Modeling the hydrologic influence of subsurface tile drainage using the National Water Model

Prasanth Valayamkunnath[1], David J Gochis[1], Fei Chen[1], Micheal Barlage[2], and Kristie J. Franz[3]

[1]National Center for Atmospheric Research (UCAR)
[2]NCAR RAL
[3]Iowa State University

November 21, 2022

Abstract

Subsurface tile drainage (TD) is a dominant agriculture water management practice in the United States (US) to enhance crop production in poorly-drained soils. Assessments of field- or watershed-level (<50 km^2) hydrologic impacts of tile drainage are becoming common; however, a major gap exists in our understanding of regional (>105 km^2) impacts of tile drainage on hydrology. The National Water Model (NWM) is a distributed 1-km resolution hydrological model designed to provide accurate streamflow forecasts at 2.7 million reaches across the US. The current NWM lacks tile drainage representation which adds considerable uncertainty to streamflow forecasts in tile-drained areas. In this study, we quantify the performance of the NWM with a newly incorporated tile drainage scheme over the heavily tile-drained Midwestern US. Implementing a tile drainage scheme enhanced the uncalibrated model performance by about 20% to 50% of the calibrated NWM (*Calib*). The calibrated NWM with tile drainage (*CalibTD*) showed enhanced accuracy with higher event hit rates and lower false alarm rates than *Calib*. *CalibTD* showed better performance in high-flow estimations as tile drainage increased streamflow peaks (14%), volume (2.3%), and baseflow (11%). Regional water balance analysis indicated that tile drainage significantly reduced surface runoff (-7% to -29%), groundwater recharge (-43% to -50%), evapotranspiration (-7% to -13%), and soil moisture content (-2% to -3%). However, infiltration and soil water storage potential significantly increased with tile drainage. Overall, our findings highlight the importance of incorporating the tile drainage process into the operational configuration of the NWM.

Modeling the hydrologic influence of subsurface tile drainage using the National Water Model

Prasanth Valayamkunnath[a], David J. Gochis[a], Fei Chen[a], Michael Barlage[a], and Kristie J. Franz[b]

[a]National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA, 80301

[b]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa, USA, 50011

*Correspondence: Prasanth Valayamkunnath (prasanth@ucar.edu)

Key Points:

1. A new subsurface tile drainage module is incorporated into the National Water Model (NWM) to predict streamflow over the tile-drained areas

2. NWM with a tile drainage module can predict high-flows and streamflow peaks better than the original NWM over heavily tile-drained areas

3. Incorporating tile drainage into the NWM considerably enhanced the streamflow event hit rates and reduced false alarm rates

Abstract

Subsurface tile drainage (TD) is a dominant agriculture water management practice in the United States (US) to enhance crop production in poorly-drained soils. Assessments of field- or watershed-level (<50 km^2) hydrologic impacts of tile drainage are becoming common; however, a major gap exists in our understanding of regional (>105 km^2) impacts of tile drainage on hydrology. The National Water Model (NWM) is a distributed 1-km resolution hydrological model designed to provide accurate streamflow forecasts at 2.7 million reaches across the US. The current NWM lacks tile drainage representation which adds considerable uncertainty to streamflow forecasts in tile-drained areas. In this study, we quantify the performance of the NWM with a newly incorporated tile drainage scheme over the heavily tile-drained Midwestern US. Implementing a tile drainage scheme enhanced the uncalibrated model performance by about 20% to 50% of the calibrated NWM (*Calib*). The calibrated NWM with tile drainage (*CalibTD*) showed enhanced accuracy with higher event hit rates and lower false alarm rates than *Calib*. *CalibTD* showed better performance in high-flow estimations as tile drainage increased streamflow peaks (14%), volume (2.3%), and baseflow (11%). Regional water balance analysis indicated that tile drainage significantly reduced surface runoff (-7% to -29%), groundwater recharge (-43% to -50%), evapotranspiration (-7% to -13%), and soil moisture content (-2% to -3%). However, infiltration and soil water storage potential significantly increased with tile drainage. Overall, our findings highlight the importance of incorporating the tile drainage process into the operational configuration of the NWM.

1. Introduction

Agriculture management practices such as irrigation, fertilizer and pesticide application, and tillage are generally employed to enhance crop productivity and are crucial for global food production and food security. Agriculture subsurface drainage, often known as subsurface tile drainage, is a widely-used agriculture water management practice to improve crop growth in regions with shallow water tables or poorly drained soils. According to the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Census of Agriculture 2017, about 22.48 million hectares (Mha) of croplands in the US are tile-drained, and 83.80% of the total tile-drained croplands of the US are concentrated in six Midwestern states (USDA-NASS, 2017) (Figure 1a), which is one of the world's most productive areas in terms of food and bioenergy, and it is located in the headwater regions of the Mississippi River (Gaunter et al., 2014; Ray et al., 2013).

In general, tile drains are buried under the crop root zone to extract saturation water (or free water) from the soil, improve root-zone soil aeration and soil quality, reduce crop root diseases and soil erosion, allow for earlier planting and enhance crop yield (Figure 1b) (Du et al., 2005; Fausey, 2005; Fausey et al., 1987; Kornecki and Fouss, 2001). Furthermore, tile drainage is known to have a significant impact on watershed hydrology (Blann et al., 2009; King et al., 2014; Rahman et al., 2014; Thomas et al., 2016), because it depletes the free water from the root-zone soil layer, resulting in enhanced infiltration and reduced surface runoff, peak flows, and flooding (Golmohammadi et al., 2017; Rahman et al., 2014; Robinson and Rycroft, 1999; Skaggs et al., 1994). Tile drainage may also increase the watershed baseflow, annual runoff volume, instream pollutant concentrations, the timing and shape of the hydrograph, and the local and regional climate by modifying energy and water flux from croplands to the atmosphere (Blann et al., 2009; Eastman et al., 2010; Guo et al., 2018; Khand et al., 2017; King et al., 2014; Magner et al., 2004; Schilling et al., 2012; Schilling and Helmers, 2008; Schilling and Libra, 2003; Schottler et al., 2014; Thomas et al., 2016; Yang et al., 2017). However, the intensity and direction of the tile-drainage impact on hydrology depend on several field-specific factors such as soil properties, antecedent soil moisture storage, climatic conditions, topography, design of the tile drainage system, and tillage practices (Blann et al., 2009; King et al., 2014; Robinson, 1990; Robinson and Rycroft, 1999; Skaggs et al.,1994; Thomas et al., 2016; Wiskow and van der

Ploeg, 2003). The above findings on the hydrologic impact of tile drainage are based on field-level or small watershed-scale (<50 km^2) studies. A comprehensive understanding of regional-scale hydrology of tile-drainage is a major knowledge gap (Hansen et al., 2013; King et al., 2014; Thomas et al., 2016). Accurate modeling of tile drainage impacts on the continental or regional water cycle is a daunting challenge due to the lack of continental-scale high-resolution tile drainage data and an efficient, fully distributed, continental-scale hydrology model with a tile drainage scheme.

In the recent decade, the flood frequency and intensity have increased over the continental United States (CONUS), especially over the Central US (Mallakpour and Villarini, 2015). To provide flash flood forecasts and other hydrologic guidance with longer lead time and less uncertainties, the National Weather Service (NWS) Office of Water Prediction (OWP) of the National Oceanic and Atmospheric Administration (NOAA) developed a hydrologic modeling framework, the National Water Model (NWM), to simulate observed and forecast streamflow for about 2.7 million stream reaches of the CONUS. However, the NWM has considerable uncertainties in the streamflow prediction over the Midwestern US (Dugger et al., 2017; Karki et al., 2021). One of the reasons for the underperformance of the NWM can be the lack of representation of subsurface tile drainage hydrology in the NWM (Hansen et al., 2013). Field-level studies have already highlighted the importance of defining tile drainage within the hydrologic models to achieve accuracy in simulated water budget components over heavily tile-drained regions (Green et al., 2006; Hansen et al., 2013).

To address these shortfalls, in this study, we investigate the regional impact of tile drainage on the NWM performance in simulating streamflow over the upper Midwestern US by developing a new tile drainage scheme and implementing it into the NWM. We evaluate the NWM model performance with tile drainage regarding the streamflow simulation with and without NWM parameter calibration, and explore the influence of tile drainage on regional water budget and regional hydrology. In these simulations, we use the recently developed 30-meter resolution Agriculture Tile drainage data for the US (AgTile-US) (Valayamkunnath et al., 2020) to explicitly define the tile-drained croplands within the NWM.

In section 2, we describe the details of the study area, process descriptions of the NWM and the new tile drainage scheme, introduction to the input and evaluation data, calibration and regionalization of model parameters, and details of model simulation experiments. Details of hydrological and statistical analysis used in this study to evaluate the model performance are presented in section 2.8. The results on the model performance evaluation, the impact of tile drainage on energy and water balance components, comparison with parallel works, perspectives, and limitations of the study are discussed in section 3.

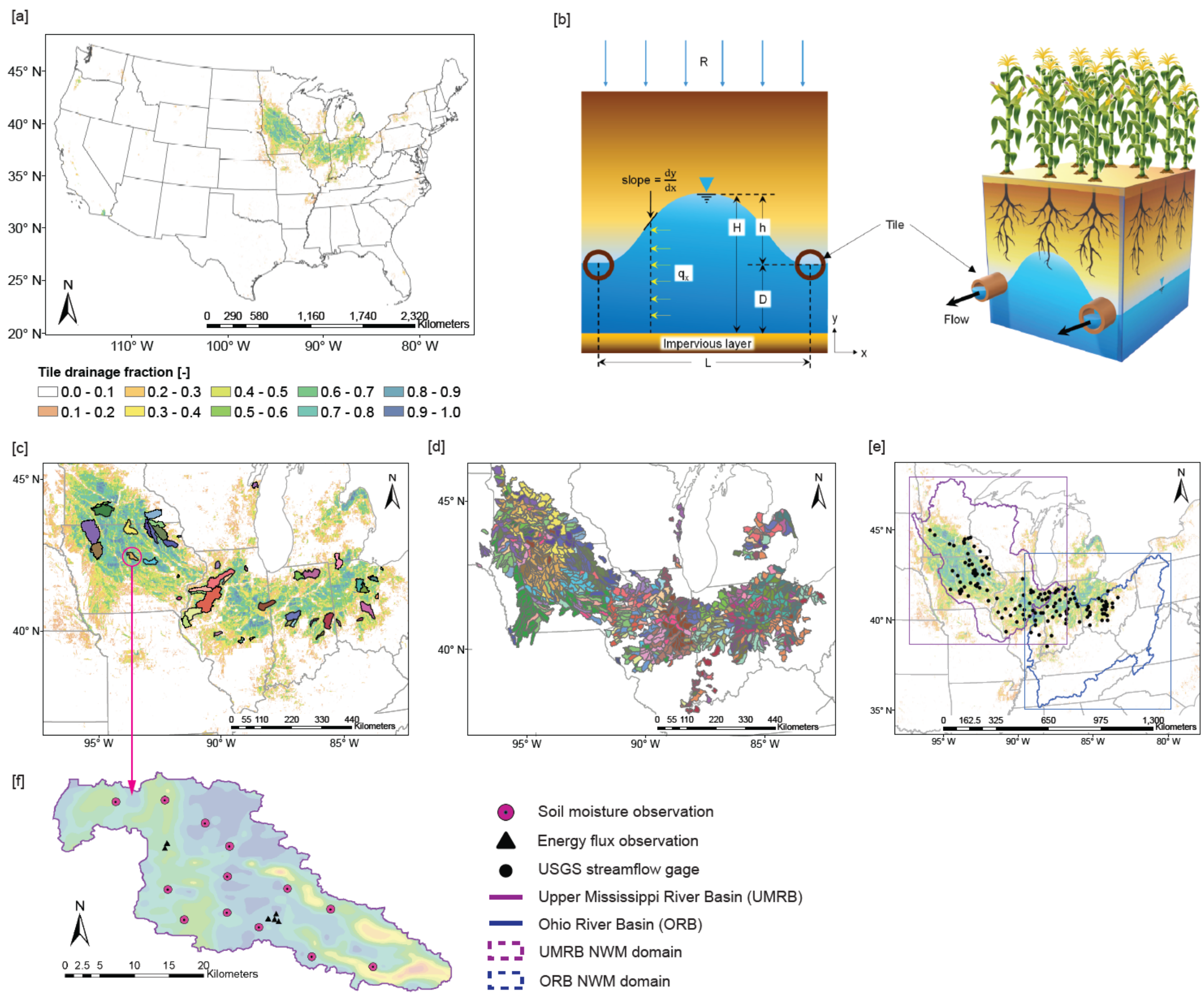

Figure 1. The Study area. (a) The spatial distribution of tile drainage over the CONUS. The color grading in (a) indicate the tile drainage area fraction on a 1-km NWM grid. (b) Schematic representation of tile drainage and parameters of Hooghoudt's tile drainage equation, (c) NWM tile drainage calibration basins, (d) spatial distribution of regionalization HUC10s. In (d), color represent corresponding donor basin for the NWM parameters in (c). (e) Represents the two HUC2 basins identified for the regional NWM simulations. (f) The spatial distribution of soil moisture and energy flux observations in the South Fork Iowa River watershed, Iowa.

2. Study area, modeling approach, and evaluation data

2.1 Study area description

Our investigation on the influence of tile drainage on the NWM performance and regional hydrology is based on the extensively tile-drained croplands of the upper Midwestern US (Figure 1 and S1). Considering computational-resource constraints, we focus on two subdomains with extensive installations of tile drainage: The Upper Mississippi River Basin (UMRB) and the Ohio River Basin (ORB) (Figure 1e). According to the AgTile-US tile drainage data (Valayamkunnath et al., 2020), nearly 50% of total tile-drained croplands of the US are in the UMRB, which accounts for 24.58% of the geographical area of the UMRB and 48% of the total cropland area of the UMRB (Figure S1). Tile-drained croplands of ORB is about 17.2% of the total tile-drained area of the US. Approximately 41.27% of the ORB croplands are tile-drained, which covers 8.79% of the geographical area of the ORB. Together, UMRB and ORB account for nearly 67% of the total tile drainage area of the US. Generally, the croplands of the upper Midwestern region are characterized by moderately to very poorly drained soils and shallow water tables (Barlage et al., 2021; Valayamkunnath et al., 2020). During the 2013-2019 period, the annual average precipitation over UMRB and ORB are 1150 mm and 1370 mm, respectively. Both basins receive the majority of the annual rainfall during the summer (June-August) season.

2.2 The National Water Model (NWM)

The NWM is a joint development between National Center for Atmospheric Research (NCAR) and NOAA NWS to provide water prediction capabilities to advance resilience to water risks. The core of the NWM is the NCAR Weather Research and Forecasting Hydrologic (WRF-Hydro) model (Gochis et al., 2018). WRF-Hydro is a parallelized distributed hydrologic model that is designed to simulate the land surface hydrology and energy states at relatively high spatial resolution (usually 1-km or less). The NWM can either be forced offline (uncoupled) using prescribed atmospheric forcing variables or coupled to the Advanced Research version of the WRF (WRF-ARW) atmospheric model (Skamarock et al., 2008). Atmospheric forcing data required for the model operation include incoming shortwave radiation (Wm^{-2}), incoming longwave radiation (Wm^{-2}), specific humidity ($kg\ kg^{-1}$), air temperature (K), surface pressure (Pa), liquid water precipitation rate ($mm\ s^{-1}$), and near-surface wind (both u and y components, $m\ s^{-1}$).

The NWM uses the Noah-MP land surface model (Niu et al. 2011) to resolve land surface processes and vertical fluxes of energy (sensible and latent heat, net radiation) and water (canopy interception, infiltration, infiltration-excess, deep percolation) within the soil column on a 1-km grid every 60 minutes. Infiltration excess, ponded water depth, and soil moisture are subsequently disaggregated from a 1-km Noah-MP grid to a high-resolution, 250-m, NWM routing grid using a time-step weighted method, and are then used in the subsurface and overland flow terrain-routing modules (Gochis et al., 2018).

Prior to the overland flow routing, the NWM subsurface flow module computes the subsurface lateral flow and resulting changes in the water table depth in the 2-m deep soil column using Dupuit–Forcheimer assumptions (Gochis et al. 2018). If subsurface lateral flow fully saturates a model grid, exfiltration is computed and added to the infiltration excess estimated by the Noah-MP and routed as surface runoff. Overland flow is calculated at a 10-seconds time-step using a fully unsteady, spatially explicit, diffusive wave routing formulation based on the steepest gradient around each grid point (Julien et al. 1995). See Gochis et al. (2018) for more details of the surface and subsurface routing schemes of NWM. As the surface flow reaches the grid identified as a channel, it is mapped to the vector channel network and routed downstream using Muskingum-Cunge channel routing formulation. In the NWM, vector channel networks are defined using National Hydrography Dataset (NHD) Plus Version 2 (NHDPlusV2) channel networks. A conceptual exponential bucket model is used to account for the contribution of baseflow to total streamflow in the NWM. Aggregated drainage from the Noah-MP soil column is mapped to a groundwater catchment corresponding to the NHDPlusV2 channel reach or catchment topology. Using an exponential storage-discharge function NWM estimates groundwater discharge for each NHDPlusV2 channel reach/catchment pair at hourly time steps (Gochis et al. 2018).

Table 1. Calibrated NWM parameters in V2.0. ('×' in the values denote that the calibration parameter is a multiplier on the default value)

Parameter name	Description	Unit	Calibration value ranges (Minimum, Maximum)
BEXP	Pore size distribution index	dimensionless	(×0.40, ×1.90)
SMCMAX	Saturation soil moisture content (i.e., porosity)	volumetric fraction	(×0.80, ×1.20)
DKSAT	Saturated hydraulic conductivity	$m\ s^{-1}$	(×0.20, ×10.00)
RSURFEXP	Exponent in the resistance equation for soil evaporation	dimensionless	(1.00, 6.00)
REFKDT	Surface runoff parameter. Increasing REFKDT decreases surface runoff	unitless	(0.10, 4.00)
SLOPE	Linear scaling of "openness" of bottom drainage boundary	0-1	(0.00, 1.00)
RETDEPRTFAC	Multiplier on retention depth limit	unitless	(0.10, 20000.00)
LKSATFAC	Multiplier on lateral hydraulic conductivity (controls anisotropy between vertical and lateral conductivity)	unitless	(10.00, 10000.00)
Zmax	Maximum groundwater bucket depth	mm	(10.00, 250.00)
Expon	Exponent controlling rate of bucket drainage as a function of depth	dimensionless	(1.00, 8.00)
CWPVT	Canopy wind extinction parameter for canopy wind profile formulation	m^{-1}	(×0.50, ×2.00)
VCMX25	Maximum carboxylation at 25°C	$umol\ m^{-2}\ s^{-1}$	(×0.60, ×1.40)
MP	Slope of Ball-Berry conductance-to-photosynthesis relationship	unitless	(×0.60, ×1.40)
MFSNO	Melt factor for snow depletion curve; larger value yields a smaller snow cover fraction for the same snow height	dimensionless	(×0.25, ×2.00)
TD_SPAC	Tile drain spacing	m	(×0.25, ×2.00)

In this study, we use NWM version 2.0 (V2.0). The NWM has parameters that can be input into the model as tables and grids and can be tuned or calibrated depending on the research requirements. The list of important NWM V2.0 parameters identified by the NCAR to regionally calibrate NWM (Dugger et al., 2017; Gochis et al., 2019) are listed in Table 1.

2.3 Tile drainage scheme

The current NWM lacks the representation of subsurface tile drainage. To compute tile drainage runoff in the NWM, we implemented a simple analytic solution for subsurface flow to drains based on Hooghoudt's tile-drainage model (Hooghoudt 1940; Ritzema, 1994). Hooghoudt's model computes steady-state flow into the tile by applying Dupuit-Forchheimer assumptions for horizontal flow in an unconfined aquifer and Darcy's Equation. The Hooghoudt's tile-drainage model is computationally simple, and therefore is commonly used to compute the tile drainage runoff in other models, especially in the DRAINMOD model (Skaggs, 1980) and Soil and Water Assessment Tool (SWAT) model (Arnold et al., 1999; Guo et al., 2018; Moriasi et al., 2012). Hooghoudt's steady-state equation that is implemented in the NWM is represented by Equation 1.

$$q = \frac{8KDh + 4Kh^2}{L^2} \tag{1}$$

Where, q is the drainage discharge (m d^{-1}), K is the hydraulic conductivity of the soil (m d^{-1}), L is the distance between tile drains, h is mid-point water table height above the tile drains (m) and D is the height of tile drain from the bottom impervious layer (m) (Figure 1b). If the tile drain is sufficient distance above the impervious layer, the streamlines will converge towards the tile drain and thus no longer be horizontal. This results in longer flowlines and extra head loss. To meet the Dupuit-Forchheimer assumptions of vertical equipotential lines and horizontal flow streamlines and to correct for convergence head loss near the tile drains, D in Equation (1) is replaced with the equivalent depth term (d_e) (Moody, 1967). The equivalent depth (d_e) represents the imaginary thinner soil layer through which the same amount of water will flow per unit time as in the actual situation (Ritzema, 1994). The value of d_e can be obtained using the analytical equations developed from Hooghoudt's solutions as a function of L, D, and radius (r) of tile drain (Moody, 1967) that are provided in Ritzema (1994).

Hooghoudt's model is a suitable option for the NWM framework because it considers most factors determining subsurface flow into tiles: *K*, *L*, *D*, soil profile depth, and water table elevation. Parameter K is already defined in the NWM. Default values of D, *r* and L are prescribed based on values reported by previous studies (Guo et al., 2018; Huffman et al., 2011; Moriasi et al., 2012; Panuska 2020; Schilling and Helmers 2008; Singh et al. 2006; 2007; Singh and Helmers 2008). The water table depth term, *h* is diagnosed at each model time-step using the degree of soil saturation simulated by Noah-MP. The tile drainage estimated by the Noah-MP at 1-km is then disaggregated onto a 250-m routing grid. In the NWM channel routing module, the lateral tile drainage runoff is mapped to the nearest vector channel network and routed downstream using Muskingum-Cunge channel routing formulation. We used the 30-meter resolution AgTile-US (Valayamkunnath et al., 2020) tile drainage map re-gridded to a 1-km NWM grid to define the tile-drained area within the model (Figure 1a).

2.4 Data

2.4.1 Observations

The study used hourly streamflow measurements from 188 United States Geological Survey (USGS) streamflow gages spanning across the heavily tile-drained croplands of the Upper Midwestern US (Figure 1c and 1e). These gages are selected from a list of USGS gages over the study area based on two criteria: 1) if the missing data in the streamflow time series is less than 20%, and 2) tile drainage fraction within the catchment is greater than 10%. To further examine the influence of tile drainage on evapotranspiration and soil moisture, we used *in-situ* measurements from the South Fork Iowa River watershed collected by the Agriculture Research Service of the United States Department of Agriculture (Coopersmith et al., 2015; 2021) (Figure 1f), including six sites with hourly flux measurements (latent and sensible heat fluxes) and 12 sites with daily soil moisture measurements. To validate the NWM simulated energy fluxes, we used daytime (9 am - 5 pm local time) hourly flux measurements.

2.4.2 Forcings for NWM

To drive the NWM, we used Analysis of Record for Calibration (AORC) high-resolution (1-km), near-surface, hourly meteorological forcing data (Kitzmiller et al., 2018) is available from 1979 to the present for the CONUS. The AORC delivers hourly accumulated precipitation and other meteorological surface parameters on a 0.0083° grid mesh. It provides superior temperature and precipitation data than the widely-used National Land Data Assimilation System Version 2 (NLDAS2) meteorological forcings (Feng et al., 2019; Xia et al., 2012). The AORC is being used as the primary source of forcing data for the calibration of the operational NWM by NCAR and OWP (Feng et al., 2019). To derive high-resolution hourly precipitation, the AORC used different sources of precipitation data such as Livneh (Livneh et al., 2013), NLDAS2 (Xia et al., 2012), Stage IV (Lin and Mitchell, 2005), radar inputs, CMORPH (Joyce et al., 2004), and Climate Forecast System Reanalysis (CFSR) (Saha et al., 2014). For temperature, Livneh, NLDAS2, and Parameter Regression on Independent Slopes Method (PRISM) (Daly et al., 2002) data were used. See Kitzmiller et al. (2018) for more details on the AORC meteorological forcings. Other variables in AORC, including specific humidity, 10-m above ground wind components, terrain-level pressure, surface downward shortwave (solar) radiation flux, and longwave (infrared) radiation flux, were derived from NLDAS2.

Additional static data used for the NWM simulations include NLCD land cover (reclassified on to USGS 27-class, 30-arc second), Hybrid STATSGO/FAO Soil Texture (19-class, 30-arc second), and AgTile-US tile drainage map (30-m).

2.5 Calibration of the NWM with a tile drainage scheme

The key elements of an automated calibration workflow are the calibration data, objective function, and the optimization algorithm employed to optimize the objective function in order to minimize the model error (Gupta et al., 1998; Singh and Woolhiser 2002; Tolson and Shoemaker 2007). Following the actual NWM calibration procedure (Gochis et al., 2019), we calibrated NWM against the USGS hourly streamflow data. The objective function used for the calibration is provided in Equation 2. The standard Nash–Sutcliffe Efficiency (NSE) emphasizes the high flow performance of the model due to squared error terms. However, combining NSE of log-transformed streamflow with standard NSE provides an additional emphasis on low flows to account for background model bias. During calibration, the objective function will be minimized.

$$Objective\ function\ = 1 - \frac{(NSE + NSE_{LOG})}{2} \quad (2)$$

Here, NSE is the Nash-Sutcliffe Efficiency and NSE_{LOG} is the Log-transformed NSE (see Table 2 for more details).

Table 2. Evaluation metrics used for the performance evaluation of the NWM.

Metrics	Equation	Description
Pearson's Correlation (COR)	$r = \frac{\sum_{i=1}^{n}(m_i - \bar{m})(o_i - \bar{o})}{\sum_{i=1}^{n}(m_i - \bar{m})^2(o_i - \bar{o})^2}$	Here, m_i and $\bar{m}$ are the i^{th} value and mean of NWM simulated streamflow, respectively. o_i and $\bar{o}$ are same as above but for the observation, and n is the length of streamflow series. Values greater than 0.5 are considered acceptable levels of performance. COR is used to capture the flow timing (Benesty et al., 2009; Moriasi et al., 2007). (Optimal value = 1)
Root mean squared error (RMSE)	$RMSE = \sqrt{\sum_{i=1}^{n}(m_i - o_i)^2 / n}$	All terms have same meaning as above. But RMSE is used to capture the flow magnitude. (Optimal value = 0)
Percent bias (Bias)	$Bias = \sum_{i=1}^{n}(m_i - o_i) \times 100 / \sum_{i=1}^{n} o_i$	All terms have same meaning as above. But Bias is used to capture the flow magnitude. (Optimal value = 0)
Nash-Sutcliffe Efficiency (NSE)	$NSE = 1 - \left[\sum_{i=1}^{n}(o_i - m_i)^2 / \sum_{i=1}^{n}(o_i - \bar{o})^2\right]$	All terms have same meaning as above. Values between 0.0 and 1.0 are generally viewed as acceptable levels of performance. NSE can capture the flow timing and magnitude errors of the high flows (Moriasi et al., 2007; Nash and Sutcliffe, 1970). (Optimal value = 1)
Log-transformed Nash-Sutcliffe Efficiency (NSE_{LOG})	$NSE_{LOG} = 1 - \left[\sum_{i=1}^{n}(\log(o_i) - \log(m_i))^2 / \sum_{i=1}^{n}\left(\log(o_i) - \overline{\log(o_i)}\right)^2\right]$	All terms have same meaning as above. Values between 0.0 and 1.0 are generally viewed as acceptable levels of performance. NSE_{LOG} can capture the flow timing and magnitude errors of the low flows (Moriasi et al., 2007). (Optimal value = 1)
Weighted NSE (NSE_{WT})	$NSE_{WT} = (NSE + NSE_{LOG}) / 2$	All terms have same meaning as above. Values between 0.0 and 1.0 are generally viewed as acceptable levels of performance. NSE_{WT} is used to capture flow timing and magnitude errors for low flows and high flows. (Moriasi et al., 2007). (Optimal value = 1)
Kling-Gupta Efficiency (KGE)	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_m}{\sigma_o} - 1\right)^2 + \left(\frac{\bar{m}}{\bar{o}} - 1\right)^2}$	Here, σ_m and σ_o are standard deviations in simulated and observed streamflow, respectively and other terms have same meaning as above. The range $-0.41 < KGE \leq 1$ could be considered as reasonable levels model performance. KGE is used to capture timing and magnitude errors. (Gupta et al., 2009; Knoben et al., 2019)

As in the official calibration strategy of the NWM V2.0, the Dynamically Dimensioned Search (DDS) algorithm (Tolson and Shoemaker, 2007) is used in this study to optimize the objective function. The algorithm is designed to scale the search in model parameter space to the user-defined maximum number of iterations. The algorithm searches globally in its initial iterations and then localizes the searches as the iterations approach the user-defined limit. The transition from global to local search is attained by dynamically and probabilistically reducing the search dimension in the neighborhood. See Tolson and Shoemaker (2007) for more details on DDS. In this study, the maximum number of iterations is set to 300 for the NWM calibration.

Since the NWM simulations are data-, time-, and computationally-intensive, calibrating it for the large river basins of the US in a single experiment is a cumbersome task. According to Feng et al. (2019), about 1469 basins across the CONUS are identified from USGS GAGES II reference basins, California Department of Water Resources (CADWR) basins, and NOAA NWS River Forecast Centers (RFC) basins for the CONUS-scale calibration of the NWM V2.0. Calibration basins are selected based on basin size, completeness of the streamflow observation record, distribution within ecoregions level III (Omernik JM. 1995) and hydrograph characteristics in comparison to other basins in the region. A basin is selected if the basin area is between 10 km^2 and 20,000 km^2, streamflow data completeness is at least 50% for the calibration period, and the basin has minimal human interventions (i.e., dams, road density, etc.) (Feng et al., 2019). To calibrate NWM for the UMRB and ORB, we used a subset of 49 basins from V2.0 calibration basins that have the tile-drainage area greater than or equal to 10% of the basin area (Figure 1c).

Before performing the calibration, we spin-up NWM for the selected 49 basins, separately, from October 1, 2007, through October 1, 2019 period using the default model parameters. Using the model state of October 1, 2019, as the “warm start,” we executed the model calibration from October 1, 2007, through October 1, 2013. A separate 1-year spin-up from October 1, 2007, through September 30, 2008, is considered for each iteration to match the model state to current conditions and suppress most instabilities from parameter changes. The critical parameters of the NWM (V2.0) related to soil, vegetation, runoff, snow, and groundwater and their description are provided in Table 1 along with the most sensitive tile-drainage model parameter, the tile spacing (L) parameter (Moriasi et al., 2012; Sammons et al., 2015; Guo et al., 2018). Using the best parameters determined by the DDS algorithm, we ran the NWM from October 1, 2007, through

October 1, 2019. Model outputs for the water years 2007-2013 are discarded as spin-up and calibration periods, and then we evaluated the model for all the 49 basins over the period October 1, 2013, to October 1, 2019.

2.6 Regionalization of calibrated NWM parameters

The total area of the calibrated basins is less than 10% of the area of UMRB and ORB combined. To compare the NWM performance with tile drainage and to quantify impacts of tile drainage on regional hydrology, regional NWM simulation experiments are necessary. To execute the NWM for regional domains presented in Figure (1e), appropriate parameters are required to be assigned for each 1-km model grid cell in the study domain. The purpose of the parameter regionalization is to transfer parameters from the calibration basins (donors) to the uncalibrated basins or 1-km model grids (receiver) (Beck et al., 2016; He et al., 2011; Hrachowitz et al., 2013; Razavi and Coulibaly, 2013). The most critical parts of the parameter regionalization process are identifying donor basins for uncalibrated areas and choosing an optimal regionalization approach. We used the regionalization based on maximum hydrological similarity (or minimum hydrologic distance) to identify donor basins for uncalibrated areas (Beck et al., 2016; Garambois et al., 2015; Sellami et al., 2014; Singh et al., 2014; Wallner et al., 2013). It is reasonable to assume that basins with similar climate, topography, vegetation, geology and soil properties have identical NWM parameters and produce similar hydrological responses. The hydrologic similarity or hydrologic distance is measured by the Gower's distance metric (Gower, 1971).

To calculate the Gower's distance between donor and receiver basins, we considered several attributes (see Table 3) based on the Hydrological Landscape Region (HLR) concept (Liu et al., 2008; Winter, 2001; Wolock et al., 2004). Before using the Gower's distance metric, we conducted a principal component analysis (PCA) to remove potential correlation between the basin attributes. Each basin attribute is scaled to [-1, 1] by subtracting the mean and then dividing by the standard deviation before the PCA. We used the following equation to quantify the Gower's distance,

$$S_{ij=} \frac{\sum_{k=1}^{n} s_{ijk}\,\delta_{ijk}}{\sum_{k=1}^{n} \delta_{ijk}} \tag{3}$$

Table 3. Basin attributes used for characterizing hydrologic similarity in NWM 2.0 with tile drainage scheme

Category	Attribute	Notes
Landform	Percent flatland (total)	Total percent cover of flatland in the basin; flatland refers to areas with a slope of less than 0.01
	Percent flatland (upland)	Upland refers to areas above the middle elevation of the basin
	Percent flatland (lowland)	Lowland refers to areas below the middle elevation of the basin
	Relief	Difference between the highest and lowest elevations
	Circularity index	The ratio of the basin’s area over the area of a circle with the same length of perimeter as the basin
Soil and geology	Percent sand	Mean percentage of sand in the soil column (upper 2m)
	Percent clay	Mean percentage of clay in the soil column (upper 2m)
	Depth to bedrock	Average thickness of soil
Land cover	Percent forest	Percent cover of forest (all types) in the basin
	Percent cropland	Percent cover of cropland (all types) in the basin
	Percent urban	Percent cover of urban areas in the basin
	Percent tile drainage	Percent cover of tile drained cropland in the basin
Climate	Feddema moisture index (FMI)	1-(PET/P) (if P>=PET) or (P/PET)-1 (if P<PET), where P & PET are annual mean precipitation and potential evapotranspiration, respectively. See Feddema (2005) and Leibowitz et al. (2016) for more details.

Where, S_{ijk} is the distance for variable *k* between a donor (*i*) and a receiver (*j*) and δ_{ijk} is the weight on variable *k*. For numerical variables, values of S_{ijk} are estimated as the absolute difference in the values of variable k between *i* and *j*, normalized by the range of variable *k* over all observations. For categorical variables, S_{ijk} is assigned to 1 if *i* and *j* are equal on variable *k* and 0 if they are not. The variables used in Equation (3) are the scores of the principal components and weights (δ_{ijk}) are calculated based on the percentages of the total variance explained by individual principal components. The receiver basins depicted in Figure (1d) are extracted from USGS 10-digit Hydrologic Unit Code (HUC10) dataset. We selected 939 HUC10 basins over the upper Midwestern US with at least 10% tile drainage (i.e., 10% tile drainage based on the total basin area) to regionalize the calibrated NWM parameters. For each HUC10 basin, we calculated Gower's distance from all the 49 calibration basins, identify a donor basin based on minimum Gower's distance (i.e., maximum hydrologic similarity) and spatial distance from the HUC10 basin, and finally transferred all the parameters to the HUC10 basins from their respective donor basin. Using the shapefile of HUC10 basins and the NWM 1-km geogrid, we mapped the parameters to the 1-km model domain. For areas with no tile drainage, we used the parameters from the official NWM V2.0 calibration experiment by NCAR and OWP.

2.7 Simulation experiments

To examine the impact of tile drainage on the NWM performance and land surface hydrology, we conducted the following NWM simulations for the UMRB and ORB regional domains.

a. *Default*: default NWM V2.0 without parameter calibration

b. *DefaultTD*: as in *Default*, but including the tile-drainage model

c. *Calib*: NWM V2.0 with calibrated parameters, mimicking the operational NWM

d. *CalibTD*: as in *Calib* but using the tile-drainage model with calibrated tile-space parameter.

Similar to the calibration experiment, we spin-up all the four regional NWM experiments from October 1, 2012, through October 1, 2019, before performing the analysis run. Using October 1, 2019 model state as the initial condition, we re-run the model from October 1, 2012, through October 1, 2019. The first water year (i.e., the water year 2012) model outputs are discarded from the analysis as we use this as an additional model spin-up period. Simulated streamflow from model outputs is extracted for 139 USGS gage locations (Figure 1e). The results presented in this study for the UMRB and ORB regional domains are only for October 1, 2013, through October 1, 2019 period.

2.8 Analysis

The analyses conducted in this study to evaluate the model performance include hydrograph analysis and statistical analysis using various statistical performance metrics provided in Table 2. We evaluated the model simulated high flows, low flows, and streamflow events with observations using hydrograph analysis. We derived high flows and low flows based on observed streamflow quantiles. We split the observed and model estimated streamflow time series into 99 segments based on streamflow quantiles ranging from 1 to 100% for every observation. Low flow is defined as streamflow below the median (50th quantile), and high flow is streamflow above the median (see Figure S2 in the supporting information for graphical explanations). For each quantile segment of the streamflow series, we estimated the model performance using metrics listed in Table 3. To identify streamflow events, we use a recently developed R package called "RNWMStat" (https://github.com/NCAR/RNWMStat) (Valayamkunnath et al., 2020). RNWMStat can detect and match streamflow events from the observed and simulated streamflow series.

The event detection algorithm in the RNWMStat follows a two-step procedure: first, the algorithm smooths the streamflow time series (simulated or observed) using the local weighted regression smoothing (LOESS) technique to remove high-frequency noises in the hydrographs; second, it determines the start, peak, and endpoints of streamflow events from the first derivative (i.e., rate of change) of smoothed streamflow series and remapped on to the original streamflow series. We matched a simulated streamflow event with an observed event if the simulated peak of an event is within the observed event (i.e., between the start and endpoints of an observed event). For the matched events, we estimate peak bias (%), timing error of peak streamflow (hours), event hit rate (%), and false alarm rate (%). Hit rate indicates the percentage of observed events that the model predicts, and false alarm rate is the percentage of model events that are not observed. For the event-based analysis, we used only the events with their peak greater than or equal to the 90^{th} percentile of streamflow. We used the Wilcoxon signed-rank test at 5%

significance level to quantify the statistical significance of the median changes in the NWM performance. The estimated p-values are provided in Table S1 to Table S3.

3. Results

3.1 NWM calibration and parameter estimation

The distributions of 14 sensitive parameters (Dugger et al., 2017; Gochis et al., 2019) from the *Default*, *Calib*, and *CalibTD* are presented in Figure 2. The physical meanings of these parameters are presented in Table 1. The new tile drainage scheme substantially altered the distributions of the NWM parameters. In *CalibTD,* the soil column is relatively water-absorbing or wetter than *Default* and *Calib*, because of its higher median values of pore size distribution index (BEXP) and soil porosity (SMCMAX). We observed a significant reduction in direct soil evaporation (RSURFEXP) and increase in infiltration (REFKDT) and surface water retention depth (RETDEPRTFAC) in *CalibTD* ($p < 0.05$). Additionally, the degree of anisotropy in the soil saturated hydraulic conductivity (LKSATFAC) is greatly reduced ($p < 0.05$) in *CalibTD* compared to *Calib*. However, the estimated LKSATFAC for *CalibTD* is significantly higher compared to *Default* ($p < 0.05$). Furthermore, the degree of openness in the bottom drainage boundary (SLOPE) is slightly higher in *CalibTD* compared to *Calib*.

Based on STATSGO2 soil data, the dominant soil types of the study region are loam, silty clay loam, and silt loam (Miller and White, 1998; USDA-NRCS, 2012). Overall, the *CalibTD* parameters ranges are acceptable for the study region with a managed agriculture and above-listed soil types (Clapp and Hornberger 1978; Lipiec et al., 2006; Livneh et al., 2015; Ma et al., 2007; Miller and White, 1998). The distributions of the NWM parameters presented in Figure 2 suggest that *CalibTD* creates favorable conditions for low surface runoff rates, high infiltration rates, a saturated soil column, and a shallow water table compared to *Calib* (Kalita et al., 2007).

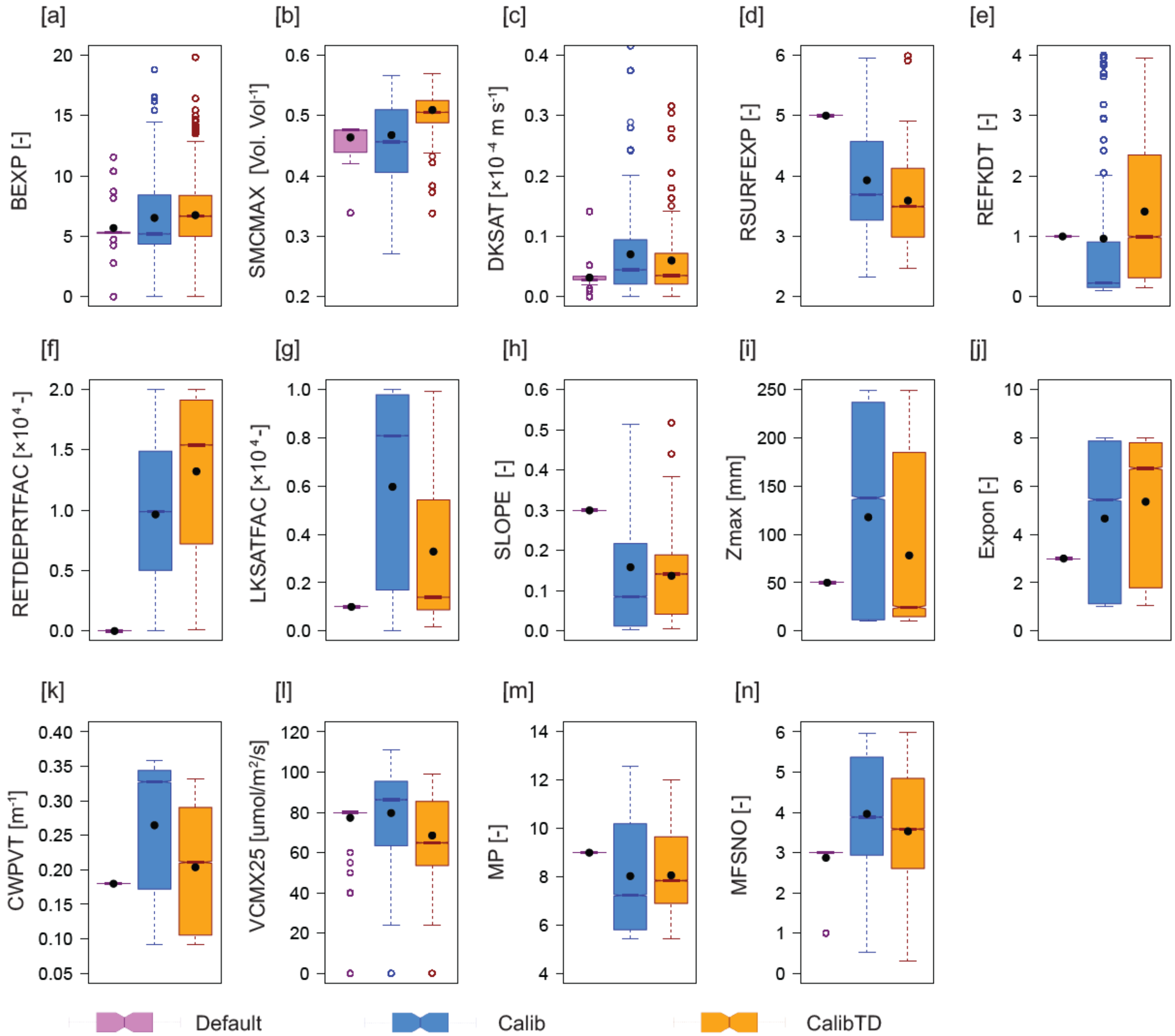

Figure 2. The distributions of the NWM parameters from *Default*, *Calib* and *CalibTD* experiments.

3.2 NWM performance evaluation: calibration and validation periods

Seasonal distributions of NWM performance evaluation metrics for calibration and validation periods are depicted in Figure 3. Representing the tile drainage process in the NWM improves the model performance during the calibration period (Figure 3a-f). Examining the *DefaultTD* model evaluation metrics indicated significant improvements in COR, NSE, NSE_{WT}, and KGE during all seasons than *Default* ($p<0.05$). Furthermore, the median and spread of RMSE are considerably reduced in *DefaultTD* during all seasons than *Default*. There are no considerable differences in the estimated Bias between *Default* and *DefaultTD*. Overall, *DefaultTD* performance is halfway between *Default* and *Calib*. That is, incorporating tile-drainage modeling into NWM using default parameters (i.e., *DefaultTD*) enhanced the NWM performance by 20% to 50% of the improvements attained by the fully-calibrated NWM (or *Calib*) from *Default* (e.g., for spring, the median NSE improved from 0.22 (*Default*) to 0.55 (*Calib*) in the non-tiled model, and from 0.22 to 0.33 in the *Default* versus *DefaultTD*). The improvement seen in the *DefaultTD* emphasizes the benefit of incorporating more physical process representation into hydrologic models, rather than relying on calibration to compensate for model deficiencies, which ultimately leads to uncertainty in model reliability across time (Andréassian, 2012; Gharari et al., 2014; Ljung, 1999).

Compared to *Default*, the biggest improvement was brought by the *Calib* based on all the metrics we considered (Figure 3a-f and Table S1). However, examining NSE, NSE_{WT}, and KGE indicated that *Calib* has considerable discrepancies in the simulated streamflow over many calibration basins. Based on the valid ranges of evaluation metrics presented in Table 2, the performance of *Calib* is unacceptable in about 18%, 6%, 20%, and 30% of the calibration basins during winter, spring, summer, and fall, respectively (Figure 3d-f). In *CalibTD*, these underperforming basin percentage is reduced to 4%, 2%, 0%, and 6%, respectively for winter, spring, summer, and fall. Additionally, we observed higher metrics medians with lower variabilities for the *CalibTD*. Seasonal analysis indicated that the NWM performance is best during summer and fall. It is due to the high amount of precipitation and streamflow during these seasons. Overall, calibration of the NWM with a tile drainage scheme (i.e., *CalibTD*) significantly improved the model performance than other model experiments ($p<0.05$) (Figure 3a-f and Table S1). Despite the improvements seen in the *DefaultTD*, it was necessary to calibrate to attain improved model performance.

Using the best parameters identified by the optimization algorithm, we executed the model for the validation period. As shown in Figure 3g-i, the *DefaultTD* outperformed *Default*. The improvements in NSE, NSE_{WT}, KGE, COR for the *DefaultTD* are significant ($p<0.05$) during winter and spring compared to *Default*. Similarly, *CalibTD* performed better than *Calib* during the validation period (Figure 3g-I and Table S2), especially during summer and fall. Examining, COR, NSE, and KGE indicated that *CalibTD* performed slightly worse during winter and spring because it failed to reproduce the flow timings and peaks accurately. Biases in the timing and intensity of snowmelt can be another reason (Suzuki and Zupanski, 2018). Overall, incorporating the tile drainage process into the NWM substantially enhanced the accuracy of the NWM over heavily tile-drained basins in the upper Midwest.

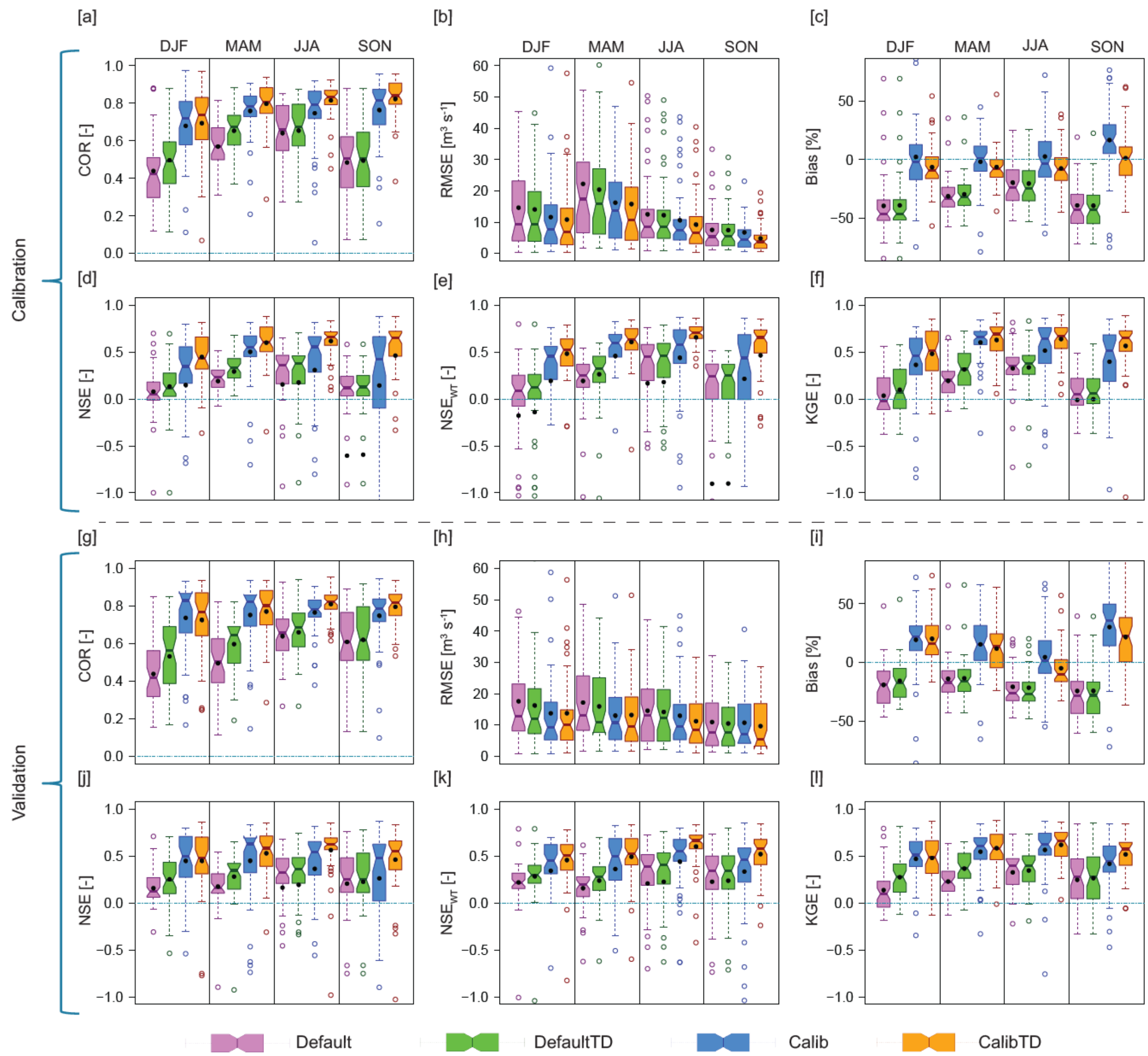

Figure 3. The NWM performance evaluation over 49 calibration basins for the calibration and validation periods. Comparison of the distribution of six evaluation metrics estimated based on the four NWM parameter experiments for the calibration (a-f) and validation (g-l) periods. Here, DJF=winter, MAM=spring, JJA=summer and SON=fall. Detailed descriptions of these metrics are provided in Table 2.

3.3 NWM performance evaluation: Regional Simulation experiments

By employing the regionalized parameters, we conducted the same set of four NWM simulations (see section 2.7) to quantify the influence of tile drainage on the NWM performance over the heavily tile-drained UMRB and ORB. The distributions of model evaluation metrics estimated using 139 USGS streamflow observations are provided in Figure 4. As mentioned earlier, *DefaultTD* is able to attain more than 50% of the improvement brought by the fully calibrated NWM from *Default* over the regional domain. It substantially enhanced the ability of NWM to capture the timing, peaks, and quantity of observed streamflow. The estimated RMSE for the *DefaultTD* is 3% to 17% less than that of the *Default*. The improvements we observed in NSE, NSE_{WT}, and KGE for the *DefaultTD* are significant ($p<0.05$) compared to *Default* in all seasons except fall (Figure 4 and Table S3). Except for RMSE in all seasons, NSE_{WT} during summer and fall, and NSE during fall, all the model evaluation metrics for the *Calib* showed significant improvements from *Default* ($p<0.05$) (Figure 4 and Table S3).

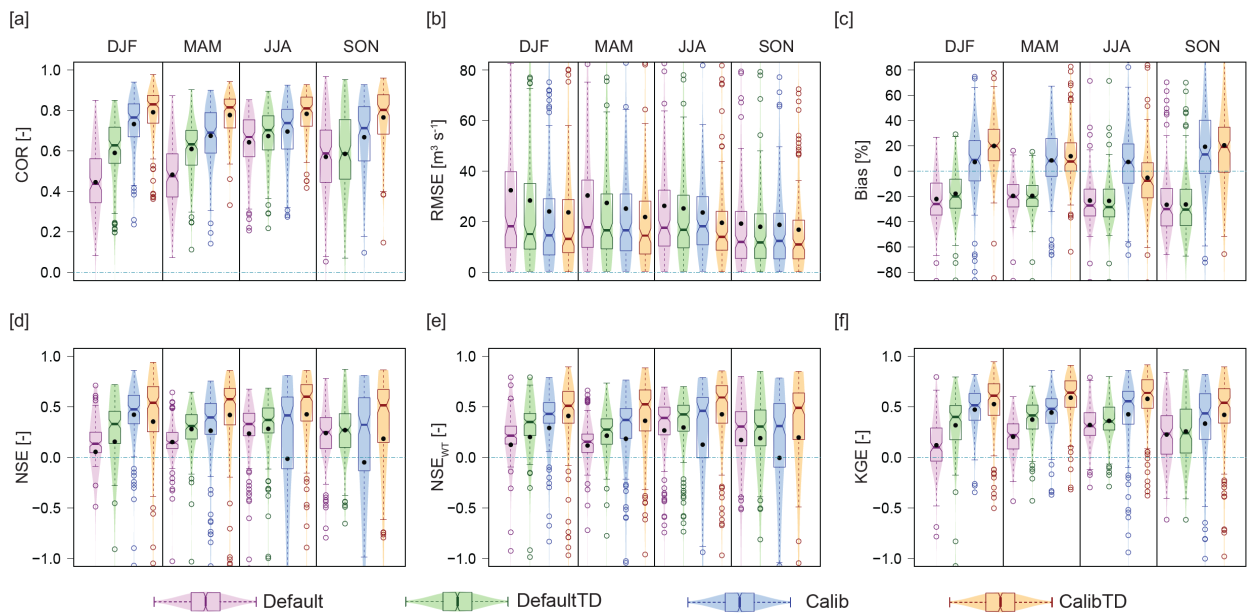

Figure 4. Seasonal NWM performance evaluation over the two HUC2 regional domains based on 139 USGS streamflow observations. Comparison of the distribution of six evaluation metrics estimated based on the four NWM parameter experiments for the regional simulation period (a-f). In (a-f), the color shading behind the boxplot indicate the data distribution density.

One of the main focuses of this study is to quantify the impact of the tile drainage scheme on calibrated NWM performance over the regional domain, and Figure 4a-f clearly shows a better performance of the *CalibTD* than *Calib*. Seasonal distributions of the model evaluation metrics showed significant ($p<0.05$) improvements in the *CalibTD* performance in reproducing the flow time, quantity, variance, and dynamics in the observed streamflow than in other model experiments. RMSE in *CalibTD* is considerably reduced by 9% to 23% compared to *Calib* (Figure 4b). However, *CalibTD* slightly overestimated (underestimated) streamflow during winter (summer) compared to observation and *Calib*, but there are no significant differences between them for spring and fall (Figure 4c and Table S3).

3.3.1 Hydrograph analysis

To understand the causes of discrepancies in the NWM simulated streamflow (mainly Bias and RMSE), we conducted hydrograph analysis using the NWM simulated streamflow from four experiments and observations. Results of the high-flow and low-flow hydrograph analysis are presented in Figure 5. The median values of performance metrics estimated for the low-flows are almost the same for *Default* and *DefaultTD* (Figure 5a, c, e, g, i, and k). The median low-flow Bias estimated for *Calib* is twice that of *Default* (Figure 5e). Even though *CalibTD* reduced low-flow biases compared to *Calib*, it still overestimated low-flows by 50%. Analyzing the distributions of NSE (Figure 5g), NSE_{WT} (Figure 5i), and KGE (Figure 5k) indicated that the NWM, in general, failed to reproduce observed low-flow accurately, consistent with previous studies assessing the NWM performance in estimating low-flows have reported similar findings (Hansen et al., 2019; Jachens et al., 2021; Karki et al., 2021). One of the reasons for the overestimation of low-flows can be the high groundwater recharge (deep percolation loss) rate in the NWM (Karki et al., 2021). The existing groundwater scheme in the NWM represents surface water–groundwater connectivity using a one-way connection from the underlying aquifer to the stream channel and omitted the influences of the stream on groundwater, and ignoring the two-way stream–aquifer fluxes in the NWM lead to overprediction of low flows (Jachens et al., 2021). Our results indicate significant reductions in the low-flow Bias and RMSE in *CalibTD* compared to *Calib*. Because tile drainage substantially reduced the groundwater recharge and rerouted the saturated soil water into the stream directly (see section 3.4 for more detailed discussion).

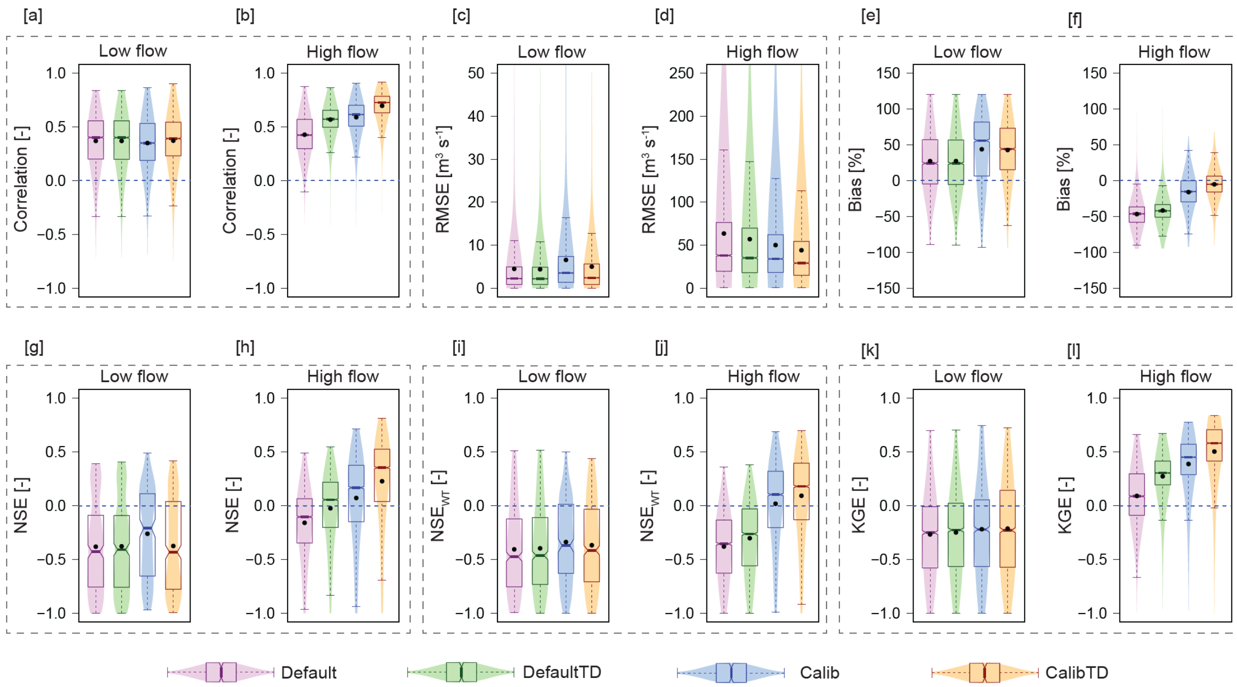

Figure 5. Evaluation of the NWM simulated high-flows and low-flows based on regional simulation. The model performance metrics are calculated by comparing the NWM estimates with 139 USGS streamflow observations. In (a-l), the color shading behind the boxplot indicate the data distribution density.

Results on high-flows revealed considerable improvements in the *DefaultTD* and *CalibTD* performance over the regional domain (Figure 5b, d, f, h, j, and l). As we highlighted before, *DefaultTD* significantly ($p<0.05$) improved the high-flow performance of the NWM compared to *Default* by increasing COR by 0.15, NSE by 0.16, and KGE by 0.22. Furthermore, *DefaultTD* is able to reduce RMSE by -2.84 m^3s^{-1} and improve the Bias by 4.2%. The variability in the model performance metrics is considerably lower in *DefaultTD* compared to *Default*. *Calib* substantially enhanced performance in reproducing the observed high-flow characteristics than *Default*. Analyzing the evaluation metrics of *Calib* indicated a significant ($p<0.05$) increase in COR by 0.19, NSE by 0.27, NSE_{WT} by 0.46, and KGE by 0.36 than in *Default*. *Calib* can better capture the timing and magnitude of observed high-flows with reduced mean error compared to *Default*. *CalibTD* further enhanced the accuracy in estimating the observed high-flow characteristics by significantly increasing COR by 0.11, NSE by 0.19, and KGE by 0.13 in *CalibTD* compared to *Calib* (Figure 5b, h, and l). Furthermore, *CalibTD* reduced the mean error by 4.88 m^3s^{-1} and Bias by 10% (Figure 5d and f). Overall, the NWM with *CalibTD* is able to better capture the timing, magnitude, and dynamics of observed high-flows very well compared to other experiments.

3.3.2 Event-based evaluation

One important goal of the NWM is to provide flash flood forecasts with longer lead times and reduced uncertainties. Thus, we analyzed the performance of NWM to capture the different characteristics of observed streamflow events using 139 USGS gage measurements. Event-based metrics estimated for different NWM experiments are presented in Figure 6. *Default* is able to reproduce about 44% of the observed streamflow events (Figure 6a). The *DefaultTD* significantly increased the event hit rate by 47% ($p<0.001$) than *Default*, and also reduced the variability in the hit rate. *Calib* significantly enhanced the hit rate of NWM by 67% ($p<0.001$) compared to *Default*. Among the four NWM experiments considered, *CalibTD* showed the highest streamflow event hit rate. The estimated hit rate in *CalibTD* is 78%, which is 7% higher than *Calib*. Moreover, the spread in the hit rate estimated for *CalibTD* is considerably lower than that of *Calib* (Figure 6a). The median false alarm rate in *Calib* is 22.5%. But in *CalibTD*, the false alarm rate is substantially reduced to 17.5% (Figure 6b).

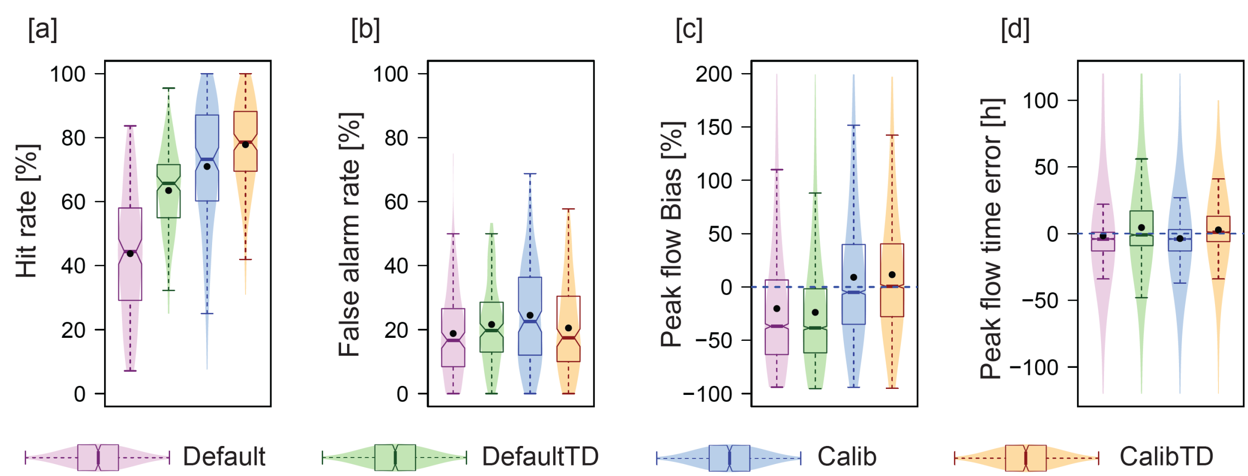

Figure 6. Event-based evaluation of the NWM based on regional simulation. The event-based statistics are calculated by comparing the NWM estimates with 139 USGS streamflow observations. In (a-d), the color shading behind the boxplot indicate the data distribution density.

Tile drainage can significantly impact the peaks and timings of streamflow events, with an earlier peak of greater magnitude (Rahman et al., 2014; Robinson et al., 1985), so we also quantified the NWM's ability to capture the peak flows, and timing of peak flows for each streamflow event. The estimated peak flow bias (%) and peak flow timing error (h) from different NWM experiments are presented in Figures 6c and 6d, respectively. There is no considerable difference between *Default* and *DefaultTD* in the estimated peak flow bias. However, *CalibTD* outperformed *Calib* and produced a lower peak flow bias of 0.57% compared to 5% in *Calib*. The median values of the estimated peak flow timing error are -3h, 0h, 4h, and 2h for *Default*, *DefaultTD*, *Calib*, and *CalibTD*, respectively. Overall, the event-based streamflow analysis indicated that NWM with *CalibTD* outperformed other NWM experiments over the heavily tile-drained UMRB and ORB. Our findings are consistent with previous studies in that the model performance to simulate streamflow over a heavily tile-drained watershed was considerably improved when they incorporated tile drainage into the model (Green et al., 2006; Hansen et al., 2013; Robinson et al., 1985; Wiskow and van der Ploeg, 2003).

3.3.3 Soil moisture evaluation

In addition to streamflow, tile drainage modifies the soil water storage. We evaluated the NWM performance using soil moisture measurements (volumetric) from 12 sites in the South Fork Iowa River watershed (Figure 1f). Using the soil moisture measurements from three different depths and NWM estimates at three model levels, we estimated COR, RMSE, and Bias in the model estimated soil moisture (Figure 7). The NWM performance in estimating the soil moisture using *Default* and *DefaultTD* is nearly identical regarding the medians of COR, RMSE. Both *Default* and *DefaultTD* showed higher median COR (0.68) and zero median Bias for the first soil layer (0-10 cm) of the NWM. A lower COR (0.60) and Bias (8%) and higher RMSE (0.062%) are estimated for the third soil layer of the NWM. Calibration substantially impacted the performance of the NMW to estimate soil moisture. For instance, *Calib* significantly reduced the NWM performance compared to *Default* by degrading COR, increasing RMSE, Bias, and their variance. This is not surprising, because the model was calibrated to optimize streamflow prediction. Although *CalibTD* underperformed compared to *Default* and *DefaultTD*, it produced better estimates of soil moisture compared to *Calib*. Also, the medians of COR, RMSE, and Bias

572 are significantly improved, and their variances are reduced when NWM employed *CalibTD* instead of *Calib*.

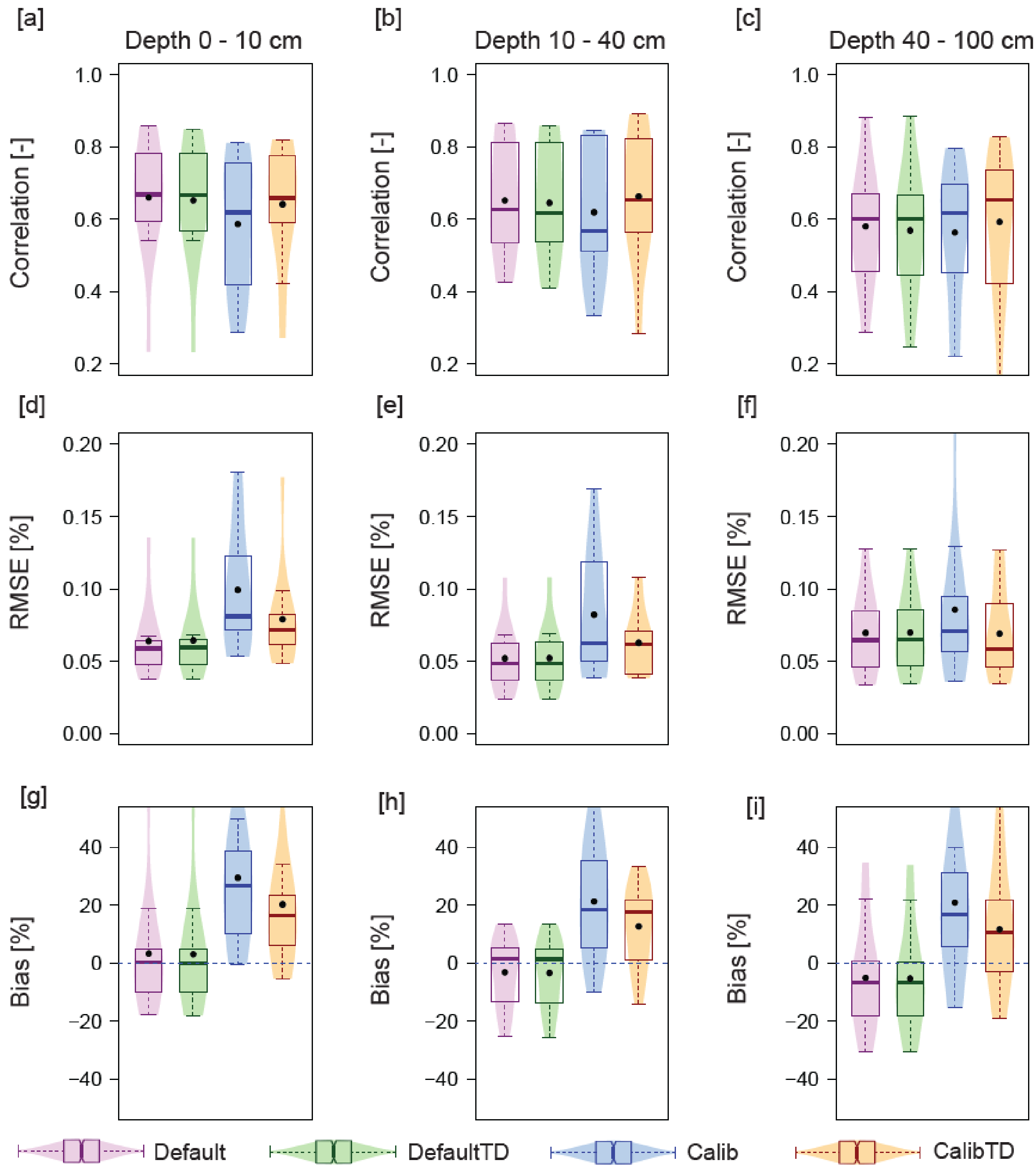

Figure 7. Evaluation of the NWM simulated soil moisture with field measurements. In (a-i), the color shading behind the boxplot indicate the data distribution density.

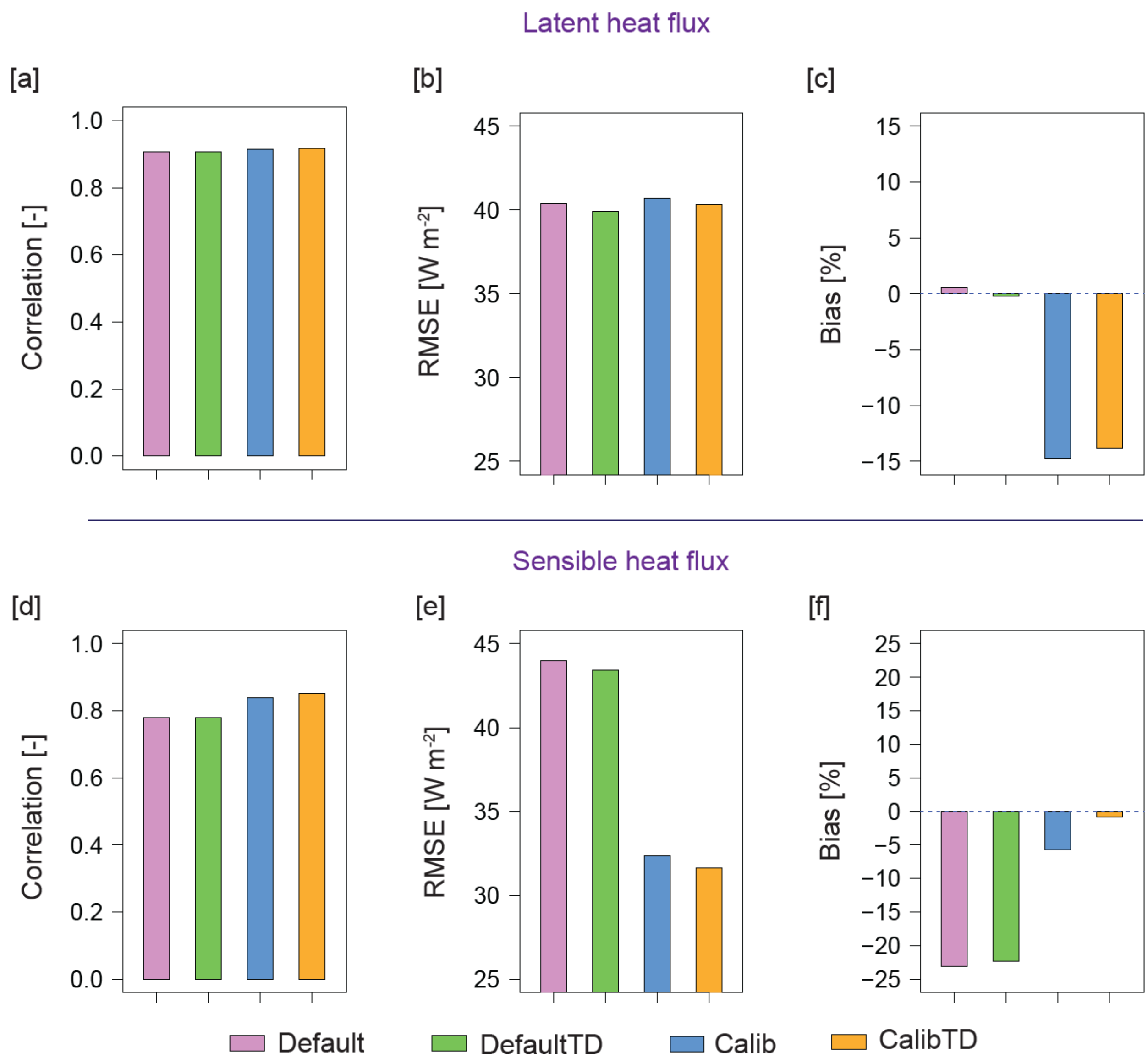

Figure 8. Accuracy assessment of NWM simulated energy balance components. (a-c) Represent the evaluation of NWM simulated latent heat fluxes (evapotranspiration), (d-f) same as (a-c), but for sensible heat fluxes.

3.3.4 Energy flux evaluation

Using the eddy covariance flux measurements from seven sites in the South Fork Iowa River watershed (Figure 1f), we evaluated the NWM simulated hourly sensible heat (SH) fluxes and latent heat (LH) fluxes (equivalent to evapotranspiration). Results of the energy flux analysis are presented in Figure 8. The results shown in Figure 8 are the averaged values of evaluation metrics estimated for the observation sites. The estimated COR and RMSE of LH for all the four NWM experiments are almost identical. Despite high correlation, the NWM estimated LH incurred a high mean error (~ 40 Wm^{-2}) (Figure 8b). NWM with *Default* and *DefaultTD* produced better estimates of LH with Bias equal to ± 1%. However, *Calib* and *CalibTD* noticeably underestimated LH by -15% and -14%, respectively. In the case of SH, *CalibTD* outperforms other NWM experiments with higher COR (0.83) and lower RMSE (32 W m^{-2}) and Bias (1%). *Calib* considerably enhanced the NWM performance in SH estimation compared to *Default* and *DefaultTD*. However, *Calib* slightly underperformed compared to *CalibTD*. Even though there are discrepancies in the NWM estimated SH and LH, our results of LH and SH indicate that the performance of the NWM is acceptable (see Table 2 for metrics ranges).

3.4. Effect of tile drainage on regional hydrology

To quantify the effects of tile drainage on regional hydrology, we analyzed land surface water balance. For this purpose, we conducted one additional NWM simulation with *CalibTD* parameters and deactivated the tile drainage scheme. This simulation with a deactivated tile drainage scheme is designated as "No tile drainage," (which is not equal to *Calib* as it uses *CalibTD* parameter set) and the NWM with *CalibTD* is defined as "With tile drainage" in this section. The results of the seasonal water balance analysis are presented in Figure 9. The results shown in Figure 9a-d are the averaged values of water balance components estimated for the tile-drained grids of the NWM within UMRB and ORB. The maximum amount of tile drainage over UMRB and ORB occurred during spring (117 ± 50 mm) followed by summer (85 ± 32 mm), winter (71 ± 40 mm), and fall (40 ± 20 mm) (Figure 9a-d). Values in the parenthesis indicate mean and one spatial standard deviation. The ratio of tile-drained water (T_D) to precipitation (P) is highest during spring (0.46), followed by winter (0.41), summer (0.20), and fall (0.12).

The results shown in Figure 9e-j are the distributions of percentage changes in the average values of water balance components that are calculated for each tile-drained grid of the NWM within UMRB and ORB. Analyzing seasonal distributions of surface runoff (S_R) changes indicated a significant decrease in S_R due to tile drainage (Figure 9e), which is consistent with previous studies (Natho-Jina et al., 1987; Robinson et al., 1985; Robinson and Rycroft, 1999; Skaggs et al., 1994). Following the seasonal tile drainage pattern, the highest decline in S_R is estimated for spring (-29%), followed by winter (-24%), summer (-14%), and fall (-7%). Tile drainage significantly decreased subsurface runoff or groundwater recharge (U_R) for all the seasons we considered (Figure 9f). This is similar to the findings of Golmohammadi et al. (2017). However, a maximum decrease is identified during spring (-50%) and summer (-50%). During winter and fall, U_R decreased by -43% and -39%, respectively. The impact of tile drainage on S_R is higher than U_R because tile drainage increases infiltration. However, all the saturation water from the infiltration are not removed by the tile drainage and a considerable amount of saturation water (5% to 10%) is still available to U_R.

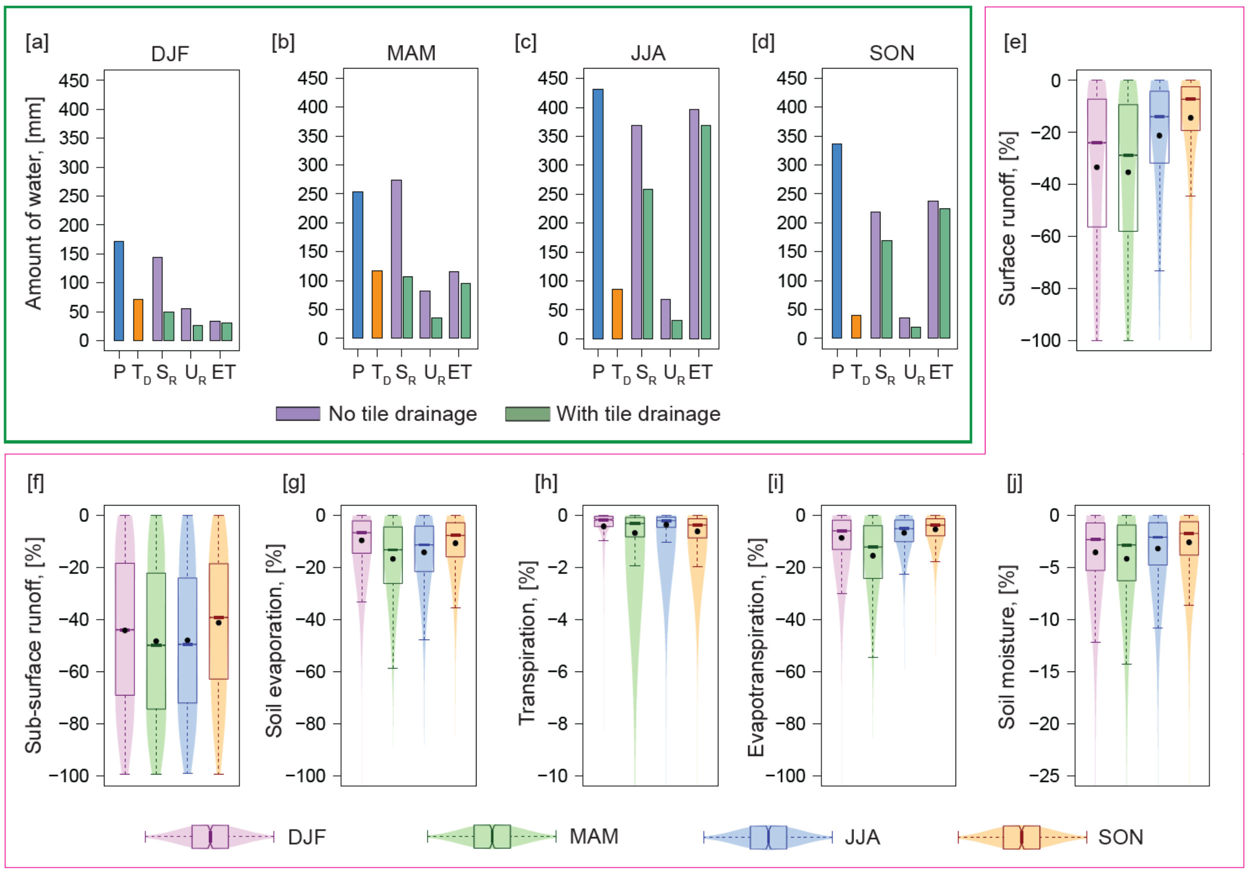

Figure 9. Impact of tile drainage on the NWM water balance components. (a-d) The seasonal totals of precipitation (P), tile drainage (T_D), surface runoff (S_R), underground runoff or groundwater recharge (U_R), and evapotranspiration (ET). The values represented in (a-d) are the averages of all the NWM tile-drained grids in the UMRB and ORB. (e-j) The changes in water balance components due to tile drainage. The results presented in (e-j), are estimated as "with tile drainage" minus "no tile drainage". In (e-j), the color shading behind the boxplot indicate the distribution density.

The main components of evapotranspiration (ET) are direct soil evaporation, transpiration, and canopy evaporation. Our analysis indicated that tile drainage significantly impacted soil evaporation (Figure 9g). The seasonal distributions of soil evaporation changes showed a more significant decrease in spring (-13%) and summer (-11%) ($p<0.05$). The reduction in soil evaporation estimated for winter and fall are -7% and -8%, respectively. Since the results on transpiration indicated minimal changes (<1%) due to tile drainage, the estimated seasonal changes in ET are almost equal to soil evaporation (Figure 9i). Studies of Khand et al. (2017), Kjaersgaard et al. (2014), and Yang et al. (2017) based on remote sensing and eddy covariance ET measurements from tile-drained croplands of the US reported similar findings on ET changes. Furthermore, we also evaluated the impact of tile drainage on root-zone soil moisture. Our results indicate that the soil moisture considerably decreased by 2% to 3% due to tile drainage. Similar findings were previously reported by many studies (Fausey 2005; Fraser and Flemming, 2001; King et al., 2014).

Additionally, we quantified the impact of tile drainage on streamflow by comparing "No tile drainage" with "With tile drainage". Tile drainage substantially altered the streamflow events by increasing peaks by 14%, increasing volume by 2.3%, delaying event start time by 2 hours, and reducing the end time by 7 hours. As indicated by previous studies, tile drainage is responsible for more short-term flashy streamflow events (De Schepper, 2017; Miller and Lyon, 2021; Rahman et al., 2014; Robinson et al., 1985). Our results indicated a considerable increase in seasonal streamflow volume due to tile drainage. The highest increase is estimated for winter (17%), followed by spring (13%), fall (13%), and summer (2.8%). Moreover, our analysis found that tile drainage enhanced the baseflow volume by 11.52%, which consistent with findings from previous studies (King et al., 2014; Moore and Larson, 1980; Schilling and Libra, 2003). However, the baseflow index is estimated as the ratio of total baseflow to the total streamflow is decreased by -9.10%. In other words, the impact of tile drainage on direct runoff (or quick flow) is more substantial compared to baseflow (Miller and Lyon, 2021). Overall, tile drainage has significant effects on most of the water balance components in the study domain.

4. Conclusion

The purpose of the study is to quantify the impacts of representing subsurface tile drainage on the National Water Model's simulated regional hydrology. We implemented Hooghoudt's tile drainage scheme into the NWM V2.0 and used 30-m resolution AgTile-US to identify tile-drained grids within the model domain. We followed the operational NWM calibration approach and calibrated 14 sensitive NWM parameters (Dugger et al., 2017; Gochis et al., 2019) along with tile spacing. Overall, the changes in these parameters suggested a water-absorbing soil column with higher infiltration rates and moisture storage potential. The calibration results also indicated reduced surface runoff and evapotranspiration over the tile-drained croplands.

Representing the tile drainage process in the NWM significantly improved its performance in estimating streamflow over the UMRB and ORB. More interestingly, the NWM with uncalibrated parameters but including a tile drainage scheme (i.e., *DefaultTD*) attained 20% to 50% of the improvements brought by the calibrated NWM (*Calib*) from *Default*. The *CalibTD* outperformed other experiments with reduced RMSE, Bias, and increased NSE, COR, and KGE. Furthermore, *CalibTD* accurately captured the dynamics in magnitude, timing, and variability of observed streamflow, especially the high-flows and low-flows. Tile drainage substantially increased peak flows, baseflow, and event volume. This significantly enhanced accuracy of the NWM to simulate high-flows in *CalibTD*. Even though *CalibTD* produced better estimates of low-flows than *Calib*, there is considerable uncertainty in the estimated low-flow timings and magnitudes. The overestimation of low-flows by the NWM can be caused by high groundwater recharge rates or lack of realism in the groundwater scheme in the NWM. Despite these discrepancies, NWM with a tile drainage scheme better estimates soil moisture, latent heat fluxes (or evapotranspiration), and sensible heat fluxes for the tile-drained croplands.

We quantified the impact of tile drainage on different water balance components, and our results indicated a significant decrease in the surface runoff, underground runoff or groundwater recharge, and evapotranspiration over UMRB and ORB. The impact of tile drainage on direct runoff (or quick flow) is more profound than on baseflow. The drainage of saturated water from the soil column by the subsurface tiles reduced the deep percolation of free water into the groundwater reservoir (Golmohammadi et al., 2017). Tile drainage removed saturated water from the soil column above the tiles and increased soil storage potential (Rahman et al., 2014). The decrease in ET over the tile drained croplands is mainly due to reduced direct soil evaporation resulting from low soil water content (Moriasi et al., 2012; Rahman et al., 2014).

Overall, tile drainage has a significant impact on regional hydrology. The representation of tile drainage process in the NWM can enhance the model's accuracy to estimate the dynamics of streamflow mainly, the timing, peaks, and volume of streamflow over a heavily tile-drained basin. Thus, our findings demonstrate the importance of incorporating tile drainage into the operational NWM for accurate flood forecasts.

Data and Code Availability Statement

All data used to generate the major figures are publicly available. The AORC data are accessed from https://hydrology.nws.noaa.gov/pub/aorc-historic/. The USGS streamflow data are available at: https://waterdata.usgs.gov/nwis/inventory/. The NLCD land cover data are available at: https://www.mrlc.gov/data/. The AgTile-US 30-m tile drainage map is available at: https://figshare.com/articles/dataset/AgTile-US/11825742/. NHDPlusV2 data can be accessed from https://nhdplus.com/NHDPlus/NHDPlusV2_data.php. The South Fork Iowa River watershed soil moisture and flux data are obtained from Coopersmith et al. (2015; 2021) (https://hrsl.ba.ars.usda.gov/southfork/index.html). The NWM source code used in this study is publicly available at: https://github.com/NCAR/wrf_hydro_nwm_public/. The RNWMStat R-Package is available at: https://github.com/NCAR/RNWMStat/.

Acknowledgements

We would like to acknowledge the support from the NOAA OAR grant NA18OAR4590381, NCAR Water System, USDA NIFA grants 2015-67003-23508 and 2015-67003-23460, and the NSF INFEWS grant 1739705. We would like to acknowledge Dr. Brian Cosgrove, Office of Water Prediction, NOAA/NWS, Silver Spring, MD, for his unconditional support in various stages of this work. Simulations are conducted using computing resources from Cheyenne High Performance Computing (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the NSF. We are grateful to three anonymous reviewers for their valuable comments.

References

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Allen, P. M. (1999). Continental scale simulation of the hydrologic balance. *JAWRA Journal of the American Water Resources Association*, *35*(5), 1037-1051. https://doi.org/10.1111/j.1752-1688.1999.tb04192.x

Barlage, M., Chen, F., Rasmussen, R., Zhang, Z., & Miguez-Macho, G. (2021). The Importance of Scale-Dependent Groundwater Processes in Land-Atmosphere Interactions Over the Central United States. *Geophysical Research Letters*, *48*(5), e2020GL092171. https://doi.org/10.1029/2020GL092171

Beck, H.E., van Dijk, A.I., De Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., & Bruijnzeel, L.A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, *52*(5), pp.3599-3622. https://doi.org/10.1002/2015WR018247

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1-4). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00296-0_5

Blann, K. L., Anderson, J. L., Sands, G. R., & Vondracek, B. (2009). Effects of agricultural drainage on aquatic ecosystems: a review. *Critical reviews in environmental science and technology*, *39*(11), 909-1001. https://doi.org/10.1080/10643380801977966

Clapp, R.B., & Hornberger, G.M. (1978). Empirical equations for some soil hydraulic properties. *Water resources research*, *14*(4), pp.601-604. https://doi.org/10.1029/WR014i004p00601

Coopersmith, E.J., Cosh, M.H., Petersen, W.A., Prueger, J., & Niemeier, J.J. (2015). Soil moisture model calibration and validation: An ARS watershed on the South Fork Iowa River. *Journal of Hydrometeorology*, *16*(3), pp.1087-1101. https://doi.org/10.1175/JHM-D-14-0145.1

Coopersmith, E.J., Cosh, M.H., Starks, P.J., Bosch, D.D., Holifield Collins, C., Seyfried, M., Livingston, S., & Prueger, J. (2021). Understanding temporal stability: a long-term analysis of USDA ARS watersheds. *International Journal of Digital Earth*, pp.1-12. https://doi.org/10.1080/17538947.2021.1943550

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate research*, *22*(2), pp.99-113. http://dx.doi.org/10.3354/cr022099

De Schepper, G., Therrien, R., Refsgaard, J. C., He, X., Kjaergaard, C., & Iversen, B. V. (2017). Simulating seasonal variations of tile drainage discharge in an agricultural catchment. *Water Resources Research*, *53*(5), 3896-3920. https://doi.org/10.1002/2016WR020209

Du, B., Arnold, J. G., Saleh, A., & Jaynes, D. B. (2005). Development and application of SWAT to landscapes with tiles and potholes. *Transactions of the ASAE*, *48*(3), 1121-1133. http://dx.doi.org/10.13031/2013.18522

Dugger, A. L., Gochis, D. J., Yu, W., Barlage, M., Yang, Y., McCreight, J., Karsten, L., Rafieeinasab, A., & Sampson, K. (2017). Learning from the National WaterModel: Regional improvements in streamflow prediction through experimental parameter and physics updates to the WRF-Hydro Community Model. 31st Conf. on Hydrology, Seattle, WA, Amer.Meteor. Soc.,6A.3.

Eastman, M., Gollamudi, A., Stämpfli, N., Madramootoo, C. A., & Sarangi, A. (2010). Comparative evaluation of phosphorus losses from subsurface and naturally drained agricultural fields in the Pike River watershed of Quebec, Canada. *Agricultural Water Management*, *97*(5), 596-604. http://dx.doi.org/10.1016/j.agwat.2009.11.010

Fausey, N. R. (2005). Drainage management for humid regions. *International Agricultural Engineering Journal*, *14*(4), 209-214.

Fausey, N. R., Doering, E. J., Palmer, M. L., & Pavelis, G. A. (1987). Purposes and benefits of drainage. *Economic Research Service, US Department of Agriculture: Washington*, *1455*, 48-51.

Feddema, J. J. (2005). A revised Thornthwaite-type global climate classification. *Physical Geography*, *26*(6), 442-466. https://doi.org/10.2747/0272-3646.26.6.442

Feng, X., Rafieeinasab, A., Karsten, L., Wu, W., Kitzmiller, D., Liu, Y., Cosgrove, B., Read, L., Dugger, A.L., Zhang, Y., & FitzGerald, K. (2019). Calibrating the National Water Model V2.

1 over the Contiguous United States. In *AGU Fall Meeting Abstracts* (Vol. 2019, pp. H43I-2134).

Fraser, H., Fleming, R., & Eng, P. (2001). Environmental benefits of tile drainage. *Prepared for: LICO. Land Improvement Contractors of Ontario. Ridgetown College, University of Guelph*.

Garambois, P.A., Roux, H., Larnier, K., Labat, D., & Dartus, D. (2015). Parameter regionalization for a process-oriented distributed model dedicated to flash floods. *Journal of Hydrology*, *525*, pp.383-399. https://doi.org/10.1016/j.jhydrol.2015.03.052

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & Savenije, H.H.G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, *18*(12), pp.4839-4859. https://doi.org/10.5194/hess-18-4839-2014

Gochis, D., Yates, D., Sampson, K., Dugger, A., McCreight, J., Barlage, M., RafieeiNasab, A., Karsten, L., Read, L., Zhang, Y., & McAllister, M. (2019). Overview of national water model calibration general strategy & optimization. *National Center for Atmospheric Research*.

Gochis, D.J., Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., McCreight, J., Mills, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., & Yu, W. (2018). The WRF-Hydro modeling system technical description, (Version 5.0). NCAR Technical Note. 107 pages. Available online at: https://ral.ucar.edu/sites/default/files/public/WRFHydroV5TechnicalDescription.pdf.

Golmohammadi, G., Rudra, R., Prasher, S., Madani, A., Youssef, M., Goel, P., & Mohammadi, K. (2017). Impact of tile drainage on water budget and spatial distribution of sediment generating areas in an agricultural watershed. *Agricultural Water Management*, *184*, 124-134. http://dx.doi.org/10.1016/j.agwat.2017.02.001

Gower, J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, pp.857-871. https://doi.org/10.2307/2528823

Green, C. H., Tomer, M. D., Di Luzio, M., & Arnold, J. G. (2006). Hydrologic evaluation of the soil and water assessment tool for a large tile-drained watershed in Iowa. *Transactions of the ASABE*, *49*(2), 413-422. https://doi.org/10.13031/2013.20415

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.E., & Moran, M.S. (2014). Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proceedings of the National Academy of Sciences*, *111*(14), pp.E1327-E1333. https://doi.org/10.1073/pnas.1320008111

Guo, T., Gitau, M., Merwade, V., Arnold, J., Srinivasan, R., Hirschi, M., & Engel, B. (2018). Comparison of performance of tile drainage routines in SWAT 2009 and 2012 in an extensively tile-drained watershed in the Midwest. *Hydrology and Earth System Sciences*, *22*(1), 89-110. https://doi.org/10.5194/hess-22-89-2018

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, *377*(1-2), 80-91. https://doi.org/10.1016/J.JHYDROL.2009.08.003

Gupta, H.V., Sorooshian, S., & Yapo, P.O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), pp.751-763. https://doi.org/10.1029/97WR03495

Hansen, A. L., Refsgaard, J. C., Christensen, B. S. B., & Jensen, K. H. (2013). Importance of including small-scale tile drain discharge in the calibration of a coupled groundwater-surface water catchment model. *Water Resources Research*, *49*(1), 585-603. http://dx.doi.org/10.1029/2011WR011783

He, Y., Bárdossy, A., & Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, *15*(11), pp.3539-3553. https://doi.org/10.5194/hess-15-3539-2011

Hooghoudt, S. B. (1940). General consideration of the problem of field drainage by parallel drains, ditches, watercourses, and channels. *Contribution to the knowledge of some physical parameters of the soil*, *7*.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., & Fenicia, F. (2013). A decade of Predictions in Ungauged Basins. *Hydrolog. Sci. J.*, p.803183. https://doi.org/10.1080/02626667.2013.803183

Huffman, R. L., Fangmeier, D. D., Elliot, W. J., Workman, S. R., & Schwab, G. O. (2011). *Soil and water conservation engineering*. St. Joseph: American Society of Agricultural and Biological Engineers. https://doi.org/10.13031/swce.2013

Jachens, E. R., Hutcheson, H., Thomas, M. B., & Steward, D. R. (2021). Effects of Groundwater-Surface Water Exchange Mechanism in the National Water Model over the Northern High Plains Aquifer, USA. *JAWRA Journal of the American Water Resources Association*, *57*(2), 241-255. https://doi.org/10.1111/1752-1688.12869

Joyce, R.J., Janowiak, J.E., Arkin, P.A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of hydrometeorology*, *5*(3), pp.487-503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Julien, P. Y., Saghafian, B., & Ogden, F. L. (1995). Raster-based hydrologic modeling of spatially-varied surface runoff. *JAWRA Journal of the American Water Resources Association*, *31*(3), 523-536. https://doi.org/10.1111/j.1752-1688.1995.tb04039.x

Kalita, P.K., Cooke, R.A., Anderson, S.M., Hirschi, M.C., & Mitchell, J.K. (2007). Subsurface drainage and water quality: The Illinois experience. *Transactions of the ASABE*, *50*(5), pp.1651-1656. https://doi.org/10.13031/2013.23963

Karki, R., Krienert, J.M., Hong, M., & Steward, D.R. (2021). Evaluating Baseflow Simulation in the National Water Model: A Case Study in the Northern High Plains Region, USA. *JAWRA Journal of the American Water Resources Association*, *57*(2), pp.267-280. https://doi.org/10.1111/1752-1688.12911

Khand, K., Kjaersgaard, J., Hay, C., & Jia, X. (2017). Estimating impacts of agricultural subsurface drainage on evapotranspiration using the Landsat imagery-based METRIC model. *Hydrology*, *4*(4), 49. https://doi.org/10.3390/hydrology4040049

King, K. W., Fausey, N. R., & Williams, M. R. (2014). Effect of subsurface drainage on streamflow in an agricultural headwater watershed. *Journal of hydrology*, *519*, 438-445. https://doi.org/10.1016/j.jhydrol.2014.07.035

Kitzmiller, D.H., Wu, W., Zhang, Z., Patrick, N., & Tan, X. (2018). The Analysis of Record for Calibration: A High-Resolution Precipitation and Surface Weather Dataset for the United States. In *AGU Fall Meeting Abstracts* (Vol. 2018, pp. H41H-06). https://ui.adsabs.harvard.edu/#abs/2018AGUFM.H41H..06K/abstract

Kjaersgaard, J., Khand, K., Hay, C., & Jia, X. (2014). Estimating Evapotranspiration from Fields with and without Tile Drainage Using Remote Sensing. In *World Environmental and Water Resources Congress 2014* (pp. 1745-1753).

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323-4331. https://doi.org/10.5194/hess-23-4323-2019

Kornecki, T. S., & Fouss, J. L. (2001). Quantifying soil trafficability improvements provided by subsurface drainage for field crop operations in Louisiana. *Applied Engineering in Agriculture*, *17*(6), 777. http://dx.doi.org/10.13031/2013.6846

Leibowitz, S. G., Comeleo, R. L., Wigington Jr, P. J., Weber, M. H., Sproles, E. A., & Sawicz, K. A. (2016). Hydrologic landscape characterization for the Pacific Northwest, USA. *JAWRA Journal of the American Water Resources Association*, *52*(2), 473-493. https://doi.org/10.1111/1752-1688.12402

Lin, Y. & Mitchell, K.E. (2005). 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications. In *Proceedings of the 19th Conference Hydrology, American Meteorological Society, San Diego, CA, USA* (Vol. 10).

Lipiec, J., Kuś, J., Słowińska-Jurkiewicz, A., & Nosalewicz, A. (2006). Soil porosity and water infiltration as influenced by tillage methods. *Soil and Tillage research*, *89*(2), pp.210-220. https://doi.org/10.1016/j.still.2005.07.012

Liu, Y.Q., Durcik, M., Gupta, H.V., & Wagener, T. (2008). Developing distributed conceptual hydrological models from geospatial databases. *Groundwater-surface water interaction: process understanding, conceptualization and modelling*, pp.94-102.

Livneh, B., Kumar, R., & Samaniego, L. (2015). Influence of soil textural properties on hydrologic fluxes in the Mississippi river basin. *Hydrological Processes*, *29*(21), pp.4638-4655. https://doi.org/10.1002/hyp.10601

Livneh, B., Rosenberg, E.A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K.M., Maurer, E.P., & Lettenmaier, D.P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *Journal of Climate*, *26*(23), pp.9384-9392. http://dx.doi.org/10.1175/JCLI-D-12-00508.s1

Ljung, L. (1999). System identification: theory for the user. *PTR Prentice Hall, Upper Saddle River, NJ*, *28*.

Ma, L., Malone, R.W., Heilman, P., Ahuja, L.R., Meade, T., Saseendran, S.A., Ascough II, J.C., & Kanwar, R.S. (2007). Sensitivity of tile drainage flow and crop yield on measured and calibrated soil hydraulic properties. *Geoderma*, *140*(3), pp.284-296. http://dx.doi.org/10.1016/j.geoderma.2007.04.012

Magner, J. A., Payne, G. A., & Steffen, L. J. (2004). Drainage effects on stream nitrate-N and hydrology in south-central Minnesota (USA). *Environmental monitoring and assessment*, *91*(1), 183-198. https://doi.org/10.1023/B:EMAS.0000009235.50413.42

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*(3), 250-254. https://doi.org/10.1038/nclimate2516

Miller, D.A., & White, R.A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth interactions*, *2*(2), pp.1-26. https://doi.org/10.1175/1087-3562(1998)002%3C0001:ACUSMS%3E2.3.CO;2

Miller, S. A., & Lyon, S. W. (2021). Tile drainage causes flashy streamflow response in Ohio watersheds. *Hydrological Processes*, *35*(8), e14326. https://doi.org/10.1002/hyp.14326

Moody, W. T. (1967). Closure to “Nonlinear Differential Equation of Drain Spacing”. *Journal of the Irrigation and Drainage Division*, *93*(3), 265-269. https://doi.org/10.1061/JRCEA4.0000511

Moore, I. D., & Larson, C. L. (1980). Hydrologic impact of draining small depressional watersheds. *Journal of the Irrigation and Drainage Division*, *106*(4), 345-363. https://doi.org/10.1061/JRCEA4.0001324

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900. http://dx.doi.org/10.13031/2013.23153

Moriasi, D. N., Rossi, C. G., Arnold, J. G., & Tomer, M. D. (2012). Evaluating hydrology of the Soil and Water Assessment Tool (SWAT) with new tile drain equations. *Journal of soil and water conservation*, *67*(6), 513-524. https://doi.org/10.2489/jswc.67.6.513

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282-290. http://dx.doi.org/10.1016/0022-1694(70)90255-6

Natho-Jina, S., Prasher, S. O., Madramootoo, C. A., & Broughton, R. S. (1987). Measurements and analysis of runoff from subsurface drained farmlands. *Canadian agricultural engineering*.

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., & Tewari, M. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D12). https://doi.org/10.1029/2010JD015139

Panuska, J. (2020). The Basics of Agricultural Tile Drainage: Basic Engineering Principals 2. *University of Wisconsin-Maddison Extension. Accessed*.

Rahman, M. M., Lin, Z., Jia, X., Steele, D. D., & DeSutter, T. M. (2014). Impact of subsurface drainage on streamflows in the Red River of the North basin. *Journal of Hydrology*, *511*, 474-483. https://doi.org/10.1016/j.jhydrol.2014.01.070

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS one*, *8*(6), e66428. https://doi.org/10.1371/journal.pone.0066428

Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of hydrologic engineering*, *18*(8), pp.958-975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690

Ritzema, H. P. (1994). Subsurface flow to drains. *Drainage principles and applications*, *16*, 263-304.

Robinson, M. (1990). *Impact of improved land drainage on river flows*. Institute of Hydrology.

Robinson, M., & Rycroft, D. W. (1999). The impact of drainage on streamflow. *Agricultural drainage*, *38*, 767-800. https://doi.org/10.2134/agronmonogr38.c23

Robinson, M., Ryder, E. L., & Ward, R. C. (1985). Influence on streamflow of field drainage in a small agricultural catchment. *Agricultural water management*, *10*(2), 145-158. https://doi.org/10.1016/0378-3774(85)90003-4

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.T., Chuang, H.Y., Iredell, M., & Ek, M. (2014). The NCEP climate forecast system version 2. *Journal of climate*, *27*(6), pp.2185-2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Sammons, R.J., Mohtar, R.H., & Northcott, W.J. (2005). Modeling subsurface drainage flow of a tile-drained small watershed using DRAINMOD. *Applied Engineering in Agriculture*, *21*(5), pp.815-834. http://dx.doi.org/10.13031/2013.19709

Schilling, K. E., & Helmers, M. (2008). Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds: Exploratory hydrograph analysis. *Hydrological Processes: An International Journal*, *22*(23), 4497-4506. https://doi.org/10.1002/hyp.7052

Schilling, K. E., & Libra, R. D. (2003). Increased baseflow in Iowa over the second half of the 20th century. *JAWRA Journal of the American Water Resources Association*, *39*(4), 851-860. https://doi.org/10.1111/j.1752-1688.2003.tb04410.x

Schilling, K. E., Jindal, P., Basu, N. B., & Helmers, M. J. (2012). Impact of artificial subsurface drainage on groundwater travel times and baseflow discharge in an agricultural watershed, Iowa (USA). *Hydrological Processes*, *26*(20), 3092-3100. https://doi.org/10.1002/hyp.8337

Schottler, S. P., Ulrich, J., Belmont, P., Moore, R., Lauer, J. W., Engstrom, D. R., & Almendinger, J. E. (2014). Twentieth century agricultural drainage creates more erosive rivers. *Hydrological processes*, *28*(4), 1951-1961. https://doi.org/10.1002/hyp.9738

Sellami, H., La Jeunesse, I., Benabdallah, S., Baghdadi, N., & Vanclooster, M. (2014). Uncertainty analysis in model parameters regionalization: a case study involving the SWAT model in Mediterranean catchments (Southern France). *Hydrology and Earth System Sciences*, *18*(6), pp.2393-2413. https://doi.org/10.5194/hess-18-2393-2014

Singh, R., & Helmers, M. J. (2008). Improving crop growth simulation in the hydrologic model DRAINMOD to simulate corn yields in subsurface drained landscapes. In *2008 Providence, Rhode Island, June 29–July 2, 2008* (p. 1). American Society of Agricultural and Biological Engineers. http://dx.doi.org/10.13031/2013.24598

Singh, R., Archfield, S.A., & Wagener, T. (2014). Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments–A comparative hydrology approach. *Journal of Hydrology*, *517*, pp.985-996. http://dx.doi.org/10.1016/j.jhydrol.2014.06.030

Singh, R., Helmers, M. J., & Qi, Z. (2006). Calibration and validation of DRAINMOD to design subsurface drainage systems for Iowa's tile landscapes. *Agricultural water management*, *85*(3), 221-232. https://doi.org/10.1016/j.agwat.2006.05.013

Singh, R., Helmers, M. J., Crumpton, W. G., & Lemke, D. W. (2007). Predicting effects of drainage water management in Iowa's subsurface drained landscapes. *Agricultural water management*, *92*(3), 162-170. https://doi.org/10.1016/j.agwat.2007.05.012

Singh, V.P., & Woolhiser, D.A. (2002). Mathematical modeling of watershed hydrology. *Journal of hydrologic engineering*, *7*(4), pp.270-292. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270)

Skaggs, R. W. (1980). Combination surface-subsurface drainage systems for humid regions. *Journal of the Irrigation and Drainage Division*, *106*(4), 265-283. https://doi.org/10.1061/JRCEA4.0001318

Skaggs, R. W., Breve, M. A., & Gilliam, J. W. (1994). Hydrologic and water quality impacts of agricultural drainage∗. *Critical reviews in environmental science and technology*, *24*(1), 1-32. https://doi.org/10.1080/10643389409388459

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of computational physics*, *227*(7), 3465-3485. https://doi.org/10.1016/j.jcp.2007.01.037

Suzuki, K., & Zupanski, M. (2018). Uncertainty in solid precipitation and snow depth prediction for Siberia using the Noah and Noah-MP land surface models. *Frontiers of Earth Science*, *12*(4), pp.672-682. https://doi.org/10.1007/s11707-018-0691-2

Thomas, N. W., Arenas, A. A., Schilling, K. E., & Weber, L. J. (2016). Numerical investigation of the spatial scale and time dependency of tile drainage contribution to stream flow. *Journal of Hydrology*, *538*, 651-666. http://dx.doi.org/10.1016/j.jhydrol.2016.04.055

Tolson, B.A., & Shoemaker, C.A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1).

USDA-NASS. (2017). *2017 Census of Agriculture*. https://doi.org/10.1029/2005WR004723

USDA-NRCS, (2012). Digital general soil map of the United States (STATSGO2). Soil Survey Staff, Natural Resources Conservation Service, *United States Department of Agriculture*.

Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D. J., & Franz, K. J. (2020). Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach. *Scientific data*, *7*(1), 1-10. https://doi.org/10.1038/s41597-020-00596-x

Valayamkunnath, P., Liu, Y., McDaniel, R., & Barlage, M. (2020). RNWMStat:A community-contributed R package for evaluating the National Water Model performance in simulating streamflow. Release V.0.1 (Version V.0.1). Zenodo. http://doi.org/10.5281/zenodo.3903720

Wallner, M., Haberlandt, U., & Dietrich, J. (2013). A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. *Journal of hydrology*, *494*, pp.59-71. http://dx.doi.org/10.1016/j.jhydrol.2013.04.022

Winter, T.C. (2001). The concept of hydrologic landscapes. *JAWRA Journal of the American Water Resources Association*, *37*(2), pp.335-349. http://dx.doi.org/10.1111/j.1752-1688.2001.tb00973.x

Wiskow, E., & van der Ploeg, R. R. (2003). Calculation of drain spacings for optimal rainstorm flood control. *Journal of Hydrology*, *272*(1-4), 163-174. http://dx.doi.org/10.1016/S0022-1694(02)00262-7

Wolock, D.M., Winter, T.C., & McMahon, G. (2004). Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses. *Environmental management*, *34*(1), pp.S71-S88. https://doi.org/10.1007/s00267-003-5077-9

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., & Livneh, B. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3). https://doi.org/10.1029/2011JD016048

Yang, Y., Anderson, M., Gao, F., Hain, C., Kustas, W., Meyers, T., Crow, W., Finocchiaro, R., Otkin, J., Sun, L., & Yang, Y. (2017). Impact of tile drainage on evapotranspiration in South Dakota, USA, based on high spatiotemporal resolution evapotranspiration time series from a multisatellite data fusion system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(6), pp.2550-2564. https://doi.org/10.1109/JSTARS.2017.2680411

Water Resources Research

Supporting Information for

Modeling the hydrologic influence of subsurface tile drainage using the National Water Model

Prasanth Valayamkunnath[a,*], David J. Gochis[a], Fei Chen[a], Michael Barlage[a], and Kristie J. Franz[b]

[a]National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA, 80301

[b]Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa, USA, 50011

* *Correspondence: Prasanth Valayamkunnath (prasanth@ucar.edu)

Contents of this file

Figures S1 to S2
Tables S1 to S3

Introduction

In this supporting information, we provide two figures and three tables to support the manuscript. The spatial distributions of land use (i.e., croplands and tile drainage) in the Upper Mississippi River Basin and Ohio River Basin are presented in Figure S1. A graphical explanation for the high-flow and low-flow identification from the streamflow timeseries is presented in Figure S2. All the p-values from the statistical significance tests conducted for the manuscript are provided in Table S1 to S3.

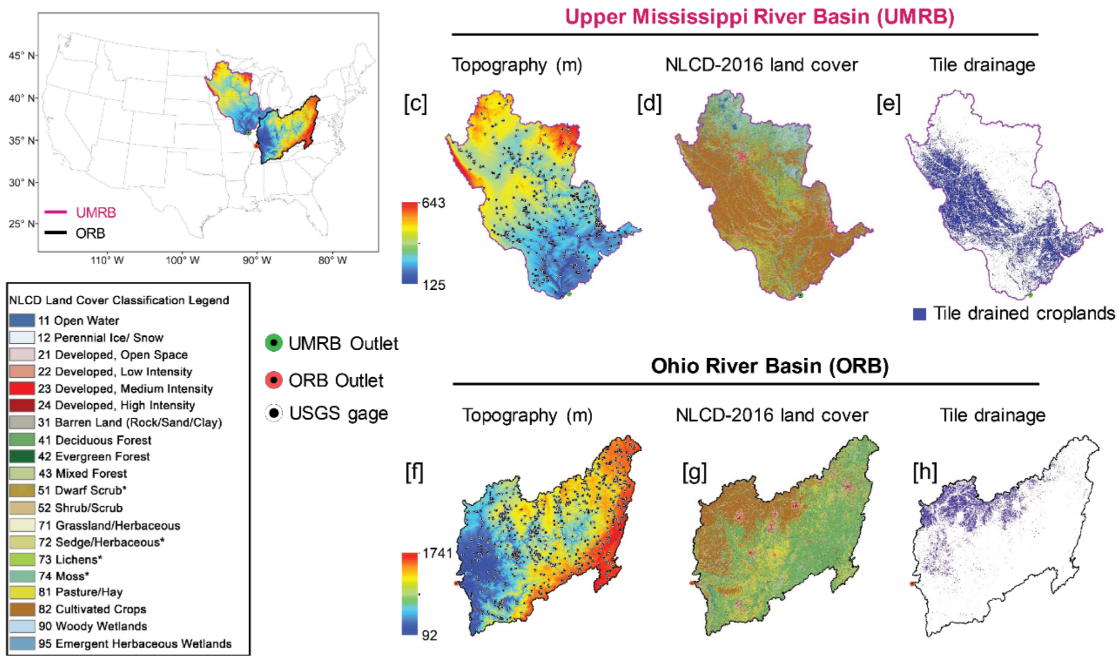

Figure S1. The spatial distributions of land use land cover in UMRB and ORB.

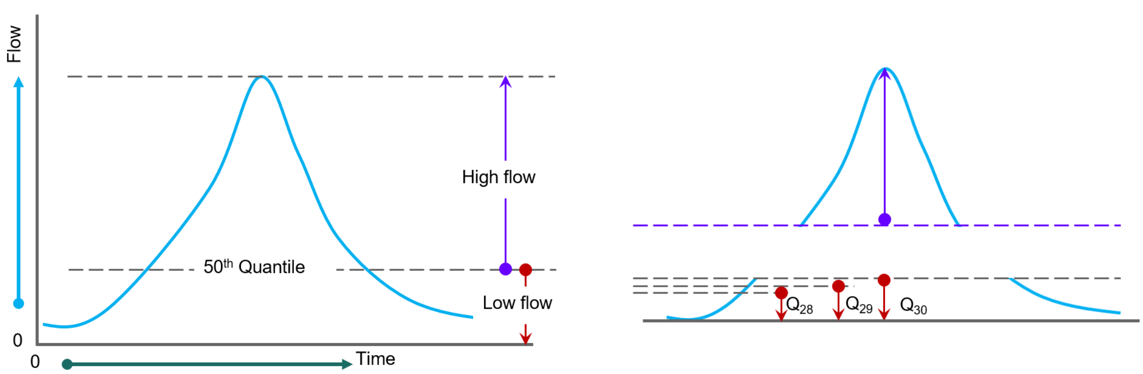

Figure S2. Schematic representation of high flow and low flow definitions used in this study.

Table S1. The statistical significance (p-value) of the NWM performance change between experiments estimated using Wilcox signed rank test for the calibration period. Red font indicates the changes are significant at 0.05 significance level.

Period	**Season**	**Metric**	**Default [V2.0] and Default [V2.0+TD]**	**Default [V2.0] and Calibrated [V2.0]**	**Default [V2.0] and Calibrated [V2.0+TD]**	**Calibrated [V2.0] and Calibrated [V2.0+TD]**
Calibration	DJF	COR	0.0721	0.0000	0.0000	0.4404
Calibration	DJF	RMSE	0.8375	0.3610	0.1896	0.6254
Calibration	DJF	PBIAS	0.8708	0.0000	0.0000	0.1306
Calibration	DJF	NSE	0.1039	0.0000	0.0000	0.0605
Calibration	DJF	NSEWT	0.4115	0.0000	0.0000	0.0557
Calibration	DJF	KGE	0.2201	0.0000	0.0000	0.0951
Calibration	MAM	COR	0.0006	0.0000	0.0000	0.0768
Calibration	MAM	RMSE	0.6055	0.1288	0.1438	0.6006
Calibration	MAM	PBIAS	0.5665	0.0000	0.0000	0.0221
Calibration	MAM	NSE	0.0010	0.0000	0.0000	0.0699
Calibration	MAM	NSEWT	0.0411	0.0000	0.0000	0.0557
Calibration	MAM	KGE	0.0018	0.0000	0.0000	0.2509
Calibration	JJA	COR	0.6105	0.0001	0.0000	0.0189
Calibration	JJA	RMSE	0.8430	0.3390	0.0909	0.3499
Calibration	JJA	PBIAS	0.7664	0.0000	0.0002	0.1101
Calibration	JJA	NSE	0.5618	0.0014	0.0000	0.0035
Calibration	JJA	NSEWT	0.7664	0.0018	0.0000	0.0059
Calibration	JJA	KGE	0.7718	0.0000	0.0000	0.1183
Calibration	SON	COR	0.7990	0.0000	0.0000	0.0994
Calibration	SON	RMSE	0.9831	0.1872	0.0243	0.2912
Calibration	SON	PBIAS	0.9099	0.0000	0.0000	0.0010
Calibration	SON	NSE	0.7881	0.0025	0.0000	0.0049
Calibration	SON	NSEWT	0.9436	0.0006	0.0000	0.0110
Calibration	SON	KGE	0.9492	0.0000	0.0000	0.0189

Table S2. The statistical significance (p-value) of the NWM performance change between experiments estimated using Wilcox signed rank test for the validation period. Red font indicates the changes are significant at 0.05 significance level.

Period	Season	Metric	Default [V2.0] and Default [V2.0+TD]	Default [V2.0] and Calibrated [V2.0]	Default [V2.0] and Calibrated [V2.0+TD]	Calibrated [V2.0] and Calibrated [V2.0+TD]
Validation	DJF	COR	0.0217	0.0000	0.0000	0.9718
Validation	DJF	RMSE	0.5908	0.0909	0.1644	1.0000
Validation	DJF	PBIAS	0.2509	0.0000	0.0000	0.3799
Validation	DJF	NSE	0.0124	0.0000	0.0000	0.9380
Validation	DJF	NSEWT	0.0135	0.0000	0.0000	0.4531
Validation	DJF	KGE	0.0021	0.0000	0.0001	0.9549
Validation	MAM	COR	0.0008	0.0000	0.0000	0.8763
Validation	MAM	RMSE	0.5859	0.1288	0.1824	0.8486
Validation	MAM	PBIAS	0.8597	0.0000	0.0000	0.2912
Validation	MAM	NSE	0.0031	0.0000	0.0000	0.6868
Validation	MAM	NSEWT	0.0496	0.0000	0.0000	0.3462
Validation	MAM	KGE	0.0005	0.0000	0.0000	0.7025
Validation	JJA	COR	0.3838	0.0000	0.0000	0.0152
Validation	JJA	RMSE	0.8375	0.3799	0.1518	0.5570
Validation	JJA	PBIAS	0.8320	0.0000	0.0000	0.0149
Validation	JJA	NSE	0.4531	0.0003	0.0000	0.0093
Validation	JJA	NSEWT	0.6558	0.0002	0.0000	0.0067
Validation	JJA	KGE	0.5288	0.0000	0.0000	0.5570
Validation	SON	COR	0.8375	0.0000	0.0000	0.1419
Validation	SON	RMSE	0.8875	0.6204	0.4573	0.5908
Validation	SON	PBIAS	0.9718	0.0000	0.0000	0.0332
Validation	SON	NSE	0.8100	0.1458	0.0003	0.0805
Validation	SON	NSEWT	0.8875	0.1039	0.0001	0.0073
Validation	SON	KGE	0.8154	0.0015	0.0000	0.0548

Table S3. The statistical significance (p-value) of the NWM performance change between experiments estimated using Wilcox signed rank test for the regional simulation. Red font indicate the changes are significant at 0.05 significance level.

Domain	**Season**	**Metric**	***Default and DefaultTD***	***Default* and *Calib***	***Default* and *CalibTD***	***Calib* and *CalibTD***
Regional	DJF	COR	0.0000	0.0000	0.0000	0.0000
Regional	DJF	RMSE	0.3787	0.0508	0.1708	0.8287
Regional	DJF	PBIAS	0.0266	0.0000	0.0000	0.0001
Regional	DJF	NSE	0.0000	0.0000	0.0000	0.1617
Regional	DJF	NSEWT	0.0000	0.0000	0.0000	0.0024
Regional	DJF	KGE	0.0000	0.0000	0.0000	0.0008
Regional	MAM	COR	0.0000	0.0000	0.0000	0.0000
Regional	MAM	RMSE	0.4326	0.2935	0.1369	0.2161
Regional	MAM	PBIAS	0.8638	0.0000	0.0000	0.5096
Regional	MAM	NSE	0.0000	0.0000	0.0000	0.0000
Regional	MAM	NSEWT	0.0000	0.0000	0.0000	0.0001
Regional	MAM	KGE	0.0000	0.0000	0.0000	0.0000
Regional	JJA	COR	0.0731	0.0005	0.0000	0.0000
Regional	JJA	RMSE	0.6772	0.8345	0.0641	0.0276
Regional	JJA	PBIAS	0.9227	0.0000	0.0000	0.0000
Regional	JJA	NSE	0.0409	0.0409	0.0000	0.0000
Regional	JJA	NSEWT	0.1897	0.1200	0.0000	0.0000
Regional	JJA	KGE	0.0339	0.0000	0.0000	0.0000
Regional	SON	COR	0.5707	0.0000	0.0000	0.0000
Regional	SON	RMSE	0.8369	0.9239	0.3731	0.2500
Regional	SON	PBIAS	0.8790	0.0000	0.0000	0.5231
Regional	SON	NSE	0.5348	0.5328	0.0006	0.0027
Regional	SON	NSEWT	0.6718	0.5840	0.0016	0.0003
Regional	SON	KGE	0.4136	0.0000	0.0000	0.0124