Enhancing the Structure of the WRF-Hydro Hydrologic Model for Semiarid Environments⌗

TIMOTHY M. LAHMERS, HOSHIN GUPTA, AND CHRISTOPHER L. CASTRO

Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, Arizona

DAVID J. GOCHIS, DAVID YATES, AND AUBREY DUGGER

National Center for Atmospheric Research, Boulder, Colorado

DAVID GOODRICH

Southwest Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture, Tucson, Arizona

PIETER HAZENBERG

Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, Arizona

ABSTRACT

In August 2016, the National Weather Service Office of Water Prediction (NWS/OWP) of the National Oceanic and Atmospheric Administration (NOAA) implemented the operational National Water Model (NWM) to simulate and forecast streamflow, soil moisture, and other model states throughout the contiguous United States. Based on the architecture of the WRF-Hydro hydrologic model, the NWM does not currently resolve channel infiltration, an important component of the water balance of the semiarid western United States. Here, we demonstrate the benefit of implementing a conceptual channel infiltration function (from the KINEROS2 semidistributed hydrologic model) into the WRF-Hydro model architecture, configured as NWM v1.1. After calibration, the updated WRF-Hydro model exhibits reduced streamflow errors for the Walnut Gulch Experimental Watershed (WGEW) and the Babocomari River in southeast Arizona. Model calibration was performed using NLDAS-2 atmospheric forcing, available from the NOAA National Centers for Environmental Prediction (NCEP), paired with precipitation forcing from NLDAS-2, NCEP Stage IV, or local gauge precipitation. Including channel infiltration within WRF-Hydro results in a physically realistic hydrologic response in the WGEW, when the model is forced with high-resolution, gauge-based precipitation in lieu of a national product. The value of accounting for channel loss is also demonstrated in the Babocomari basin, where the drainage area is greater and the cumulative effect of channel infiltration is more important. Accounting for channel infiltration loss thus improves the streamflow behavior simulated by the calibrated model and reduces evapotranspiration bias when gauge precipitation is used as forcing. However, calibration also results in increased high soil moisture bias, which is likely due to underlying limitations of the NWM structure and calibration methodology.

1. Introduction

From 1984 to 2013, flood events in the United States cost the nation $7.95 billion per year and resulted in an average of 85 fatalities per year (National Weather Service 2014), making flooding one of the deadliest and costliest hazards for the United States. Flash floods occur when atmospheric conditions are favorable for the occurrence of sustained (often convective) heavy precipitation, where sufficient ascent and moisture are both present (e.g., Doswell et al. 1996). Flash flood events are

difficult to predict in semiarid environments, such as the southwestern United States, due to the difficulty of having good estimates of antecedent soil moisture, and due to the relatively short durations and highly localized character of extreme precipitation events. Initial land surface soil moisture conditions can be highly uncertain due to limited soil moisture observations, while elevated terrain hampers accurate estimation of precipitation using weather radar (e.g., Zamora et al. 2014). Therefore, flash flood prediction in these semiarid domains is particularly challenging and is further complicated by the dry antecedent conditions of ephemeral channel beds, resulting in large transmission losses (Goodrich et al. 1997). Previous efforts to improve the physical process representation in hydrologic forecasts within these environments have focused on the implementation of spatially distributed hydrologic models, making use of gridded high-resolution precipitation data (e.g., Blöschl et al. 2008; Looper and Vieux 2012; Broxton et al. 2014; Hardy et al. 2016; Reed et al. 2007; Gourley et al. 2017). In the present study, the added value of accounting for channel infiltration in a spatially distributed hydrologic model is evaluated for two basins in the southwest United States.

To ameliorate the uncertainties associated with hydrologic forecasts, and to provide longer forecast lead times for flash flood events, the National Weather Service (NWS) Office of Water Prediction (OWP), of the National Oceanic and Atmospheric Administration (NOAA), recently implemented the operational National Water Model (NWM), based on architecture of the Weather Research and Forecasting (WRF) WRF-Hydro hydrologic model (Gochis et al. 2015). The NWM is a continental-scale distributed hydrologic model that produces streamflow forecasts for 2.7 million stream reaches across the contiguous United States (CONUS) based on observed and forecasted precipitation. For the semiarid southwestern United States, the model is subject to significant errors (Dugger et al. 2017) and, accordingly, the present study focuses on improving the performance of the WRF-Hydro model (configured as NWM v1.1) for this region.

In the southwest United States, much of the streamflow that occurs in ephemeral channels is due to surface runoff associated with convective rainfall occurring during the North American monsoon season (NAM) (e.g., Maddox et al. 1995; McCollum et al. 1995; Adams and Comrie 1997). During the NAM, precipitation can either be phase-locked to the high terrain or propagate into the low deserts as squall lines or mesoscale convective systems (MCSs; e.g., Pytlak et al. 2005; Bieda et al. 2009; Finch and Johnson 2010; Newman and Johnson 2012; Seastrand et al. 2015; Lahmers et al. 2016;

Luong et al. 2017), resulting in significant amounts of precipitation and surface runoff.

Throughout southeastern Arizona, the groundwater levels tend to be deep, so that the surface runoff that flows through the channel network infiltrates into the soil, where it becomes a source for groundwater recharge (e.g., Blasch et al. 2004). Goodrich et al. (2004) note that the depth to groundwater in the U.S. Department of Agriculture (USDA) Agricultural Research Service (USDA-ARS) Walnut Gulch Experimental Watershed (WGEW; Moran et al. 2008), a tributary to the San Pedro River in southeast Arizona with no perennial channels, ranges from approximately 50 to 145 m in the lower and middle basin, respectively. However, local recharge from channel infiltration can make up a significant portion of the local water balance, as recharge estimates (based on groundwater modeling, surface water balance measurements, and evaluation of deep groundwater storage using microgravity measurements) suggest that channel infiltration accounted for 15%–40% of groundwater recharge in the San Pedro basin during wetter than average monsoon seasons (Goodrich et al. 2003, 2004). Virtually all of the runoff generated in the WGEW is the result of monsoon convective precipitation (Goodrich et al. 1997; Stone et al. 2008).

As currently implemented, WRF-Hydro does not permit water to exit the channel network once it enters a channel from the terrain routing grid (Gochis et al. 2015). Consequently, the operational version of the NWM does not account for channel infiltration losses. Because of the importance of channel infiltration processes in semiarid regions like the southwest, we modified the WRF-Hydro structure to include channel infiltration (e.g., Goodrich et al. 2004). Specifically, we used a scheme similar to the one implemented in the KINEROS2 distributed watershed model (Goodrich et al. 2012), which was originally developed for the southwestern United States and has been widely used for hydrologic forecasting in semiarid domains. KINEROS2 has been tested and used extensively (e.g., Goodrich et al. 2012; Yatheendradas et al. 2008) in the WGEW and other basins in southern Arizona.

Previous work has attempted to quantify and estimate the nature and magnitude of channel infiltration at various scales. Lane (1983) used an ordinary differential equation to approximate the rate of change in runoff volume with channel flow distance; observed inflow–outflow discharge from ephemeral reaches were then used to compute equation coefficients via regression analysis. Noorduijn et al. (2014) modeled infiltration from an artificial channel using Philip's equation, accounting for both the gravity and pressure terms in the sorption equation, while Callegary et al. (2007) quantified the recharge

potential of ephemeral channel reaches in the southwest CONUS using field data. Redistribution of water across the land surface caused by infiltration may also affect the local water balance; Zampieri et al. (2012) showed, using a modified version of the Community Land Model (CLM), that parameterization of the redistribution of water through infiltration out of the channel network was needed for the model to reproduce Oklahoma Mesonet soil moisture observations.

In addition to incorporating a representation of channel infiltration, the present study investigates the model calibration problem; this is important due to the inherent uncertainty in the model parameter estimates and their impact on the simulated hydrologic response (e.g., Yatheendradas et al. 2008). For spatially distributed hydrologic models (potentially having tunable parameters at each grid point), calibration is complicated by the high dimensionality of the model (e.g., Smith and Gupta 2012). This problem can be ameliorated using spatial regularization (Gupta et al. 2008, 2009; Samaniego et al. 2010), wherein model parameters are associated with a priori nonlinear functions of observed surface data. Doing so results in a smaller number of hyperparameters (i.e., parameters controlling the degree of modification to the actual model parameters through transfer functions) that must be calibrated to ensure that the original model parameters are physically consistent with catchment properties (e.g., Gupta et al. 2008, 2009). For example, Pokhrel et al. (2012) used a spatial regularization approach to calibrate the NWS Hydrology Laboratory Research Distributed Hydrological Model (HL-RDHM), and Vergara et al. (2016) used spatial climate and land data to derive a priori estimates of routing parameters for the kinematic wave routing model across the CONUS to be used for flash flood forecasts. Advanced spatial regularization techniques that have been proposed include the multiscale parameter regionalization (MPR) method, which uses spatial transfer functions to compute model parameters based on high-resolution spatial data that account for spatial heterogeneity across grid cells (e.g., Samaniego et al. 2010).

In this study, we followed the approach presented by Pokhrel et al. (2012) and calibrated a select set of WRF-Hydro parameters (determined by sensitivity analysis). This methodology is consistent with the calibration methodology developed for the NWM v1.1 by the National Center for Atmospheric Research (NCAR).

We evaluate the combined impacts of accounting for channel infiltration and parameter calibration on the simulated hydrologic response of WRF-Hydro for the WGEW and the Babocomari River basin, which are both tributaries to the San Pedro River. This work was focused primarily on the WGEW (149 km^2), where

relatively accurate precipitation forcing data are available from a network of 88 weighing rain gauges. The same calibration approach was extended to the larger Babocomari River basin, where the impacts of channel infiltration are greater due to its large area, despite precipitation forcing uncertainties. The NWM model structure and datasets used are described in section 2 of this paper, and the calibration approach is outlined in section 3. Calibration results are presented in section 4. Results from the model calibration and evaluation and possible future changes to the NWM are further discussed in section 5, and conclusions are included in section 6.

2. WRF-Hydro model and user modifications

a. NWM WRF-Hydro model structure and configuration

WRF-Hydro (Gochis et al. 2015) is a parallelized distributed hydrologic model that can either be forced offline using prescribed atmospheric forcing variables, or coupled to the Advanced Research version of the WRF (WRF-ARW) atmospheric model (Skamarock et al. 2008). A simplified schematic of WRF-Hydro is shown in Fig. 1. Atmospheric forcing data needed to execute WRF-Hydro offline include incoming shortwave radiation, incoming longwave radiation, specific humidity, air temperature, surface pressure, and near surface wind (both u and v components). The NWM is a particular configuration of WRF-Hydro, which uses the Noah-MP land surface model (LSM; Niu et al. 2011) to resolve vertical fluxes within the soil column and exchanges with the atmosphere. Noah-MP is configured using gridded NWM soil parameters, which govern the drainage of water through the soil column, with 1-km grid resolution. NWM WRF-Hydro resolves horizontal surface and subsurface fluxes on a 250-m grid resolution routing grid (with 10-s and 60-min time steps, respectively). Since the routing grid cells are 4 times smaller than the Noah-MP grid cells, spatially varying quantities on the 250-m routing grid are aggregated back to the 1-km grid during model time steps when Noah-MP is called (every 60 min) and disaggregated back to the 250-m routing grid.

The subsurface flow module on the WRF-Hydro grid computes changes to the water table in the 250-m soil column (which is assumed to be 2 m deep everywhere), using Dupuit–Forcheimer assumptions. This assumes that the hydraulic gradient is based on differences in the groundwater table depth along the steepest gradient in eight possible directions around a routing grid point (Gochis et al. 2015). If subsurface flow causes a model grid point to become saturated, exfiltration is computed and this resultant water ponding is combined with
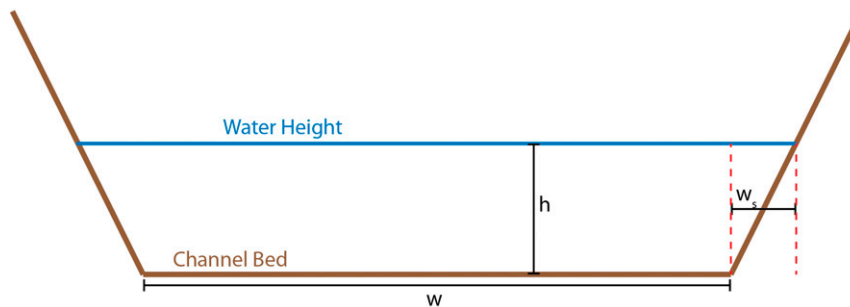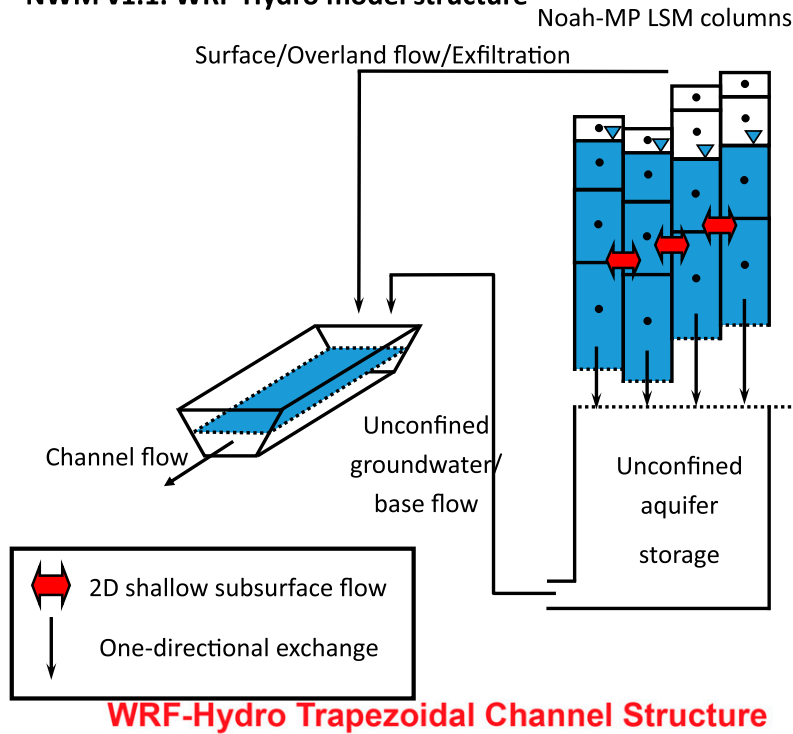
NWM v1.1. WRF-Hydro model structure

FIG. 1. Illustration of the uncoupled WRF-Hydro hydrologic model structure. The Noah-MP LSM and routing grid columns are shown at the top right. Baseflow from the bottom of the Noah-MP LSM is passed to the baseflow bucket model (shown middle right), labeled as unconfined aquifer storage. Water from the baseflow bucket model and surface runoff from the terrain routing grid are both returned to the channel network shown at left. Channel infiltration is a sink in the model structure (assumed to be deep groundwater recharge). The vertical cross section of a trapezoidal channel used by the WRF-Hydro Muskingum–Cunge routing scheme is also shown at the bottom. The sides of the channel are assumed infinite. The channel width w is the BtmWdth parameter, and the channel side slope h/w_s is the ChSlp parameter.

infiltration excess and routed as surface runoff. Surface flow is computed using diffusive wave routing based on the steepest gradient around each grid point (Julien et al. 1995; Ogden 1997). Details of the surface and subsurface routing schemes of WRF-Hydro are discussed in detail in Gochis et al. (2015).

When surface flow reaches a grid cell that is designated as a channel, it is mapped to the vector channel network and routed downstream (with a 5-min time step). The NWM channel network is based on the National Hydrography Dataset (NHD) Plus Version 2 (NHDPlusV2) (McKay et al. 2012). In the present study, we added a channel infiltration parameter that is physically representative of the channel bed conductivity ($m\,s^{-1}$) (ChannK; Table 1). Flow in the channels is computed using an iterative Muskingum–Cunge function for each reach.

TABLE 1. NWM parameters considered for calibration by NCAR (Dugger et al. 2017) and channel parameters considered as part of the present study. Parameters are organized based on areas of the WRF-Hydro model structure.

Name	Description	Units
	Soil parameters	
BEXP	Pore size distribution index	Dimensionless
SMCMAX	Saturation soil moisture content (i.e., porosity)	Volumetric fraction
DKSAT	Saturated hydraulic conductivity	$m\,s^{-1}$
	Runoff parameters	
REFKDT	Surface runoff parameter; REFKDT is a tunable parameter that significantly impacts surface infiltration and hence the partitioning of total runoff into surface and subsurface runoff. Increasing REFKDT decreases surface runoff.	Unitless
SLOPE	Linear scaling of "openness" of bottom drainage boundary	0–1
RETDEPRTFAC	Multiplier on retention depth limit	Unitless
LKSATFAC	Multiplier on lateral hydraulic conductivity (controls anisotropy between vertical and lateral conductivity)	Unitless
	Groundwater parameters	
Zmax	Maximum groundwater bucket depth	mm
Expon	Exponent controlling rate of bucket drainage as a function of depth	Dimensionless
	Vegetation parameters	
CWPVT	Canopy wind parameter for canopy wind profile formulation	m^{-1}
VCMX25	Maximum carboxylation at 25°C	μmol m^{-2} s^{-1}
MP	Slope of Ball–Berry conductance relationship	Unitless
HVT	Canopy top height	m
	Snow parameters	
MFSNO	Melt factor for snow depletion curve; larger value yields a smaller snow cover fraction for the same snow height	Dimensionless
	Channel parameters	
ChannK	Channel bed conductivity (for channel infiltration function)	$m\,s^{-1}$
BtmWdth	Bottom width of channel	m
ChSlp	Channel side slope	Unitless
N	Manning's N	$s\,m^{-1/3}$

This vector routing scheme is more computationally efficient than other WRF-Hydro channel routing configurations (e.g., 1D diffusive wave routing) and has the capability to be mapped to specific rivers and reaches that may be of interest to emergency managers and stakeholders; however, it cannot resolve backwater flow (Gochis et al. 2015).

Deep baseflow (below the soil column) in WRF-Hydro is computed using a conceptual exponential bucket model. All water that drains out of the Noah-MP LSM 2-m soil column is mapped to a groundwater catchment, which corresponds to the NHDPlusV2 channel reach/catchment topology. There are a few relatively small reaches in the NWM domain that are not mapped directly to an NHD bucket. Water from this conceptual groundwater bucket is gradually returned to the channel reach that directly corresponds to its underlying catchment, where the rate of discharge is controlled by three tuning parameters in an exponential equation.

WRF-Hydro has parameters that may be specified through input tables and grids, and these parameters may be adjusted or calibrated depending on the study region and hydrologic behavior of interest. Priority Noah-MP and WRF-Hydro parameters, that were selected by the NCAR to regionally calibrate NWM version 1.1 (Dugger et al. 2017), are shown in Table 1. For all modeling described herein, model code and a priori parameters are based on NWM v1.1, except where modifications are noted. A priori Noah-MP soil parameter values in WRF-Hydro are based on Rawls et al. (1982), from the STATSGO2 soil texture dataset (available at http://websoilsurvey.nrcs.usda.gov/). Channel parameters, which correspond to individual reaches based on stream order and other channel characteristics (with the exception of the channel infiltration parameter; see next section) and that were evaluated after the first round of calibration in the WGEW (see section 3a), are shown in Table 1.

b. Model domains

In the present study, the effects of modeled channel infiltration are evaluated for the WGEW and the Babocomari River basin. Figure 2 shows the elevation grid for the 250-m NWM routing grid for the entire

San Pedro basin; however, the simulations performed in this study use cutout grids for the Babocomari basin and WGEW.

For all calibration simulations, atmospheric forcing is derived from NLDAS-2 atmospheric variables that are regridded to the model domains; however, the model is calibrated with both regridded 1/8° NLDAS-2 precipitation and NCEP 4-km Stage IV precipitation, which is based on WSR-88D radar and gauge precipitation (Lin and Mitchell 2005). To reduce the influence of possible forcing uncertainty and bias, WRF-Hydro in WGEW was forced with gauge-based precipitation (88 gauges in the 149 km^2 watershed area) interpolated to the WGEW-domain model grid. WGEW is useful for analyzing the performance of WRF-Hydro because it includes 20 soil moisture measurement sites, two AmeriFlux (Department of Energy 2018) flux towers, and subwatersheds with supercritical flumes for measuring runoff (Fig. 3).

Figure 4 demonstrates the profound role of channel infiltration in the lowermost reach of the WGEW, where streamflow during two runoff events from the two upstream flumes that come together to form a single channel (runoff gauges 2 and 7), is highly attenuated by the time it reaches runoff flume 1, situated 7 km downstream, defining the basin outlet. Model topography and two initial soil parameters in WGEW, which are a function of soil type, are shown in Fig. 3. This figure shows that soil saturated hydraulic conductivity based on NWM v1.1 grids in WGEW is relatively homogeneous, as it is derived from STATSGO2 soil texture class which is homogenous in this domain. Higher-resolution SSURGO soil texture data (available at https://websoilsurvey.nrcs.usda.gov/) in WGEW shows considerably more variability, which is considered in our discussion (section 5); however, this study uses the NWM STATSGO2 grids to stay consistent with the NWM v1.1 configuration. DKSAT, according to the NWM a priori parameters, is constant at 3.37×10^{-6} m s^{-1} everywhere in WGEW, except in the extreme upper basin. Vegetation across WGEW includes primarily desert shrubs and grasslands; however, the town of Tombstone, Arizona, is partially urbanized. WGEW has oak forests in its extreme upper reaches.

The Babocomari River basin characteristics are shown in Fig. 5. This basin has a larger drainage area, and much of the headwaters are at high elevations. Like WGEW, NWM STATSGO soil characteristics likely underestimate the variability of the soils, per the SSURGO dataset. The Babocomari basin has a flux tower and three soil moisture observation sites with reliable data through the calibration period (Fig. 5), so model states at these sites are also evaluated.
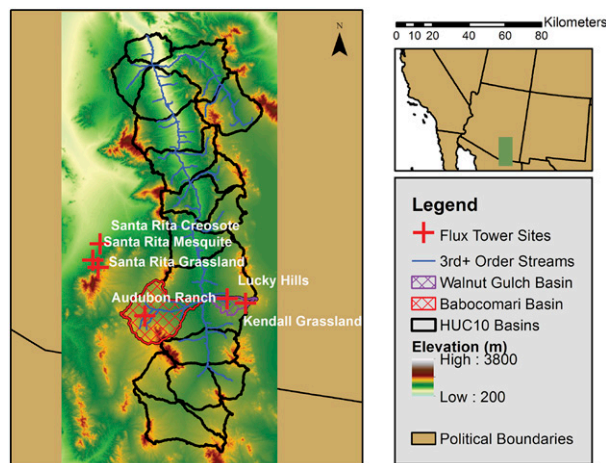

FIG. 2. The WRF-Hydro NWM 250-m routing grid (displayed as elevation) for the San Pedro River basin is shown. The Walnut Gulch and Babocomari River calibration basins are in the hatched areas. Spatial reference information for the San Pedro routing grid is shown in the top right corner of the figure.

c. Baseflow bucket parameter modifications

The current WRF-Hydro conceptual bucket model, which assumes one-way direct connection between attenuated baseflow from a groundwater basin and the overlaying channel, provides a poor representation of baseflow in event-driven, ephemeral channels in semiarid environments. Depth to groundwater is often substantial in the southwest CONUS, such that surface water–groundwater processes can become decoupled. In addition, water from channels often infiltrates to recharge the local aquifer (e.g., Blasch et al. 2004), and this source of recharge is currently not represented in the NWM.

To prevent unrealistic simulations of baseflow in the channel network, we disabled the baseflow bucket model everywhere in the model domain by setting the Noah-MP bottom drainage scaling (SLOPE) parameter to zero, imposing a no-flow condition at the bottom of the Noah-MP LSM in both the WGEW (Fig. 3b) and the Babocomari basin (Fig. 5b). This assumption is consistent with NWM calibrated parameters for this region, where a small SLOPE parameter effectively eliminates deep baseflow to the channel. Neither WGEW nor the Babocomari basins contain perennial channels fed from deep groundwater, and there are no perennial channels shown in either basin per the NHDPlusV2 dataset attributes, so they are not good candidates for applying the conceptual baseflow bucket module in its current form.

d. Channel infiltration loss function

The WRF-Hydro channel routing scheme assumes a trapezoidal channel geometry, and the length and slope of specific reaches is specified in the NHDPlusV2
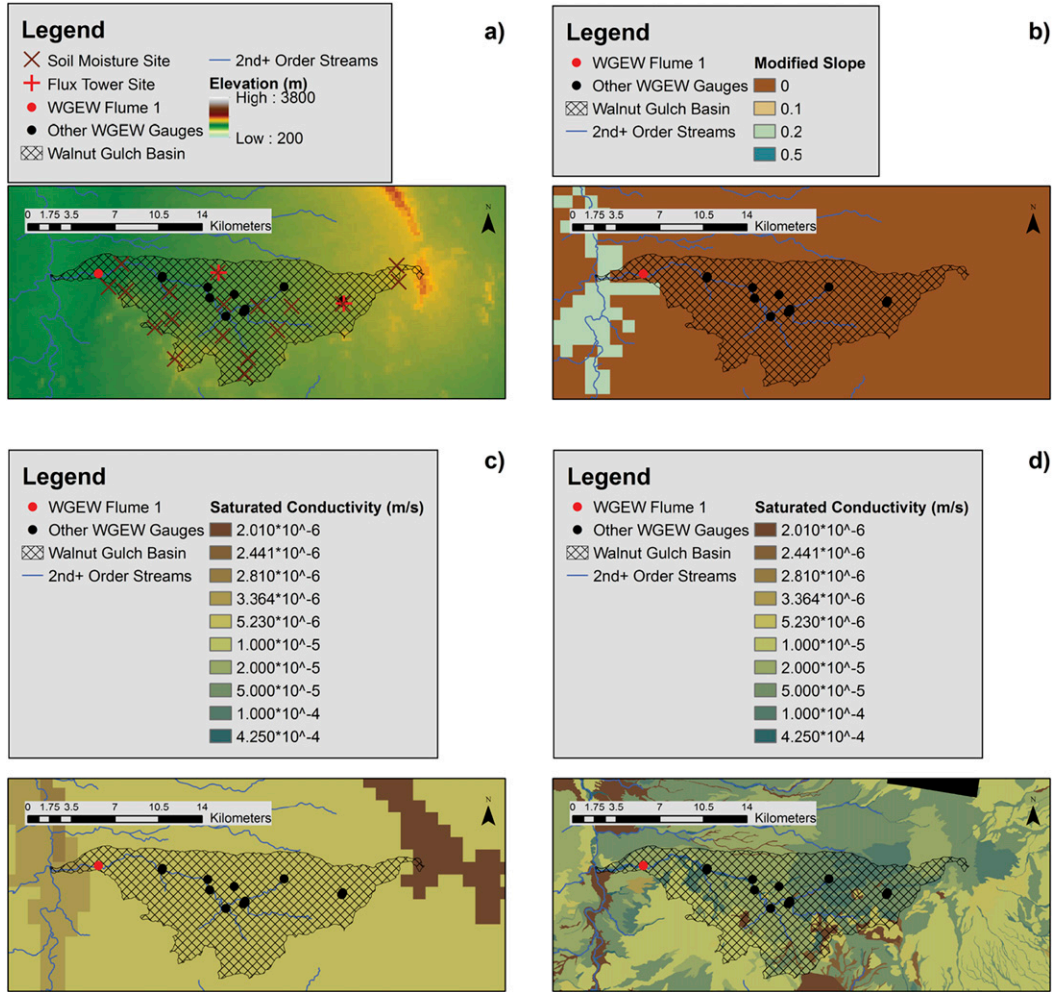
FIG. 3. (a) WGEW elevation, (b) modified bottom drainage scaling factor (Noah-MP SLOPE parameter), (c) saturated soil conductivity (Noah-MP DKSAT parameter), and (d) saturated soil conductivity (from layer-averaged 2014 SSURGO data) are plotted. WGEW soil moisture and flux tower sites are also shown in (a). WGEW gauges plotted refer to runoff gauges. All gridded data are based on the NWM in the southwest CONUS, and SSURGO conductivity is derived by the USGS (https://water.usgs.gov/GIS/metadata/usgswrd/XML/ds866_ssurgo_variables.xml).

dataset. A cross section of a WRF-Hydro channel is shown in Fig. 1. If the volume of water in a reach is known, the height of the water h (m) and the wetted perimeter p (m) can be calculated by the following argument. First, the cross-sectional area a (m^2) of the water may be computed by dividing the channel volume by the length of the reach, assuming a constant height along the entire length of the reach. Based on the trapezoidal shape assumption, cross-sectional area is equivalent to

$$a = h(w + w_s). \tag{1}$$

We can compute w_s (m) as a function of the riverbank slope s (h/w_s in Fig. 1) and water height. Note that s is equivalent to the WRF-Hydro ChSlp parameter. This may be written as

$$a = h\left(w + \frac{h}{s}\right) = wh + \frac{1}{2}h^2; \quad w_s = \frac{h}{s}. \tag{2}$$

Since area can be derived for a channel volume, for a given bottom width w and bank slope s, Eq. (2) can be solved for h. It is possible to derive the wetted perimeter p from h using

$$p = w + \sqrt{h^2 + \left(\frac{h}{s}\right)^2} = w + \sqrt{\frac{h^2}{s^2}(s^2 + 1)}. \tag{3}$$

Due to the effects of uneven channel bed topography, wetted perimeter will be reduced during periods of low flow, as water will only flow within the lowest portions of an irregular channel. To account for this tendency in
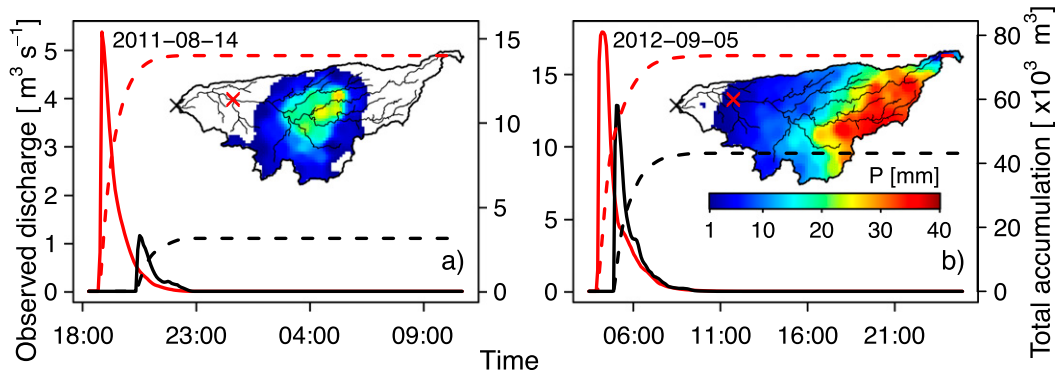
FIG. 4. (left) Observed and (right) accumulated streamflow at WGEW runoff gauges 2 and 7 (combined, red) and WGEW runoff gauge 1 (outlet, black) for two flow events are depicted on the *y* axis of each panel. The shaded spatial plots of WGEW depict storm total precipitation for each event (mm).

trapezoidal channels, *p* is reduced during periods of low flow using a conceptual model similar to what is used in KINEROS2 (Woolhiser et al. 1990). KINEROS2 computes a corrected wetted perimeter p_e for a channel using the function

$$p_e = \min\left[\frac{h}{b\sqrt{w}}, 1\right] p, \tag{4}$$

where *b* (unitless) is set to a constant value of 0.15. Channel infiltration I $(\text{m}^3\,\text{s}^{-1})$ may then be derived as
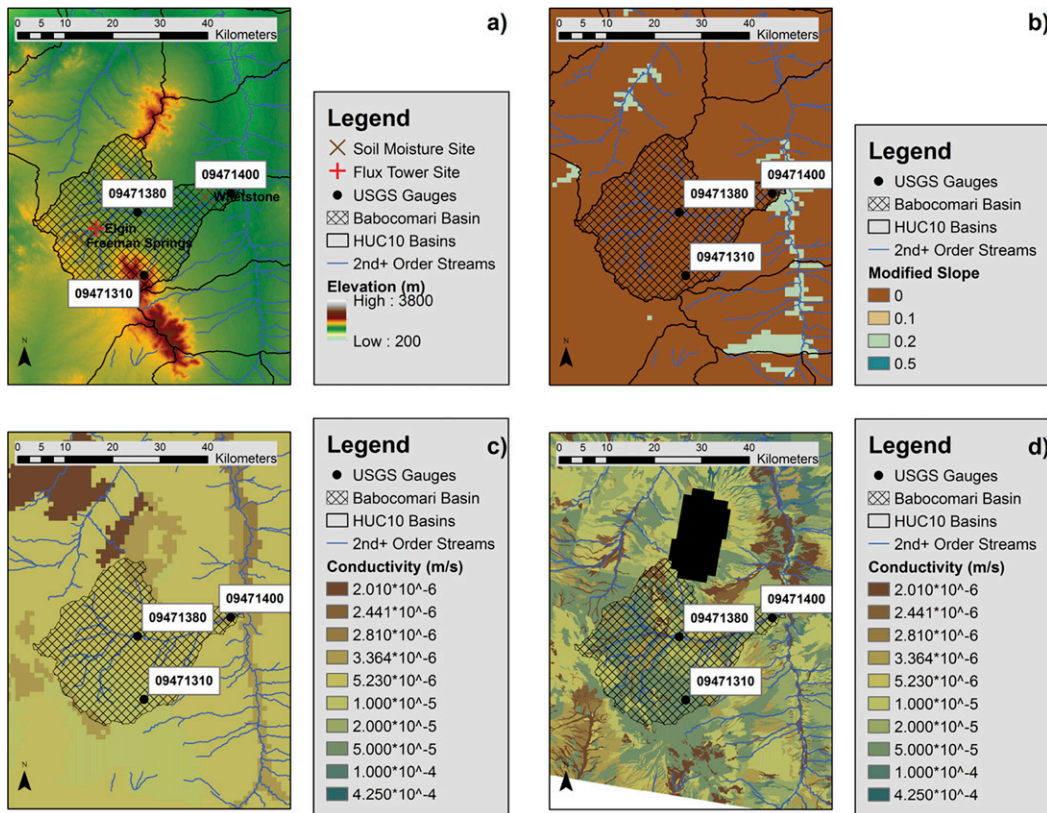
FIG. 5. (a) Babocomari basin elevation, (b) modified bottom drainage scaling factor (Noah-MP SLOPE parameter), (c) saturated soil conductivity (Noah-MP DKSAT parameter), and (d) saturated soil conductivity (from layer-averaged 2014 SSURGO data) are plotted. NOAA HMT soil moisture and the AmeriFlux flux tower site are also shown in (a). USGS gauges plotted refer to runoff gauges. All gridded data are based on the NWM in the southwest CONUS, and SSURGO conductivity is derived by the USGS (https://water.usgs.gov/GIS/metadata/usgswrd/XML/ds866_ssurgo_variables.xml).

$$I = klp_e. \tag{5}$$

In this function, k (m s^{-1}) is the saturated conductivity of the channel bed (the ChannK parameter), and l (m) is the length of a channel reach. Channel infiltration is accounted for in the Muskingum–Cunge routing scheme of WRF-Hydro, as an added sink in the iterative calculation, by deriving the effective flow height and wetted perimeter for a channel reach volume and assigning a saturated conductivity for the channel bed below. We assume b to be constant, while the saturated conductivity of the channel bed (ChannK) is initially set equivalent to the saturated soil conductivity of Noah-MP (DKSAT). Water that infiltrates out of the channels is assumed to contribute to deep groundwater recharge and is therefore removed from the model. During calibration, initial ChannK is adjusted by a scalar multiplier, a simple form of spatial regularization (e.g., Pokhrel et al. 2012).

3. WRF-Hydro calibration intervals

We calibrated the NWM with three different configurations of increasing complexity to evaluate: 1) the importance of channel parameters following the addition of the channel infiltration function, 2) the added value of the channel infiltration function in an ephemeral catchment with a relatively simple hydrologic response (i.e., WGEW), and 3) the calibrated model performance with two nationally available forcing products (Table 2).

Optimization of daily streamflow was performed using the dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker 2007), which is capable of converging to near optimal parameter sets with fewer (approximately 100–500) iterations (e.g., Lespinas et al. 2017) than the widely used Shuffled Complex Evolution function (e.g., Duan et al. 1992) that can require ~10 000 iterations to converge to an optimal solution. This makes DDS better suited for calibration of computationally expensive, distributed, physically based, models like the WRF-Hydro model. For all calibration exercises, except the first (section 3a), we use 500 iterations of DDS with the updated NWM v1.1 utilizing channel infiltration. Lespinas et al. (2017) show that improvements to model skill are greatest between 100 and 500 DDS iterations and that less value is added beyond 500. As the first exercise (section 3a) is simply intended to bring the NWM parameters to an acceptable state so that channel parameters can be evaluated, we only performed 250 iterations for this step. For all calibration simulations, NWM v1.1 with channel infiltration is calibrated to daily streamflow to stay consistent with

TABLE 2. NWM WRF-Hydro parameter calibration configurations and parameter ranges. For each calibration, 500 iterations were used in both basins, except in Stage 1. All parameters marked with an ''m'' were adjusted spatially by a scalar multiplier, and all parameters marked with an ''a'' were adjusted spatially by a scalar addition constant. An asterisk (*) indicates that the model is calibrated both with and without channel loss.

Basin/gauge	Forcing	Iterations	Calibration period	SMCMAX (m)	DKSAT (a)	DKSAT (m)	REFKDT (const)	ChannK (m)	ChSlp (m)
Stage 1: Channel parameter sensitivity; NWM used 15-min (instead of 60-min) Noah-MP time step									
Walnut Gulch	WGEW-gauge	250	WY2009–2011	0.8–1.2	-1×10^6 to 1×10^6	0.5–2.0	0.1–4.0	0.001–1.0	0.1–1.5
Stage 2: Evaluation of channel infiltration in WGEW									
Walnut Gulch	WGEW-gauge	500	WY2011–2013	0.8–1.2	-1×10^6 to 1×10^6	0.5–2.0	0.1–4.0	0.001–1.0*	0.1–1.5
Stage 3: Calibration with Stage IV and NLDAS-2 precipitation									
Walnut Gulch	Stage IV	500	WY2011–2013	0.8–1.2	-1×10^6 to 1×10^6	0.5–2.0	0.1–4.0	0.001–1.0*	0.1–1.5
Babocomari River	Stage IV	500	WY2009–2011	0.8–1.2	-1×10^6 to 1×10^6	0.5–2.0	0.1–4.0	0.001–1.0*	0.1–1.5
Walnut Gulch	NLDAS2	500	WY2011–2013	0.8–1.2	-1×10^6 to 1×10^6	0.5–2.0	0.1–4.0	0.001–1.0*	0.1–1.5
Babocomari River	NLDAS2	500	WY2009–2011	0.8–1.2	-1×10^6 to 1×10^6	0.5–2.0	0.1–4.0	0.001–1.0*	0.1–1.5

TABLE 3. Model hourly correlation coefficient with varying values of ChSlp (labeled as s), N (labeled as n), and BtmWdth (labeled as w). Model parameters are adjusted by either dividing or multiplying a priori N, ChSlp, or BtmWdth parameters by a constant.

ChSlp (s)	$s = s/5$	$s = s/2.5$	Control	$s = s \times 2.5$	$s = s \times 5$
	0.1966	0.3838	0.6349	0.6463	0.5867
Manning's N (n)	$n = n/2$	$n = n/1.5$	Control	$n = n \times 1.5$	$n = n \times 2$
	0.2242	0.4175	0.6349	0.6784	0.5146
BtmWdth (w)	$w = w - 4$	$w = w - 2$	Control	$w = w + 2$	$w = w + 4$
	0.5830	0.6241	0.6349	0.6826	0.6974

NWM v1.1 calibration methods, and we consider hourly streamflow when evaluating the NWM in WGEW since most streamflow events occur at the hourly time scale.

All of the calibrations reported here are based on optimization of the Kling–Gupta efficiency (KGE) performance metric (Gupta et al. 2009), which equally weights correlation, water balance, and variance errors. Like the Nash–Sutcliffe efficiency (NSE), KGE is optimal when equal to 1, and negative values of KGE are considered to have low skill. KGE is optimal when the Euclidian distance from an ideal point for the ratio of modeled to observed standard deviation α, ratio of modeled to observed mean β, and correlation coefficient r are minimized (Gupta et al. 2009):

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}. \quad (6)$$

For all basins, initial model states were derived by executing WRF-Hydro with default parameters with their calibration precipitation forcing (WGEW-gauge, Stage IV, or NLDAS-2) for an 8-yr spinup period during water years (WY) 2007–15. The model state at the end of this period was used as the initial "warm" state for calibration, consistent with the practices of the NCAR WRF-Hydro NWM development team. This ensures a long-term model spinup of multiple years to allow state variables to reach equilibrium. Then for each calibration iteration, model parameters were given one year of

additional spinup to equilibrate before the calibration and evaluation periods (described in detail below).

a. Stage 1: Channel parameter sensitivity

To evaluate the sensitivity of the channel routing parameters, WRF-Hydro was first calibrated to eliminate water balance errors using WGEW gauge precipitation forcing. This step was only performed for the WGEW basin. As bias is a component of KGE, this objective function permitted us to reduce water balance errors and simulate an otherwise realistic hydrologic response (since correlation and variance errors are also accounted for). To select parameters for this calibration, sensitivity analysis on daily KGE was first performed on the updated NWM, using linear adjustments applied to default prior estimates (e.g., DKSAT shown in Fig. 3c) of the model parameters (Table 1). Selected calibration parameters included ChannK (channel conductivity), DKSAT (soil conductivity), REFKDT (infiltration scaling), and SMCMAX (soil porosity). These parameters were chosen because sensitivity analysis revealed that they had the greatest impact on model KGE. ChannK, SMCMAX, and DKSAT were computed by multiplying initial NWM parameters by a constant, and REFKDT was assumed constant in the whole model domain. DKSAT was adjusted by an addition constant (before being adjusted by a multiplicative constant).

TABLE 4. Final parameter adjustment constants for each WRF-Hydro NWM calibration. Note that REFKDT is constant throughout the model domain, so the values for this parameter in the table are directly equivalent to the NWM parameter. As in Table 2, "m" and "a" indicate multiplication or addition adjustment constants, respectively.

| Basin/gauge	Forcing	SMCMAX (m)	DKSAT (a)	DKSAT (m)	REFKDT (const)	ChannK (m)	ChSlp (m)
Walnut Gulch	WGEW gauge	1.006	-2.30×10^{-7}	1.031	3.692	0.387	1.499
Walnut Gulch (without loss)	WGEW gauge	1.109	5.30×10^{-7}	1.014	2.785	—	1.419
Walnut Gulch	Stage IV	0.940	4.49×10^{-7}	1.643	1.198	0.621	0.695
Walnut Gulch (without loss)	Stage IV	1.075	4.09×10^{-7}	1.539	1.551	—	1.154
Babocomari River	Stage IV	1.004	-6.19×10^{-7}	0.940	1.491	0.998	0.252
Babocomari River (without loss)	Stage IV	0.877	9.09×10^{-7}	0.707	3.998	—	0.984
Walnut Gulch	NLDAS-2	1.028	1.33×10^{-7}	0.526	1.994	0.988	1.488
Walnut Gulch (without loss)	NLDAS-2	1.045	2.15×10^{-7}	0.965	1.196	—	1.499
Babocomari River	NLDAS-2	0.842	8.28×10^{-7}	0.501	1.067	0.348	0.138
Babocomari River (without loss)	NLDAS-2	0.800	-1.95×10^{-7}	0.697	2.277	—	0.593
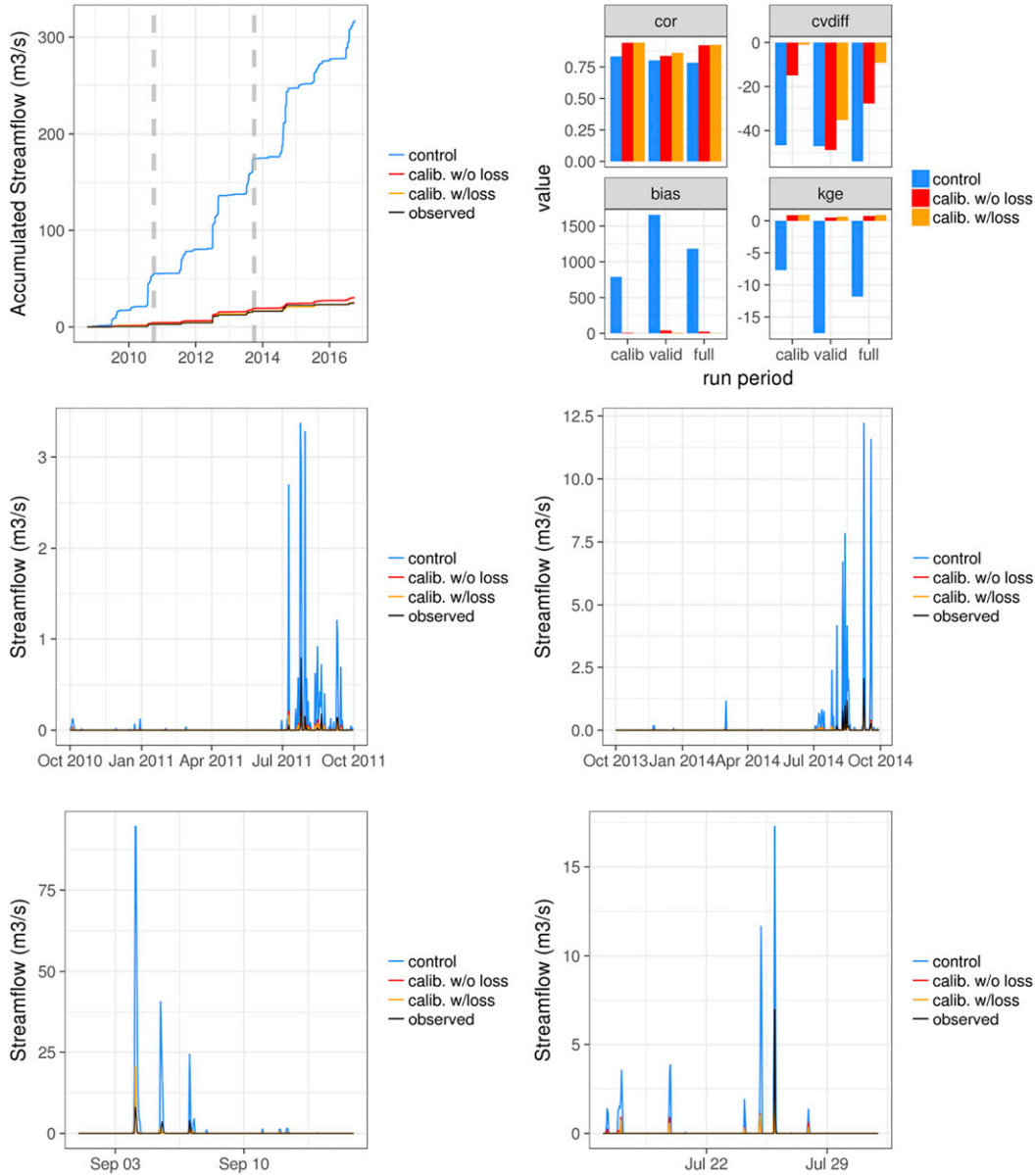
FIG. 6. Example of control (blue), calibrated (red), and calibrated with channel loss (orange) NWM streamflow for Walnut Gulch runoff gauge 1, where the model is forced with gauge precipitation. (top left) Accumulated streamflow and (top right) model skill scores. Annual daily streamflow for (middle left) WY2011 and (middle right) WY2014. Hourly streamflow for (bottom left) 1–15 Sep 2012 and (bottom right) 16–31 Jul 2013.

For this first step of calibration, the Noah-MP time step was reduced from 60 min (the standard for the operational NWM) to 15 min, so that the model produced outputs, including streamflow, at 15-min temporal resolution. This permitted us to critically analyze the sensitivity of the channel infiltration function at fine resolution (not shown). This calibration improved WGEW model KGE (with daily temporal resolution) from −1.57 to 0.80 for WY2009–11 (the calibration period; spinup period, WY2008) by reducing water

balance errors that caused excessive runoff and low ET biases (not shown).

Following initial calibration, a simple one-at-a-time parameter sensitivity analysis was performed on the WRF-Hydro channel parameters, including BtmWdth (channel width), ChSlp (outer channel slope), and *N* (Manning's *N*; see Table 1). For each parameter, initial parameter values in the WGEW domain were adjusted by either addition or multiplication to the initial parameter with a constant. A priori parameter values were

TABLE 5. NWM WRF-Hydro daily streamflow skill scores after calibration, including: correlation coefficient (COR), coefficient of variation (CV) percent bias, percent bias, and KGE. Skill scores with an asterisk symbol (*) indicate where calibration reduced the model skill. The evaluation period encompasses all years in the WY2009–2016 period outside of the specified calibration period. Note that calibration period years indicate water years for that period.

Basin/gauge	Forcing precipitation	Calibration years	COR	CV percent bias	Percent bias	KGE
Calibration period						
Walnut Gulch	WGEW gauge	2011–13	0.9442	−0.9797	−0.0994	0.9431
Walnut Gulch (without loss)	WGEW gauge	2011–13	0.9420	−15.0190	9.8052	0.8679
Walnut Gulch	Stage IV	2011–13	0.9436	2.0405	−0.0043	0.9400
Walnut Gulch (without loss)	Stage IV	2011–13	0.9392	−6.8931	3.1255	0.9209
Babocomari River	Stage IV	2009–11	0.7661	−3.6493	2.6121	0.7643
Babocomari River (without loss)	Stage IV	2009–11	0.6527*	−45.8413*	67.6339	0.2342
Walnut Gulch	NLDAS-2	2011–13	0.6081	−4.9914	0.3365	0.6053
Walnut Gulch (without loss)	NLDAS-2	2011–13	0.6012	−25.8172*	11.7025	0.5504
Babocomari River	NLDAS-2	2009–11	0.4239	−11.0953	7.3702	0.4174
Babocomari River (without loss)	NLDAS-2	2009–11	0.4221	−38.0215*	18.2809	0.3377
Evaluation period						
Walnut Gulch	WGEW gauge	2011–13	0.8626	−35.1647	7.6912	0.6596
Walnut Gulch (without loss)	WGEW gauge	2011–13	0.8385	−48.8372*	40.8510	0.4794
Walnut Gulch	Stage IV	2011–13	0.5248	−38.2333	18.7505	0.4238
Walnut Gulch (without loss)	Stage IV	2011–13	0.4864	−50.8070*	22.4752	0.3128
Babocomari River	Stage IV	2009–11	0.2918*	16.6644*	88.2720	−0.6469
Babocomari River (without loss)	Stage IV	2009–11	0.2787*	−48.4689*	265.3276	−1.8878
Walnut Gulch	NLDAS-2	2011–13	0.4708	−6.9326	59.8780	0.0637
Walnut Gulch (without loss)	NLDAS-2	2011–13	0.4666	−34.9931	119.4619	−0.3761
Babocomari River	NLDAS-2	2009–11	0.2706	40.3256*	324.8557	−4.9753
Babocomari River (without loss)	NLDAS-2	2009–11	0.2366	12.7921	421.1274	−5.4893

assigned by the NCAR NWM WRF-Hydro development team. This exercise permitted evaluation of the model hydrologic response to perturbations of the channel routing parameters. The main evaluation metric of interest for this analysis was the hourly correlation coefficient because its accuracy is a proxy for errors in the timing of streamflow events (e.g., Gupta et al. 2009). This analysis showed that ChSlp and N, which affect the transit time of streamflow, impact the model correlation coefficient (Table 3), with correlation coefficients increasing when these parameters were multiplied by 2.5 and 1.5, respectively. However, in further analyses, we chose to only vary ChSlp, while keeping Manning's channel roughness N constant. We could have also varied this latter parameter, as both parameters (ChSlp and N) affect the model correlation coefficient (Table 3). However, as both parameters have a similar impact (though, not necessarily identical) on NWM streamflow transit time, it was decided that including an additional parameter to calibrate would have been somewhat redundant.

b. Stage 2: Evaluation of channel infiltration in WGEW

To demonstrate the added value of channel infiltration, NWM WRF-Hydro was recalibrated for WGEW with the same parameters used for Stage 1 of calibration (see last section) but with the addition of the

ChSlp parameter, for WY2011–13 (WY2010 for spinup) with channel infiltration either active or disabled. The change to the calibration period from Stage 1 was intended to include both wet and dry years in the calibration. The ChannK parameter multiplication constant was constrained between 0.0 and 1.0, based on the assumption that saturated conductivity (DKSAT) sets an upper limit for channel infiltration. This implicitly assumes that the channel bed properties for a region will tend to be consistent with local soil characteristics, with soil conductivity being maximized when saturated. The Noah-MP time step for this calibration was set to 60 min, consistent with the operational NWM v1.1.

All adjustment constants for selected parameters and their ranges are shown in Table 2. For parameters where a priori values are adjusted by multiplication or addition constants (e.g., DKSAT), the ranges of the adjustment constants are shown. Note that these adjustment factors are not the parameters themselves. For example, if SMCMAX has an adjustment factor of 0.8, the a priori values of SMCMAX will be multiplied by 0.8 everywhere to compute the new parameters, thus preserving the spatial patterns of the parameters as they are calibrated. REFKDT is left constant throughout the model domain, so the possible ranges of the actual values for this parameter are shown in Table 2. The final parameter adjustment factors and calibration

configuration for these simulations are shown in Table 4. As noted above, 500 iterations of DDS were used, and daily streamflow KGE was the objective function. To evaluate whether the calibration of the NWM improved the entire physical model state and not just simulated streamflow, soil moisture and evapotranspiration (ET) fluxes within WGEW were evaluated after calibration, with and without channel infiltration.

c. Stage 3: Calibration with Stage IV and NLDAS-2 precipitation

To demonstrate the potential application of the afore-mentioned channel infiltration function and calibration methods in the operational NWM, WRF-Hydro was also calibrated using NLDAS-2 and NCEP Stage IV precipitation as forcing instead. Calibration periods were 3 years long, with one extra prior year for spinup (i.e., WY2011–13 with spinup WY2010 and WY2009–11 with spinup WY2008 for the WGEW and Babocomari basin, respectively) and selected to include both wet and dry years for each basin (Table 2). To demonstrate the effects of channel infiltration, the model was calibrated with and without channel loss.

4. Results

Following calibration, WRF-Hydro, configured as NWM v1.1 with channel loss, was run with both default uncali-brated parameters (as a benchmark), calibrated parameters without channel infiltration, and calibrated parameters with channel infiltration for WY2009–16. WY2008 was used as spinup, and 3 years of this 8-yr period were calibration years. The results from this evaluation are presented herein.

a. Calibrated model performance and added value of channel infiltration in WGEW

In WGEW, using rain gauge observations, calibration eliminated water balance errors, largely by eliminating spurious flashy peaks in the uncalibrated simulation, and performed better when channel infiltration was included, yielding slightly higher correlation coefficients (Fig. 6). The resulting cumulative simulated streamflow for WY2009–16 is less biased, demonstrating the added value of calibration with channel loss for reducing water balance errors. Adding channel loss reduced the negative percent bias of the modeled versus observed coefficient of variation (labeled as "cvdiff" in Fig. 6), suggesting that the model with channel loss had less tendency to un-derestimate the variance of the streamflow. To quantify the amount of channel infiltration in the calibrated model, we executed the NWM with the same parameters as the calibrated solution with channel infiltration, but with channel infiltration disabled. Adding channel in-filtration reduces the output at the outlet of WGEW by

TABLE 6. NWM WRF-Hydro skill scores for Noah-MP level 1 (0–10 cm) soil moisture simulations (compared to areal averages and 5-cm Walnut Gulch soil moisture measurements) and simu-lated ET fluxes (compared to flux tower ET at the Lucky Hills and Kendall Grassland sites). Soil moisture (ET) skill scores are based on hourly (daily) data.

Evaluation metric	Control (with loss)	Calibration (no loss)	Calibration (with loss)
Soil moisture (basin average)			
Percent bias	93.1831	105.2521	102.5220
Correlation coef	0.8587	0.8800	0.8543
KGE	0.0269	−0.1427	−0.0834
Lucky Hills 5-cm soil moisture			
Percent bias	83.5882	94.4479	91.7655
Correlation coef	0.8466	0.8585	0.8352
KGE	0.0372	−0.1649	−0.0775
Kendall Grassland 5-cm soil moisture			
Percent bias	64.7340	74.5087	71.5879
Correlation coef	0.8569	0.8656	0.8525
KGE	−0.1439	−0.4362	−0.2611
Lucky Hills ET			
Percent bias	−10.822	−1.5496	−0.5983
Correlation coef	0.890 77	0.889 22	0.891 72
KGE	0.733 37	0.843 73	0.834 77
Kendall Grassland ET			
Percent bias	−15.464	−6.6524	−5.9558
Correlation coef	0.865 09	0.863 55	0.865 89
KGE	0.688 57	0.802 24	0.793 92

21.73% over the model with channel infiltration dis-abled (see figure in the online supplemental material), indicating the fraction of water that becomes channel infiltration. Table 5 shows that when WRF-Hydro is calibrated with channel infiltration, the model KGE is 0.9431, and the correlation coefficient is 0.9442 for daily streamflow during the calibration period. KGE was re-duced to 0.8679 when the model was calibrated without channel loss. These values show slightly less skill when zero values of modeled and observed streamflow are omitted from the analysis (see supplemental material). As would be expected, model skill decreases in the eval-uation period (WY2009–10 and WY2014–16), with a no-tably lower KGE (0.4794 versus 0.6596) and higher bias (40.85% versus 7.69%) when the model was calibrated without channel loss, versus with channel loss (Table 5). For hourly resolutions, the calibrated model results with channel loss had a KGE of 0.83 (0.55) and a correlation coefficient of 0.83 (0.80) for the calibration (evaluation) periods (not shown here). For two sample hydrographs, the bottom panels of Fig. 6 show that calibration improves modeled hydrograph shape even though the model strug-gles to represent specific hourly events. The calibrated model without channel loss produces slightly more intense spurious runoff events than the model with loss.
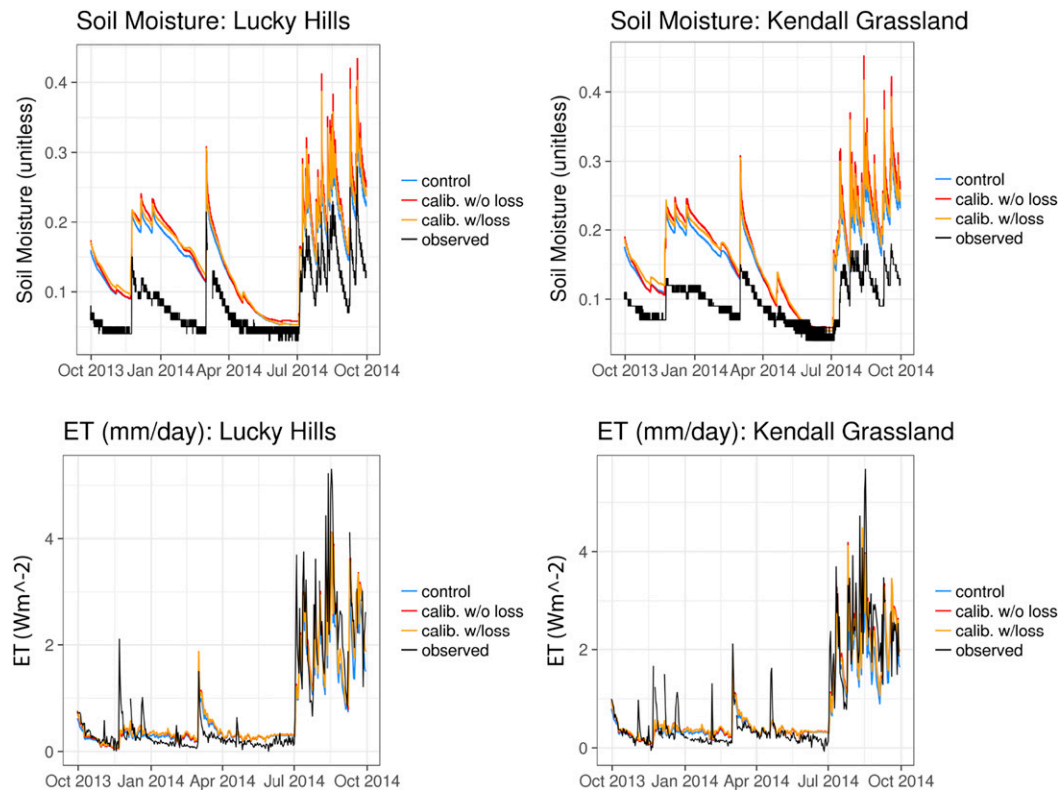
FIG. 7. WY2014 NWM Noah-MP level 1 soil volumetric water content from the control (blue), calibrated (red), and calibrated with channel loss (orange) compared to 5-cm observations (black) for (top left) Lucky Hills and (top right) Kendall Grassland. (bottom) Modeled ET from the control (blue), calibrated (red), and calibrated with channel loss (orange) solutions for the same period and field sites are plotted. Flux tower ET measurements are plotted in black.

b. Soil moisture and flux tower ET evaluation in WGEW

To go beyond runoff only and consider the physical process representation of WRF-Hydro, model soil moisture and ET were evaluated against observations in WGEW starting in WY2009 and ending in April 2015. WGEW 5-cm soil moisture observations were compared to the area average of Noah-MP 0–10-cm soil moisture, based on 20 soil moisture sites in WGEW including the two AmeriFlux flux tower sites (i.e., Kendall Grassland and Lucky Hills). At these flux tower sites, observed hourly soil moisture is compared to model soil moisture averages from the Noah-MP grid point closest to the site and the two points surrounding it in all directions (25 points total), to minimize the impact of spurious model simulations for a single grid point. Calibration increased the positive bias of near-surface soil moisture, averaged throughout the basin; however, including channel infiltration in the calibration did not increase the soil moisture bias by as much (Table 6). Calibrating without channel loss nominally increased correlation coefficients for the basin

average and two sites, and adding channel infiltration had little effect on the correlation coefficients after calibration. Soil moisture from WY2014, which had a wetter than average NAM season, is shown in Fig. 7. This figure also indicates that the modeled soil moisture solutions have a greater bias after calibration. We computed KGE for model soil moisture, which also reflects the increased bias after calibration more than the nominal gains in correlation coefficient. These results show that the calibrated WRF-Hydro model (with and without channel infiltration) increases the total amount of water inside the upper soil layer by 4.8% (computed basin average soil moisture in control vs calibration with infiltration, with WGEW gauge forcing). This increase in soil moisture in desert catchments is not observed in the observation datasets. All modeled solutions have a positive bias and slow dry-down during dry periods, which might be from limiting the bottom drainage scaling factor to eliminate deep groundwater baseflow. Increased values of REFKDT from calibration in all the basins likely also increased soil moisture, by increasing infiltration. Analysis of 15-cm soil moisture data (not
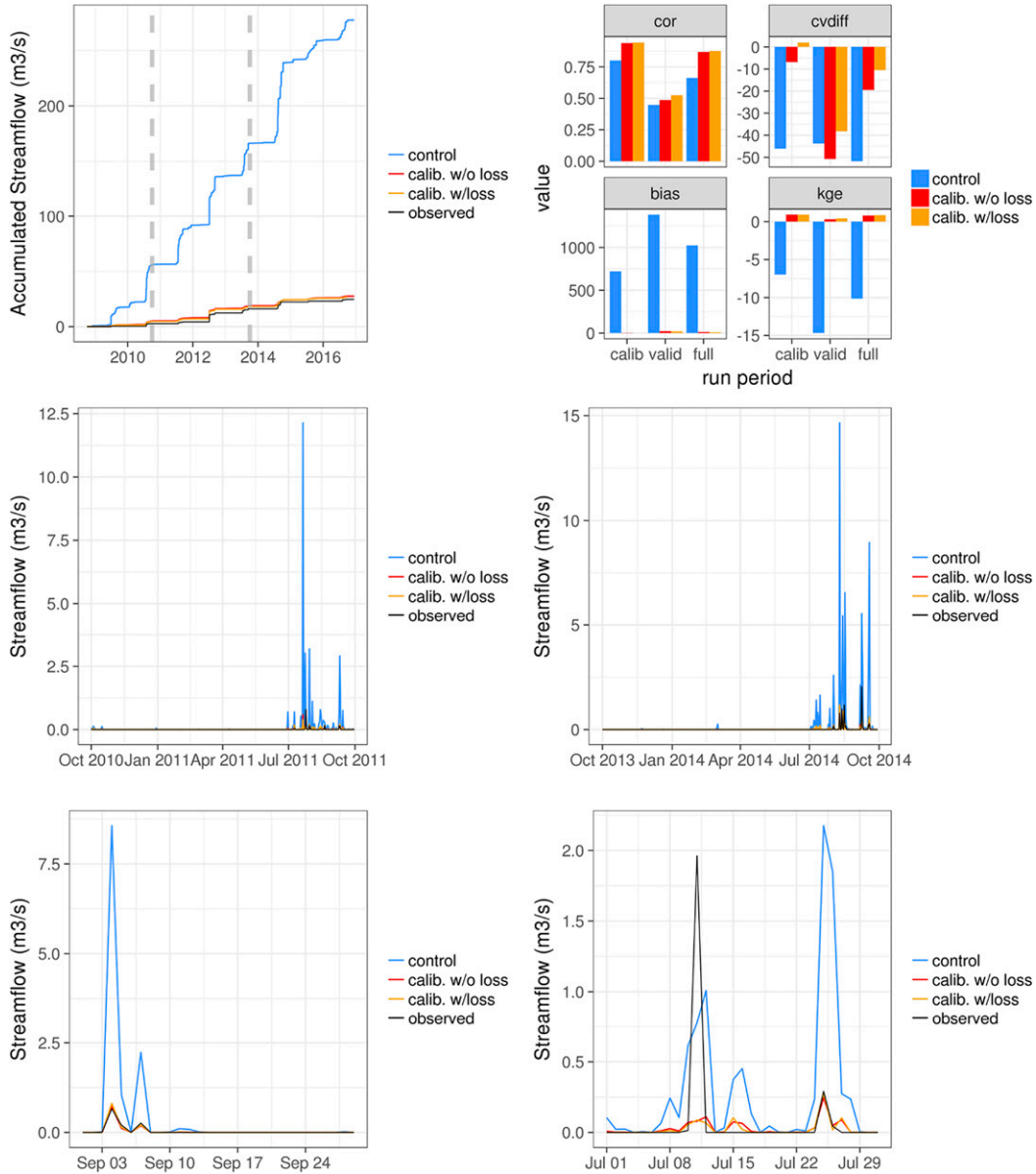
FIG. 8. Example of control (blue), calibrated (red), and calibrated with channel loss (orange) NWM streamflow for Walnut Gulch runoff gauge 1, where the model is forced with NCEP Stage IV precipitation. (top left) Accumulated streamflow and (top right) model skill scores. Annual daily streamflow for (middle left) WY2011 and (middle right) WY2014. Daily streamflow for (bottom left) September 2012 and (bottom right) July 2013.

shown) at Kendall Grassland and Lucky Hills reveals that soil moisture is less biased compared to observations, which may suggest that the Noah-MP 0–10-cm layer might be better representative of deeper soils. We acknowledge that representativeness errors in the soil moisture observation sites may also be possible.

Benefits of calibration for the water balance can be identified when considering observed daily ET fluxes from the Kendall Grassland and Lucky Hills AmeriFlux flux tower sites from WY2009 to December 2015 (Table 6),

compared to averages of corresponding model grid points (same as for soil moisture). WRF-Hydro, without calibration has a negative ET bias at Lucky Hills and Kendall Grassland. This negative ET bias is consistent with the uncalibrated model's tendency to overestimate streamflow, indicating that too little precipitation infiltrates at the land surface (due to low REFKDT and DKSAT values), forcing too much water to flow into the channel network as overland flow. Calibration of WRF-Hydro with channel infiltration, which eliminates the positive

streamflow bias, further reduces the magnitude of the ET bias at both Kendall Grassland and Lucky Hills (compared to the calibrated simulation without channel loss). This implies that when WRF-Hydro is executed with accurate forcing precipitation, calibration can improve the accuracy of the simulated streamflow and reduce the ET bias, despite increasing soil moisture bias, to improve the water balance. Figure 7 shows that the high soil moisture bias might be associated with too much ET during the dry season, while ET is less during the wet season. This tendency may be related to the soil moisture bias discussed above.

c. Extension of calibration methods with NLDAS-2 and Stage IV precipitation

WRF-Hydro was calibrated with NLDAS-2 and Stage IV precipitation in WGEW, and streamflow bias errors were consistently reduced with calibration. However, correlation coefficients demonstrate little improvement outside the calibration period regardless of whether channel loss is added (Fig. 8, Table 5). This is true with both Stage IV and NLDAS-2 forcing. NLDAS-2 forced results are shown in the supplemental material. Bias also increases outside of the calibration period. Adding channel infiltration does reduce coefficient of variation errors and produces a realistic hydrograph, despite timing errors. The fact that correlation coefficients and bias for the same study area are degraded when Stage IV or NLDAS-2 precipitation is used suggests that precipitation forcing may be a source of uncertainty for hydrologic models in this region. Table 7 implies that soil moisture bias errors persist regardless of precipitation forcing, but correlation coefficients remain high regardless of calibration. As with the gauge forcing, KGE values seem to reflect bias errors. Model ET (at Kendall Grassland and Lucky Hills) has a positive bias, which increases with calibration with both datasets (Table 7), implying that calibration might be compensating for precipitation biases and the distribution of Stage IV precipitation events (see discussion). Given the low runoff/rainfall ratios common in the semiarid Southwest, small errors in rainfall forcing typically result in large errors in runoff predictions (Goodrich et al. 2012).

Calibrating WRF-Hydro without channel loss did improve the KGE and bias metrics in the Babocomari basin (regardless of forcing), but it did not yield a realistic hydrologic response (see Fig. 9 for Stage IV results and supplemental material for NLDAS-2 results). The same figure shows that calibrating with channel loss did produce a realistic hydrologic response. Correlation coefficients were low during the evaluation period (WY2012–16), outside the calibration period (WY2009–11), consistent with WGEW. Cumulative streamflow in

TABLE 7. As in Table 6, using Stage IV or NLDAS-2 precipitation forcing.

Evaluation metric	Control (with loss)	Calibration (no loss)	Calibration (with loss)
Lucky Hills 5-cm soil moisture (Stage IV)			
Percent bias	102.8033	113.5252	107.3987
Correlation coef	0.7752	0.7792	0.7241
KGE	−0.1492	−0.3142	−0.1785
Kendall Grassland 5-cm soil moisture (Stage IV)			
Percent bias	80.5597	91.1242	84.2638
Correlation coef	0.7870	0.7876	0.7515
KGE	−0.2739	−0.5128	−0.2376
Lucky Hills ET (Stage IV)			
Percent bias	16.0258	25.3312	25.5819
Correlation coef	0.8397	0.8394	0.8383
KGE	0.77291	0.68363	0.694
Kendall Grassland ET (Stage IV)			
Percent bias	5.6164	14.8450	14.7850
Correlation coef	0.8367	0.8368	0.8376
KGE	0.8219	0.775 64	0.779 78
Lucky Hills 5-cm soil moisture (NLDAS-2)			
Percent bias	102.6702	108.2907	107.4980
Correlation coef	0.7920	0.8010	0.7984
KGE	−0.1740	−0.2744	−0.2563
Kendall Grassland 5-cm soil moisture (NLDAS-2)			
Percent bias	77.9496	82.7476	82.0299
Correlation coef	0.8056	0.8108	0.8088
KGE	−0.3001	−0.4431	−0.4138
Lucky Hills ET (NLDAS-2)			
Percent bias	15.0899	20.2991	20.2678
Correlation coef	0.839 46	0.841 89	0.8429
KGE	0.779 65	0.738 55	0.739 66
Kendall Grassland ET (NLDAS-2)			
Percent bias	2.3197	6.6863	6.6654
Correlation coef	0.805 36	0.804 22	0.803 67
KGE	0.780 21	0.785 66	0.784 89

Fig. 9 shows that the later years (WY2014–16) were wetter and had higher bias, suggesting that the hydrologic response may not be stationary (i.e., having constant statistical properties over time; Wilks 2006) over the calibration and evaluation periods. The poor hydrologic response in these later years may also be due to errors in the Stage IV precipitation. Despite the limitations of the NLDAS-2 and Stage IV precipitation products in the Babocomari basin, the added value of channel infiltration is clear, as calibrating the model without channel loss cannot completely eliminate the positive bias (Fig. 9, Table 5). The calibrated model without channel loss produces excess flow during dry periods, as the flashy runoff peaks are reduced by increasing infiltration into the soil column, leading to excessive baseflow entering the stream network, which
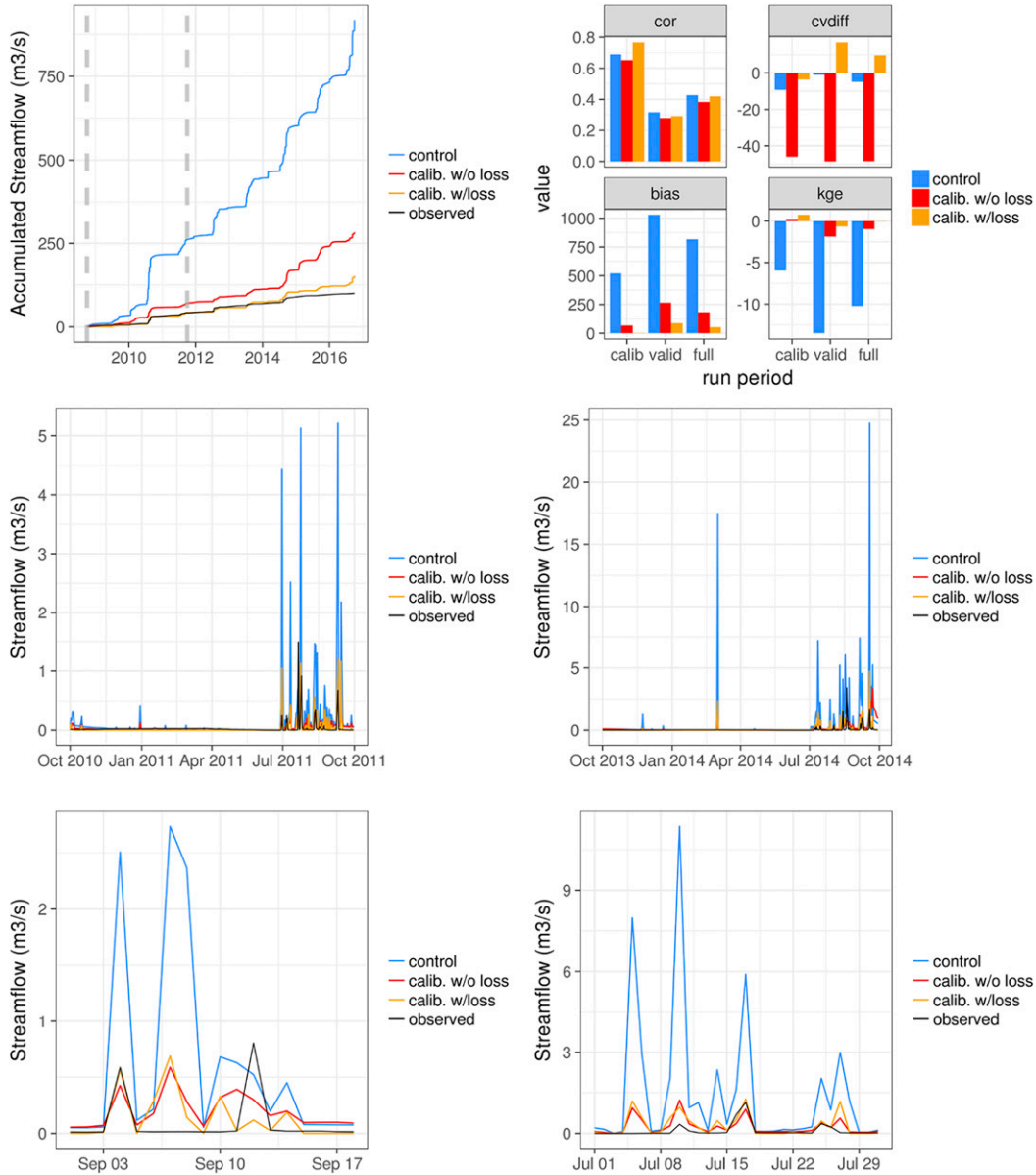
FIG. 9. As in Fig. 8, but for the Babocomari River, with NCEP Stage IV forcing.

may be associated with a low coefficient of variation bias (possibly implying more frequent flow with less variance).

Calibrating WRF-Hydro with Stage IV forcing with channel infiltration increases Noah-MP level 1 soil moisture bias at three soil moisture sites operated by the NOAA Hydrometeorology Testbed (HMT; Table 8) and has little effect on model soil moisture correlation coefficient (Fig. 10, Table 8). When channel infiltration is disabled, calibration improves soil moisture bias, but this is clearly at the expense of hydrologic response (see above). Observed ET at the Audubon

Research Ranch AmeriFlux Site from WY2009 to December 2011 is also compared to model ET, which generally reflects the positive ET bias, also shown in WGEW with Stage IV precipitation (Fig. 10; Table 8). For deeper soil levels (20 and 50 cm), the model underestimates soil moisture variability and has a consistent high bias (see Fig. 11 and supplemental material) at all three HMT sites. This high bias decreases the soil moisture KGE. This is consistent with the soil moisture bias in the upper layers, and it suggests that too much water is entering the soil moisture column, to compensate for the runoff biases in the NWM. The filling

TABLE 8. As in Table 7, for the Babocomari basin with NOAA HMT soil moisture sites and flux tower observations from the Audubon Research Ranch.

Evaluation metric	Control (with loss)	Calibration (no loss)	Calibration (with loss)
Elgin, AZ 5-cm soil moisture (Stage IV)			
Percent bias	234.8663	231.5843	241.3407
Correlation coef	0.8282	0.7906	0.8233
KGE	−1.4053	−1.3373	−1.4761
Freeman Springs, AZ 5-cm soil moisture (Stage IV)			
Percent bias	166.5594	162.2021	171.2363
Correlation coef	0.7728	0.7542	0.7725
KGE	−0.6714	−0.6198	−0.7201
Whetstone, AZ 5-cm soil moisture (Stage IV)			
Percent bias	24.1429	23.2152	25.5160
Correlation coef	0.8215	0.7773	0.8180
KGE	0.6905	0.6681	0.6740
Audubon Research Ranch ET (Stage IV)			
Percent bias	16.9774	24.3583	20.9221
Correlation coef	0.8467	0.8409	0.8461
KGE	0.768 06	0.7044	0.739 94
Elgin, AZ 5-cm soil moisture (NLDAS-2)			
Percent Bias	234.3116	213.9251	218.0668
Correlation Coef.	0.7956	0.7556	0.7666
KGE	−1.4121	−1.1544	−1.2018
Freeman Springs, AZ 5-cm soil moisture (NLDAS-2)			
Percent bias	160.3989	144.7427	147.7776
Correlation coef.	0.7617	0.7384	0.7434
KGE	−0.6213	−0.4568	−0.4825
Whetstone, AZ 5-cm soil moisture (NLDAS-2)			
Percent bias	29.6126	22.9163	24.3761
Correlation coef	0.7809	0.7366	0.7490
KGE	0.6200	0.6055	0.6280
Audubon Research Ranch ET (NLDAS-2)			
Percent bias	15.5984	16.7333	15.6011
Correlation coef	0.822 93	0.814 11	0.817 04
KGE	0.7342	0.744 88	0.751 87

of the soil columns is also likely due to the imposing of no-flow conditions at the base of Noah-MP to shut off the baseflow bucket model or due to other limits of Noah-MP parameters.

5. Discussion

a. Calibration and precipitation forcing errors

Calibration with channel loss improves streamflow skill in the Babocomari basin with NCEP Stage IV and NLDAS-2 precipitation forcing, while calibrating the NWM without channel loss cannot produce a realistic hydrologic response in the same basin, unlike in WGEW. To compensate for the lack of channel infiltration, the optimal value for the REFKDT parameter increased in the Babocomari basin, to enable increased soil infiltration before the water reached the channels (Table 4). This relationship was less consistent in WGEW. Despite this compensating effect, the model without channel loss still had higher bias (Table 5), particularly in the evaluation period. Thus, it appears that parameter compensation can partially mask out the biases from channel infiltration in WGEW; however, this is not possible in a larger basin with more drainage area (and more area for water to enter the channels that must eventually infiltrate) like the Babocomari basin.

In the San Pedro basin, where streamflow in excess of small base flows is primarily driven by warm season convection, the calibrated model with NCEP Stage IV and NLDAS-2 forcing is associated with lower correlation coefficients, particularly during the evaluation period. As Stage IV remains the more reliable product (in the present study), this discussion centers on assessing the errors associated with only Stage IV precipitation. Figure 12 demonstrates the likely cause of the poor streamflow correlation, as observed precipitation from a NOAA HMT gauge at Elgin, Arizona, sometimes poorly matches Stage IV precipitation, despite the climatology of the two datasets being similar. Figure 12 shows that the evaluation period was wetter than the calibration period, which may partly explain the high model bias after 2013 in the WGEW and the Babocomari basins. Similar results from two other Babocomari basin rain gauges are shown in supplemental material.

One possible reason for this correlation error is that precipitation estimation by radar in the southwest United States is affected by elevated terrain (causing beam blockage and thus leading to the need for high measurement elevations; e.g., Zamora et al. 2014). Radar and available gauge observations are used to derive NCEP Stage IV precipitation (Lin and Mitchell 2005). Limitations in the quality of Stage IV data may cause the reduced streamflow correlation coefficients in the Babocomari River and WGEW. A detailed analysis of differences in WGEW gauge and WSR-88D rainfall data can be found in Morin et al. (2003, 2005).

Regridded 4-km Stage IV and 1/8° NLDAS2 forcing may also spatially spread precipitation over a larger area, therefore buffering precipitation over the landscape and reducing localized high-intensity events. As surface runoff only occurs when there is sufficient precipitation intensity to exceed the infiltration capacity of the soil, spreading of precipitation could reduce surface runoff that might otherwise occur over a small area associated with locally heavier precipitation
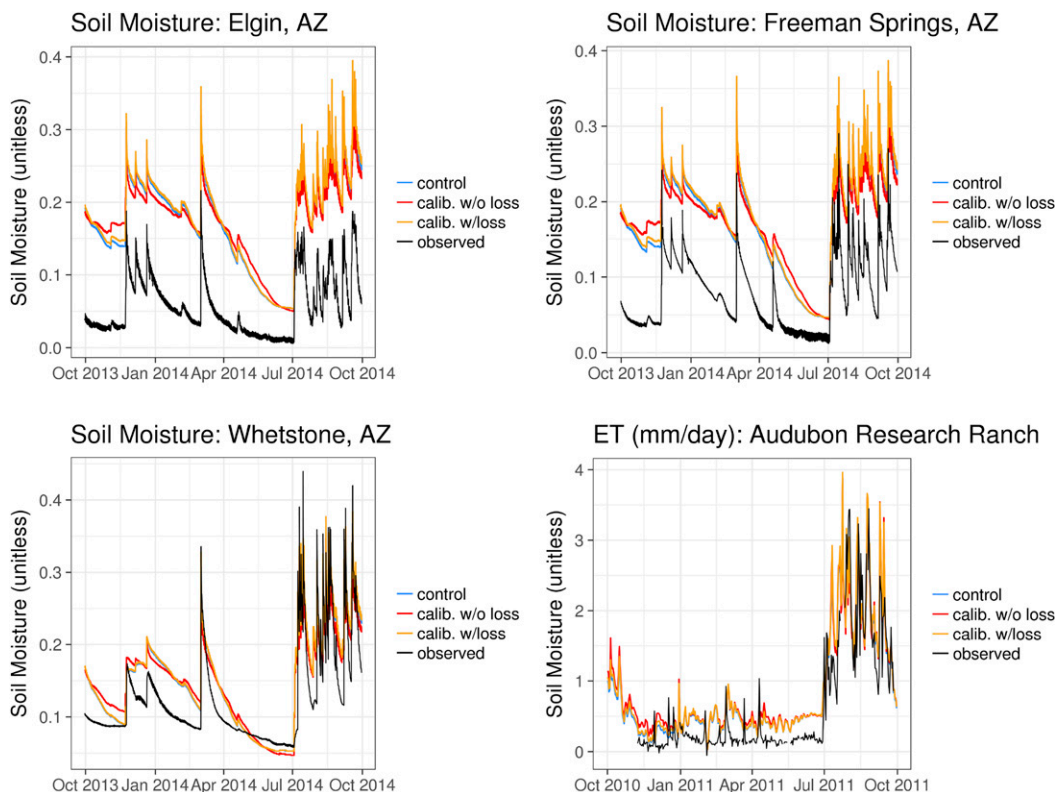
FIG. 10. WY2014 Noah-MP level 1 soil volumetric water content from the control (blue), calibrated (red), and calibrated with channel loss (orange) NWM compared to 5-cm observations (black) for (top left) Elgin, AZ, (top right) Freeman Springs, AZ, and (bottom left) Whetstone, AZ. (bottom right) WY 2011 Modeled ET from the control (blue), calibrated (red), and calibrated with channel loss (orange) solutions for the same period at the Audubon Research Ranch flux tower site. Flux tower ET measurements are plotted in black in this panel. Note that WY2014 ET was not available at this site.

intensity. This might explain why Stage IV forcing produces more precipitation and less streamflow over the WGEW domain (Fig. 12). Spatial spreading of precipitation, and subsequent reduction of surface runoff, may also be part of the cause of the high soil moisture and ET bias.

b. Suggestions for future work

The addition of channel infiltration and subsequent calibration of WRF-Hydro enables it to produce a more realistic hydrologic response with reduced water balance errors in two basins in the southwest United States. Calibration of WRF-Hydro in WGEW with gauge precipitation produces hourly streamflow with KGEs of 0.83. While this calibration does increase the positive bias of near surface soil moisture in WGEW, it does maintain the soil moisture correlation coefficients and reduces ET bias. Stage IV or NLDAS-2 forcing, which is derived by WSR-88D radar data and gauge precipitation or model and satellite remote sensing products, respectively, is subject to greater error and uncertainty

than gauge precipitation, which reduces the skill of the model, irrespective of the channel infiltration function. Given the limitations of the NWM with channel infiltration and its calibration routines described in this work, we make several suggestions for future development of the NWM and other distributed hydrologic models used for forecasting.

Short-term solutions that may ameliorate the calibration challenges for the NWM presented in our results include:

• New precipitation forcing datasets: NOAA is now developing the high-resolution (~1 km) Analysis of Record for Calibration (AORC) precipitation and temperature dataset. While testing is still ongoing, this dataset may add some value over the NLDAS-2 and NCEP Stage IV forcing products, and partially ameliorate some soil moisture and ET bias, described above.

• Additional calibration parameters and metrics: The increased soil moisture bias, despite reduction in streamflow errors from calibration, may indicate potential deficiencies in the Noah-MP structure, including some
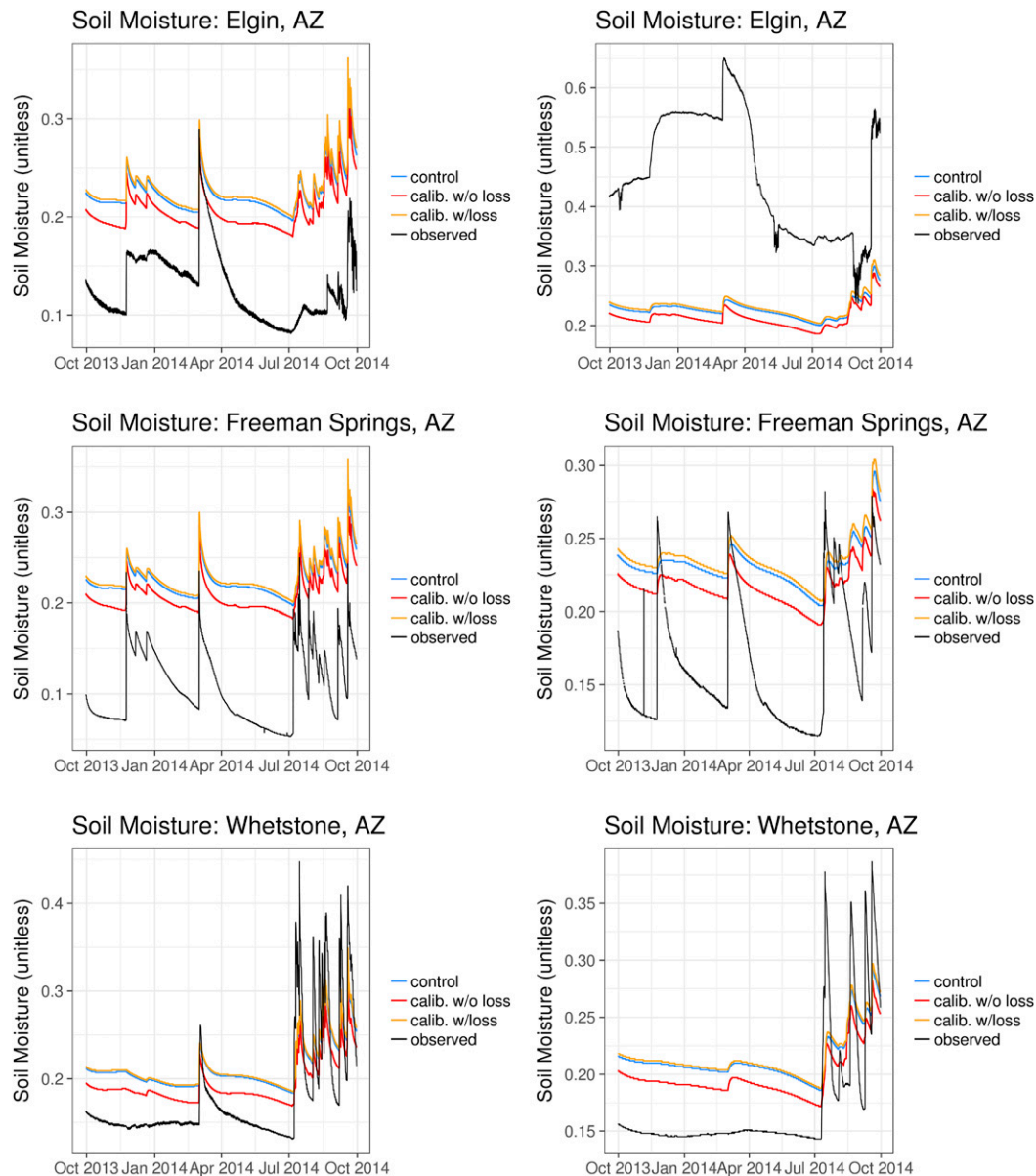
FIG. 11. WY 2014 NWM Noah-MP soil volumetric water content from the control (blue), calibrated (red), and calibrated with channel loss (orange) compared to observations (black) for (top) Elgin, AZ, (middle) Freeman Springs, AZ, and (bottom) Whetstone, AZ. (left) Noah-MP level 2 (10–40 cm) soil moisture compared to observed 20-cm soil moisture, and (right) Noah-MP level 3 (40–100 cm) soil moisture compared to observed 50-cm soil moisture.

hard-coding of parameters (e.g., Mendoza et al. 2015). Cuntz et al. (2016) showed that the water partitioning of Noah-MP was dependent upon several hard-coded parameters. Efforts are now ongoing to update the Noah-MP code to allow for more multicomponent calibration and adjustment of these hard-coded parameters, to improve its physical process representation.

The limitations of the NWM and its calibration approach may also be further addressed with more permanent based solutions in the long term, including:

• Multivariable calibration: Addressing the soil moisture biases may additionally require calibrating the model to quantities other than streamflow, including soil moisture. Basins around the United States with soil moisture observations, including those maintained by NOAA HMT in Arizona and California (Zamora et al. 2011), would be good candidates for
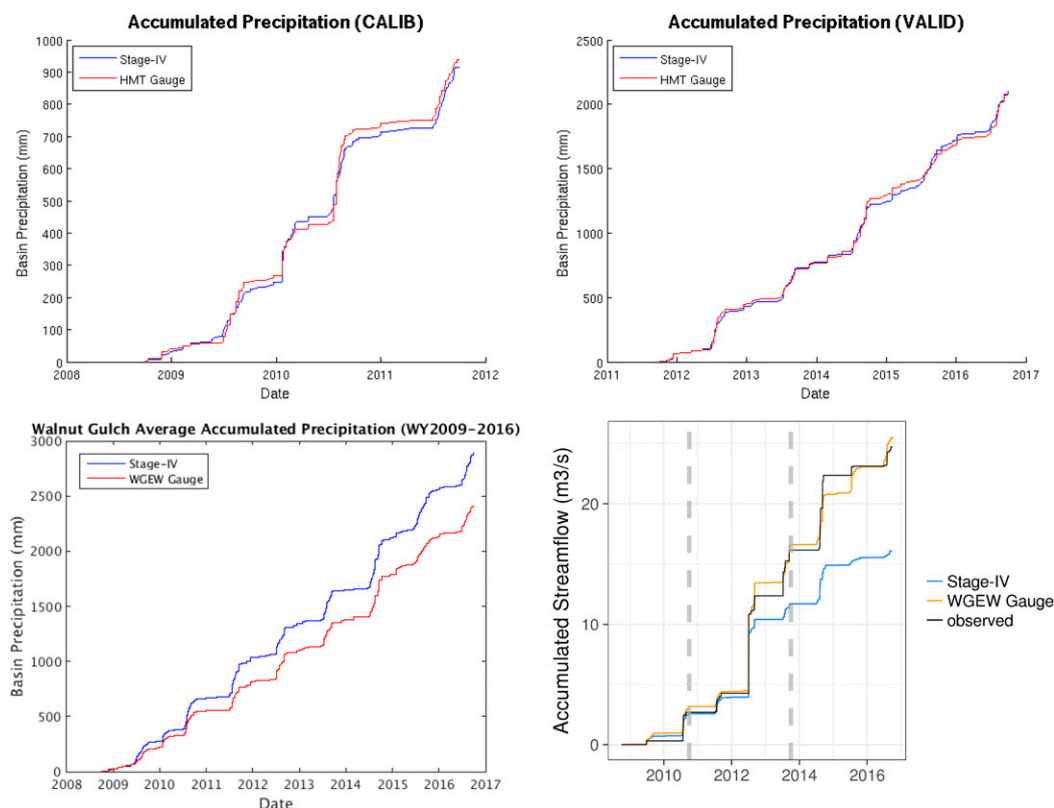
FIG. 12. Elgin, AZ accumulated gauge precipitation (red) and NCEP Stage IV precipitation at the same point (blue) (top left) for the calibration period (WY2009–11) and (top right) the evaluation period (WY2012–16). (bottom left) WY2009–16 area average accumulated WGEW gauge (red) and NCEP Stage IV (blue) precipitation in WGEW. (bottom right) Accumulated NWM WRF-Hydro streamflow (for the same period) at the basin outlet with WGEW forcing (orange) and Stage IV forcing (blue). Parameters in the accumulated streamflow plot are optimized for gauge precipitation.

this. Calibrating WRF-Hydro to soil moisture may permit researchers to reduce soil moisture errors, thus making the model's representation of streamflow and ET more physically consistent.

- Land and soil datasets: More detailed soil datasets (e.g., SSURGO) could also be used to better constrain soil parameters, as Figs. 3 and 5 indicate that a priori NWM parameters likely underestimate soil heterogeneity compared to the SSURGO data shown in the same figures. SSURGO data, which are based on local soil surveys, were used to analyze output from the Noah LSM in Zamora et al. (2014). Use of a SSURGO dataset may also reduce the soil moisture errors and biases shown earlier in this study.

- Alternatives to the baseflow bucket model: Soil moisture errors in the Babocomari basin in deeper layers may also have been caused by modifying the free drainage scaling parameter to essentially a no-flow boundary on the bottom of the Noah-MP LSM. The current baseflow bucket model's use of NHD catchments keeps baseflow relatively local by returning it to

the same channel. This is not realistic in basins with groundwater decoupled from surface water, so it was disabled for this work. A WRF-Hydro configuration that is coupled to a physically based groundwater model, such as ParFlow (e.g., Maxwell et al. 2015) could be a possible solution to this.

- Further enhancements to channel infiltration: Another area for future analysis is to make the channel infiltration function more representative of physical processes, like suction from dry soil (Smith and Goodrich 2000; Smith et al. 2002). The infiltration scheme implemented in the NWM for the present study only accounts for the impact of gravity. Soil suction results in an increase of infiltration into the dry soil at the start of a runoff event (Smith and Parlange 1978; Parlange et al. 1982); however, its impact has not yet been considered. Accounting for this process in a future version of the model, may yield further improvements to WRF-Hydro channel infiltration. Another potential improvement would be if the channel infiltration function were coupled back into the Noah-MP LSM,

permitting the model to resolve the effects of channel infiltration on the surface fluxes in riparian areas, even when water is not flowing. A saturated soil column could also limit channel infiltration during wetter periods.

6. Conclusions

The operational NWM is subject to systematic errors in the southwest CONUS, in part due to lack of a representation of ephemeral channel infiltration. These errors reduce the skill of the NWM as an operational tool in the southwest CONUS by the NWS. Addition of a channel infiltration function and subsequent calibration permits the model to produce a more realistic hydrologic response in low-elevation regions that are fed by convective rainfall. To fully realize the benefits of this infiltration scheme, future work is needed to eliminate the soil moisture biases of the Noah-MP LSM in WRF-Hydro. This might be accomplished through direct calibration to soil moisture, but may also require changes to the Noah-MP structure, and modification or elimination of the WRF-Hydro baseflow bucket scheme to permit deep groundwater recharge in arid regions. Spatial spreading of precipitation by the Stage IV precipitation product likely also contributes to soil moisture and ET biases.

Analysis of modeled streamflow and associated forcing precipitation data suggests that uncertainties associated with national precipitation datasets will need to be considered in future model calibration efforts. A logical step for evaluating model forcing uncertainty is to use the NOAA AORC forcing dataset for calibration of the NWM. Heterogeneity of the soil column depth is also not yet resolved in WRF-Hydro. These limitations of WRF-Hydro could be addressed in subsequent releases of the operational NWM. The observed improvements to modeled ET fluxes, when WRF-Hydro was forced with gauge data and calibrated, suggests that the model could potentially resolve surface fluxes in a realistic manner. Calibration in the larger Babocomari basin, clearly demonstrates the added value of channel infiltration, as calibration without it yields an unrealistic hydrologic response.

REFERENCES

Adams, D. K., and A. C. Comrie, 1997: The North American Monsoon. *Bull. Amer. Meteor. Soc.*, **78**, 2197–2213, https://doi.org/10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2.

Bieda, S. W., C. L. Castro, S. L. Mullen, A. C. Comrie, and E. Pytlak, 2009: The relationship of transient upper-level troughs to variability of the North American Monsoon System. *J. Climate*, **22**, 4213–4227, https://doi.org/10.1175/2009JCLI2487.1.

Blasch, K., T. P. Ferré, J. Hoffmann, D. Poll, M. Baily, and J. Cordova, 2004: Processes controlling recharge beneath ephemeral streams in southern Arizona. *Groundwater Recharge in a Desert Environment: The Southwestern United States*, J. F. Hogan, F. M. Phillips and B. R. Scanlon, Eds., Water Science and Application Series, Vol. 9, Amer. Geophys. Union, 69–76.

Blöschl, G., C. Reszler, and J. Komma, 2008: A spatially distributed flash flood forecasting model. *Environ. Modell. Software*, **23**, 464–478, https://doi.org/10.1016/j.envsoft.2007.06.010.

Broxton, P., P. A. Troch, M. Schaffner, C. Unkrich, and D. Goodrich, 2014: An all-season flash flood forecasting system for real-time operations. *Bull. Amer. Meteor. Soc.*, **95**, 399–407, https://doi.org/10.1175/BAMS-D-12-00212.1.

Callegary, J. B., J. M. Leenhouts, N. V. Paretti, and C. A. Jones, 2007: Rapid estimation of recharge potential in ephemeral-stream channels using electromagnetic methods, and measurements of channel and vegetation characteristics. *J. Hydrol.*, **344**, 17–31, https://doi.org/10.1016/j.jhydrol.2007.06.028.

Cuntz, M., J. Mai, L. Samaniego, M. Clark, V. Wulfmeyer, O. Branch, S. Attinger, and S. Thober, 2016: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *J. Geophys. Res. Atmos.*, **121**, 10 676–10 700, https://doi.org/10.1002/2016JD025097.

Department of Energy, 2018: AmeriFlux Data Use Policy. http://ameriflux.lbl.gov/data/data-policy/.

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: An ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581, https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.

Douglas, M. W., R. A. Maddox, K. Howard, and S. Reyes, 1993: The Mexican monsoon. *J. Climate*, **6**, 1665–1677, https://doi.org/10.1175/1520-0442(1993)006<1665:TMM>2.0.CO;2.

Duan, Q., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.*, **28**, 1015–1031, https://doi.org/10.1029/91WR02985.

Dugger, A. L., and Coauthors, 2017: Learning from the National Water Model: 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, https://ams.confex.com/ams/97Annual/webprogram/Paper314352.html.

Finch, Z. O., and R. H. Johnson, 2010: Observational analysis of an upper-level inverted trough during the 2004 North American Monsoon Experiment. *Mon. Wea. Rev.*, **138**, 3540–3555, https://doi.org/10.1175/2010MWR3369.1.

Gochis, D. J., W. Yu, and D. N. Yates, 2015: The WRF-Hydro model technical description and user's guide, version 1.0. NCAR Tech. Doc., 120 pp., https://www.ral.ucar.edu/projects/wrf_hydro.

Goodrich, D. C., L. J. Lane, R. A. Shillito, S. N. Miller, K. H. Syed, and D. A. Woolhiser, 1997: Linearity of basin response as a function of scale in a semi-arid watershed. *Water Resour. Res.*, **33**, 2951–2965, https://doi.org/10.1029/97WR01422.

——, D. G. Williams, C. L. Unkrich, R. L. Scott, K. R. Hultine, D. Pool, A. Coes, and J. Hogan, 2003: Multiple approaches to estimate ephemeral channel recharge. *Proc. First Interagency Conf. on Research in the Watersheds*, Tucson, AZ, U.S. Department of Agriculture, 118–124, http://www.tucson.ars.ag.gov/icrw/Proceedings/Goodrich.pdf.

——, and Coauthors, 2004: Comparison of methods to estimate ephemeral channel recharge, Walnut Gulch, San Pedro River Basin, Arizona. *Groundwater Recharge in a Desert Environment: The Southwestern United States*, J. F. Hogan, F. M. Phillips and B. R. Scanlon, Eds., Water Science and Application Series, Vol. 9, Amer. Geophys. Union, 77–99.

——, and Coauthors, 2012: KINEROS2/AGWA: Model use, calibration, and evaluation. *Trans. ASABE*, **55**, 1561–1574, https://doi.org/10.13031/2013.42264.

Gourley, J. J., and Coauthors, 2017: The FLASH Project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372, https://doi.org/10.1175/BAMS-D-15-00247.1.

Gupta, H. V., T. Wagener, and Y. Liu, 2008: Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrol. Processes*, **22**, 3802–3813, https://doi.org/10.1002/hyp.6989.

——, H. Kling, K. Yilmaz, and G. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *J. Hydrol.*, **377**, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hardy, J., J. J. Gourley, P.-E. Kirstetter, Y. Hong, F. Kong, and Z. L. Flamig, 2016: A method for probabilistic flash flood forecasting. *J. Hydrol.*, **541**, 480–494, https://doi.org/10.1016/j.jhydrol.2016.04.007.

Julien, P. Y., B. Saghafian, and F. L. Ogden, 1995: Raster-based hydrological modeling of spatially-varied surface runoff. *J. Amer. Water Resour. Assoc.*, **31**, 523–536, https://doi.org/10.1111/j.1752-1688.1995.tb04039.x.

Lahmers, T. M., C. L. Castro, D. K. Adams, Y. L. Serra, J. J. Brost, and T. Luong, 2016: Long-term changes in the climatology of transient inverted troughs over the North American monsoon region and their effects on precipitation. *J. Climate*, **29**, 6037–6064, https://doi.org/10.1175/JCLI-D-15-0726.1.

Lane, L. J., 1983: Transmission losses. *SCS National Engineering Handbook*, U.S. Government Printing Office, 19-1–19-21.

Lespinas, F., A. Dastoor, and V. Fortin, 2017: Performance of the dynamically dimensioned search algorithm: influence of parameter initialization strategy when calibrating a physically based hydrological model. *Hydrol. Res.*, **49** (4), 971–988.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, http://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.

Looper, J. P., and B. E. Vieux, 2012: An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model. *J. Hydrol.*, **412–413**, 114–132, https://doi.org/10.1016/j.jhydrol.2011.05.046.

Luong, T. M., C. L. Castro, H. Chang, T. Lahmers, D. K. Adams, and C. A. Ochoa-Moya, 2017: The more extreme nature of North American monsoon precipitation in the southwestern United States as revealed by a historical climatology of simulated severe weather events. *J. Appl. Meteor. Climatol.*, **56**, 2509–2529, https://doi.org/10.1175/JAMC-D-16-0358.1.

Maddox, R. A., D. M. McCollum, and K. W. Howard, 1995: Large-scale patterns associated with severe summertime thunderstorms over central Arizona. *Wea. Forecasting*, **10**, 763–778, https://doi.org/10.1175/1520-0434(1995)010<0763:LSPAWS>2.0.CO;2.

Maxwell, R. M., L. E. Condon, and S. J. Kollet, 2015: A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geosci. Model Dev.*, **8**, 923–937, https://doi.org/10.5194/gmd-8-923-2015.

McCollum, D. M., R. A. Maddox, and K. W. Howard, 1995: Case study of a severe mesoscale convective system in central Arizona. *Wea. Forecasting*, **10**, 643–665, https://doi.org/10.1175/1520-0434(1995)010<0643:CSOASM>2.0.CO;2.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012: NHDPlus version 2: User guide. U.S. EPA Doc., 179 pp., ftp://ftp.horizon-systems.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.

Mendoza, P. A., M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta, 2015: Are we unnecessarily constraining the agility of complex process-based models? *Water Resour. Res.*, **51**, 716–728, https://doi.org/10.1002/2014WR015820.

Moran, M. S., and Coauthors, 2008: Preface to special section on Fifty Years of Research and Data Collection: U.S. Department of Agriculture Walnut Gulch Experimental Watershed. *Water Resour. Res.*, **44**, W05S01, https://doi.org/10.1029/2007WR006083.

Morin, E., W. F. Krajewski, D. C. Goodrich, X. Gao, and S. Sorooshian, 2003: Estimating rainfall intensities from weather radar data: The scale-dependency problem. *J. Hydrometeor.*, **4**, 782–797, https://doi.org/10.1175/1525-7541(2003)004<0782:ERIFWR>2.0.CO;2.

——, R. A. Maddox, D. C. Goodrich, and S. Sorooshian, 2005: Radar *Z–R* relationship for summer monsoon storms in Arizona. *Wea. Forecasting*, **20**, 672–679, https://doi.org/10.1175/WAF878.1.

National Weather Service, 2014: United States Flood Loss Report - Water Year 2014. https://www.nws.noaa.gov/os/water/Flood%20Loss%20Reports/WY14%20Flood%20Loss%20Summary.pdf.

Newman, A., and R. H. Johnson, 2012: Mechanisms for precipitation enhancement in a North American monsoon upper-tropospheric trough. *J. Atmos. Sci.*, **69**, 1775–1792, https://doi.org/10.1175/JAS-D-11-0223.1.

Niu, G. Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Noorduijn, S. L., M. Shanafield, M. A. Trigg, G. A. Harrington, P. G. Cook, and L. Peeters, 2014: Estimating seepage flux from ephemeral stream channels using surface water and

groundwater level data. *Water Resour. Res.*, **50**, 1474–1489, https://doi.org/10.1002/2012WR013424.

Ogden, F. L., 1997: CASC2D Reference Manual. Dept. of Civil and Environmental Engineering U-37, University of Connecticut, 106 pp.

Parlange, J.-Y., I. Lisle, R. D. Braddock, and R. E. Smith, 1982: The three-parameter infiltration equation. *Soil Sci.*, **133**, 337–341, https://doi.org/10.1097/00010694-198206000-00001.

Pokhrel, P., K. K. Yilmaz, and H. V. Gupta, 2012: Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *J. Hydrol.*, **418–419**, 49–60, https://doi.org/10.1016/j.jhydrol.2008.12.004.

Pytlak, E., M. Goering, and A. Bennett, 2005: Upper-tropospheric troughs and their interaction with the North American monsoon. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., JP2.3, https://ams.confex.com/ams/pdfpapers/85393.pdf.

Rawls, W. J., D. L. Brakensiek, and K. E. Saxton, 1982: Estimation of soil water properties. *Trans. ASABE*, **25**, 1316–1320, https://doi.org/10.13031/2013.33720.

Reed, S., J. Schaake, and Z. Zhang, 2007: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.*, **337**, 402–420, https://doi.org/10.1016/j.jhydrol.2007.02.015.

Samaniego, L., R. Kumar, and S. Attinger, 2010: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.*, **46**, W05523, https://doi.org/10.1029/2008WR007327.

Seastrand, S., Y. Serra, C. L. Castro, and E. Ritchie, 2015: The dominant synoptic-scale modes of North American monsoon precipitation. *Int. J. Climatol.*, **35**, 2019–2032, https://doi.org/10.1002/joc.4104.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Smith, R. E., and J.-Y. Parlange, 1978: A parameter-efficient hydrologic infiltration model. *Water Resour. Res.*, **14**, 533–538, https://doi.org/10.1029/WR014i003p00533.

——, and D. C. Goodrich, 2000: Model for rainfall excess patterns on randomly heterogeneous areas. *J. Hydrol. Eng.*, **5**, 355–362, https://doi.org/10.1061/(ASCE)1084-0699(2000)5:4(355).

Smith, M., and H. V. Gupta, 2012: The Distributed Model Intercomparison Project (DMIP) – Phase 2 experiments in the Oklahoma region, USA. *J. Hydrol.*, **418-419**, 1–2, https://doi.org/10.1016/j.jhydrol.2011.09.036.

Smith, R. E., K. R. J. Smettem, P. Broadbridge, and D. A. Woolhiser, 2002: *Infiltration Theory for Hydrologic Applications.* Water Resources Monogr., Vol. 15, Amer. Geophys. Union, 213 pp.

Stone, J. J., M H. Nichols, D. C. Goodrich, and J. Buono, 2008: Long-term runoff database, Walnut Gulch Experimental Watershed, Arizona, United States. *Water Resour. Res.*, **44**, W05S05, https://doi.org/10.1029/2006WR005733.

Tolson, B. A., and C. A. Shoemaker, 2007: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.*, **43**, W01413, https://doi.org/10.1029/2005WR004723.

Vergara, H., P.-E. Kirstetter, J. Gourley, Z. Flamig, Y. Hong, A. Arthur, and R. Kolar, 2016: Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the conterminous United States. *J. Hydrol.*, **541A**, 421–433, https://doi.org/10.1016/j.jhydrol.2016.06.011.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Woolhiser, D. A., R. E. Smith, and D. C. Goodrich, 1990: A kinematic runoff and erosion manual: Documentation and user manual. USDA Rep. ARS 77, 141 pp., https://www.tucson.ars.ag.gov/unit/publications/PDFfiles/703.pdf.

Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner, and A. Stewart, 2008: Understanding uncertainty in distributed flash flood forecasting for semiarid regions. *Water Resour. Res.*, **44**, W05S19, https://doi.org/10.1029/2007WR005940.

Zamora, R. J., F. M. Ralph, E. Clark, and T. Schneider, 2011: The NOAA Hydrometeorology Testbed soil moisture observing networks: Design instrumentation, and preliminary results. *J. Atmos. Oceanic Technol.*, **28**, 1129–1140, https://doi.org/10.1175/2010JTECHA1465.1.

——, E. P. Clark, E. Rogers, M. B. Ek, and T. M. Lahmers, 2014: An examination of meteorological and soil moisture conditions in the Babocomari River basin before the flood event of 2008. *J. Hydrometeor.*, **15**, 243–260, https://doi.org/10.1175/JHM-D-12-0142.1.

Zampieri, M., E. Serpetzoglou, E.N. Anagnostou, E.I. Nikolopoulos, and A. Papadopoulos, 2012: Improving the representation of river–groundwater interactions in land surface modeling at the regional scale: Observational evidence and parameterization applied in the Community Land Model. *J. Hydrol.*, **420–421**, 72–86, https://doi.org/10.1016/j.jhydrol.2011.11.041.