialized at 0000 UTC 1 May 2010 was used.

Previous studies suggest that precipitation simulation is more sensitive to microphysics and cumulus parameterization schemes than parameterizations for other processes in the model (Del Genio et al. 2005; Pennelly et al. 2014; Zhang and McFarlane 1995). Here, three microphysics parameterization schemes for mixed phase clouds were tested, including (1) Morrison double moment scheme (coded as "Morrison" here), (2) New Thompson scheme ("Thompson"), and (3) WSM-5 scheme ("WSM5"). Three cumulus parameterization schemes were also evaluated, including the (1) Kain-Fritsch scheme (coded as "KF" here), (2) Grell-Devenyi scheme ("GD"), and (3) Grell-Freitas scheme ("GF"). In the nested runs (g5 and g2), the cumulus scheme is used only in the 15-km domain because convection is explicitly resolved at the 5- and 2-km resolutions. Grell and Freitas (2014) noted that at coarser resolutions, the GF scheme functions as a cumulus parameterization to represent the unresolved deep convection, but at this solution of a few kilometers, deep convection is explicitly resolved and the GF scheme mainly represents shallow convection. Thus, another set of simulations are designed to test the scale-aware GF scheme in which the GF cumulus scheme is applied to all the domains (15 km, 5 km, 2 km) in the nested runs (g5 and g2). Other schemes are fixed in all the experiments, and they are: RRTM long wave radiation scheme, Dudhia shortwave radiation scheme, revised MM5 surface layer scheme, Yonsei University (YSU) planetary boundary layer scheme, Noah land surface scheme.

The total number of combinations of the different options in grid sizes (3), IC/BCs (3), microphysics schemes (3), and cumulus schemes (3 in the g15 runs, 2 in the g5 and g2 runs) amounts to 63 WRF runs designed and conducted for this study.

Table 1. Binary Results Indices for Spatial Coverage Evaluation Metric

	Observed		
Simulated	Yes	No	Sum
Yes	Hits (YY)	False alarms (YN)	$YY + YN$
No	Misses (NY)	Correct rejections (NN)	$NY + NN$
Sum	$YY + NY$	$YN + NN$	Total $= YY + YN + NY + NN$

Framework Evaluation Metrics

Independent precipitation observation data are required for the assessment of the storm simulations. One option is gauge data because they provide the most accurate estimate of rainfall amount and duration. In some cases, gauge data may not be available due to either the age of the storm or the gauges having stopped working (such as the Nashville international airport station in the Nashville 2010 storm event); the gridded data can be used to validate model results. Here, the NEXRAD Stage IV precipitation dataset (Fig. 1) is used as the reference in selecting the optimal model configuration, given its high accuracy and good spatial coverage. Cumulative 48-h rainfall is evaluated by the spatial correlation coefficient between the simulated and Stage IV 48-h total rainfall. This reveals how the model performs in capturing the rainy area and the spatial heterogeneity of total rainfall. For extreme rainfall events used in engineering analysis, it is important that the numerical model capture the core precipitating areas as accurately as possible. The validation steps used the Livneh daily CONUS near-surface gridded meteorological data (Livneh et al. 2013). This dataset is developed from gauge observations, and it provides an estimation of daily precipitation. By validating the results using a different reference, reference-dependent conclusions can be avoided.

Additional metrics employed include: probability of detection (POD), false alert ratio (FAR), frequency bias (Bias), Heidke skill score (HSS), critical success index (CSI, or TS), and Gilbert skill score (GSS, or ETS). They are defined as statistics of the binary result indices in Table 1. Table 2 shows the definitions of these metrics, as well as the ranges of their values. These metrics measure only the accuracy in the coverage of the rainy versus nonrainy area. Therefore, when the magnitude of precipitated water matters a lot, it would be better to use the correlation or root mean square error (RMSE) between observed rainfall and simulated rainfall for the period of interest (e.g., 6, 24, 48, and 72 h in the PMP design).

Table 2. Definition of Evaluation Metrics on Storm Performance in Spatial Coverage Using Metrics from Table 1

Metric	Definition	Best score	Worst score
POD	$\frac{YY}{YY+NY}$	1	0
FAR	$\frac{YN}{YY+YN}$	0	1
Bias	$\frac{YY+YN}{YY+NY}$	1	0 or ∞
HSS	$\frac{2\times(YY\cdot NN-YN\cdot NY)}{(YY+NY)(NY+NN)+(YY+YN)(YN+NN)}$	1	$-\infty$
TS	$\frac{YY}{YY+NY+YN}$	1	0
ETS	$\frac{YY-YY_{\mathrm{rand}}}{YY+NY+YN-YY_{\mathrm{rand}}}$, where $YY_{\mathrm{rand}}=\frac{(YY+YN)(YY+NY)}{\mathrm{Total}}$	1	$-1/3$

Note: YY (hits) means both simulation and observation indicate rainfall at the grid or station; YN (false alarm) means only simulation indicates rainfall at the grid/station; NY (misses) means only observation indicates rainfall at the grid or station; NN (correct rejection) means neither observation nor simulation indicates rainfall at the grid or station. More details are in Table 1. All these metrics are monotonous, except for bias.

This can be done using either station data or gridded data. Other terms worth considering are the storm duration (start time and end time) and peak rainfall (to classify the storm severity). The Nash-Sutcliffe model efficiency coefficient (NS) is also used to quantify the simulated precipitation. When applied to a map, this coefficient can be defined by Eq. (1), where N = the total number of grid points in the map; P_o = the observed precipitation; and P_m is the simulated precipitation. The range of NS is from $-\infty$ to 1, and 1 is a perfect score. A higher NS indicates stronger capacity of the model. These metrics quantitatively evaluate the model performance; thus, the recommendations given by these metrics can be applied to engineering practice with confidence (Bennett et al. 2013)

$$NS = 1 - \frac{\sum_{n=1}^{N} (P_o^n - P_m^n)^2}{\sum_{n=1}^{N} (P_o^n - \overline{P_o})^2} \quad (1)$$

These metrics measure different aspects of model performance and provide different recommendations for the best combination of parameterizations to support different applications. The POD metric and storm duration are more useful if the successful forecast of the rainy area is more important, such as the search for possible shelter areas. The FAR metric should be weighted more if the cost of emergency relocation is high, in which case unnecessary effort related to areas that are actually not rainy should be avoided. In the infrastructure design practice, the total amount of rainfall and peak rainfall would be more important. If simulated rainfall data is being used as input to other models (such as hydrological models for stream flow forecasting), then a high spatial correlation or Nash-Sutcliffe coefficient between simulated and observed rainfall would be more desired.

This paper takes multiple metrics into consideration as a set when assessing model performance because no single metric captures all the pertinent performance features. For example, a good numerical model configuration should produce a high probability of detection for rain as well as a high critical success index, but a low false alert ratio. Several metrics were combined and a unified score (US) created. The US is defined by Eq. (2), in which POD_n, FAR_n and CSI_n are normalized metrics defined by Eqs. (3)–(5). By combining different aspects of model performance into the score, the unified score is used to identify the best combinations for the overall performance reflected by the multidimensional metrics that appeal to the engineering infrastructure community (Sikder and Hossain 2016)

$$\mathrm{US} = \mathrm{POD}_n^2 - \mathrm{FAR}_n^2 + \mathrm{CSI}_n^2 \quad (2)$$

$$\mathrm{POD}_n = \frac{\mathrm{POD} - \min(\mathrm{POD})}{\max(\mathrm{POD}) - \min(\mathrm{POD})} \quad (3)$$

$$\mathrm{FAR}_n = \frac{\mathrm{FAR} - \min(\mathrm{FAR})}{\max(\mathrm{FAR}) - \min(\mathrm{FAR})} \quad (4)$$

$$\mathrm{CSI}_n = \frac{\mathrm{CSI} - \min(\mathrm{CSI})}{\max(\mathrm{CSI}) - \min(\mathrm{CSI})} \quad (5)$$

Evaluation of Reconstruction of the Nashville 2010 Extreme Storm

Fig. 4 shows the observed and simulated 48-h total rainfall between UTC 0000 1 May 2010 and UTC 0000 3 May 2010. Panel 4(a) is the NEXRAD observation; panel 4(b) is from the WRF simulation using the g5 grids (15 km–5 km nested grids), NAM IC/BC, Morrison microphysics, and KF cumulus parameterization schemes. This is one of the best simulations suggested by the

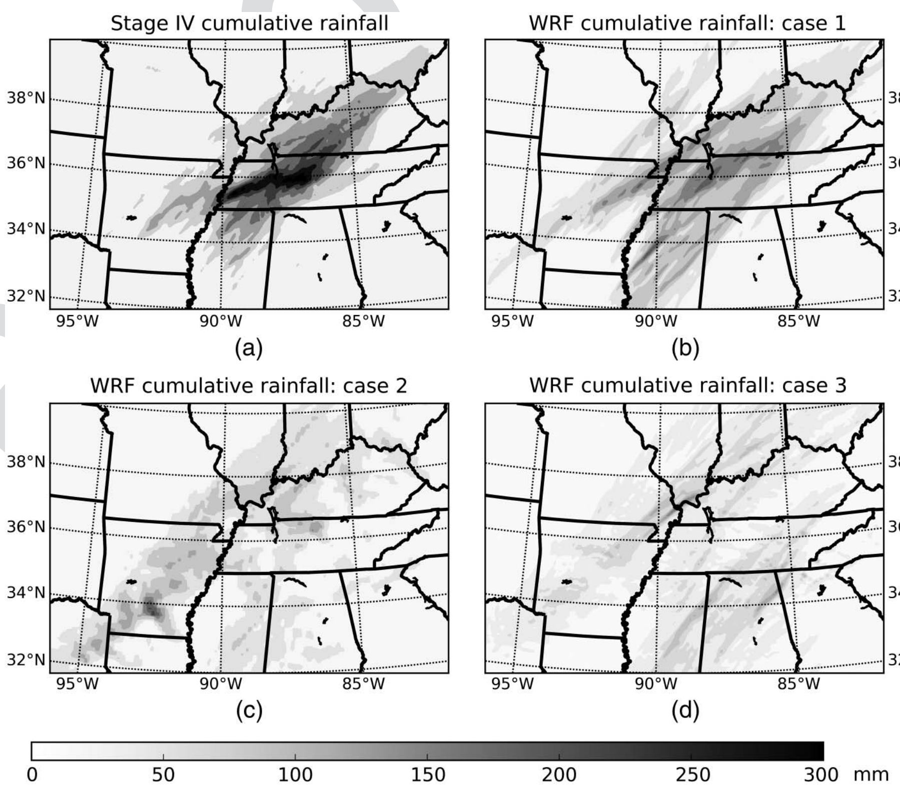


Fig. 4. Stage IV observed and WRF simulated 48-h (0000 UTC 1 May–0000 UTC 3 May, 2010 total rainfall during Nashville 2010 storm event

Table 3. Evaluation of Averaged 48-h Total Rainfall Simulated in the Evaluation Area (Normalized Using Stage IV Observed 48-h Total) in the Nashville 2010 Storm Event

	NCEP2			NNRP			NAM		
MP	KF	GD	GF	KF	GD	GF	KF	GD	GF
				15-km grids					
Morrison	0.857	0.727	0.708	0.745	0.662	0.661	0.855	0.678	0.684
Thompson	**0.921**	0.774	0.744	0.797	0.711	0.707	**0.879**	0.719	0.719
WSM-5	**0.866**	0.754	0.740	0.753	0.705	0.698	0.855	0.712	0.718
				5-km grids					
Morrison	0.874	—	0.766	0.695	—	0.680	0.890	—	0.759
Thompson	0.856	—	0.766	0.676	—	0.707	**0.905**	—	0.766
WSM-5	**0.892**	—	0.787	0.707	—	0.706	**0.899**	—	0.793
				2-km grids					
Morrison	0.827	—	0.780	0.692	—	0.663	**0.882**	—	0.816
Thompson	0.773	—	0.723	0.636	—	0.603	0.855	—	0.781
WSM-5	**0.841**	—	0.794	0.683	—	0.648	**0.898**	—	0.829

Note: Bold numbers are the top three scores with the best performance within each grid resolution.

evaluation. Comparison of panel4(b) with panel 4(a) indicates that this model setup is able to reconstruct the heavy rainfall area in midwest Tennessee. The rainfall amount gradient is properly described by this model setup. Also, the big southwest–northeast pattern of the 48-h total rainfall is clearly captured. Panel 4(c) shows a simulation with moderate scores under evaluation, and panel 4(d) shows one of the worst simulations. Though all the simulations captured the northeast–southwest–oriented rain band, the detailed rainfall distributions from various model configurations differ a lot; thus, evaluation based on the purpose of the modeling framework is necessary. The detailed evaluation is shown as a demo of using different metrics to establish the extreme storm events modeling framework.

The Stage IV data and simulation results were all conservatively regridded to the 1/16-degree grids within the d03 domain for the following analysis. All the metrics were computed using the results within the box of lat (31°N, 40°N), lon (95°W, 84°W), which is referred to here as "evaluation area."

The total rainfall amount in the event reveals the potential magnitude of the successive flood and suggests how destructive the storm would be. To evaluate the WRF simulated results, the ratios of simulated total rainfall to the Stage IV total rainfall over the evaluation area were calculated and shown in Table 3. Numbers in this table are all normalized using the observed Stage IV 48-h total rainfall; thus, the closer to 1, the more accurately the model reconstructs this event. For each grid size, the top three combinations are highlighted in bold in the table.

All these combinations tend to underestimate the total rainfall in the evaluation area. However, the best results (such as g15-NCEP2-Thompson-KF and g5-NAM-Thompson-KF) are fairly close to the observed amount, with the difference within 10%. Also, the performance of NCEP2 is comparable to those of NAM IC/BC, both of which are significantly better than NNRP IC/BC. The simulated total rainfall amount is sensitive to the cumulus scheme because the difference in KF results from NCEP2 and NAM IC/BC is less than 7%, whereas the difference due to cumulus schemes is larger than 10%. The best results come from coarser resolutions. Thus, for total rainfall estimation, the optimal framework would go up to only 5 km resolution.

Table 4 shows the spatial correlations and RMSEs between the simulated 48-h total rainfall maps and the Stage IV total rainfall map. Values in parentheses are the RMSE results. For each grid size, the top three combinations in the correlation coefficient are highlighted in bold in the table. Similarly, the top three combinations in RMSE are also bolded in the table, and they are exactly those deriving the best spatial correlation. At all three grid scales,

Table 4. Spatial Correlation and RMSE between Simulated and Stage IV Reference 48-h Cumulative Rainfall Distribution in the Nashville 2010 Storm Event

	NCEP2			NNRP			NAM		
MP	KF	GD	GF	KF	GD	GF	KF	GD	GF
				15-km grids					
Morrison	0.364 (65.0)	0.344 (66.4)	0.231 (69.4)	0.259 (69.4)	0.345 (66.8)	0.139 (71.6)	**0.597 (55.4)**	0.488 (62.2)	0.471 (62.7)
Thompson	0.359 (65.5)	0.368 (65.2)	0.254 (68.4)	0.249 (69.8)	0.344 (66.2)	0.125 (71.6)	**0.606 (54.7)**	0.516 (60.5)	0.516 (60.5)
WSM-5	0.365 (64.9)	0.362 (65.6)	0.271 (68.1)	0.261 (69.4)	0.361 (65.8)	0.122 (71.8)	**0.589 (55.8)**	0.485 (61.8)	0.418 (64.2)
				5-km grids					
Morrison	0.455 (62.1)	—	0.171 (71.3)	0.311 (69.1)	—	0.154 (71.1)	**0.773 (43.8)**	—	0.500 (60.6)
Thompson	0.335 (68.1)	—	0.216 (69.4)	0.334 (68.5)	—	0.159 (70.7)	**0.698 (49.2)**	—	0.509 (60.2)
WSM-5	0.322 (68.9)	—	0.220 (70.0)	0.337 (68.0)	—	0.172 (70.3)	**0.700 (49.2)**	—	0.537 (58.7)
				2-km grids					
Morrison	0.596 (55.6)	—	0.527 (59.3)	0.289 (70.1)	—	0.293 (70.2)	**0.766 (44.4)**	—	**0.705 (49.4)**
Thompson	0.490 (61.0)	—	0.380 (65.8)	0.277 (70.1)	—	0.302 (69.3)	0.697 (49.6)	—	0.644 (53.6)
WSM-5	0.482 (61.4)	—	0.435 (63.8)	0.318 (68.4)	—	0.313 (68.7)	**0.708 (48.8)**	—	0.623 (54.6)

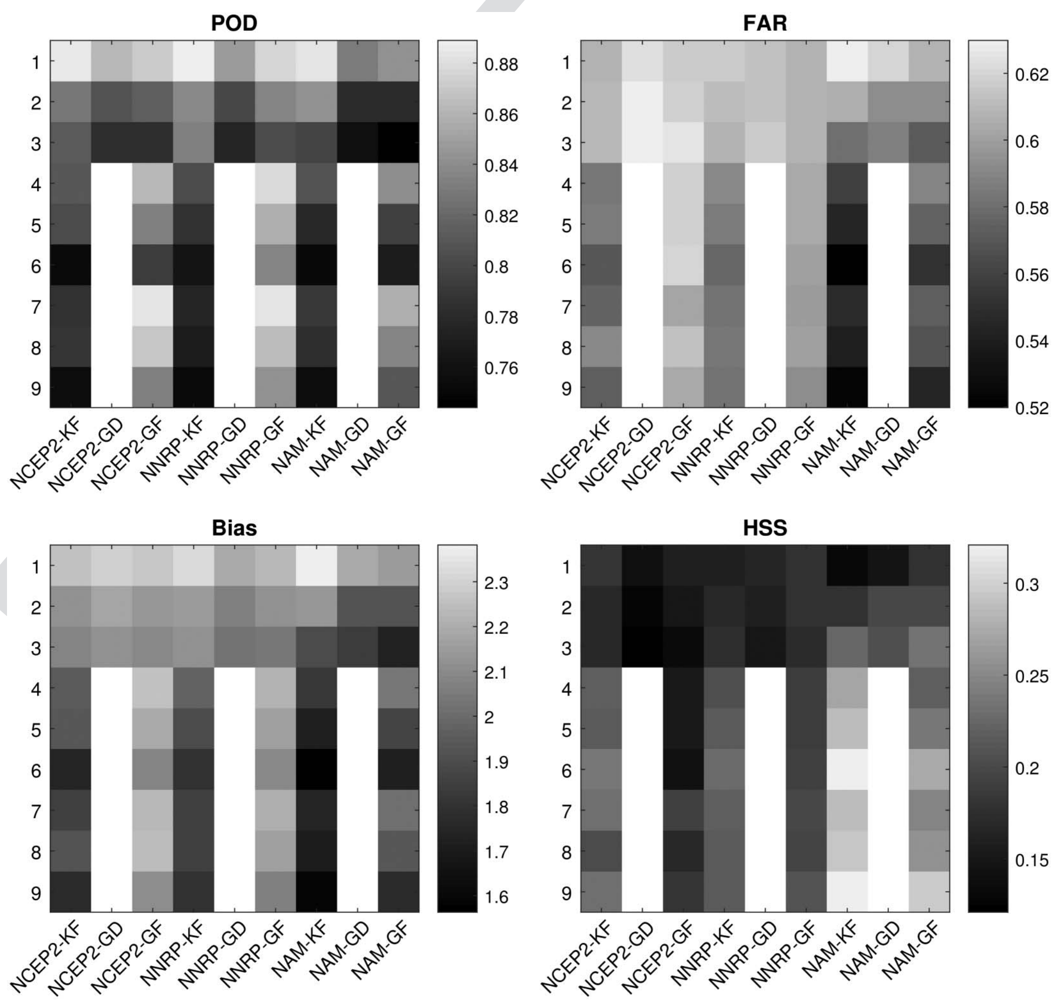
Note: Values in parentheses are RMSE (unit: mm/day). Bold numbers are the top three scores with the best performance (highest correlation or lowest RMSE) within each grid resolution.

NAM provides the best estimates of the 48-h total rainfall. Within each IC/BC category, the difference from different microphysics schemes is not huge (usually only within 20% of the score), but different cumulus parameterization schemes have significant impacts on the precipitation simulation quality. This is especially notable in the simulations driven by the NCEP2 IC/BC, where the spatial correlation ranges from ~0.2 (GF scheme) to ~0.6 (KF scheme), and the correlations with the KF scheme are always higher than those with the GF scheme. Also, the g5-NAM-Morrison-KF case [Fig. 4(b)] produced the best spatial correlation among all the tested cases. Based on Table 3, NAM IC/BC and the KF cumulus scheme are recommended for storm reconstructions that address the spatial distribution of the cumulative rainfall (such as PMP design). However, because this result is based only on the cumulative rainfall, it does not reveal temporal evolution information.

At the 5- and 2-km grid scales, all the combinations produce stronger correlations. As can be seen in the following analysis, NAM often produces the best quantitative evaluation values in the finer grids. The top combinations for the 5- and 2-km grids are similar. The difference among the best correlation results at the three different grid scales is not significant. In general, higher-resolution simulations are able to capture finer-scale features, although the improvement from 5 to 2 km is marginal.

In certain types of engineering infrastructure analyses, it is important to know both the location and period of the storm event. A better picture of the spatial-temporal structure of the storm would help make better operation plans for the drainage systems, for example. To better evaluate the simulated spatial–temporal structures of the storm, quantitative scores were computed for the 63 simulations. Unlike the calculation of spatial correlation using rainfall total, the computation here used hourly rainfall data. Fig. 5 visualizes the evaluation on the spatial coverage of hourly rainfall simulated by WRF. Blank panels mean the corresponding combination was not tested (similar to "-" in Table 3). Fig. 5(a) shows the POD, with greater values representing more skillful simulations. Similarly, Fig. 5(b) shows the FAR (lower values are better). POD reflects the probability of rainfall grid points being successfully simulated as "rainy" by the numerical model. FAR evaluates the simulation accuracy of nonrainy regions, so combining it with POD can provide a better assessment of the simulation quality.

The general information from Figs. 5(a) and 5(b) suggests that as the numerical model takes advantage of the finer grids, the simulation quality usually improves. The g15 grid shows somewhat better POD than some of the g5 and g2 results, which is possible because POD measures only how completely the observed rainfall area is covered by the simulation. Fig. 5(a) suggests that the Morrison microphysics scheme tends to overestimate rainfall coverage, and this is supported by the higher FAR values in Fig. 5(b). Compared with the g15 grid, finer grid simulations are able to reduce the likelihood of false alert: The range of the best three FAR scores in the g15 grid is [0.571, 0.588], which is less skillful than



F5:1 **Fig. 5.** Evaluation of spatial coverage simulated by WRF the g5 results of [0.520, 0.551]. Similar to the findings from the spatial correlation and total rainfall analyses, the biggest difference in the FAR comes from the choice of IC/BCs: NAM outperforms others at both coarser and finer grids. Also, the WSM-5 scheme tends to produce less spatial extent of rainfall, so it performs better for the FAR score.

Fig. 5(c) shows the frequency bias scores. A bias score higher than 1 means the model overestimates the rainfall coverage, and a score less than 1 suggests an underestimation. As WRF is applied in the finer grids, the bias scores steadily converge to 1. All microphysics schemes benefit from the use of the finer grids. All of the bias scores are larger than 1, which indicates that all models overestimate the rainfall area. Because the total rainfall amount analysis suggests that all models underestimate the total rainfall amount, the simulated picture is most likely to be an expanded rainy area with a rain rate smaller than the observed rate. This is confirmed by comparing Fig. 4(b) to Fig. 4(a). Fig. 5(d) presents the HSS, with higher scores indicating better simulations. For a simulation with nonzero capability in forecasting and simulation, the HSS must be greater than 0. Fig. 5(d) shows that all 63 simulations have some capabilities for forecasting/simulation. Similar to the FAR scores, NAM IC/BC performs best at both coarser and finer grids. The improvement from the g15 to g5 grid is significant (an approximately 20% increase), but the even finer g2 grid does not provide further improvement. Thus, the 5 km grid is an acceptable compromise for PMP simulation because it does not compromise simulation quality at the expense of reduced computational burden. In terms of microphysics schemes, WSM-5 is best for both the finer and coarse grids. In the coarse grid, the KF cumulus scheme is also a good choice when combined with the Morrison or new Thompson cumulus schemes.

Fig. 6 shows the evaluation based on metrics that consider multiple aspects of the rainfall simulation quality. Fig. 6(a) shows the CSI grades (the higher the better). Any skillful forecast or simulation should have greater than 0 grades. Fig. 6(c) shows the GSS grades (the higher the better). GSS improves CSI grades by taking into account the randomness of the observation, and it also requires a positive grade for the simulation to be considered skillful. The largest differences come from the choice of the IC/BC data source, and it is obvious that WSM-5 is the winning microphysics scheme at various grids.

As shown in the previous figures, different metrics usually yield differing recommendations. They are helpful for specific purposes, but a better metric would be desired to simultaneously evaluate

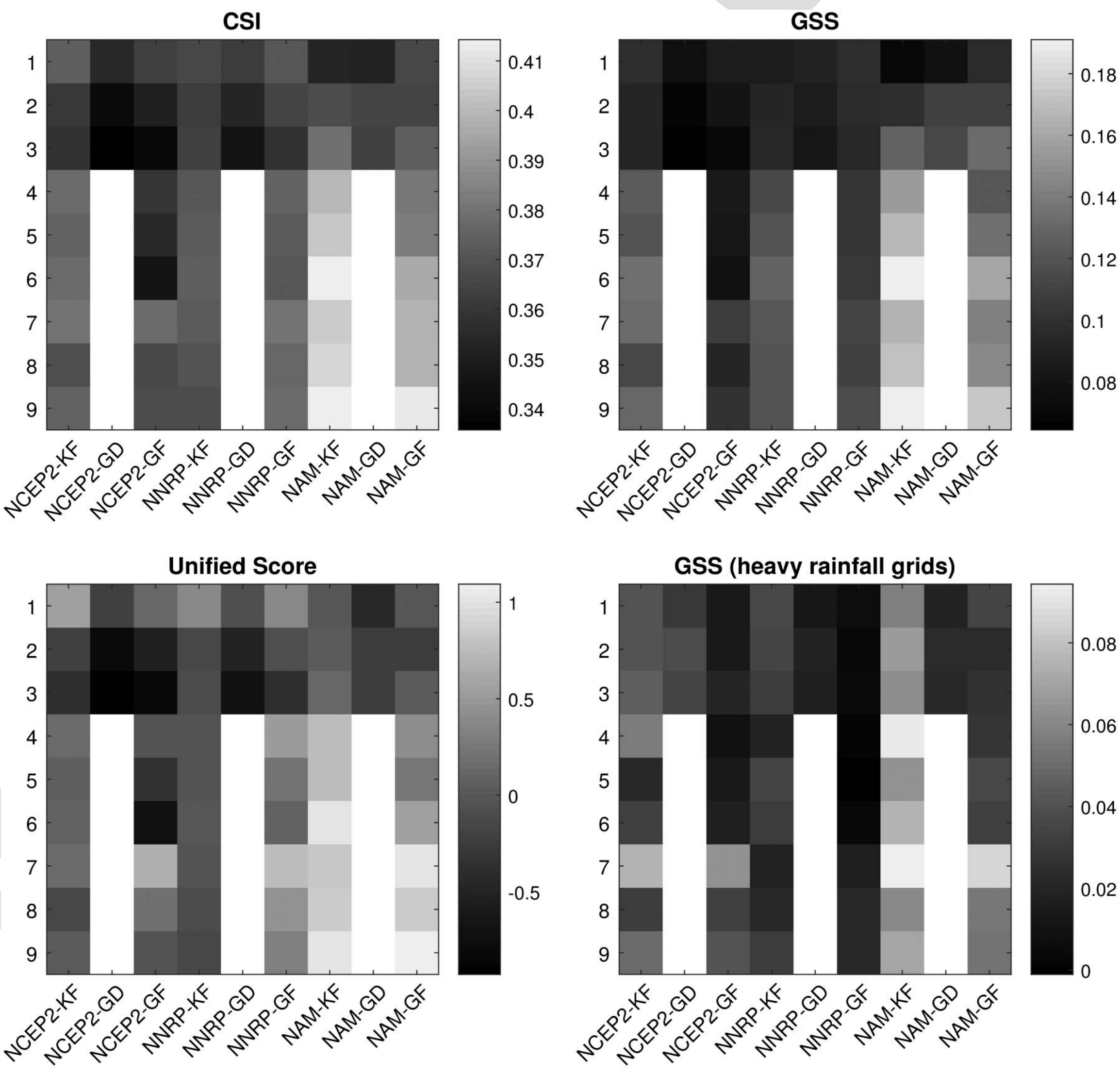


Fig. 6. Evaluation involving multiple aspects of rainfall simulation quality

Table 5. Nash-Sutcliffe Metric between Simulated and Stage IV and Livneh Reference Rainfall Map in the Nashville 2010 Storm Event

	NCEP2			NNRP			NAM		
MP	KF	GD	GF	KF	GD	GF	KF	GD	GF
				15-km grids					
Morrison	0.09 (−0.20)	0.05 (−0.04)	−0.03 (−0.09)	−0.03 (−0.43)	0.04 (−0.06)	−0.10 (−0.24)	0.34 (−0.02)	0.17 (−0.08)	0.16 (0.01)
Thompson	0.08 (−0.28)	0.09 (−0.01)	0.00 (−0.05)	−0.05 (−0.48)	0.06 (−0.10)	−0.10 (−0.28)	0.36 (0.00)	0.21 (−0.01)	0.21 (−0.01)
WSM-5	0.09 (−0.18)	0.08 (−0.09)	0.01 (−0.11)	−0.03 (−0.48)	0.07 (−0.09)	−0.11 (−0.31)	0.33 (−0.01)	0.18 (−0.02)	0.12 (−0.01)
				5-km grids					
Morrison	0.17 (−0.16)	—	−0.09 (−0.46)	−0.02 (−0.38)	—	−0.08 (−0.29)	0.59 (0.43)	—	0.21 (0.01)
Thompson	0.00 (−0.69)	—	−0.04 (−0.37)	−0.01 (−0.46)	—	−0.07 (−0.33)	0.48 (0.18)	—	0.22 (0.03)
WSM-5	−0.02 (−0.59)	—	−0.05 (−0.51)	0.01 (−0.52)	—	−0.06 (−0.34)	0.48 (0.15)	—	0.26 (0.04)
				2-km grids					
Morrison	0.34 (0.09)	—	0.24 (0.07)	−0.06 (−0.48)	—	−0.06 (−0.48)	0.58 (0.30)	—	0.47 (0.15)
Thompson	0.20 (−0.25)	—	0.07 (−0.46)	−0.06 (−0.56)	—	−0.04 (−0.44)	0.47 (0.11)	—	0.38 (−0.05)
WSM-5	0.19 (−0.41)	—	0.13 (−0.46)	0.00 (−0.62)	—	−0.01 (−0.51)	0.49 (−0.01)	—	0.36 (−0.20)

Note: Numbers in parentheses are with Livneh reference. Stage IV 48-h total precipitation map is used to evaluate the simulated 48-h total precipitation. Livneh gridded daily precipitation data on May 1, 2010, is used to evaluate the simulated total precipitation on this day.

multiple aspects of the modeling framework. For this purpose, the unified scores [see Eq. (2)] were calculated and are shown in Fig. 6(c). At a coarser grid (15 km), the Morrison microphysics scheme provides the best results. With the NCEP2 IC/BC, the KF scheme yields the highest scores in the g15 domain setup [Fig. 2(a)] group. As the model is run in the finer grids, the NCEP2 results produce lower scores, even negative sometimes. At the finer grids (5 and 2 km), however, NAM yields the best detailed estimates of rainfall. NNRP gives the worst results in both coarse and fine grids, and the scores degrade further in the finer grids. With NAM providing IC/BC, the 2-km simulations are more skillful than the 5-km simulations and less sensitive to the parameterizations used, though the extra improvement is marginal. Also, the GF cumulus scheme produces the best US score in the g5 and g2 domain setups. This implies the GF scheme is scale aware, and it does not double-count the deep convection along with the rainfall that is resolved by the microphysics process.

For extreme events, it is sometimes more useful to analyze the areas with heavy rainfall because they tend to result in the heaviest human and economic losses. NOAA's definition of a heavy storm is an event with an hourly rain rate larger than 7.6 mm. Using this threshold to filter out nonheavy rainfall area, the model performance over the heavy rain area can be evaluated. Fig. 6(d) shows the GSS for the Nashville 2010 event with only heavy (> 7.6 mm/h) rainfall cells and timesteps are treated as rainy cells.

Unlike Fig. 6(b), the KF cumulus scheme tends to work best in the heavy rain area. It is obvious that the KF cumulus scheme is a winning option at various scales. At coarser resolutions, the WSM-5 microphysics scheme tends to work better, whereas Morrison is dominantly better at finer resolutions. The best modeling frameworks recommended by general GSS scores (Fig. 6b) are quite different from those highlighted by the heavy rain–area GSS scores; thus, it is necessary to identify the specific objectives of the modeling framework and choose the corresponding evaluation metrics.

In applications where storm magnitude is important (e.g., when used as input to hydrological or hydraulic models), it is necessary to quantify the simulated rainfall in both spatial extent and amount. Here, the simulated 48-h rainfall maps are compared to Stage IV and Livneh data under Nash-Sutcliffe coefficient. The results are shown in Table 5, where the values in parentheses are the results between the modeled data and the Livneh reference. The higher the Nash-Sutcliffe coefficient is, the better the model predicts the rainfall pattern. When compared to Stage IV data, Morrison microphysics and KF cumulus schemes outperform others in both coarser and finer grids. The same holds true when they are compared against the Livneh reference, suggesting that their superiority is independent from the choice of reference. In terms of IC/BC, NAM produced the best results, followed by simulations with NCEP2 data.

To check the statistics of simulated rainfall intensity, the hourly rainfall intensities can be plotted as histograms in Fig. 7. In these panels, the *x*-axis shows the hourly rainfall intensity and the *y*-axis shows the total count of such hourly rainfall intensities across the evaluation area in the 48-h duration. Each panel in the figure shows a combination of IC/BC, microphysics, and cumulus schemes. The black lines are the histograms from Stage IV data, the blues lines are those using g15 grids, red lines are those using g5 grids, and green lines are from g2 grids. The biggest difference comes from grid size, where g15 results are often biased away from observation. In most cases, g5 results are closer to the observation, and the improvement from g5 to g2 is not significant. This confirms that g5 is a balance between accuracy and computing burden. Again, Morrison microphysics and the KF cumulus schemes produced better results here.

This evaluation suggests that different demands are best met with different model options for extreme storm simulation. These options have their own strengths and weaknesses. However, collectively, they can be used to generate a multiphysics ensemble forecast, which is useful for providing an "envelope" at a certain confidence level for an engineering application. For example, a range of possible PMP estimates can be much more useful for risk management than a single deterministic value. The results show that the width of the envelope is largely determined by uncertainty in the IC/BCs, followed by sensitivity to grid resolution. The use of the scale-aware GF scheme tends to reduce model sensitivity to resolution as intended and consistently yields high unified scores regardless of the microphysics parameterizations used. These results demonstrate the possibilities of capturing the full range of the envelope using fewer but carefully tested configurations of the end members for design PMP estimates.

In summary, NAM is better for finer-grid simulation, whereas NCEP2 is also a good choice at coarser grids for extreme storms. At finer grids, Morrison or WSM-5 is often a winning option. At the coarse grid scale, the results from different microphysics and

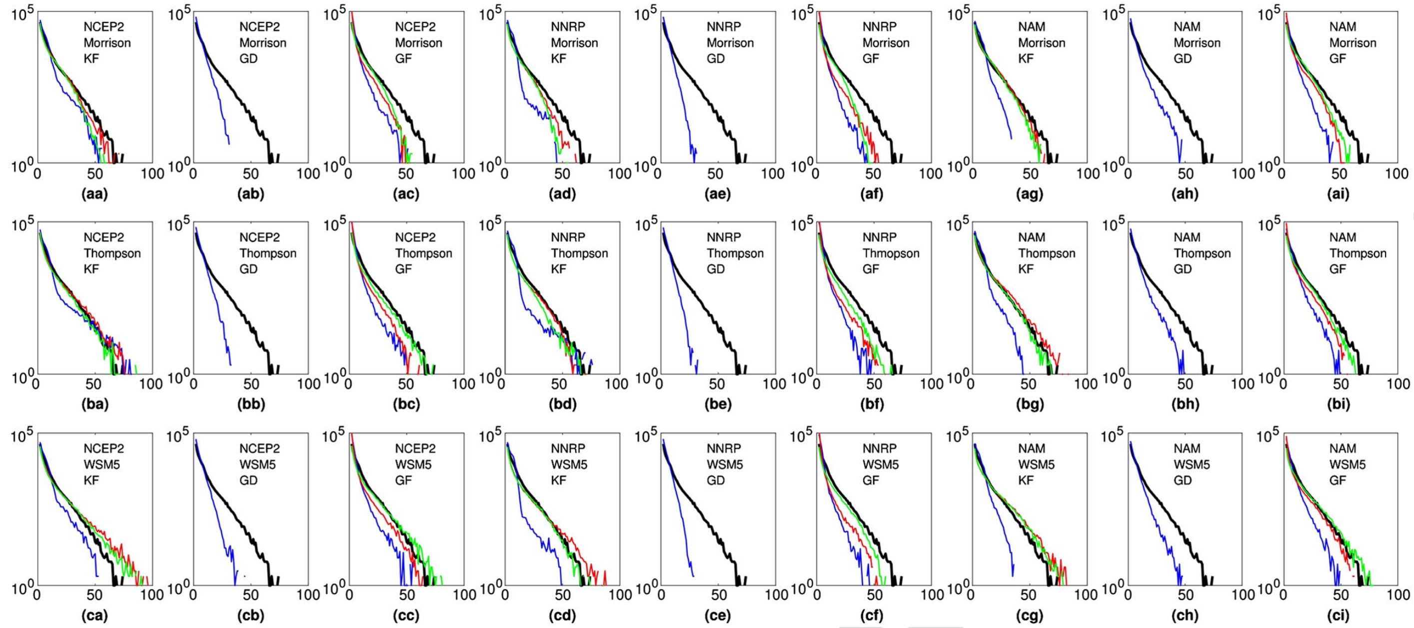


Fig. 7. Evaluation of hourly rainfall intensity histogram simulated by WRF

cumulus schemes are mixed. Combinations that better resolve the spatial-temporal structure of the storm are: g15-NAM-Morrison-KF, g5-NAM-Morrison-KF, g15-NCEP2-Thompson-GF, and g5-NAM-WSM5 (with KF or GF cumulus parameterization scheme). The improvement from g5-NAM-WSM5 to g2-NAM-WSM5 is insignificant (e.g., CSI changed from 0.40 to 0.41), so given the larger computing requirements, the g2 option is not recommended here. For general purposes, NAM-Morrison-KF is recommended as a starting choice. With enough computing capacity, the g5 grid is recommended, but g15 is also acceptable when running with this configuration.

For the Nashville 2010 storm reconstruction, the recommendation emerging from the application of the framework differs from previous studies. For example, Mahoney (2013) recommended the 4 km–1.3 km nested grids, and the NAM forecast IC/BC with the new Thompson microphysics and no cumulus schemes. The maximum 48-h total rainfall captured by Mahoney (2013) was 260 mm, whereas in this study, it is 239 mm from the 1.6 km grid. However, the WSM-5 scheme was not tested by Mahoney (2013). In this study, the 300-mm 48-h total rainfall isohyet was captured by using the WSM-5 microphysics scheme. Although these estimates are smaller than the maximum 48-h total rainfall from the Stage IV reference precipitation data (330 mm), the use of WSM-5 represents an advance in capturing the high-precipitation area.

This study includes the four major factors that affect atmospheric model performance. However, there are still some other factors that can be fine-tuned as needed, such as land surface process, planetary boundary scheme, and land use condition. Following the same methodology outlined in this study, these factors can be added into this evaluation framework to achieve even better simulation quality, if desired by the engineering community.

Validation of Optimal Model Configuration

This study proved that the recommended model configuration is independent from reference choice. The representativeness of this finding for other storms remains a question. Therefore, the optimal WRF configuration was applied to other two storm events, the 1997 January 1–3 storm in the American River watershed, California (denoted as the 1997CA event), and the 1980 December 24–26 storm in the Pacific Northwest region (1980PNW event). Because of data availability, these two events are reconstructed using the NCEP2 IC/BC.

The 1997CA event happened in northern California and caused a severe flood in Sacramento in the following days. The observed maximum 24-h precipitation was 284 mm, which made it one of the greatest storms in this area. The 1980PNW event happened in the Washington and Oregon, and the observed maximum 24-h precipitation was 234 mm. This storm is one of the big storms used in the HMR for PMP in the Pacific Northwest region (HMR57).

These storm events were reconstructed using the optimal model configuration (15 km–5 km nested grids, Morrison microphysics, and KF cumulus schemes) that was obtained through the Nashville 2010 study. The simulated 3-day total rainfall of these two events, plus the simulated 1-day rainfall of the Nashville 2010 event, are shown in Fig. 8. Figs. 8(a and c) show the model reconstructed 3-day precipitation, and Figs. 8(b and d) show the Livneh reference. The third column shows the difference as WRF-Livneh. It shows that this model configuration depicts the heavy rainy area in both spatial extent and magnitude: In the 1997CA simulation, the model captures the storm center along the Sierra Nevada; in the 1980PNW simulation, it captures the heavy rainy band along the coast.

To quantity the performance of these reconstructions, all nine model configurations in 15-km grids, and six configurations in 15 km–5 km grids were also tested. The evaluation of the Nash-Sutcliffe coefficient on the simulated maximum 3-day rainfall is shown in Table 6. In the 1997CA simulations, this optimal configuration (based on the Nashville 2010 storm) produced the best result. In the 1980PNW simulations, the performance of this optimal model configuration is within the top three among all the experiments. This confirms the capability of the optimal model configuration in reconstructing other severe storms, and this is independent of the choice of reference data.

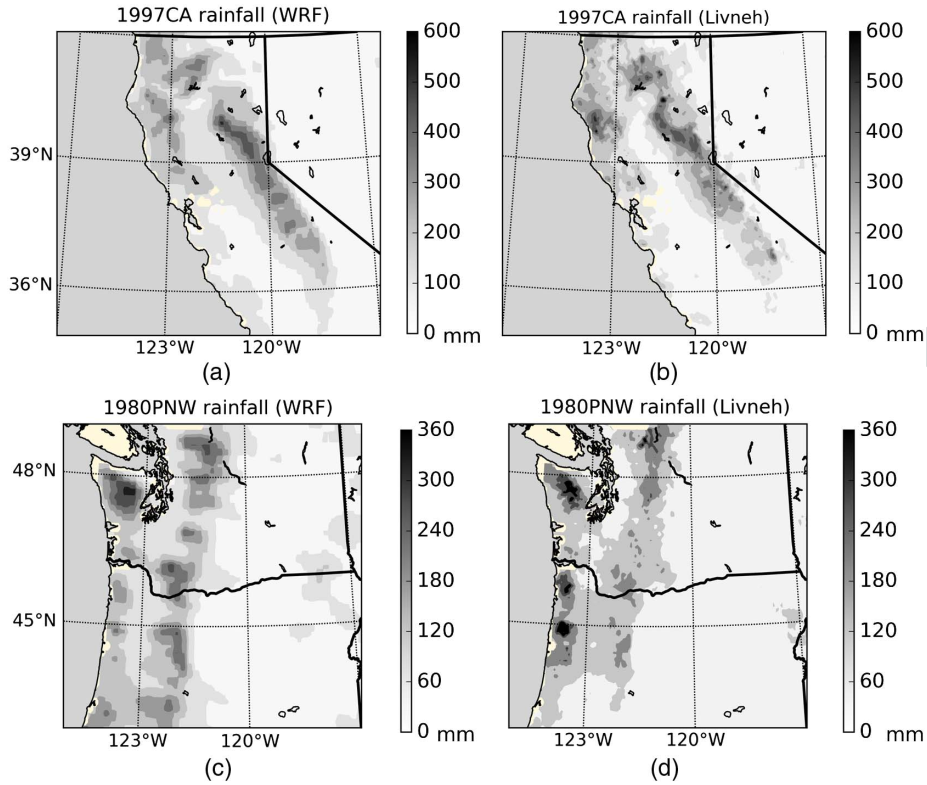


Fig. 8. Evaluation of 1997CA and 1980PNW storms using optimal WRF configuration

Table 6. Nash-Sutcliffe Metric between Simulated and Livneh Reference 3-Day Cumulative Rainfall Map in the 1997CA and 1980PNW Storm Events

	1997CA			1980PNW		
MP	KF	GD	GF	KF	GD	GF
15-km grids						
Morrison	**0.69**	0.67	**0.69**	**0.53**	**0.54**	**0.52**
Thompson	0.61	0.59	0.62	0.50	0.50	0.50
WSM-5	0.59	0.57	0.63	0.41	0.41	0.40
5-km grids						
Morrison	0.64	—	**0.68**	0.49	—	0.49
Thompson	0.57	—	0.63	0.46	—	0.47
WSM-5	0.52	—	0.61	0.31	—	0.33

Note: Livneh gridded precipitation data is used as references.

Conclusions

In this study, an approach to establishing an optimal WRF-based framework for extreme storm event simulation was investigated. The goal was to introduce a more physically based method to the engineering design and analyses community currently engaged in large water-management infrastructure issues of today and tomorrow. This framework takes into consideration the uncertainties coming from various IC/BC data sources, grid resolutions, cloud microphysics, and cumulus parameterization schemes. These are the major contributors to the final model performance.

In the demonstration, a WRF-based modeling framework was established for extreme storm events in the CONUS region based on the Nashville 2010 storm and validated it using two other storms in California and the Pacific Northwest. Based on the engineering intent, the best model configuration can be different. For general purposes, it is recommended that the WRF model be configured as: 15 km or nested 15 km–5 km grids, the NCEP2 or NAM boundary condition, and the Morrison microphysics scheme with the Kain-Fritsch cumulus scheme. This configuration is either the optimal configuration or a starting point that leads to quick convergence to the final optimal configuration.

For future studies, the authors hope to complete application and validation of the optimal WRF modeling framework for a large number of storms that were maximized for PMP estimation in HMR reports. Because the use of atmospheric numerical models for engineering infrastructure analyses is gaining popularity among the infrastructure community, future studies should focus on improving current design and practice among engineers. Some examples are: (1) exploration of physics-based probable maximum flood (PMF, the flood due to PMP), (2) impact of land use and land cover change and global warming on PMP and PMF during extreme storms, (3) improving streamflow forecast and thus improving reservoir and dam operation during extreme storm events, (4) multiphysics ensemble–based analyses of numerical model output for risk management.

Acknowledgments

This study is motivated by an American Society of Civil Engineers (ASCE) Task Committee on "Infrastructure Impacts of Landscape-driven Weather change" chaired by the second author. Therein, model-based PMPs are currently being explored for modernizing current engineering practice of PMPs based on storms many decades old. The first author was supported by NASA grant NNAX15AC63G. Dr. Ruby Leung acknowledges the support from the U.S. Department of Energy Office of Science Biological and Environmental Research as part of the Regional and Global Climate Modeling Program. The Pacific Northwest National

Laboratory is operated for DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO1830.

References

Abbs, D. J. (1999). "A numerical modeling study to investigate the assumptions used in the calculation of probable maximum precipitation." *Water Resour. Res.*, 35(3), 785–796.

Bennett, N. D., et al. (2013). "Characterising performance of environmental models." *Environ. Modell. Software*, 40, 1–20. [6]

Casagli, N., Dapporto, S., Ibsen, M. L., Tofani, V., and Vannocci, P. (2006). "Analysis of the landslide triggering mechanism during the storm of 20th–21st November 2000, in Northern Tuscany." *Landslides*, 3(1), 13–21.

Chang, H. I., Kumar, A., Niyogi, D., Mohanty, U. C., Chen, F., and Dudhia, J. (2009). "The role of land surface processes on the mesoscale simulation of the July 26, 2005 heavy rain event over Mumbai, India." *Global Planetary Change*, 67(1), 87–103.

Chen, X., and Hossain, F. (2016). "Revisiting extreme storms of the past 100 years for future safety of large water management infrastructures." *Earth's Future*, 4(7), 306–322.

Cong, X., Ni, G., Hui, S., Tian, F., and Zhang, T. (2006). "Simulative analysis on storm flood in typical urban region of Beijing based on SWMM." *Water Resour. Hydropower Eng.*, 4, 64–67. [8] [7]

Del Genio, A. D., Kovari, W., Yao, M. S., and Jonas, J. (2005). "Cumulus microphysics and climate sensitivity." *J. Clim.*, 18(13), 2376–2387.

Durkee, J. D., et al. (2012). "A synoptic perspective of the record 1–2 May 2010 mid-south heavy precipitation event." *Bull. Am. Meteorol. Soc.*, 93(5), 611–620.

Evans, J. E., Mackey, S. D., Gottgens, J. F., and Gill, W. M. (2000). "Lessons from a dam failure." *Ohio J. Sci.*, 100(5), 121–131. [9]

Giannaros, T. M., Kotroni, V., and Lagouvardos, K. (2016). "WRF-LTNGDA: A lightning data assimilation technique implemented in the WRF model for improving precipitation forecasts." *Environ. Modell. Software*, 76, 54–68. [10]

Grell, G. A., and Freitas, S. R. (2014). "A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling." *Atmos. Chem. Phys.*, 14(10), 5233–5250.

Hershfield, D. M. (1965). "Method for estimating probable maximum rainfall." *J. Am. Water Works Assoc.*, 57(8), 965–972. [11]

Huschke, R. E. (1959). *Glossary of meteorology*, American Meteorological Society, Boston.

Kumar, A., Dudhia, J., Rotunno, R., Niyogi, D., and Mohanty, U. C. (2008). "Analysis of the 26 July 2005 heavy rain event over Mumbai, India using the Weather Research and Forecasting (WRF) model." *Q. J. R. Meteorol. Soc.*, 134(636), 1897–1910.

Kunkel, K. E., et al. (2013). "Probable maximum precipitation and climate change." *Geophys. Res. Lett.*, 40(7), 1402–1408.

Laprise, R. (1992). "The Euler equations of motion with hydrostatic pressure as an independent variable." *Monthly Weather Rev.*, 120(1), 197–207.

Liu, J., Bray, M., and Han, D. (2012). "Sensitivity of the weather research and forecasting (WRF) model to downscaling ratios and storm types in rainfall simulation." *Hydrol. Process.*, 26(20), 3012–3031.

Livneh, B., et al. (2013). "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions." *J. Clim.*, 26(23), 9384–9392.

Mahoney, K. (2013). "The impact of model physics and upstream moisture sources on the May 2010 Tennessee flooding event: An examination of precipitation and surface hydrology." [12]

Moore, B. J., Neiman, P. J., Ralph, F. M., and Barthold, F. E. (2012). "Physical processes associated with heavy flooding rainfall in Nashville, Tennessee, and vicinity during 1–2 May 2010: The role of an atmospheric river and mesoscale convective systems." *Monthly Weather Rev.*, 140(2), 358–378.

Morrison, H., Thompson, G., and Tatarskii, V. (2009). "Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes." *Monthly Weather Rev.*, 137(3), 991–1007.

Pennelly, C., Reuter, G., and Flesch, T. (2014). "Verification of the WRF model for simulating heavy precipitation in Alberta." *Atmos. Res.*, 135–136, 172–192. [13]

Rajeevan, M., Kesarkar, A., Thampi, S. B., Rao, T. N., Radhakrishna, B., and Rajasekhar, M. (2010). "Sensitivity of WRF cloud microphysics to simulations of a severe thunderstorm event over Southeast India." *Ann. Geophys.*, 28(2), 603–619.

Rao, Y. V. R., Hatwar, H. R., Salah, A. K., and Sudhakar, Y. (2007). "An experiment using the high resolution ETA and WRF models to forecast heavy precipitation over India." *Pure Appl. Geophys.*, 164(8-9), 1593–1615. [14]

Sikder, S., and Hossain, F. (2016). "Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven river basins." *J. Adv. Model. Earth Syst.*, 8(3), 1210–1228.

Skamarock, W. C., et al. (2008). "A description of the advanced research WRF version 3. NCAR Tech. Note." Tech Report. [15]

Stensrud, D. J. (2009). *Parameterization schemes: keys to understanding numerical weather prediction models*, Cambridge University Press, Cambridge, UK.

Stratz, S. A., and Hossain, F. (2014). "Probable maximum precipitation in a changing climate : Implications for dam design." *J. Hydrol. Eng.*, 10.1061/(ASCE)HE.1943-5584.0001021, 6014006. [16]

Tan, E. (2010). "Development of a methodology for probable maximum precipitation estimation over the American River watershed using the WRF model." Univ. of California, Davis, CA.

Vaidya, S. S., and Kulkarni, J. R. (2007). "Simulation of heavy precipitation over Santacruz, Mumbai on 26 July 2005, using mesoscale model." *Meteorol. Atmos. Phys.*, 98(1-2), 55–66.

Winters, A. C., and Martin, J. E. (2014). "The role of a polar/subtropical jet superposition in the May 2010 Nashville flood." *Weather Forecasting*, 29(4), 954–974.

WMO (World Meteorological Organization). (1986). "Manual for estimation of probable maximum precipitation." *Operational Hydrology Rep. 1.* [17]

Zhang, G. J., and McFarlane, N. A. (1995). "Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model." *Atmos.-Ocean*, 33(3), 407–446.