2012

The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments

Michael B. Smith
NOAA/NWS, michael.smith@noaa.gov

Victor Koren
NOAA National Weather Service

Seann Reed
NOAA National Weather Service

Ziya Zhang
NOAA National Weather Service

Yu Zhang
NOAA National Weather Service

See next page for additional authors

Michael B. Smith, Victor Koren, Seann Reed, Ziya Zhang, Yu Zhang, Fekadu Moreda, Zhengtao Cui, Naoki Mizukami, Eric A. Anderson, and Brian A. Cosgrove

Journal of Hydrology 418–419 (2012) 3–16

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments

Michael B. Smith [a,*], Victor Koren [a], Seann Reed [a], Ziya Zhang [a], Yu Zhang [a], Fekadu Moreda [b], Zhengtao Cui [a], Naoki Mizukami [a], Eric A. Anderson [a], Brian A. Cosgrove [a]

[a] *Office of Hydrologic Development, NOAA National Weather Service, 1325 East West Highway, Silver Spring, MD 20910, United States*
[b] *Water and Ecosystems Management, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, United States*

A R T I C L E I N F O

Article history:
Available online 16 September 2011

Keywords:
Distributed hydrologic model
Rainfall runoff
Hillslope routing
Channel routing
Streamflow simulation
Lumped model

S U M M A R Y

The Office of Hydrologic Development (OHD) of the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) conducted the second phase of the Distributed Model Intercomparison Project (DMIP 2). After DMIP 1, the NWS recognized the need for additional science experiments to guide its research-to-operations path towards advanced hydrologic models for river and water resources forecasting. This was accentuated by the need to develop a broader spectrum of water resources forecasting products (such as soil moisture) in addition to the more traditional river, flash flood, and water supply forecasts. As it did for DMIP 1, the NWS sought the input and contributions from the hydrologic research community.

DMIP 1 showed that using operational precipitation data, some distributed models could indeed perform as well as lumped models in several basins and better than lumped models for one basin. However, in general, the improvements were more limited than anticipated by the scientific community. Models combining so-called conceptual rainfall-runoff mechanisms with physically-based routing schemes achieved the best overall performance. Clear gains were achieved through calibration of model parameters, with the average performance of calibrated models being better than uncalibrated models. DMIP 1 experiments were hampered by temporally-inconsistent precipitation data and few runoff events in the verification period for some basins. Greater uncertainty in modeling small basins was noted, pointing to the need for additional tests of nested basins of various sizes.

DMIP 2 experiments in the Oklahoma (OK) region were more comprehensive than in DMIP 1, and were designed to improve our understanding beyond what was learned in DMIP 1. Many more stream gauges were located, allowing for more rigorous testing of simulations at interior points. These included two new gauged interior basins that had drainage areas smaller than the smallest in DMIP 1. Soil moisture and routing experiments were added to further assess if distributed models could accurately model basin-interior processes. A longer period of higher quality precipitation data was available, and facilitated a test to note the impacts of data quality on model calibration. Moreover, the DMIP 2 calibration and verification periods contained more runoff events for analysis. Two lumped models were used to define a robust benchmark for evaluating the improvement of distributed models compared to lumped models. Fourteen groups participated in DMIP 2 using a total of sixteen models. Ten of these models were not in DMIP 1.

This paper presents the motivation for DMIP 2 Oklahoma experiments, discusses the major project elements, and describes the data and models used. In addition, the paper introduces the findings, which are covered in a companion results paper (Smith et al., this issue). Lastly, the paper summarizes the DMIP 1 and 2 experiments with commentary from the NWS perspective. Future papers will cover the DMIP 2 experiments in the western USA mountainous basins.

Published by Elsevier B.V.

1. Introduction

1.1. Background

The Office of Hydrologic Development (OHD) of the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) conducted the second phase of the Distributed Model Intercomparison Project (DMIP 2). The first phase of

* Corresponding author. Tel.: +1 301 713 0640x128; fax: +1 301 713 0963.
E-mail address: michael.smith@noaa.gov (M.B. Smith).

0022-1694/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jhydrol.2011.08.055

DMIP (hereafter called DMIP 1) proved to be a landmark comparison of lumped and distributed models in the southern Great Plains of the USA (Smith et al., 2004a; Reed et al., 2004). Twelve groups participated in DMIP 1, including researchers from China, Denmark, Canada, New Zealand, and universities and institutions in the USA. Models ranged from conceptual representations of the soil column applied in various computational elements, to more comprehensive physically-formulated models based on highly detailed triangulated representations of the terrain. Results from DMIP 1 activities were published in a special issue of the Journal of Hydrology in October, 2004 (Smith et al., 2004a).

DMIP 1 provided valuable guidance to the NWS research-to-operations program for improved hydrologic models for river and water resources forecasting. For example, follow-on work from DMIP 1 led to the deployment in February, 2007 of the first distributed model for operational forecasting in the NWS (Schmidt et al., 2007; Cooper, 2004; Shultz and Corby, 2004). The NWS distributed model has shown cases of improved operational river forecasts (e.g., Jones et al., 2009). At the same time, the results of DMIP 1 allowed the NWS to implement the distributed model with realistic expectations regarding distributed model improvement compared to lumped models. In addition, the NWS and other participants used DMIP 1 to identify model shortcomings and improve their models (e.g., Mascaro et al., 2010; Ivanov et al., 2008; Gassman et al., 2007; Koren et al., 2010, 2006; Di Luzio and Arnold, 2004). The DMIP 2 Oklahoma (OK) experiments were conducted between 2005 and 2007, concluding with a participants' workshop at NWS headquarters in September, 2007.

The purpose of this paper is to introduce the overall scope of DMIP 2 and provide background information for other papers in this special issue. Additional details can be found on the DMIP 2 website: http://www.nws.noaa.gov/oh/hrl/dmip/2/. While DMIP 2 encompassed two geographic domains, this paper focuses on the research-to-operations questions and experiments in the Oklahoma region basins. Subsequent papers will address the issues and experiments in the western DMIP 2 basins.

This paper is organized as follows. Section 2 introduces the need for DMIP 2 from an NWS perspective followed by a discussion of science issues and knowledge gaps. Also highlighted are the differences compared to DMIP 1. The pertinent science-to-operations questions are listed in Section 3. The test basins are described in Section 4. Data for DMIP 2 are described in Section 5. Section 6 presents the modeling experiments. The models used in DMIP 2 are described in Section 7. Section 8 introduces the results covered in Smith et al. (this issue) and then Section 9 summarizes the paper and discusses NWS perspectives.

2. Need for DMIP 2

2.1. NWS motivation

As with DMIP 1, the NWS realized the need for an accelerated phase of science experiments to guide its implementation of advanced hydrologic models for river, flash flood, and water resources forecasting. This was accentuated by the need to develop a growing list of water resources forecasting products in addition to the more traditional river, flash flood and water supply forecasting mission (NWS, 2004; McEnery et al., 2004). The need for water resources forecasting is based on end-user requests and the recommendations of the National Research Council (NRC) that point to hydrologic forecasting as one of the ten 'grand challenges' in environmental sciences in the next generation (NRC, 2000). For example, the NWS is very interested in producing high spatial resolution soil moisture products (Koren et al., 2006; Moreda et al., 2005) which have been shown to be economically beneficial (e.g., Georgakakos and Carpenter, 2006; Torell et al., 2011). To this end, the NWS sought input from the hydrologic research community.

Smith et al. (2004a) listed a set of initial requirements for NWS operational distributed modeling. We expand that list here as additional background for the DMIP 2 Oklahoma experiments.

(a) The distributed model should be computationally feasible in real time. Any model used at NWS offices for operational forecasting must run efficiently.
(b) The model should be amenable to manual and/or automatic data assimilation to keep model states on track.
(c) It should be amenable to uncertainty analysis via ensembles or other means. The NWS is actively implementing approaches to quantify the uncertainty of their lumped-model river forecasts (e.g., Seo et al., 2010). Distributed models will need to fit into such a framework.
(d) The distributed model should have effective parameter estimation and calibration schemes that expedite model implementation. Efficient schemes are necessary given that the NWS must implement models for river, flash flood, and water resources prediction for the entire Nation. Efficient schemes also enhance the use of the model calibration process as an effective step in training operational hydrologic forecasters (Smith et al., 2003).
(e) The distributed model should perform at least as well in an overall sense as the current operational lumped model, while providing improvements in basin outlet simulations in cases of pronounced spatial variability of precipitation and basin features.
(f) The distributed models should provide accurate hydrologic information at ungauged points. For example, distributed models calibrated at the basin outlet should provide accurate estimates of soil moisture at interior ungauged locations.

After DMIP 1, the NWS developed a version of its research distributed model that could be used in operations and made it available to NWS River Forecast Centers (RFCs) in February 2007 for river stage forecasting. To date, the distributed model has been used in the non-snow southern portions of the USA. Expanded operational deployment of distributed models into other parts of the USA requires that the NWS investigate issues not covered in DMIP 1 such as high spatial resolution snow accumulation and melt, sparse data networks, orographically enhanced precipitation, rapidly varying terrain features, and others.

2.2. Scientific background

While DMIP 1 served as a successful comparison of lumped and distributed models, it also highlighted problems, knowledge gaps, and topics that needed to be investigated. Moving forward after DMIP 1, OHD believed there was a continued need to provide the academic community with an opportunity to test research models using operational quality data, providing a means to identify techniques that may be suitable for operational forecasting. Even beyond DMIP 2, OHD intends to maintain the data availability so that DMIP 2 participants and others may use them for future research and development.

The new aspects of the Oklahoma experiments in DMIP 2 were designed to advance our knowledge beyond what was learned from DMIP 1. In DMIP 2, ten models participated that were not in DMIP 1, providing the scientific and operational community with an expanded set of results on the comparison of lumped and distributed models (see Smith et al., this issue). Moreover, DMIP 1 was limited by a relatively short data record containing only a few significant rainfall–runoff events during the verification period from which statistics could be computed and inferences made. Thus, the need remained for further DMIP 1-like testing in order to more rigorously evaluate the hypotheses related to lumped and distributed modeling. At the launch of DMIP 2, several years of more recent data were available to support such comparisons. Also, DMIP 1 was somewhat hampered by the quality of the multisensor estimates of observed precipitation. These data problems led some DMIP 1 participants to report problems with calibration (Reed et al., 2004). The quality of these multisensor data has been much studied in the DMIP 2 basins and elsewhere (e.g., Westcott et al., 2008; Xie et al., 2006; Jayakrishnan et al., 2004; Stellman et al., 2001; Young et al., 2000; Wang et al., 2000; Smith et al., 1999; Johnson et al., 1999) and researchers have identified problems such as underestimation and non-stationarity resulting from changes in the raw data processing algorithms (Young et al., 2000; see also ‘About the Multisensor (NEXRAD and gauge) Data’, http://www.nws.noaa.gov/oh/hrl/dmip/2/docs/about_multisensor.pdf). These known deficiencies were exacerbated by the typically short period of record of the multisensor precipitation products. Avoiding these problem-prone periods often leaves an insufficient period of high-quality data for model calibration. To alleviate these problems, work is underway to generate a consistent high-quality reanalysis of the radar multisensor precipitation estimates (Nelson et al., 2010, 2006). DMIP 2 did not use precipitation data from three early years with known underestimation problems but included data from more recent years.

One of the greatest challenges of distributed modeling is the prediction of hydrologic variables over a range of spatial scales and at ungauged interior locations. To address this challenge, a distributed model should reasonably well represent the heterogeneities of watershed properties through its modeled processes, structure and parameters. Unfortunately, limitations in the availability of spatial data often reduce model evaluation to a simple comparison of modeled and observed streamflow at the gauged outlet (Reed et al., 2004) and greatly impede an evaluation of the spatial correctness of model parameters and outputs.

If distributed models can reliably represent processes at basin interior points, then these models can be used to generate products such as flash-flood forecasts or spatially variable information such as soil moisture estimates for agriculture (e.g., Georgakakos and Carpenter, 2006; Torell et al., 2011). Alternatively, success at modeling interior points provides confidence that the models are producing the right answer at a stream gauge for the right reasons upstream (Kirchner, 2006). For example, Koren et al. (2008) demonstrated that calibration using soil moisture observations in addition to streamflow can result in more confidence in the *a posteriori* model parameters because more basin processes are being represented.

DMIP 1 attempted to address this challenge through blind simulations of nested and basin interior observed discharges at a limited number of sites. Reed et al. (2004) reported that some DMIP 1 models had success at predicting interior streamflow without specific calibration at that point. DMIP 2 revisited this question but enhanced the investigation in two ways. First, more interior stream gauges were located in the study basins. Table 1 shows the new stream gauges for DMIP 2. The gauge on the Illinois River south of Siloam Springs, Arkansas (AR) allowed us to forego using the data from the USGS gauge at Watts, OK used in DMIP 1. The Watts gauge is downstream of the ruins of a dam and small lake known as Lake Frances on the Illinois River. Analysis of the streamflow at the gauges upstream (Siloam Springs) and downstream (Watts) of this dam showed regulation effects which may have complicated the analyses of model results (see Section 4.2.2).

Moreover, DMIP 2 expanded the analysis of interior process representations to include spatial comparisons of simulated and observed soil moisture. Investigations using soil moisture data have typically been hampered by a lack of reliable observations organized at a high spatial resolution. While much work has been done to estimate soil moisture from satellites, these methods are currently limited by observations of only the top few centimeters of the soil surface. The test basins in DMIP 1 are mostly contained in Oklahoma, offering an opportunity to use the soil moisture observations from the Oklahoma Mesonet (Brock et al., 1995; Illston et al., 2004). In this network, over 100 soil moisture sensors were installed at depths of 5, 25, 60, and 75 cm. These depths were selected to enhance agricultural and meteorological modeling, facilitate drought monitoring, and to generate research-quality data sets. Appendix A presents more details regarding the sensors and data. Recent work has shown the validity of using these data to detect droughts (Illston et al., 2008; Illston et al., 2004, 2003; Illston and Basara, 2002), evaluate distributed model performance (Koren et al., 2006), and for multivariable calibration of lumped models (Koren et al., 2008) as well as for other major studies (e.g. NLDAS, Mitchell et al., 2004). Koren et al. (2006) presents a comparison of computed and observed soil moisture using the Mesonet data. Fortin (1998) provides a good example of such experiments with the Sacramento model. Schaake et al. (2004) inter-compared NLDAS model-generated soil moisture fields with each other and with available observations. The NLDAS soil moisture estimates were generated on a 1/8th degree grid, which is too coarse for the planned NWS water resources forecast products.

Table 1
USGS stream gauges and basin drainage areas for the Oklahoma region basins. The italized areas denote additional gauges that were not used in DMIP 1.

No	USGS no	Name	DMIP-2 ID	Latitude (°)	Longitude (°)	Area (km^2)
1	7332500	Blue R. nr Blue, OK	BLUO2	33.99694	−96.24803	1233
1a	*7332390*	*Blue R. near Connerville, OK*	*CONNR*	*34.38333333*	*−96.60027778*	*419.6*
2	7196500	Illinois River near Tahlequah OK	TALO2	35.92286889	−94.9235658	2484
3	7197000	Baron Fork at Eldon OK	ELDO2	35.92120028	−94.8385633	795
4	7196973	Peacheater Creek at Christie OK	PEACH	35.95480806	−94.6963369	65
5	7196000	Flint Creek near Kansas OK	KNSO2	36.1864725	−94.7068914	285
6	7195500	Illinois River near Watts OK	WTTO2	36.13008	−94.57216	1645
7	7194800	Illinois River at Savoy AR	SAVOY	36.10313567	−94.34437763	433
8	7189000	Elk River near Tiff City Mo	TIFM7	36.63146139	−94.5868886	2258
9	*7188653*	*Big Sugar Creek near Powell MO*	*POWEL*	*36.615872*	*−94.182222*	*365*
10	*7188885*	*Indian Creek near Lanagan MO*	*LANAG*	*36.599275*	*−94.44965*	*619*
11	*7194880*	*Osage Creek near Cave Springs AR*	*CAVES*	*36.28146623*	*−94.22798384*	*90*
12	*7195000*	*Osage Creek near Elm Springs AR*	*ELMSP*	*36.22202302*	*−94.28854149*	*337*
13	*7195430*	*Illinois River South of Siloam Springs AR*	*SLOA4*	*36.10869244*	*−94.53355206*	*1489*
14	*7195800*	*Flint Creek at Springtown AR*	*SPRIN*	*36.25563475*	*−94.43394*	*37*
15	*7195865*	*Sager Creek near West Siloam Springs OK*	*WSILO*	*36.2017483*	*−94.6052206*	*49*
16	*7196900*	*Baron Fork at Dutch Mills AR*	*DUTCH*	*35.880092*	*−94.486606*	*105*

Observed soil moisture data were taken from the Illinois State Water Survey. These data were collected twice per month. Schaake et al. (2004) found better agreement between observed and simulated ranges of water storage variability than between observed and simulated amounts of total water storage. Despite the spatial density limitations of the Oklahoma Mesonet, (e.g., one sensor per county) and other issues (Illston et al., 2004; Basara and Crawford, 2000) it is prudent to perform experiments to understand the real value of these currently available data and work towards developing requirements for future sensor deployment.

Reed et al. (2004) noted the uncertainty in the performance of DMIP 1 models in the smallest basin (65 km^2) and called for tests with nested basins of various sizes. We addressed this need in DMIP 2 by locating two interior basins with drainage areas of 39 and 45 km^2.

Continued research is necessary to develop and refine distributed models and their parameter estimation and calibration schemes. Effective schemes are especially critical for operational deployment of distributed models for real-time forecasting. We consider parameter estimation and calibration as distinct but linked processes (Reed et al., 2004; Koren et al., 2003b; Madsen, 2003; Refsgaard, 1997). For the DMIP 2 experiments, we define parameter estimation as the derivation of *a priori* estimates of model parameters from physical properties of the basin such as soil texture. Calibration is the process of refining the *a priori* (or other initial) parameters so that an acceptable level of error is achieved between simulated and observed hydrologic variables. In DMIP 2, participants were asked to generate uncalibrated and calibrated simulations at basin outlets and at blind interior points. Participants were free to use any parameter estimation/calibration scheme they desired, resulting in a wide array of approaches. The improvement gained by model calibration is quantified in the discussion of DMIP 2 results (Smith et al., this issue). The following paragraphs highlight some of the issues and recent advances in these areas.

Parameter estimation approaches for lumped models tend to be hydrograph driven with some physical reasoning (e.g. Anderson (2002) for the SAC-SMA). On the other hand, efforts to parameterize distributed models tend to put more emphasis on physical reasoning, but researchers have found that results can still be improved through hydrograph-driven calibration (e.g., Reed et al., 2004).

Model parameter estimation has received a great deal of attention in recent years, aided by the development of soil texture and other data sets of physical basin attributes. Examples here include the derivation of *a priori* estimates of the parameters for the SAC-SMA model (Anderson et al., 2004; Koren et al., 2000), the Precipitation Runoff Modeling System (PRMS, Leavesley et al., 2003), the Hydrologic Research Center Distributed Hydrologic Model (HRC-DHM, Carpenter and Georgakakos, 2004), a version of the Variable Infiltration Capacity (VIC) model (Abdulla et al., 1996), a version of TOPMODEL, (Ao et al., 2004) and others. *A priori* parameter estimation schemes produce spatially consistent parameter sets (Bastidas et al., 2003; Seibert and McDonnell, 2003). Moreover, *a priori* parameterization schemes provide a cost-effective and physically realistic approach to model implementation for operational forecasting.

Calibration techniques for distributed models are less mature compared to lumped models due to the large number of parameters involved and our incomplete knowledge of the actual physical processes in the heterogeneous landscape. Calibration approaches that are both efficient and take full advantage of available physical data continue to be elusive in spite of the high level of activity in this arena (Campo et al., 2006). Approaches developed to date can be placed into one of several groups. First, one strategy is to use scalars to uniformly adjust (automatically or manually) the parameters in each grid or computational element in a watershed (e.g., McMillan et al., 2008; Francés et al., 2007; Koren et al., 2003c, 2004; Eckhardt et al., 2005; Bandaragoda et al., 2004; Leavesley et al., 2003; Giertz et al., 2006; Jinkang et al., 2007; White et al., 2003). This approach is acutely dependent on the use of effective *a priori* parameterization schemes based on physical basin characteristics to reduce the high dimensionality of the problem (Reed et al., 2004; Leavesley et al., 2003; Koren et al., 2000, 2003b, 2004; Senarath et al., 2000). The premise of using scalar adjustment factors is that there is value in preserving the spatial variation of the physical information as reflected in a model's parameters and is predicated on the ability to generate meaningful *a priori* model parameters from soils and other physical data sets. Koren et al. (2004) and Carpenter and Georgakakos (2004) advocate an additional check when calibrated lumped parameters are available. In this step, the distributed *a priori* parameters are scaled to agree with the lumped calibrated parameters.

Another approach is to calibrate models by focusing on events having spatially non-uniform rainfall (e.g., McMillan et al., 2008; Ivanov et al., 2004). In such cases, only the parameters of the 'active' computational elements (i.e., receiving rainfall) are adjusted. Similarly, Mascaro et al. (2010) and Vivoni et al. (2006) conducted studies in which a 'nested basin' calibration approach was used for the basin above the USGS gauge on Baron Fork at Eldon, OK (DMIP 2 identifier ELDO2). In these studies, sub-basin- and event-specific calibration was performed to focus on the impacts of nowcasting and ensemble forecasting. Starting with the long term basin-outlet calibration at ELDO2 from Ivanov et al. (2004) for DMIP 1, the parameters were adjusted in a 'nested basin' or 'multistream gauge' approach in which the parameters of the two sub basins were modified prior to calibrating the main basin response in two events.

Still others attempt to use multivariable and multisite measurements (e.g., hydrographs and aquifer height, Madsen and Kristensen, 2002); hydrographs and sedimentographs (Kalin and Hantush, 2006) in a multiple-objective minimization framework. Step-wise calibration of individual processes has also been developed (Vieux and Moreda, 2003; Rousseau et al., 2003). Campo et al. (2006) used automatic calibration to determine parameters to fit two objective functions: minimizing the error in simulated gauged flow and soil saturation indexes. Recent advances in performance measures include the use of pattern agreement measures such as the Hausdorf metric (Tcherednichenko et al., 2004; Bastidas and Li, 2004; Marron and Tsybakov, 1995), the earth movers distance (Kim et al., 2010) and the Information Mean Squared Error (IMSE, Wealands et al., 2004) to minimize the differences between observed and simulated time-varying spatial patterns of hydrologic variables.

Regularization is an approach to reduce the dimensionality of the calibration problem. In simple terms, regularization (sometimes called a 'bottom-up' approach) is a strategy that utilizes additional information about the model parameters to construct constraints in the form of equations to simplify an ill-posed inverse problem (e.g., Pokhrel et al., 2008; Marce et al., 2008). Doherty and Skahill (2006) used regularization in the simultaneous calibration of five subwatershed models. Their method allowed for inter-subwatershed parameter variation, and achieved better flow simulation than without such parameter variation. 'Top-down' approaches such as regionalization have also been used to parameterize distributed models (e.g., Gotzinger and Bardossy, 2007). Other studies highlight progress towards computationally feasible approaches which, when combined with effective *a priori* parameters, regularization schemes or other methods to reduce problem dimensionality, can be used to optimize distributed model parameters in practical applications (e.g., Kuzmin et al., 2008; Goswami and O'Conner, 2007; Bastidas and Li, 2004).

Yet another major need is the testing of models in a 'pseudo-forecast environment' with forecast-quality forcing data. Such tests are a logical complement to the process simulation experiments in DMIP 1. While much work has been done to evaluate the improvements realized by distributed models in historical simulation mode, the NWS also needs to investigate the potential gains when used for hydrologic forecasting. The well-documented model intercomparsion experiment of the World Meteorological Organization (WMO, 1992) highlighted the testing of models in a forecasting environment. One of the conclusions of that workshop was that good simulation (process) models are necessary for longer lead-time forecasts. In DMIP 1, process models were tested in simulation mode and thus satisfied this conclusion from the WMO experiment. Initial DMIP 2 plans called for a forecast test component as a natural complement to the process experiments in DMIP 1. Georgakakos and Smith (1990) argued for such an experiment as follow-on work to the WMO model comparisons of the 1980s, stating that the rainfall input component of the input uncertainty contributes the most to prediction uncertainty. The need for a forecast component in DMIP 2 is mentioned here but we will wait to execute this experiment and report on its findings at a future date.

Continued work is needed to address the question: can basins be identified *a priori* that would show gains from distributed models compared to lumped models for forecasting at the basin outlet. (We have already commented on the ability of some distributed models to provide useful information at interior points). Such identification procedures might help guide operational agencies in the efficient implementation of distributed models. While this question was not explicitly investigated via DMIP 1 modeling instructions, it was nonetheless a good opportunity to explore this question. Smith et al. (2004b) and Koren et al. (2003a) used the DMIP 1 observed streamflow and precipitation data in an attempt to derive diagnostic indicators to assess the potential benefit of distributed models before the model is applied. Distinct differences in precipitation spatial variability and basin behavior were identified. Yet, no threshold values of the indices could be derived. DMIP 2 addressed this question by providing several more years of observed precipitation and streamflow data to continue the types of empirical analyses performed by Smith et al. (2004b), Koren et al. (2003a) and others. Li and Sivapalan (2011) used these data to investigate the spatial variability of runoff generation in the Illinois River basin. They noted counter-intuitive behavior in that basin response times were slower under wet (saturation excess runoff) conditions than under dry (subsurface flow dominated) conditions. In light of these and other studies, continued work is necessary to understand the interaction of spatial variability of precipitation, basin features, and runoff generation to warrant the use of a distributed model.

3. Science questions

The following science questions were proposed for the Oklahoma region experiments of DMIP 2. Some of these were repeated from DMIP 1 in order to evaluate them given longer archives of higher quality data than were available in DMIP 1. The science questions are framed for the interest of the broad scientific community and in most cases include a corollary to address the distributed modeling requirements for NWS and other operational forecast agencies (e.g., Rousseau et al., 2003).

Can distributed hydrologic models provide increased simulation accuracy compared to lumped models? This question was one of the dominant questions in DMIP 1. Reed et al. (2004) found that lumped models outperformed distributed models in more cases than distributed models outperformed lumped models. The specific question for the NWS mission is: under what circumstances should NWS use distributed hydrologic models rather than (or in addition to) lumped models to provide hydrologic services?

What simulation improvements can be realized through the use of a more recent period of radar precipitation data than was used in DMIP 1? What is the impact of calibrating a distributed model with temporally inconsistent multisensor precipitation observations? One of the issues faced in DMIP 1 was the time-varying biases of the multisensor precipitation data (Reed et al., 2004) which affected the simulations in the model calibration and verification periods. DMIP 2 did not use the problematic 1993–1996 period of radar data. Simulations and analyses were based on the period starting in 1996. For the NWS, the question is whether using this later (and less bias-prone) period of data can lead to improved calibrations and simulations.

Can distributed models reasonably predict processes such as runoff generation and soil moisture re-distribution at interior locations? At what scale can soil moisture models be validated given current models and sensor networks? For the NWS, the corollary question is: can distributed models provide valuable, spatially-varied estimates and operational predictions of soil moisture, soil temperature and runoff?

In what ways do routing schemes contribute to the simulation success of distributed models? Can the differences in the rainfall–runoff transformation process be better understood by running computed runoff volumes from a variety of distributed models through a common routing scheme? One of the recommendations from DMIP 1 was to separate the comparisons of routing and rainfall runoff techniques (Reed et al., 2004). Such experiments are necessary complements to validating distributed models with interior-point flow and soil moisture observations in an attempt to generate 'the right results for the right reasons.' Lohmann et al. (2004, 1998) present large scale examples of such a test.

What is the potential for distributed models configured for basin outlet simulations to generate meaningful hydrographs at interior locations for flash flood forecasting? Inherent in this question is the hypothesis that better outlet simulations are the result of accurate hydrologic simulations at points upstream of the gauged outlet. This question is repeated from the DMIP 1 experiments. Reed et al. (2004) identified reasonable performance for small ungauged areas. For the NOAA/NWS, the question is: can distributed runoff and flow predictions for small, ungauged locations be used to improve upon the existing NWS flash flood forecasting procedure?

What combination parameter estimation schemes and calibration strategies seem to be most effective and what is the level of effort required? As in DMIP 1, the DMIP 2 modeling instructions specified that participants were to generate uncalibrated and calibrated simulations. For operational agencies, it is important to weigh any simulation improvements gained in the process of calibration against the effort required.

Two additional science questions were identified in the DMIP Science Plan. Unfortunately, we were not able to address them in DMIP 2. Nonetheless, they are included here for the interest of the scientific community.

What is the performance of (distributed) models if they are calibrated with observed precipitation data but use forecasts of precipitation? Is there a forecast lead time at which the distributed and lumped model forecasts converge? How far out into the future can distributed models provide better forecasts than currently used lumped models? Because forecast precipitation data have a lower resolution and are much more uncertain than their observed counterparts, the benefits of distributed models may diminish for longer lead times.

What combinations of physical characteristics (shape, feature heterogeneity) and rainfall variability warrant the use of distributed hydrologic models for improved basin outlet simulations compared to lumped models? Can basins be identified beforehand that would *realize gains from distributed model application?* The corollary question for the NWS is: at what river forecast points can we expect distributed models to effectively capture essential spatial variability so as to provide better simulations and forecasts than the current lumped model?

4. Description of test basins

4.1. Overview

Fig. 1 shows the two major geographic regions for the experiments conducted in DMIP Phase 2. As seen in Fig. 1, the Oklahoma region and basins in DMIP 1 were used in DMIP 2. Two neighboring basins in the western USA were selected as good candidates for distributed model tests in hydrologically complex areas. The western DMIP 2 basins are mentioned here with the intent of providing a more complete description of the western basin experiments in a planned journal paper.

4.2. Oklahoma region

As in DMIP 1, the Blue River, the Elk River, and the Illinois River basins were used for specific tests regarding lumped and distributed models. Table 1 and Fig. 2 present the additional stream gauges identified for use in DMIP 2. For tests related to soil moisture, DMIP 2 used a 'synthetic basin' encompassing the entire state of Oklahoma with its Mesonet series of soil moisture observations shown in Fig. 3 and described in Appendix A. Smith et al. (2004a) present a description of the Illinois, Elk, and Blue River basins and the rationale for their selection for lumped and distributed model comparisons. Readers are referred to Smith et al. (2004a) for the details of these basins so as to avoid undue repetition here. However, recently-obtained background information on two of the basins is provided here.

4.2.1. Blue River at Blue, OK

The Blue River basin has a unique shape and response characteristics among the Oklahoma region basins, and displayed benefits from distributed modeling in DMIP 1 (Reed et al., 2004). However, several DMIP 1 and 2 participants noted odd behavior in this basin, such as rapidly rising and falling hydrographs and water balance problems. We conjectured the presence of additional processes that complicate the response of the basins, and have since obtained information which sheds lights on the complex response (Osborn, 2009; OWRB, 2003; Fairchild et al., 1990). As shown in Fig. 4, the upper reaches of the Blue River are underlain by the Arbuckle-Simpson aquifer, and thus are affected by sinkholes and complicated by sections that gain and lose water. Moreover, the largest spring in Oklahoma (Byrd's Mill Spring, USGS gauge 07334200) removes water from the basin and discharges it to the northeast. This spring discharges about 0.57 $m^3\ s^{-1}$ and is the primary source of water for the city of Ada, OK (Osborn, 2009). Fig. 4 also shows the numerous springs above Connerville, OK. Finally, the location of the Blue River is apparently controlled by lithological variability and not by the potentiometric surface (Todd Halihan, OK State U., personal communication).

4.2.2. Illinois River at Siloam Springs, AR

The USGS gauge on the Illinois River at Watts, OK used in DMIP 1 was not used in DMIP 2 as it is downstream from a weir and ruins of a small dam. Instead, DMIP 2 took advantage of the newly installed gauge just upstream at Siloam Springs, OK to avoid any attenuating effects of the structure. The drainage area of this basin is only slightly smaller than the Illinois River at Watts, OK (1489 km^2 versus 1645 km^2). Fig. 5 compares the observed hourly streamflow at the gauges downstream (Watts) and upstream (Siloam Springs) of this weir and ruins for a storm event in June, 2000. The initial peaks on the rising limbs of the hydrograph show the effects of flow entering the main channel from tributaries downstream of the structure and just upstream of the Watts gauge. However, the main Watts hydrograph peaks lower and later than the main Siloam Springs hydrograph, indicating that the structure has at least some attenuating effects on the flow.

5. Data

OHD encouraged participation in DMIP 2 by providing data sets (or links to them), processing algorithms, test cases, and documentation. Some of these data are repeated from DMIP 1 (e.g., DEM data). The following sections highlight the changes and additions for DMIP 2.

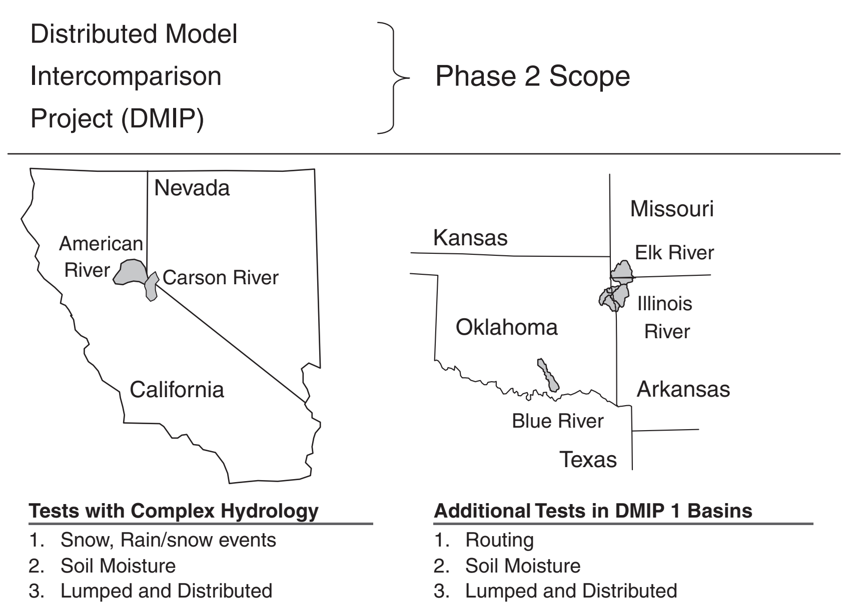

Fig. 1. The geographic scope and hydrologic investigations in DMIP 2.

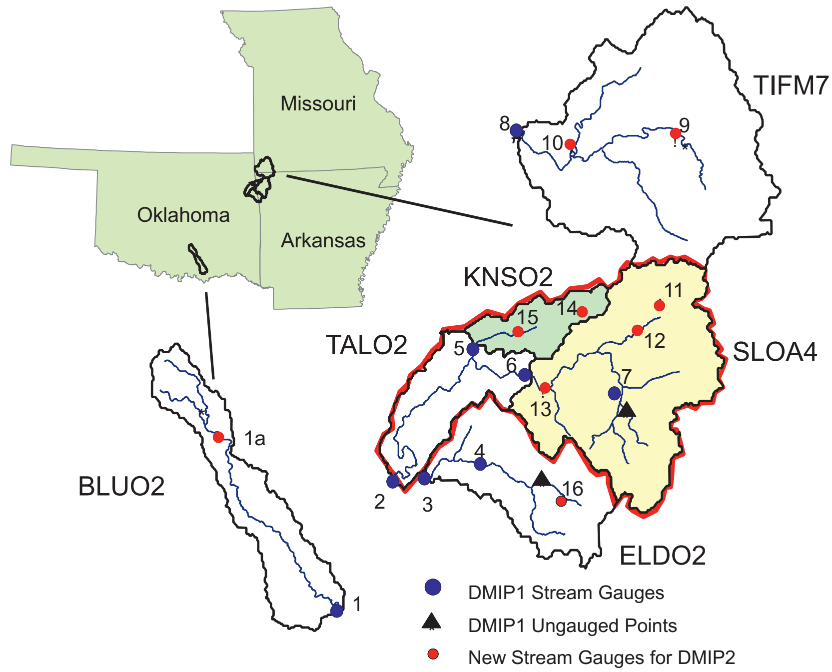

Fig. 2. Location of DMIP test basins and interior computational points in the Oklahoma, Missouri, and Arkansas area. Note that additional gages have been located for DMIP 2. The red line indicates the outline of the TALO2 basin. The yellow shaded area is the SLOA4 basin and the green shaded area is the KNSO2 basin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

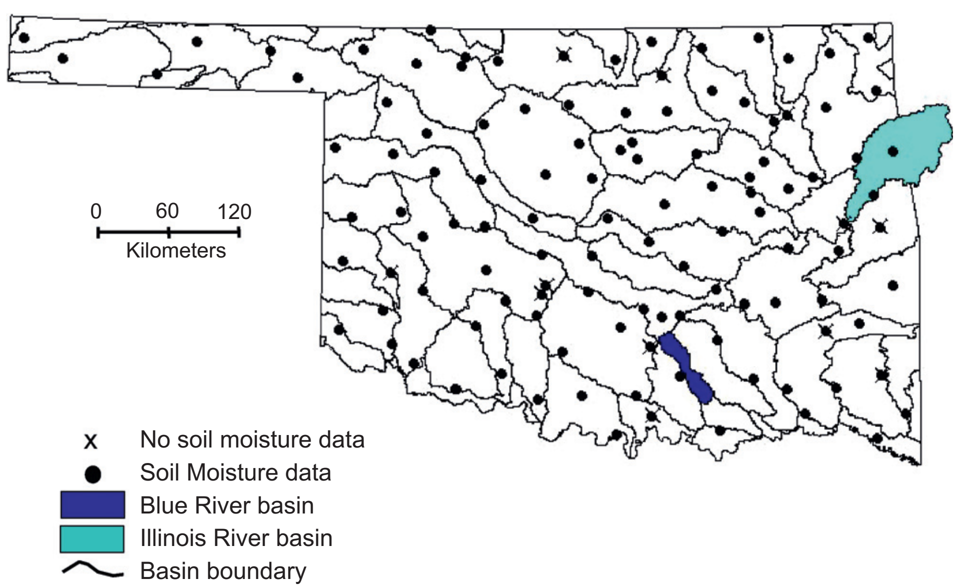

Fig. 3. Location of Oklahoma Mesonet sites as they relate to the test basins in DMIP 2.

5.1. Multisensor precipitation data

One of the goals in DMIP 2 was to provide a more consistent data set of multisensor precipitation observations. In DMIP 1, the period of data was from May, 1993 through July, 2000, encompassing an interval of known underestimation and algorithmic changes from 1993 to 1996 (Reed et al., 2007, 2004; Young et al., 2000; see also 'About the Multisensor (NEXRAD and gauge) Data', http://www.nws.noaa.gov/oh/hrl/dmip/2/docs/about_multisensor.pdf). To avoid these problematic data in DMIP 2, a later period was selected from the archive of operational radar-gauge precipitation from the NWS Arkansas-Red Basin River Forecast Center (ABRFC). Data were provided from October 1, 1995 to September 30, 1996 for a 'warm-up' period to allow participants' models to stabilize before the calibration period of October 1, 1996 to September 30, 2002. The verification period spanned from October 1, 2002 to September 30, 2006. Whereas the ABRFC relied on the Stage III algorithm for radar precipitation estimation prior to 1996, they adapted the use of a locally-developed algorithm Process1 (P1) to create the vast majority of the multisensor precipitation products starting in late 1996. In Stage III, the multisensor precipitation products for individual radars are combined near the end of the process to cover an entire RFC domain. In P1, the radar-only precipitation fields are first combined to cover the RFC area, followed by bias correction with gauge observations to derive a multisensor precipitation estimate. Young et al. (2000) provide a thorough description of the Stage III and P1 processes. The Hydrologic Rainfall Analysis project (HRAP) grid defines the spatial resolution of

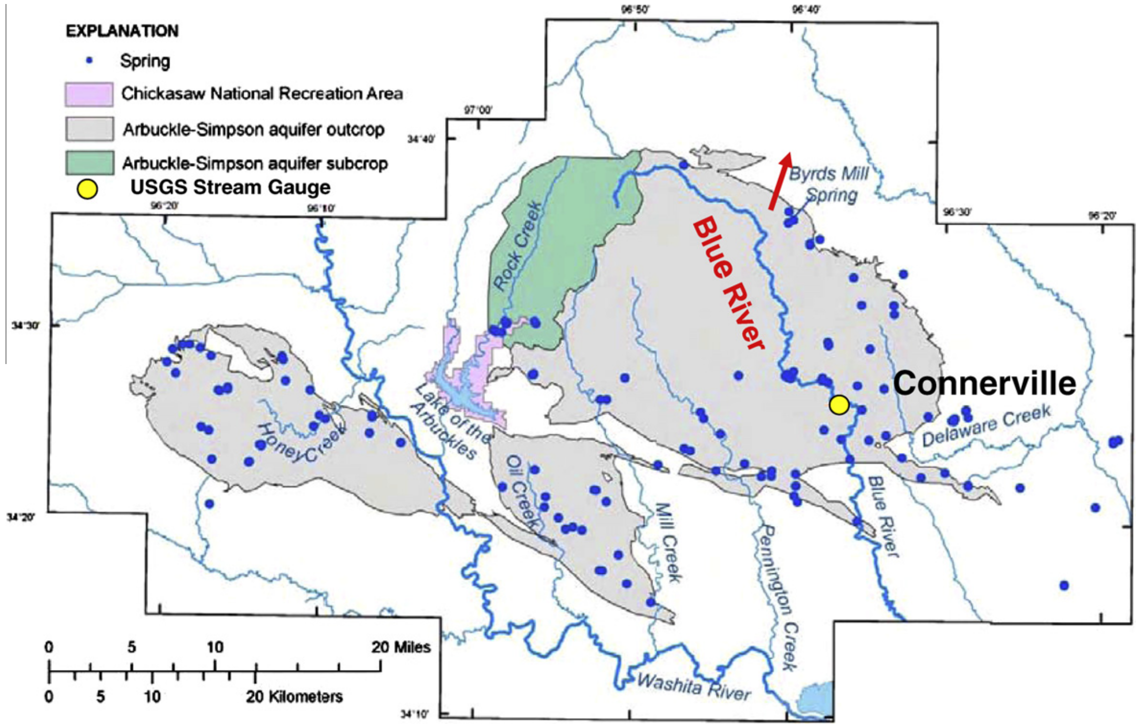

Fig. 4. The location of the Arbuckle-Simpson aquifer showing the Blue River and Byrd's Mill Spring. Blue dots indicate the location of springs. The red arrow indicates the direction of flow from Byrd's Mill Spring out of the Blue River basin. The USGS stream gauge at Connerville, OK is shown. Copyright: Oklahoma Water Resources Board. Reproduced with permission.

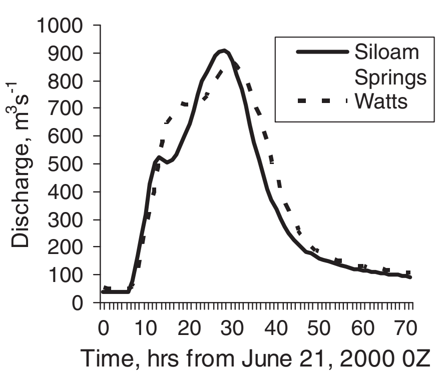

Fig. 5. Observed streamflow for June 21–23 event for the USGS gauges on the Illinois River at Siloam Springs, AR and Watts, OK.

these precipitation estimates. Nominally, the HRAP grid size is 4 km by 4 km, but the actual cell size varies with latitude. Interested readers are referred to Green and Hudlow (1982) and Reed and Maidment (1999) for more information about the HRAP grid definition.

5.2. Flow direction data

Flow direction grid files at several resolutions (or links to the agencies that provide them) were provided for the convenience of any participants who wished to use them. In these grids, each cell contains an integer indicating one of eight possible flow directions. The 30-m DEM flow direction grids were produced by the National Severe Storms Laboratory (NSSL) using the Jenson and Domingue (1988) algorithms implemented via commercial GIS software.

Although 30-m flow directions are more accurate and were used to derive the basin boundaries provided on the DMIP 2 web site, 400-m flow directions grids were also provided. The 400-m flow direction grids were derived using 400-m DEMs and digitized streamline files. Flow direction grids for the HRAP and ½ HRAP grid cells used to map multisensor precipitation data were derived using DEM data and DEM derivatives using the method described by Reed (2003).

5.3. Observed streamflow

Provisional observed hourly discharge data were obtained from USGS personnel in Oklahoma, Arkansas, Kansas, and Missouri. These data were quality controlled at OHD by checking against the published mean daily flow data from the USGS. Suspect data were set to missing. The data were converted to Greenwich Mean Time (GMT) to correspond to the multisensor precipitation data.

5.4. Cross section data

In addition to the bridge cross section data used in DMIP 1, surveyed cross section information were provided for 26.4 km on Flint Creek and 56.3 km on the Illinois River from near the Watts gauge downstream to the Tahlequah gauge. These data were obtained from a study for the Southwestern Electric Power Co. by the consulting firm *Freeze and Nichols*.

5.5. Meteorological data

Two sources of meteorological data were provided in DMIP 2 so that participants could compute daily estimates of potential evaporation if desired. The first source was the North American Regional Reanalysis (NARR) data set (Mesinger et al., 2004, 2006). This data set was selected primarily as it covered the basins in both the Oklahoma and western regions and has been used in other studies (e.g., Lei et al., 2007). The NARR data are a long term, consistent, assimilation-based climate dataset for the North American domain. The data are available at a 3 h temporal resolution and a 32 km spatial resolution. The data are generated via a 'frozen' version of the Eta regional numerical weather prediction model combined with the Noah land surface model. The data cover the 25-year period 1979–2003, and are being continued in near-real time as the Regional Climate Data Assimilation System (R-CDAS). Data assimilated into the model consist primarily of free atmospheric variables such as upper-air temperature, pressure heights, and humidity from rawinsondes, dropsondes, and satellite retrievals. Output fields include surface variables such as precipitation, temperature, wind, relative humidity, and radiative fluxes.

In addition, OHD provided a link to the surface downward short wave (SW) radiative flux data from the NOAA National Environmental Satellite, Data, and Information Service (NESDIS). These data are derived from GOES satellite observations. The NESDIS SW data are stored and distributed using a server at the University of Maryland, College Park.

As in DMIP 1, climatic monthly mean values of potential evaporation (PE) were also provided in mm/day for the study basins. These values were derived from free water surface (FWS) evaporation maps contained in NOAA Technical Report 33 (Farnsworth et al., 1982) and mean monthly station data in NOAA Technical Report 34 (Farnsworth and Thompson, 1982).

5.6. Location of Oklahoma Mesonet sites

DMIP 2 provided a list of the latitude/longitude coordinates of the Oklahoma Mesonet stations. The participants generated simulations of volumetric soil moisture for two depth ranges at these locations for analysis. While OHD obtained the observed soil moisture values from the Mesonet organizers, these were not made available to participants as they are proprietary. Appendix A contains details of the soil moisture sensors used in the Mesonet.

5.7. Soil texture information

In addition to the State Soil Geographic (STATSGO) data provided in DMIP 1, OHD provided a link to finer resolution county-level soil information called the Soil Survey Geographic (SSURGO) data set. The SSURGO data are typically available at a scale of at least 1:24,000. They are approximately ten times the resolution of STATSGO data in which the soil polygons can be on the scale of 100–200 km^2.

6. Overview of modeling experiments

DMIP 2 defined specific modeling tests to investigate the science questions discussed earlier. Participants were required to follow explicit instructions for generating the required simulations. A brief overview is provided here while the full modeling instructions can be found in http://www.nws.noaa.gov/oh/hrl/dmip/2/docs/ok_modeling_instructions.pdf. In the papers that follow in this special issue, readers can refer to this site for the interpretation of results.

In all experiments, participants were required to use the operational multisensor precipitation data provided via the DMIP 2 web site. This requirement was established to help NWS evaluate the models when forced with operational (versus research)-quality data. As in DMIP 1, no state updating was allowed. Model runs were generated in historic simulation mode in order to evaluate the basic ability of models to simulate basin processes.

6.1. Simulation experiments: lumped and distributed models

These tests essentially followed the DMIP 1 Project Design and Modeling Instructions (Smith et al., 2004a). Calibrated and un-calibrated simulations from participants' distributed models were tested against observed streamflow and corresponding lumped-model simulations. Models were to be set up so that simulations were generated at the gauged outlet and at specific gauged points in the basin interior. However, no calibration was allowed using the interior flow data in order to evaluate the predictive capability of distributed models at ungauged points. As in DMIP 1, participants' streamflow simulations were evaluated against observed hourly flow data as well as the lumped model simulations from the NWS operational Sacramento Soil Moisture Accounting Model (SAC-SMA; DMIP 2 acronym LMP) and the lumped simulations from the French participants at Centre d'etude du Machinisme Agricole du Génie Rural des Eaux et Forêts; (CEMAGREF; DMIP 2 acronym CEM). A "warm-up" period from October 1, 1995 to September 30, 1996 was specified to allow model states to stabilize. The calibration period extended from October 1, 1996 to September 30, 2002, while the verification period spanned the period of October 1, 2002 to September 30, 2006.

6.2. Comparisons of computed and observed runoff volumes and soil moisture

For this experiment, participants were instructed to set up and execute their models over an area encompassing the Oklahoma Mesonet shown in Fig. 3. Participants could execute their models at any resolution, but were required to convert soil moisture (daily average) and runoff estimates to the HRAP grid scale (4×4 km). Participants were asked to generate daily grids of these two quantities. Soil moisture content simulations were requested at the 0–25 mm and 25–75 mm depth ranges for comparison with observed data. The models were only to perform water balance computations without any routing. To simplify the application of the models over such a large domain, participants were instructed to use *a priori* parameters with no calibration. A "warm-up" period from October 1, 1995 to December 31, 1996 was allowed. The evaluation statistics were computed over the period spanning from January 1, 1997 to December 31, 2002.

Our intent was to conduct soil moisture tests using data commonly and widely available for operational hydrologic forecasting. As such, the CONUS-scale STATSGO dataset was selected for model parameter estimation. A dominant soil texture grid was derived at the HRAP resolution from the CONUS STATSGO data set. Vertically, the dominant texture is representative of the upper zone of the Sacramento model, which most often corresponds to a physical depth somewhere between 0–20 and 0–30 cm. This data set as well as the two derived soil properties (1) saturation volumetric water content (porosity) and, (2) residual volumetric water content (wilting point), were provided via the 'STATSGO Soil Data for Oklahoma Soil Moisture Experiment' link on the DMIP2 'Data' web page. The porosity and wilting point data were derived using the method of Cosby et al. (1984). Participants were requested to use these data sets for their models if at all possible.

Statistical analyses were conducted over watershed scales and not at specific points. Comparison analyses were performed on a soil moisture saturation ratio SR calculated as:

$$SR = \frac{\theta - \theta_r}{\theta_s - \theta_r}$$

where θ is volumetric water content, θ_r is residual volumetric water content (or wilting point), and θ_s is the saturation volumetric water content (or porosity).

6.3. Common channel routing scheme

This experiment was designed to address the science question in Section 3 derived from the DMIP 1 recommendation to separate the routing and rainfall–runoff comparisons (Reed et al., 2004). In this experiment, participants were asked to generate unrouted runoff depth time series that would subsequently be routed through a common routing scheme. Participants were asked to generate runoff depths (aggregated to 1 h time step) at the HRAP scale for distinct 2–3 month periods for the Blue and Tahlequah Rivers. The participants were allowed to use whatever basin discretization for their models, but were required to average the runoff volumes to the 4 km HRAP scale. The OHD distributed model using kinematic hillslope and channel routing was used to route the participants' runoff volumes. Participants' surface runoff was routed using overland routing before it entered the channel. Subsurface flow was assumed to enter directly into the channel. For models that did not have two runoff components, routing was performed as only surface or subsurface runoff depending on the participant's desire. Participants were asked to specify which routing they preferred if they submitted a combined runoff value, and were allowed to determine which of their runoff components were to be routed using hillslope or channel routing within the OHD distributed model.

6.4. Impact of inconsistent precipitation data on model calibration

This experiment was for returning DMIP 1 participants, who were asked to generate simulations using the DMIP 2 period of multisensor precipitation data but with their DMIP 1 calibrated parameters. Following this step, the returning DMIP 1 participants could re-calibrate their model parameters using the DMIP 2 data.

7. Participants and models

The NWS was very encouraged by the level of participation in DMIP 2. Some DMIP 1 participants returned with revised or new models, while several new groups chose to participate as seen in Table 2. A wide variety of models was represented in DMIP 2. CEMAGREF contributed the widely-used lumped conceptual GR4J model. As in DMIP 1, NWS/OHD used the HL-RDHM model, but the gridded SAC-SMA component was modified to include physically-based treatment of frozen ground and soil moisture (Koren et al., 2006, 2007). The Noah land surface model was used by NCEP/EMC as in DMIP 1. Noah computes a comprehensive suite of surface energy and water fluxes to provide lower boundary conditions for NWS operational numerical weather prediction models. The University of Illinois used the THREW model, which is an REW-based approach to solving mass, momentum, and heat balance equations. The Wuhan University model LL-III couples equations of two-dimensional transient subsurface flow and one-dimensional forms of unsteady overland, channel, and ground water flow. The University of Oklahoma contributed simulations from Vflo™. This model computes infiltration rate and saturation excess runoff. It solves the surface flow equations using a finite difference scheme in time and a finite element method in space. The Agricultural Research Service (ARS) used the SWAT model as in DMIP 1. SWAT combines a semi-distributed rainfall/runoff scheme with Muskingum channel routing for analysis of flow from agricultural areas. The Danish Hydraulics Institute (DHI) contributed simulations from two modeling system: MIKE SHE and MIKE 11. In both systems, DHI used conceptual rainfall–runoff models with dynamic wave channel routing. The University of California at Irvine (UCI) used a semi-distributed application of the SAC-SMA. Sub-basin runoff volumes were converted to streamflow with unit hydrographs, followed by kinematic channel routing. The University of Arizona contributed two sets of simulations. The first set was generated using a gridded SAC-SMA model with Muskingum routing in the University of Arizona Distributed Hydrologic Model (DHM-UA). A second set of simulations was generated using a version of HL-RDHM to investigate parameter calibration strategies. The University of Nebraska adopted the Hydrologic Simulation Program-FORTRAN (HSPF) model in a gridded format. The Imperial College of London used a semi-distributed model with a rainfall/runoff component based on the work of Moore (1985). A new participant in DMIP 2, the Vrije University of Brussels used the WetSpa model. WetSpa is a gridded rainfall–runoff model with diffusive wave channel routing. The University of Alberta at Edmonton contributed simulations with the semi-distributed physically-based hydrologic model DPHM-RS. The companion results paper presents more information about the participants' models as do the other papers in this Special Issue.

8. Evaluation of results and expected outcomes

A companion results paper written by all DMIP 2 participants (Smith et al., this issue) presents the results and conclusions for

Table 2
DMIP 2 participating institutions and models.

	Group and DMIP 2 acronym	Model and Primary Reference	Model in DMIP 1?
1	National Weather Service Office of Hydrologic Development (OHD)	HL-RDHM. Modified SAC-SMA with kinematic hillslope and channel routing (Koren et al., 2004)	Yes
2	DHI Water and Environment, Denmark (DH1, DH2)	1. MIKE 11 (Butts et al., 2004)	Yes
		2. MIKE SHE (Butts et al., 2004)	No
3	U. of Arizona (AZ1, AZ2)	1. DHM-UA (Pokhrel et al., 2008). Semi-distributed SAC-SMA and Muskingum routing	No
		2. HL-RDHM (Koren et al., 2004)	No
4	National Centers for Environmental Prediction Environmental Modeling Center (EMC)	Noah land surface model (Ek et al., 2003). Computes energy and moisture fluxes for numerical weather models.	Yes
5	U. of Oklahoma (UOK)	Vflo™ (Vieux, 2004) Finite element in space solution to surface water equations.	No
6	USDA Agricultural Research Service and Blackland Research and Extension Center of Texas A&M University System (ARS)	SWAT (Di Luzio and Arnold, 2004)	Yes
7	Vrije U. Brussels, Belgium (VUB)	WetSpa (Liu and De Smedt, 2004)	No
8	Hydraulic and Electrical College of Wuhan University, China (WHU)	LL-III (numerical hydrodynamic, Li, 2001)	Yes
9	U. California at Irvine (UCI)	Semi-distributed SAC-SMA (Khakbaz et al., 2011. Note: this is a revised version of the semi-distributed SAC-SMA by Ajami et al., 2004)	Yes
10	Imperial College of London (ICL)	Semi-distributed conceptual using approach of Moore (1985)	No
11	U. of Nebraska at Lincoln (NEB)	HSPF (Ryu, 2009)	No
12	U. of Illinois (ILL)	THREW (Tian et al., 2006)	No
13	CEMAGREF, France (CEM)	GR4J lumped conceptual (Perrin et al., 2003)	No
14	U. of Alberta, Canada (UAE)	DPHM-RS (Biftu and Gan, 2001)	No

the Oklahoma region experiments. As in DMIP 1, widely accepted statistical measures were used to analyze participants' simulations over a range of periods, flow intervals and events. Readers are referred to Smith et al. (2004a) for a discussion of these statistical measures. Model performance for events was also investigated. From our discussions at the DMIP 2 results workshop in September 2007, the NWS team (and the DMIP 2 participants) proposed a set of primary statistical measures to be stressed in the calibration of operational models. These are discussed in the companion DMIP 2 results paper (Smith et al., this issue).

The DMIP 2 Oklahoma experiments extended our understanding of distributed model performance in uncomplicated basins with perhaps the best quality of operational multisensor precipitation estimates available. The results strengthened the conclusions of DMIP 1 and showed that distributed models are making strides towards achieving their potential. The percentage of model-basin cases showing improvement of distributed models compared to lumped simulations at basin outlets and interior points was 18%, 24%, and 28%, for runoff volume, peak flow, and peak timing, respectively. These values correspond to 14%, 33%, and 22% respectively, in DMIP 1. While there may not seem to be much gain compared to DMIP 1 results, the DMIP 2 values were based on more precipitation-runoff events, more model-basin combinations (148 versus 51), more interior ungauged points (9 versus 3), and a benchmark comprised of two lumped model simulations. Thus, we believe that the DMIP 2 findings are more robust.

Two distributed models were able to provide reasonably good soil moisture simulations; however the streamflow simulation performance of one model was markedly better than the other. DMIP 2 also highlighted the need for consistent precipitation data for model calibration. Another important finding from DMIP 2 is that while calibration of model parameters provided improved performance for most models, calibration alone was not able to greatly improve performance beyond that achieved using *a priori* parameters. In addition, some uncalibrated models were able to out-perform some calibrated models, highlighting the strength of several model/parameterization combinations.

As with DMIP 1, participants are leveraging the DMIP 2 project and its data to investigate ideas not explicitly identified. As described earlier, Li and Sivapalan (2011) investigated the relationship between the spatial heterogeneity of runoff generation mechanisms and event runoff scaling behavior. Pokhrel and Gupta (in press) used the DMIP 2 data sets to investigate the role of precipitation and basin variability given only the basin outflow hydrograph.

9. Summary

The Oklahoma DMIP 2 project was formulated as a logical complement to the experiments in DMIP 1. The NWS recognized the need for additional science experiments to guide its research-to-operations path towards advanced hydrologic models for river and water resources forecasting. The DMIP 2 experiments were more comprehensive than those in DMIP 1. Many additional interior stream gauges and data were identified in these basins to evaluate the ability of distributed models to predict interior flow. These additional gauged points included interior basins that were smaller than in DMIP 1. We also tested the ability of distributed models to generate soil moisture fields. Fourteen groups submitted simulations using 16 models. Ten of these had not participated in DMIP 1. As in DMIP 1, the models ranged in type, complexity, spatial application scale, and parameter estimation-calibration techniques.

Taken together, the results from the Oklahoma experiments in DMIP 1 and 2 have provided a robust view of the state of distributed modeling with operational precipitation data at typical NWS basin scales. Over 20 models participated in the two phases of DMIP in the Oklahoma region.

The combined results from DMIP 1 and 2 in the Oklahoma region show that spatially distributed hydrologic modeling is advancing. In a practical way, DMIP has confirmed that spatially distributed hydrologic modeling should and will continue to play a major role in NWS river and water resources forecasting. The DMIP results also highlighted the need to have realistic performance expectations as distributed models are operationally implemented. While the improvements of distributed models compared to lumped models for basin outlet simulations may not be (yet) as great as once anticipated, the ability of distributed models to match or exceed a lumped model while providing information at interior ungauged points is nonetheless an encouraging and notable achievement.

Of course, the DMIP experiments have had limitations. We are grateful for the participation of many agencies and institutions, especially considering that specific funding support was not available through DMIP. In hindsight, however, funding may have fostered more complete participation in specific tests such as the soil moisture and routing experiments.

Moving ahead, distributed modeling results need to be explored in light of the uncertainty in model parameters, model structure, and input data. Such experiments should be performed in expanded tests such forecasting/hindcasting to note he benefits of distributed modeling at different forecast lead times. Continued work is needed to develop data assimilation approaches. Moreover, continued research, development, and testing of distributed models are needed in complex areas of the mountainous western USA. As mentioned before, two basins in the western USA were selected for DMIP 2 evaluations of lumped and distributed models in complex regions with highly variable terrain, orographic enhancement of precipitation, snow accumulation and melt, and data quantity issues. The results of these experiments will be published in forthcoming papers.

Acknowledgements

We gratefully acknowledge the Oklahoma Mesonet team for providing us with comprehensive soil moisture data over the study region. Brad Illston provided assistance with understanding the feasible range of the Mesonet soil moisture observations. John L. Rutledge of the consulting firm Freese and Nichols provided surveyed cross section data from a study commissioned by the Southwestern Electric Power Company. Valuable assistance and guidance for the Oklahoma experiments was provided by Billy Olsen, Bill Lawrence, Mike Pierce, John Schmidt, and James Paul of the ABRFC. Dr. Todd Halihan of Oklahoma State University provided information on the hydrogeology of the Blue River. David Adams and Darrell Walters of the USGS provided hourly observed streamflow. We are grateful to Noel Osborn of the Oklahoma Water Resources Board for granting permission to use Fig. 4. Without the support of the participants' agencies and institutions, DMIP 2 would not have been possible. The careful review and comments by anonymous reviewers are appreciated.

Appendix A. Soil moisture measurements from the Oklahoma Mesonet

The Oklahoma Mesonet was established in the early 1990s as an automated network of meteorologic observing stations spread throughout that state. The Mesonet consists of over 100 stations or roughly one station per county. The goals established for this network were to operate these stations in real time, observe nine meteorologic variables, transmit these observations in real time, and relay the data to a central site for collection, quality control, and storage for use and dissemination (Illston et al., 2004; Brock et al., 1995).

Soil moisture sensors were installed starting in 1996 at approximately 60 sites, followed by installation at an additional 43 sites in 1998 and 1999. Campbell Scientific 229-L soil moisture sensors (CSI 229-l, Campbell, 2010) are installed at depths of 5, 25, 60, and 75 cm. These depths were selected to enhance agricultural and meteorologic modeling, facilitate drought monitoring, and to generate research-quality data sets. The CSI-229-L consists of thermocouples as temperature sensors and a resistor to function as a heating element. The CSI-229-l measures the change in temperature before and after a heat pulse is introduced. From the measured temperature differences, soil water content, soil matric potential, and fractional water index can be calculated (Schnieder et al., 2003; Basara and Crawford, 2000).

Basara and Crawford (2000) identified discrepancies in the near-surface (5 and 25 cm) and deep layer (60 and 75 cm) soil moisture observations. They theorized that the installation method for the sensors at the two deepest layers contains a fundamental flaw that could lead to measurement errors in certain circumstances. The CSI-229-L sensors at these depths were installed at a 45° angle from the vertical, allowing soil water to flow down the instrument wire to moisten the sensor without affecting the remainder of the soil layer. However, they noted that the installation error affects less than one percent of all the soil moisture observations between 1996 and 1999. They recommended that future sensors be installed horizontally at the deeper layers.

Illston et al. (2004) compared the upper layer (5 and 25 cm) soil moisture observations from the CSI-229-L to soil core samples at 40 sites during an extended drying phase. The authors concluded that the CSI-229-L performed fairly well.

The upper bound of the soil moisture content is limited by the accuracy of the Campbell Scientific 229L sensor. Temperature observations from this sensor are converted to soil water potential using empirical relationships (Schnieder et al., 2003). Given that the lower limit of the observed values of the temperature reference is approximately 1.4 °C, the equation for computing soil water potential does not return values much moister that −0.10 bars. Thus, the range of the soil moisture observations is from 0.1 to 10 bars.

References

Abdulla, F.A., Lettenmaier, D.P., Wood, E.F., Smith, J.A., 1996. Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red basins. J. Geophys. Res. 101 (D3), 7449–7459.

Ajami, N.K., Gupta, H., Wagener, T., Sorooshian, S., 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. J. Hydrol. 298, 112–135.

Anderson, E.A., 2002. Calibration of Conceptual Hydrologic Models for Use in River Forecasting. NWS Publication. <http://www.nws.noaa.gov/oh/hrl/calb/calibration1102/main.htm>.

Anderson, R.M., Koren, V.I., Reed, S.M., 2004. Using SSURGO data to improve Sacramento model *a priori* parameter estimates. J. Hydrol. 320, 103–116.

Ao, T., Ishidaira, H., Takeuchi, K., Kiem, A.S., Yoshitari, J., Fukami, K., Magome, J., 2004. Relating BTOPMC model parameters to physical features of MOPEX basins. J. Hydrol. 320, 84–102.

Bandaragoda, C., Tarboton, D.G., Woods, R., 2004. Application of TOPNET in the distributed model intercomparison project. J. Hydrol. 298, 178–201.

Basara, J.B., Crawford, T.M., 2000. Improved installation procedures for deep-layer soil moisture measurements. Journal of Atmospheric and Oceanic Technology 17, 879–884.

Bastidas, L.A., Gupta, H.V., Hsu, K.L., Sorooshian, S., 2003. Parameter, structure, and model performance evaluation for land-surface schemes. In: Duan et al. (Eds.), Calibration of Watershed Models. Water Science and Application, vol. 6. AGU, pp. 239–254.

Bastidas, L.A., Li, S., 2004. Multi-Objective Calibration and Sensitivity Analysis of a Distributed Model using Similarity Measures. Presented at AGU Fall Meeting, Paper H13C-0419, Eos Trans. AGU, vol. 85, No. 47 (Fall Meeting Supplement).

Biftu, G.F., Gan, T.Y., 2001. Semi-distributed, physically- based, hydrologic modeling of the Paddle River Basin, Alberta using remotely sensed data. J. Hydrol. 244, 137–156.

Brock, F.V., Crawford, K.C., Elliot, R.L., Cuperus, G.W., Stadler, S.J., Johnson, H.L., Eilts, M.D., 1995. The Oklahoma Mesonet: a technical overview. J. Atmos. Oceanic Technol. 12, 5–19.

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modeling uncertainty for streamflow simulation. J. Hydrol. 298 (1–4), 242–266.

Campbell Scientific 229L Water Matric Potential Sensor User Manual, 2010. On-line Reference: <http://www.campbellsci.com/229-l> (accessed July 14.07.10).

Campo, L., Caparrini, F., Castelli, F., 2006. Use of multi-platform, multi-temporal remote-sensing data for calibration of a distributed hydrologic model: an application in the Arno basin, Italy. Hydrol. Process. 20, 2693–2712.

Carpenter, T.M., Georgakakos, K.P., 2004. Continuous streamflow simulation with the HRCDHM distributed hydrological model. J. Hydrol. 298, 61–79.

Cooper, D., 2004. An operational forecast office perspective of the National Weather Service hydrologic distributed hydrologic modeling system (HDMS). In: Proceedings of 3rd Federal Interagency Hydrologic Modeling Conference, April 2–6, 2006, Reno Nevada.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20 (6), 682–690.

Di Luzio, M., Arnold, J.G., 2004. Formulation of a hybrid calibration approach for a physically based distributed model with NEXRAD data input. J. Hydrol. 298, 136–154.

Doherty, J., Skahill, B.E., 2006. An advanced regularization methodology for use in watershed model calibration. J. Hydrol. 327, 564–577.

Eckhardt, K., Fohrer, N., Frede, H.G., 2005. Automatic model calibration. Hydrol. Process. 19, 651–658.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res. 108 (D22). doi:10:1029/2002JD003296.

Fairchild, R.W., Hanson, R.L., Davis, R.E., 1990. Hydrology of the Arbuckle Mountains Area, South-Central Oklahoma, Oklahoma Geological Survey Circular 91 ISSN 0078-4397, 117 pp.

Farnsworth, R.K., Thompson, E.S., 1982. Mean Monthly, Seasonal, and Annual Pan Evaporation for the United States. NOAA Technical Report NWS 34, Washington, DC, 82 p.

Farnsworth, R.K., Thompson, E.S., Peck, E.L., 1982. Evaporation Atlas for the Contiguous 48 United States. NOAA Technical Report NWS 33, Washington, DC.

Fortin, S.M., 1998. A Qualitative Assessment of the Soil Moisture Parameterization of the Sacramento Soil Accounting Model during the Fall of 1997 and Spring of 1998. A Non-Thesis Research Paper Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of Master of Science in Meteorology, University of Oklahoma, Tulsa.

Francés, F., Velez, J.I., Velez, J.J., 2007. Split-parameter structure for the automatic calibration of distributed hydrological models. J. Hydrol. 332 (1–2), 226–240.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research. Transactions of the American Society of Agricultural and Biological Engineers 50 (4), 1211–1250, ISSN 0001-2351.

Georgakakos, K.P., Carpenter, T.M., 2006. Potential value of operationally available and spatially distributed ensemble soil water estimates for agriculture. J. Hydrol. 328 (1–2), 177–191.

Georgakakos, K.P., Smith, G.F., 1990. On improved hydrologic forecasting–results from a WMP real-time forecasting experiment. J. Hydrol. 114, 17–45.

Giertz, S., Diekkruger, B., Steup, G., 2006. Physically-based modeling of hydrological processes in a tropical headwater catchment (West Africa – process representation and multi-criteria validation. Hydrol. Earth Syst. Sci. 10, 829–847.

Goswami, M., O'Conner, K.M., 2007. Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall–runoff model. J. Hydrol. Sci. 52 (3), 432–449.

Gotzinger, J., Bardossy, A., 2007. Comparison of four regionalization methods for a distributed hydrological model. J. Hydrol. 33, 374–384.

Green, D.R., Hudlow, M.D., 1982. Hydrometeorologic grid mapping procedures. In: AWRA International Symposium on Hydrometeorology, Denver, CO, Available on Request from NWS/OHD.

Illston, B.G., Basara, J.B., 2002. A soil moisture analysis of drought conditions using the Oklahoma Mesonet. In: Preprints, 13th Conference on Applied Climatology; American Meteorological Society, Portland, Oregon, May 13–16.

Illston, B.G., Basara, J.B., Crawford, K.C., 2003. Soil moisture observations from the Oklahoma Mesonet. GEWEX News 13 (3), 13–14.

Illston, B.G., Basara, J.B., Crawford, K.C., 2004. Seasonal to interannual variations of soil moisture measured in Oklahoma. Int. J. Climatol. 24, 1883–1896.

Illston, B.G., Basara, J.B., Fisher, D.K., Elliott, R., Fiebrich, C.A., Crawford, K.C., Humes, K., Hunt, E., 2008. Mesoscale monitoring of soil moisture across a statewide network. J. Atmos. Oceanic Technol. 25, 167–182.

Ivanov, V.Y., Bras, R.L., Vivoni, E.R., 2008. Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedback. Water Resour. Res. 44, W03429. doi:10:1029/2006WR005588.

Ivanov, V.Y., Vivoni, E.R., Bras, R.L., Entekhabi, D., 2004. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. J. Hydrol. 298, 80–111.

Jayakrishnan, R., Srinivasan, R., Arnold, J.G., 2004. Comparison of raingage and WSR-88D Stage III precipitation data over the Texas-Gulf basin. J. Hydrol. 292, 135–152.

Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. Photogramm. Eng. Remote Sens. 54, 1593–1600.

Jinkang, D., Shunping, X., Youpeng, X., Xu, C., Singh, V.P., 2007. Development and testing of a simple physically-based distributed rainfall–runoff model for storm runoff simulation in humid forested basins. J. Hydrol. 336, 334–346.

Johnson, D., Smith, M., Koren, V., Finnerty, B., 1999. Comparing mean areal precipitation estimates from NEXRAD and rain gauge networks. J. Hydrol. Eng. 4 (2), 117–124.

Jones, E.T., Roth, K., Costanza, K., 2009. A comparison of the NWS distributed versus lumped hydrologic model. In: AMS 23 Conference on Hydrology, Paper 4.3, Phoenix Arizona, January 10–15.

Kalin, L., Hantush, M.H., 2006. Comparative assessment of two distributed watershed models with application to a small watershed. Hydrol. Process. 20, 2285–2307.

Khakbaz, B., Imam, B., Hsu, K., Sorooshian, S., 2011. From lumped to distributed via semi-distributed: calibration strategies for semi-distributed hydrologic models. J. Hydrol. this issue, doi:10.1016/j.jhydrol.2009.02.021.

Kim, J.-K, Bastidas, L., Clark, E., 2010. Performance Evaluation of the Distributed RDHM Model in the Rockies. Poster H23A-1178, Fall Meeting of the AGU, San Francisco, December, 2010.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42, W03S04. doi:10.1029/2005WR004362.

Koren, V., Moreda, F., Reed, S., Smith, M., Zhang, Z., 2006. Evaluation of a grid-based distributed hydrological model over a large area: model uncertainties at different scales. Predictions in Ungauged Basins: Promise and Progress (Proceedings of Symposium S7 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005), vol. 303. IAHS Publ., pp. 47–56.

Koren, V., Smith, M., Cui, Z., Cosgrove, B., 2007. Physically-Based Modifications to the Sacramento Soil Moisture Accounting Model: Modeling the Effects of Frozen Ground on the Rainfall–Runoff Process. NOAA Technical Report NWS 52, <http://www.weather.gov/oh/hrl/hsmb/docs/hydrology/PBE_SAC-SMA/NOAA_Technical_Report_NWS_52.pdf>.

Koren, V., Smith, M., Cui, Z., Cosgrove, B., 2010. Modification of Sacramento Soil Moisture Accounting Heat Transfer Component (SAC-HT) for Enhanced Evapotranspiration. NOAA Technical Report NWS 53, <http://www.weather.gov/oh/hrl/hsmb/docs/hydrology/PBE_SAC-SMA/NOAA_Technical_Report_NWS_53.pdf>.

Koren, V., Reed, S., Smith, M., Zhang, Z., Seo, D.-J., 2004. Hydrology laboratory research modeling system (HL-RMS) of the US National Weather Service. J. Hydrol. 291, 297–318.

Koren, V., Reed, S., Zhang, Z., Seo, D.-J., Moreda, F., Kuzmin V., 2003a. Use of Spatially Variable Data in River Flood Prediction. AGU-EGS-EUG Assembly, April 9–14, 2003, Nice, France.

Koren, V., Smith, M., Duan, Q., 2003b. Use of *a priori* parameter estimates in the derivation of spatially consistent parameter sets of rainfall–runoff models. In: Duan et al. (Eds.), Calibration of Watershed Models: Water Science and Applications, vol. 6. AGU Press, pp. 239–254.

Koren, V., Reed, S., Smith, M., Zhang, Z., 2003c. Combining physically-based and conceptual approaches in the development and parameterization of a distributed system. In: Weather Radar Information and distributed Hydrological Modeling. Proceedings of Symposium HSO_3 held during IUGG2003 at Sapporo, July 2003, Publication No. 282, pp. 101–108.

Koren, V., Moreda, M., Smith, M., 2008. Use of soil moisture observations to improve parameter consistency in watershed calibration. Phys. Chem. Earth 33, 1068–1080.

Koren, V. Smith, M., Wang, D., Zhang, Z., 2000. Use of soil property data in the derivation of conceptual rainfall–runoff model parameters. In: 15th Conference on Hydrology, Long Beach, CA, AMS, pp. 103–106.

Kuzmin, V., Seo, D.-J., Koren, V., 2008. Fast and efficient optimization of hydrologic model parameters using *a priori* estimates and stepwise line search. J. Hydrol. 353, 109–128.

Leavesley, G.H., Hay, L.E., Viger, R.J., Marskstrom, S.L., 2003. Use of *a priori* parameter-estimation methods to constrain calibration of distributed-parameter models. In: Duan et al. (Eds.), Calibration of Watershed Models. Water Science and Application, vol. 6. AGU Press, pp. 255–266.

Lei, F., Koren, V., Smith, M., Moreda, F., 2007. A sensitivity study of an energy-budget snow accumulation and ablation model. In: Paper J6.4, Proceedings of the 21st Conference on Hydrology, 87th Meeting of the American Meteorological Society, 13–18 January, San Antonio, Texas.

Li, L., 2001. A physically-based rainfall–runoff model and distributed dynamic hybrid control inverse technique. Soil-vegetation-atmosphere transfer schemes and large-scale hydrological models. In: Proceedings of a Symposium held during the Sixth IAHS Scientific Assembly at Maastricht, The Netherlands, July, 2001. IAHS Pub. No. 270, pp. 135–142.

Li, H., Sivapalan, M., 2011. Effect of spatial heterogeneity of runoff generation mechanisms on the scaling behavior of event runoff responses in a natural river basin. Water Resour. Res. 47, W00H08. doi:10.1029/2010WR009712.

Liu, Y.B., De Smedt, F., 2004. WetSpa Extension: Documentation and User Manual. Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Belgium, 108 pp.

Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., 2004. Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. J. Geophys. Res. 109, D07S91. doi:10.1029/2003JD003517.

Lohmann, D., Lettenmaier, D.P., Liang, X., Wood, E.F., Boone, A., Chang, S., Chen, F., Dai, Y., Desborough, C., Dickinson, R.E., Duan, Q., Ek, M., Gusev, Y.M., Habets, F., Irannejad, P., Koster, R., Mitchell, K.E., Nasonova, O.N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A.B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z.-L., Zeng, Q.C., 1998. The project for intercomparison of land-surface parameterization schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes. Global Planet. Change 19, 161–179.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modeling using automatic calibration with multiple objectives. Adv. Water Resour. 26, 205–216.

Madsen, H., Kristensen, M., 2002. A multi-objective calibration framework for parameter estimation in the MIKE SHE integrated hydrological modeling system. In: Kovar, K., Hrkal, Z. (Eds.), ModelCARE 2002: Proceedings of the 4th International Conference on Calibration and Modeling in Groundwater Modeling, Prague, Czech Republic. Acta Univ. Carol. – Geol. 46 (2), 270–273.

Marce, R., Ruiz, C.E., Armengol, J., 2008. Using spatially distributed parameters and multi-response objective functions to solve parameterization of complex applications of semi-distributed hydrologic models. Water Resour. Res. 44. doi:10.1029/2006WR005785.

Mascaro, G., Vivoni, E.R., Deidda, R., 2010. Implications of ensemble quantitative precipitation forecast errors on distributed streamflow forecasting. J. Hydrometeorol. 11, 69–86.

Marron, J.S., Tsybakov, A.B., 1995. Visual error criteria for qualitative smoothing. J. Am. Stat. Assoc. 90 (430), 499–507 (Theory and Methods, June).

McEnery, J., Ingram, J., Duan, Q., Adams, T., Anderson, L., 2004. NOAA's advanced hydrologic prediction service: building pathways for better science in water forecasting. Bull. Am. Meteorol. Soc., 275–385. doi:10.1175/BAMS-86-3-375.

McMillan, H.K., Clark, M.P., Ibbitt, R.P., 2008. Calibration strategies for distributed hydrological models using qualitative knowledge of spatially variable process characteristics. Geophys. Res. Abstr. 10, EGU2008-02318.

Mesinger, F., DiMego, G., Kalnay, E., Shafran, Pl., Ebisuzaki, W., Jovic, D., Wollen, J., Mitchell, K., Rogers, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. (March), 343–360.

Mesinger, F., DiMego, G., Kalnay, E., Shafran, Pl, Ebisuzaki, W., Jovic, D., Wollen, J., Mitchell, K., Rogers, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2004. North American Regional Reanalysis, Paper P1.1., 15th Symposium on Climate Change and Global Variations, 84th Conference of the AMS, Seattle, Wash., 11–15, January.

Mitchell, K.E., Lohmann, D., Houser, P.R., Wood, E.F., Schaake, J.C., et al., 2004. The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distribute hydrological modeling system. J. Geophys. Res. 109, D07S90. doi:10.1029/2003JD003823.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. J. Hydrol. Sci. 30 (2), 273–297.

Moreda, F., Koren, V., Cui, Z., Reed, S., Zhang, Z., Smith, M., 2005. Hydrologic Modeling on a 4 km Grid over the Conterminous United States (CONUS). A Poster Presented at the AGU 2005 Fall Meeting December 5–9, San Francisco.

National Research Council (NRC), 2000. Grand Challenges in Environmental Sciences. Committee on Grand Challenges in Environmental Sciences, Oversight Commission for the Committee on Grand Challenges in Environmental Sciences, National Academy Press, Washington, DC, 96 pp.

Nelson, B.R., Seo, D.-J., Kim, D., 2010. Multi-sensor precipitation analysis. J. Hydrometeorol. 11, 666–682. doi:10.1175/2010JHM1210.1.

Nelson, B.R, Seo, D.J., Kim, D., Bates, J.J., 2006. Multisensor Precipitation Reanalysis. American Meteorological Society 2006 Annual Meeting, Paper 4.6, Atlanta, GA, January 29–February 2, 2006.

NWS, 2004. The NWS Integrated Water Science Plan (IWSP), Report of the IWSP team, <http://www.nws.noaa.gov/ost/climate/STIP/carter_iwsp_100605.pdf>.

Oklahoma Water Resources Board (OWRB), 2003. The Arbuckle-Simpson Hydrology Study, November, 4 pp.

Osborn, N.I., 2009. Arbuckle-Simpson Hydrology Study – Final Report to the US Bureau of Reclamation, Prepared by the Oklahoma Water Resources Board, December, Cooperative Agreement No. 03FC601814, 42 pp.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1–4), 275–289.

Pokhrel, P., Gupta, H.V., Wagener, T., 2008. A spatial regularization approach to parameter estimation for a distributed watershed model. Water Resour. Res. 44, W12419. doi:10.1029/2007WR006615.

Pokhrel, P., and Gupta, H.V., in press. On the ability to infer spatial catchment variability using streamflow hydrographs. Water Resour. Res. doi:10.1029/2010WR009873.

Refsgaard, J.C., 1997. Parameterization, calibration, and validation of distributed hydrological models. J. Hydrol. 198, 69–97.

Reed, S., Schaake, J., Zhang, Z., 2007. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. J. Hydrol. 337, 402–420.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., Participants, D.M.I.P., 2004. Overall distributed model intercomparison project results. J. Hydrol. 298 (1–4), 27–60.

Reed, S., 2003. Deriving flow directions for coarse resolution (1–4 km) gridded hydrologic modeling. Water Resour. Res. 39 (9), 4-1–4-11.

Reed, S.M., Maidment, D.R., 1999. Coordinate transformations for using Nexrad data in GIS-based hydrologic modeling. J. Hydrol. Eng. 174, 182.

Rousseau, A.N., Fortin, J.P., Villeneuve, J.P., 2003. A process-oriented multiple-objective calibration strategy accounting for model structure. Calibration of Watershed Models: Water Science and Applications, vol. 6. American Geophysical Union Press, pp. 153–164.

Ryu, J.H., 2009. Application of HSPF to the distributed model intercomparison project: case study. Journal of Hydrologic Engineering 14 (8), 847–857.

Schaake, J.C., Duan, Q., Koren, V., Mitchell, K.E., Houser, P.R., Wood, E.F., Robock, A., Lettenmaier, D.P., Lohmann, D., Cosgrove, B., Sheffield, J., Luo, L., Wiggins, R.W., Pinker, D.T., Tarpley, J.D., 2004. An Intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS). J. Geophys. Res. 109, D01S90. doi:10.1029/2002JD003309.

Schmidt, J.A., Anderson, A.J., Paul, J.H., 2007. Spatially-variable, physically-derived flash flood guidance. In: Paper 6B.2, Proceedings of the 21st Conference on Hydrology, 87th Meeting of the AMS, San Antonio, Texas, Jan 15–18.

Schnieder, J.M., Fisher, D.K., Elliott, R.L., Brown, G.O., Bahrmann, C.P., 2003. Spatiotemporal variations in soil water: first results from the ARM SGP CART network. J. Hydrometeorol. 4, 106–120.

Seibert, J., McDonnell, J.J., 2003. The quest for an improved dialog between modeler and experimentalist. In: Duan et al. (Eds.), Calibration of Watershed Models. Water Science and Application, vol. 6. AGU Press, pp. 301–316.

Senarath, U.S., Ogden, F.L., Downer, C.W., Sharif, H.O., 2000. On the calibration and verification of two-dimensional, distributed, Hortonian, continuous watershed models. Water Resour. Res. 36 (6), 1495–1510.

Seo, D.-J., Demargne, J., Wu, L., Liu, Y., Brown, J.D., Regonda, S., and Lee, H., 2010. Hydrologic ensemble prediction for risk-based water resources management and hazard mitigation. In: 2nd Joint Federal Interagency Conference, Las Vegas, NV, June 27–July 1, 2010.

Shultz, M.J., Corby, R.J., 2004. Development, calibration, and implementation of a distributed hydrologic model for use in real-time river forecasting. In: Proceedings of 3rd Federal Interagency Hydrologic Modeling Conference, April 2–6, 2006, Reno Nevada.

Smith, M.B., Koren, V., Zhang, Z., Zhang, Y., Reed, S., Cui, Z., Moreda, F., Anderson, E., Cosgrove, B., Mizukami, N., DMIP 2 participants, this issue. Results from the DMIP 2 Oklahoma experiments.

Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S., Zhang, Z., Duan, Q.-Y., Moreda, F., Cong, S., 2004a. The distributed model intercomparison project (DMIP): motivation and experiment design. J. Hydrol. 298 (1–4), 4–26.

Smith, M.B., Koren, V.I., Zhang, Z., Reed, S.M., Pan, J.-J., Moreda, F., 2004b. Runoff response to spatial variability in precipitation: an analysis of observed data. J. Hydrol. 298 (1–4), 267–286.

Smith, M.B., Koren, V., Finnerty, B., Johnson, D., 1999. Distributed Modeling: Phase 1 Results, NOAA Technical Report NWS 44, February, 250 pp. <http://www.nws.noaa.gov/oh/hrl/distmodel/data/tech%20report%2044.pdf>.

Smith, M.B., Laurine, D.P., Koren, V.I., Reed, S.M., Zhang, Z., 2003. Hydrologic model calibration in the National Weather Service. In: Duan et al. (Eds.), Calibration of Watershed Models: Water Science and Applications, vol. 6. AGU Press, pp. 133–152.

Stellman, K.M., Fuelberg, H.E., Garza, R., Mullusky, M., 2001. An examination of radar and rain gauge-derived mean areal precipitation over Georgia watersheds. J. Hydrometeorol. 16 (February), 133–144.

Tcherednichenko, I., Bastidas, L.A., Lansey, K., 2004. Model performance evaluation of distributed hydrological modeling for semi-arid regions. In: Proceedings of the 2004 World Water and Environmental Resources Congress June 27–July 1, 2004, Salt Lake City, UT. Sponsored by Environmental and Water Resources Institute (EWRI) of the American Society of Civil Engineers.

Tian, F., Hu, H., Lei, Z., Sivapalan, M., 2006. Extension of the representative elementary watershed approach for cold regions via explicit treatment of energy related processes. Hydrol. Earth Syst. Sci. 10, 619–644.

Torell, L.A., McDaniel, K.C., Koren, V., 2011. Estimating grass yield on Blue Grama range from seasonal rainfall and soil moisture. J. Rangeland Ecol. Manage. 64 (56–66), 1–2. doi:10.2111/REM-D-09-00107.1.

Vieux, B.E., 2004. Distributed Hydrologic Modeling Using GIS. Water Science Technology Series, second ed., vol. 48. Kluwer Academic Publishers, Norwell, Massachusetts, 289 pp, ISBN 1-4020-2459-2.

Vieux, B.E., Moreda, F.G., 2003. Ordered physics-based parameter adjustment of a distributed model. In: Duan et al. (Eds.), Calibration of Watershed Models: Water Science and Applications, vol. 6. AGU Press, pp. 267–281.

Vivoni, E.R., Entekhabi, D., Bras, R.L., Ivanov, V.Y., Van Horne, P., Grassotti, C., Hoffman, R.N., 2006. Extending the predictability of hydrometeorological flood events using radar rainfall nowcasting. J. Hydrometeorol. 7, 660–677.

Wang, D., Smith, M.B., Zhang, Z., Reed, S., Koren, V.I., 2000. Statistical comparison of mean areal precipitation estimates from WSR-88D, operational, and historical gage networks. In: 15th Annual Conference on Hydrology, 80th Meeting of the AMS, Long Beach, CA, January 10–14, J2.17.

Wealands, S.R., Grayson, R.B., Walker, J.P., 2004. Investigating spatial pattern comparison methods for distributed hydrological model assessment. In: Proceedings of Complexity and Integrated Resources Management, Sponsored by the International Environmental Modeling and Software Society, iEMSs 2004 International Conference, University of Osnabrück, Germany, 14–17 June.

Westcott, N.E., Knapp, H.V., Hilberg, S.D., 2008. Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. J. Hydrol. 351, 1–12.

White, L.W., Vieux, B., Armand, D., LeDimet, F.X., 2003. Estimation of optimal parameters for a surface hydrology model. Adv. Water Resour. 26, 337–348.

World Meteorological Organization (WMO), 1992. Simulated Real-time Intercomparison of Hydrological models, Operational Hydrology Report No. 38, WMO-No. 779, Secretariate of the World Meteorological Organization, Geneva, Switzerland.

Xie, H., Zhou, X., Hendricks, J.M.H., Vivoni, E.R., Guan, H., Tian, Y.Q., Small, E.E., 2006. Evaluation of NEXRAD stage III precipitation data over a semi-arid region. J. Am.Water Resour. Assoc. (January), 237–256.

Young, C.B., Bradley, A.A., Krajewski, W.F., Kruger, A., 2000. Evaluating NEXRAD multi-sensor precipitation estimates for operational hydrologic forecasting. J. Hydrometeorol. 1 (June), 241–254.