Hydrologic Model Sensitivity to Temporal Aggregation of Meteorological Forcing Data: A Case Study for the Contiguous United States

ASHLEY E. VAN BEUSEKOM,[a] LAUREN E. HAY,[a] ANDREW R. BENNETT,[a] YOUNG-DON CHOI,[b] MARTYN P. CLARK,[c] JON L. GOODALL,[b] ZHIYU LI,[d] IMAN MAGHAMI,[b] BART NIJSSEN,[a] AND ANDREW W. WOOD[e]

[a] *University of Washington, Seattle, Washington*
[b] *University of Virginia, Charlottesville, Virginia*
[c] *University of Saskatchewan, Canmore, Alberta, Canada*
[d] *University of Illinois at Urbana–Champaign, Urbana, Illinois*
[e] *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 12 June 2021, in final form 23 September 2021)

ABSTRACT: Surface meteorological analyses are an essential input (termed "forcing") for hydrologic modeling. This study investigated the sensitivity of different hydrologic model configurations to temporal variations of seven forcing variables (precipitation rate, air temperature, longwave radiation, specific humidity, shortwave radiation, wind speed, and air pressure). Specifically, the effects of temporally aggregating hourly forcings to hourly daily average forcings were examined. The analysis was based on 14 hydrological outputs from the Structure for Unifying Multiple Modeling Alternatives (SUMMA) model for the 671 Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS) basins across the contiguous United States (CONUS). Results demonstrated that the hydrologic model sensitivity to temporally aggregating the forcing inputs varies across model output variables and model locations. We used Latin hypercube sampling to sample model parameters from eight combinations of three influential model physics choices (three model decisions with two options for each decision, i.e., eight model configurations). Results showed that the choice of model physics can change the relative influence of forcing on model outputs and the forcing importance may not be dependent on the parameter space. This allows for model output sensitivity to forcing aggregation to be tested prior to parameter calibration. More generally, this work provides a comprehensive analysis of the dependence of modeled outcomes on input forcing behavior, providing insight into the regional variability of forcing variable dominance on modeled outputs across CONUS.

KEYWORDS: Forcing; Hydrology; Hydrologic models

1. Introduction

Hydrologic models are used to predict and manage water resources and provide insights into key hydrologic processes. Models are thus a leading strategy for understanding catchment dynamics and the impacts of variability in climate and land use, among other factors. There are numerous hydrologic models available (Hrachowitz and Clark 2017), ranging from simple conceptual models (e.g., the Monthly Water Balance Model; McCabe and Markstrom 2007) to more complex physically based models [e.g., the Gridded Surface Subsurface Hydrologic Analysis (GSSHA); Downer and Ogden 2004].

Regardless of the model complexity, all models are constrained by uncertainties associated with deficiencies in forcing data, model parameters, and model structure. While hydrologic model forcings are considered to be a leading source of uncertainty (Clark and Slater 2006; Kato et al. 2007; Zaitchik et al. 2010; Newman et al. 2015a), most sensitivity studies focus on the sensitivity of model output to changes in model parameters or model structure (e.g., see Wagener and Pianosi 2019). Fewer studies focus on the sensitivity of model output to uncertainties in meteorological forcing data.

The focus on model parameters is perhaps surprising given the uncertainties in meteorological forcing data. Numerous studies point to model inadequacy resulting from biases in precipitation (e.g., Magnusson et al. 2015; Sperna Weiland et al. 2015). Looking only at precipitation biases, studies have found that the choice of model configuration has an influence on the sensitivity to the precipitation biases, and the combination of configuration and forcing decisions is more influential on model output than parameter biases (Mockler et al. 2016). Systematic errors in other forcing variables such as air temperature and long- and shortwave radiation are often smaller than precipitation (Gelati et al. 2018), but may still affect model output. Thus, studies should explore the suite of input forcings beyond precipitation.

There is a gap in research and analyses that addresses the relative influence of input forcings across large geographical domains. Beven (2019) notes that a fundamental problem with hydrologic modeling is the quality of the meteorological input forcings, a finding that is also reflected by the long-standing emphasis on the development of forcing datasets in operational hydrology. At a minimum, hydrologic models require precipitation and temperature forcing inputs, with higher complexity models typically requiring additional meteorological forcing variables (such as long- and shortwave radiation, air pressure, wind speed, and humidity) at increasingly higher spatial and temporal resolutions. A comprehensive analysis of the dependence of modeled outcomes on input forcing behavior can provide valuable information to the hydrologic

Corresponding author: Ashley Van Beusekom, AVBscience@gmail.com

DOI: 10.1175/JHM-D-21-0111.1

modeling community, guiding the directions for areas of improvements.

Each forcing variable has associated uncertainties that can contribute to hydrologic model uncertainties. These may be attributed to disaggregation from daily to hourly (many variables are only measured on a daily basis), interpolation from point measurements to the areal unit of the model (e.g., a grid or a watershed), sparse and irregular station locations, the procedures used to fill in missing data, and varying record lengths and observation intervals (Clark and Slater 2006; Tang et al. 2020; Bennett et al. 2020). There are a number of contiguous United States (CONUS) wide forcing datasets available, each with their own advantages and disadvantages. Many of these are deterministic (single valued), but more recently ensemble forcing datasets such as the spatial-regression-based analysis of Newman et al. (2015a) have been introduced. For example, the Gridded Meteorological Ensemble Tool (GMET) dataset (Newman et al. 2015a) quantifies forcing dataset uncertainty but focuses on daily gridded precipitation and temperature variables and does not include the additional forcing variables or the higher spatial and temporal resolutions required by many of the more complex models in use today. While hourly forcings may be deemed necessary as input for a given model, understanding the implications of using hourly data in a model could guide the modeler when determining how to formulate their hourly datasets. It may be that temporally varying hourly forcing data have little effect on modeled outputs in certain hydroclimatic regions.

Small-scale studies (i.e., restricted to one or a few basins) have shown that no single forcing is important across all modeled variables, model sensitivity strongly depends on the output of interest, and basin-by-basin differences exist in sensitivities to forcings (Raleigh et al. 2015). In addition, a model can often be calibrated even with inaccurate forcings to reproduce streamflow, but it is unlikely that such a model will also be physically reasonable (Ghatak et al. 2018). Thus, a method for model improvement that yields a model whose performance is not highly dependent on the degree to which the "final" calibrated parameter choices are tuned, will help to support confidence in the model's physical representation. The equifinality concept in hydrology, as well as common sense, tells us that there is no single best/final model parameter choice since different model configurations and parameter sets may be acceptable in reproducing the behavior of a given system (Beven 2006). Put another way, there are often multiple nonunique sets of model configurations and parameters that match model and observed outputs equally well (Beven and Freer 2001). Therefore, rather than testing the sensitivity of modeled outputs to variation in input forcings using a single final optimized or calibrated parameter set, the parameter space for commonly calibrated parameters along with a range of model configurations and input forcings could be used. This enables exploring the interactions between input forcing, parameter space, and model configuration and their relative influences on model outputs. When applied across a large number of basins that are characterized by a wide range of hydroclimatic conditions, this type of analysis can provide insight into the varying importance of different forcing variables and whether this is dependent on model configuration and/or parameter choices.

The objective of this study was to provide insight into the dominance and regional variability of different input model forcings across the CONUS. We used the Structure for Unifying Multiple Modeling Alternatives (SUMMA) hydrologic modeling framework (Clark et al. 2015a,b,c) to configure and test a range of hydrological model alternatives (parameterizations and configurations) for the 671 CAMELS (Catchment Attributes and Meteorology for Large-Sample Studies) basins—a large-sample hydrometeorological dataset across the CONUS (Newman et al. 2015b). For each CAMELS basin, we seek to determine the extent to which hydrological model output is sensitive to temporal aggregation in individual model forcings and if the sensitivity of model outputs to temporal aggregation in input forcings depends on the set of parameter values and/or model configuration. The combination of the individual CAMELS basin results as a whole provides information useful to guide investments in future research needed to further improve meteorological forcing data.

2. Study basins

The CAMELS large-sample hydrometeorological catchment set (Newman et al. 2015b; Addor et al. 2017a,b) was selected for this study. The CAMELS dataset was developed for community use and consists of input forcings, catchment attributes, several model implementations, and relevant streamflow observations for 671 small- to medium-sized basins across the CONUS (Fig. 1) that span a wide range of hydroclimatic conditions. The 671 CAMELS basins have minimal human influence and are almost exclusively smaller, headwater-type basins (median basin size of 336 km^2).

3. Methods

The sensitivity of hydrologic model outputs in the CAMELS basins to variations in input forcings was investigated using the SUMMA hydrologic modeling framework. The SUMMA outputs were used to determine if output sensitivities to forcings were influenced by parameter values and/or the model configuration. An overview of the forcing datasets, parameter sets, and model configurations used in the SUMMA model runs is shown in Fig. 2. The following sections describe the hydrologic modeling framework (section 3a), development of the forcing datasets (section 3b), sets of model runs that test a range of forcing datasets and parameter options with a selection of model configuration options for each of the CAMELS basins (section 3c), and statistic chosen as an indicator of model output sensitivity to a change in input forcing (section 3d).

a. Hydrologic modeling framework

The SUMMA hydrologic modeling framework can be used to provide insights in the dominance of different physical processes and regional variability in the suitability of different

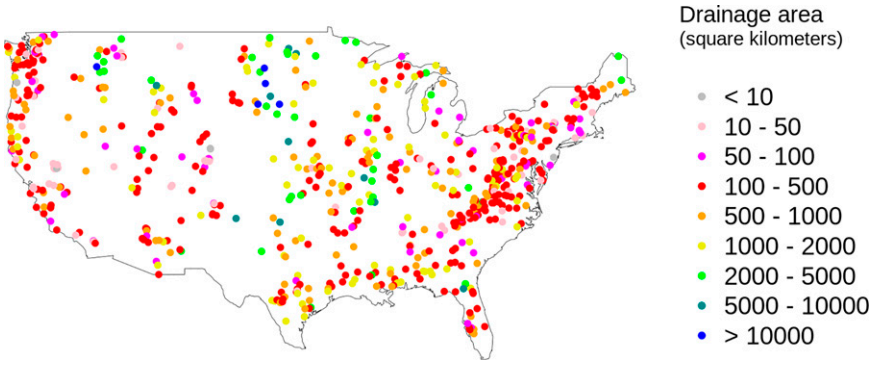

FIG. 1. Location of 671 CAMELS basins colored by drainage area.

modeling approaches (Clark et al. 2015a,b,c) making it an ideal framework in this study for systematic evaluation of alternative modeling approaches. The initial (or default) SUMMA configuration and parameterization used in this study were developed by Wood and Mizukami (2022), which established a baseline model physics selection, soil layer and aquifer configuration, attribute inputs, baseline model outputs and an a priori calibration parameter dataset. Table 1 lists the model decisions for the default SUMMA model configuration used in this study. The model decisions determine the mathematical process representations used in the simulation. Each CAMELS basin was configured to contain a single modeling unit. For each of the CAMELS basins, variations in 14 SUMMA-generated outputs, described in Table 2, were examined with respect to variations in the input forcings, under different model parameterizations and configurations.

b. *Temporal aggregation of forcing datasets*

SUMMA requires seven subdaily input forcing variables, which include precipitation rate (*ppt*), air temperature (*tmp*), specific humidity (*hum*), shortwave radiation (*swr*), longwave radiation (*lwr*), wind speed (*wnd*), and air pressure (*prs*). The SUMMA-CAMELS basin setups used in this study included hourly input SUMMA forcings derived from the hourly NLDAS-2 (North American Land Data Assimilation System; https://ldas.gsfc.nasa.gov/nldas/v2/forcing). The NLDAS-2

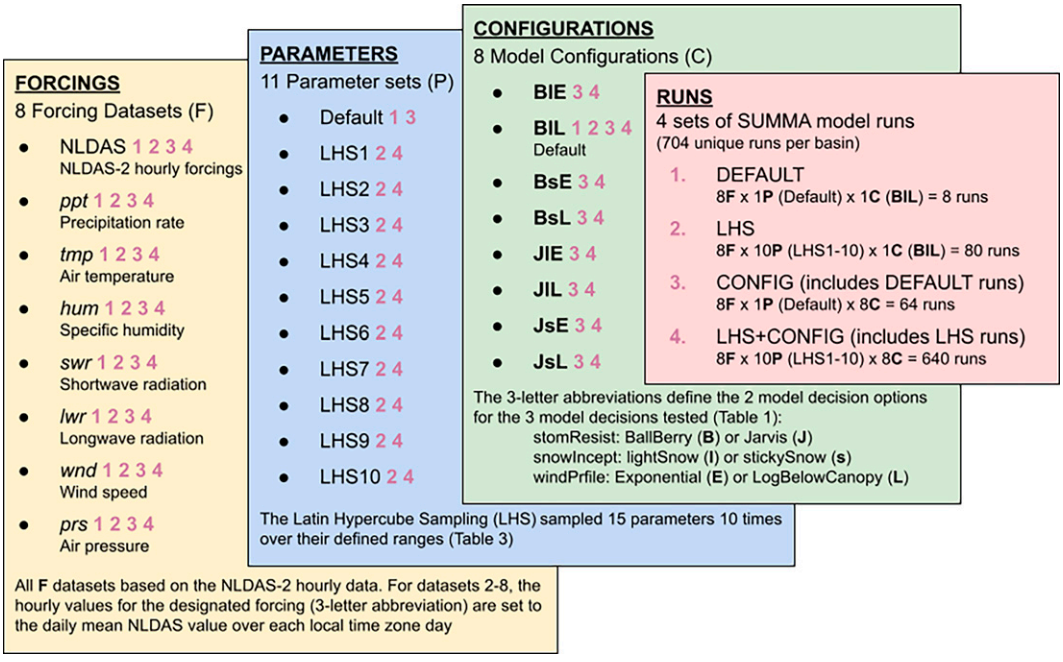

FIG. 2. An overview of the forcing datasets (FORCINGS; yellow box), parameter sets (PARAMETERS; blue box), and model configurations (CONFIGURATIONS; green box) used in the 704 SUMMA model runs (RUNS; pink box) performed for each of the 671 CAMELS basins. Note the pink numbers that follow each forcing, parameter, and configuration refers to the SUMMA model run set as numbered in the pink RUNS box (e.g., the Default parameter set in the PARAMETERS box is used with SUMMA model runs 1 and 3 in the RUNS box).

TABLE 1. SUMMA default model configuration for this study. Model decisions in bold are changed later in the paper. (https://summa.readthedocs.io/en/develop/configuration/SUMMA_model_decisions/).

Model decision	Model decision description	Chosen option
soilCatTbl	Soil-category dataset	STAS
vegeParTbl	Vegetation-category dataset	MODIFIED_IGBP_MODIS_NOAH
soilStress	Function for the soil moisture control on stomatal resistance	NoahType
stomResist	**Function for stomatal resistance**	**BallBerry**
num_method	Numerical method	itertive
fDerivMeth	Method to calculate flux derivatives	analytic
LAI_method	Method to determine LAI and SAI	monTable
f_Richards	Form of Richards equation	mixdform
groundwatr	Groundwater parameterization	bigBuckt
hc_profile	Hydraulic conductivity profile	constant
bcUpprTdyn	Type of upper boundary condition for thermodynamics	nrg_flux
bcLowrTdyn	Type of lower boundary condition for thermodynamics	zeroFlux
bcUpprSoiH	Type of upper boundary condition for soil hydrology	liq_flux
bcLowrSoiH	Type of lower boundary condition for soil hydrology	drainage
veg_traits	Vegetation roughness length and displacement height	Raupach_BLM1994
canopyEmis	Parameterization for canopy emissivity	difTrans
snowIncept	**Parameterization for snow interception**	**lightSnow**
windPrfile	**Canopy wind profile**	**logBelowCanopy**
astability	Stability function	louisinv
canopySrad	Method for canopy shortwave radiation	BeersLaw
alb_method	Albedo representation	conDecay
compaction	Compaction routine	anderson
snowLayers	Method to combine and subdivide snow layers	CLM_2010
thCondSnow	Thermal conductivity representation for snow	jrdn1991
thCondSoil	Thermal conductivity representation for soil	funcSoilWet
spatial_gw	Method for spatial representation of groundwater	localColumn
subRouting	Method for subgrid routing	timeDlay

data are on a 1/8° grid spacing and range from January 1979 to the present. The nonprecipitation land surface forcing fields for NLDAS-2 were derived from the analysis fields of the NCEP North American Regional Reanalysis (NARR; https://www.emc.ncep.noaa.gov/mmb/rreanl/). The details of the spatial interpolation, temporal disaggregation, and vertical adjustment were those employed in NLDAS-1, as presented by Cosgrove et al. (2003). To produce SUMMA model forcings, NLDAS-2 outputs were spatially averaged over each of the 671 CAMELS basins (Mizukami and Wood 2021; Wood and Mizukami 2022). NLDAS-2 is hereafter referred to as NLDAS.

The FORCINGS box in Fig. 2 describes the eight forcing datasets tested with the SUMMA-CAMELS setup. SUMMA outputs (Table 2) generated with the hourly NLDAS forcing dataset were considered the experimental benchmark (NLDAS dataset 1; FORCINGS box in Fig. 2). The temporal aggregation of the input forcings was performed by first

TABLE 2. SUMMA output variables chosen for analysis.

No.	Variable type	SUMMA variable name	Description (units)
1	Liquid water fluxes for the soil domain	SurfaceRunoff	Surface runoff (m s^{-1})
2		AquiferBaseflow	Baseflow from the aquifer (m s^{-1})
3		Infiltration	Infiltration of water into the soil profile (m s^{-1})
4		RainPlusMelt	Rain plus melt (m s^{-1})
5		SoilDrainage	Drainage from the bottom of the soil profile (m s^{-1})
6	Turbulent heat transfer	LatHeatTotal	Latent heat from the canopy air space to the atmosphere (W m^{-2})
7		SenHeatTotal	Sensible heat from the canopy air space to the atmosphere (W m^{-2})
8		SnowSublimation	Snow sublimation/frost (below canopy or nonvegetated) (kg m^{-2} s^{-1})
9	Snow	SWE	Snow water equivalent (kg m^{-2})
10	Vegetation	CanopyWat	Mass of total water on the vegetation canopy (kg m^{-2})
11	Derived	NetRadiation	Net radiation (W m^{-2})
12		TotalET	Total evapotranspiration (kg m^{-2} s^{-1})
13		TotalRunoff	Total runoff (m s^{-1})
14		TotalSoilWat	Total mass of water in the soil (kg m^{-2})

TABLE 3. Calibration parameters chosen for Latin hypercube sampling.

Parameter name	Minimum	Maximum	Default	Constraints
k_macropore	1.0×10^{-7}	0.1	0.0001	
k_soil	1.0×10^{-7}	1.0×10^{-5}	Variable	
theta_sat	0.3	0.6	Variable	>critSoilTranspire; >fieldCapacity; >theta_res
aquiferBaseflowExp	1	10	2.0	
aquiferBaseflowRate	0	0.1	0.1	
qSurfScale	1	100	50	
summerLAI	0.01	10	3	
frozenPrecipMultip	0.5	1.5	1	
heightCanopyTop	0.05	100	Variable	>heightCanopyBottom
heightCanopyBottom	0	5	Variable	
routingGammaShape	2	3	2.5	
routingGammaScale	1	100 000	20 000	
albedoRefresh	1	10	1.0	
tempCritRain	272.16	274.16	273.16	
windReductionParam	0	1	0.28	

taking the 1-h NLDAS forcings and computing a daily mean (over each 24-h period). This daily mean value was used for each of the hourly forcing values for that 24-h period (i.e., all forcings were held constant over a 24-h period, yielding hourly time series with no diurnal cycle with the exception of shortwave radiation, averaged over daytime hours only and zero at night). To test which 24-h constant forcing had the most impact on the SUMMA outputs, seven additional datasets were developed (FORCINGS box in Fig. 2; datasets 2–8) by holding each of the individual forcing variables constant over a 24-h period while the other six forcing variables contain the original hourly NLDAS values. Testing the impacts of aggregation error in this way not only represents the least informative prior which covers a range of timing errors in the forcing data, but also simultaneously covers a broader scope of magnitude errors. Magnitude error effects can be determined through differences in hourly-time step model output, since the forcings were constant for 24 h of input.

c. *SUMMA model runs*

Forcing variables influence model output to different degrees depending on the chosen model configuration and parameterization along with the basin characteristics. In this study, four sets of SUMMA model runs (DEFAULT, LHS, CONFIG, LHS+CONFIG; RUNS box in Fig. 2) were executed to test a range of forcing datasets, parameter values, and model configuration options. The four sets of SUMMA model runs resulted in 704 unique model runs for each of the 671 CAMELS basins (listed in RUNS box in Fig. 2) and are described below.

1) DEFAULT RUNS

The DEFAULT model runs were used in the initial testing of model output sensitivity to a constant daily input forcing. For each CAMELS basin, eight DEFAULT model runs were executed (RUNS box in Fig. 2) which used all eight forcing datasets (FORCINGS box in Fig. 2) with the default parameter set (PARAMETERS box in Fig. 2) and the default model configuration (BIL; CONFIGURATIONS box in Fig. 2).

2) PARAMETER SPACE

The output sensitivities to forcing variations were examined for each CAMELS basin by comparing model output that was produced through sampling across a group of 15 SUMMA parameters (Table 3). The 15 parameters have been previously identified as sensitive and have been calibrated in other studies (Wood and Mizukami 2022; A. W. Wood et al. 2021, unpublished manuscript). Here, the entire parameter space is explored, instead of calibrating each of the 15 parameters to a specific value. This was not meant to be a parameter sensitivity analysis. Traditional parameter sensitivity analysis (i.e., Morris 1991; Sobol' 1993) would have been computationally expensive and beyond the scope of the study. To test if a change in a parameter affected the forcing influence on the model outputs, the 15 parameters (listed in Table 3) were sampled 10 times over their defined ranges using Latin hypercube sampling (LHS) for a representation of the parameter space, for each CAMELS basin. The randomLHS function in R was used (https://www.rdocumentation.org/packages/lhs/versions/1.1.0/topics/randomLHS) to create unique 10×15 LHS sampling matrices for each CAMELS basin. The 671 LHS matrices were used to produce 10 parameter files per basin with random sampling of the 15 parameters while considering the parameter constraints listed in Table 3. The LHS model runs were used to test model output sensitivity to a constant daily input forcing and a range of parameters. For each CAMELS basin, 80 LHS model runs were executed (RUNS box in Fig. 2) which used all eight forcing datasets (FORCINGS box in Fig. 2) with the 10 LHS parameter sets (PARAMETERS box in Fig. 2) and the default model configuration (BIL; CONFIGURATIONS box in Fig. 2).

3) MODEL CONFIGURATION

The choice of model configuration can have a substantial impact on the magnitude and distribution of the output

uncertainty, with model configuration uncertainty greatly outweighing parameter uncertainty in practical applications (Mockler et al. 2016; Poulin et al. 2011). To reduce the subjectivity in a priori assumptions associated with model configuration, Seiller et al. (2017) considered the model set of possible model configurations as an ensemble with a jointly calibrated set of parameters. Here we focused on a less computationally expensive method to explore the effects of model configuration.

In this study, we used SUMMA to configure a range of hydrological model alternatives to determine if the regional variability in forcing dominance changes with SUMMA configuration. Eight SUMMA configurations were chosen based on the parameterization (i.e., physics) choices that showed the most differences in test cases as discussed in Clark et al. (2015b). Combinations of all three model decisions were tested (bold in Table 1) with two options for each decision, resulting in eight different SUMMA configurations.

The CONFIGURATIONS box in Fig. 2 lists the three model decisions and two related options for each decision (note the default model configuration for this study, BallBerry, lightSnow, and logBelowCanopy (BlL), is listed in bold in Table 1). As described in Clark et al. (2015a), the sensitivity of evapotranspiration to different model representations of stomatal resistance (stomResist model decision) was tested using the BallBerry (Ball et al. 1987) and Jarvis (Jarvis 1976) options, the parameterization of snow interception (snowIncept model decision) was tested using the lightSnow (Hedstrom and Pomeroy 1998) and stickySnow (Andreadis et al. 2009) options, and the canopy wind profile (windPrfile model decision) was tested using the exponential (Choudhury and Monteith 1988; Niu and Yang 2004) and logBelowCanopy (Andreadis et al. 2009; Mahat et al. 2013) options. The CONFIG model runs were used to test model output sensitivity to a constant daily input forcing and a range of model configurations. For each CAMELS basin, 64 CONFIG model runs were executed (RUNS box in Fig. 2) which used all eight forcing datasets (FORCINGS box in Fig. 2) with the default parameter set (PARAMETERS box in Fig. 2) and all eight model configurations (CONFIGURATIONS box in Fig. 2).

4) Parameter space and model configuration

A final set of model runs was executed to determine the relation between the SUMMA model configuration and the parameter space. For each CAMELS basin, 640 LHS+CONFIG model runs were executed (RUNS box in Fig. 2) which used all eight forcing datasets (FORCINGS box in Fig. 2) with the 10 LHS parameter sets (PARAMETERS box in Fig. 2) and all eight model configurations (CONFIGURATIONS box in Fig. 2).

d. Kling–Gupta efficiency test

Two of the most popular metrics of hydrological model evaluation are the Nash–Sutcliffe efficiency test (NSE; Nash and Sutcliffe 1970) and the Kling–Gupta efficiency test (KGE; Gupta et al. 2009). The KGE addresses several perceived shortcomings in the NSE (see Gupta et al. 2009). A modified and scaled version of the KGE was used as an indicator of model output sensitivity to a change in input forcing. Model outputs generated with the forcing datasets with one forcing held constant for each day while the other six forcing variables contain the original hourly NLDAS values (FORCINGS box in Fig. 2; datasets 2–8) were compared with the benchmark (NLDAS, dataset 1) on an hourly time step, using the KGE test to determine the influence of changing from an hourly varying to an hourly constant forcing value.

The modified KGE_m value was calculated as (Clark et al. 2021)

$$\mathrm{KGE}_m = 1 - \sqrt{(r-1)^2 + \left[(\mu_{\mathrm{cnst}} - \mu_b)/\sigma_b\right]^2 + \left[(\sigma_{\mathrm{cnst}} - \sigma_b)/\sigma_b\right]^2}, \tag{1}$$

where r is the correlation coefficient, μ is the mean, σ is the standard deviation, and the subscripts b and cnst refer to the benchmark and hourly-constant results, respectively. This modified KGE_m calculation avoids the amplified $\mu_{\mathrm{cnst}}/\mu_b$ values when μ_b is small and avoids the KGE value dependence on the units of measurement (as discussed by Santos et al. 2018; Clark et al. 2021).

The KGE_m values range from $-\infty$ to 1, with 1 being a perfect match with the benchmark results. Similar to Beck et al. (2020), we scaled KGE_m values to avoid the heavy influence of large negative values. The following transform was used (based on Mathevet et al. 2006):

$$\mathrm{KGE} = \mathrm{KGE}_m/(2 - \mathrm{KGE}_m). \tag{2}$$

This results in KGE values that range between -1 and 1, with lower KGE values indicating larger forcing influence on an output variable. If the KGE has a value of 1 (no change from outputs generated using the benchmark), then daily forcing datasets could replace hourly values in a given basin with no discernible change in results. KGE values less than 1 indicate that SUMMA outputs are sensitive to the forcing that is being held constant and, depending on how small the KGE value is, care should be taken when determining hourly values for this forcing variable.

SUMMA output variables (Table 2) produced using the NLDAS forcing (dataset 1; FORCINGS box in Fig. 2) were considered the benchmark (b). The 14 SUMMA output variables produced using the seven forcing datasets that hold an individual NLDAS forcing variable constant (cnst; datasets 2–8; FORCINGS box in Fig. 2) were compared with the benchmark by computing the KGE values on five years of hourly model output from 1 October 1991 to 30 September 1996. There was not a reason to choose those exact five years but initial tests showed sensitivity was not related to the particular five years chosen.

4. Results

The following sections present the KGE results for each set of SUMMA model runs outlined in the RUNS box in Fig. 2 (DEFAULT, LHS, CONFIG, and LHS+CONFIG).

a. Default runs (DEFAULT)

For each of the 14 selected SUMMA output variables, the boxplots in Fig. 3 show the range in the DEFAULT-KGE

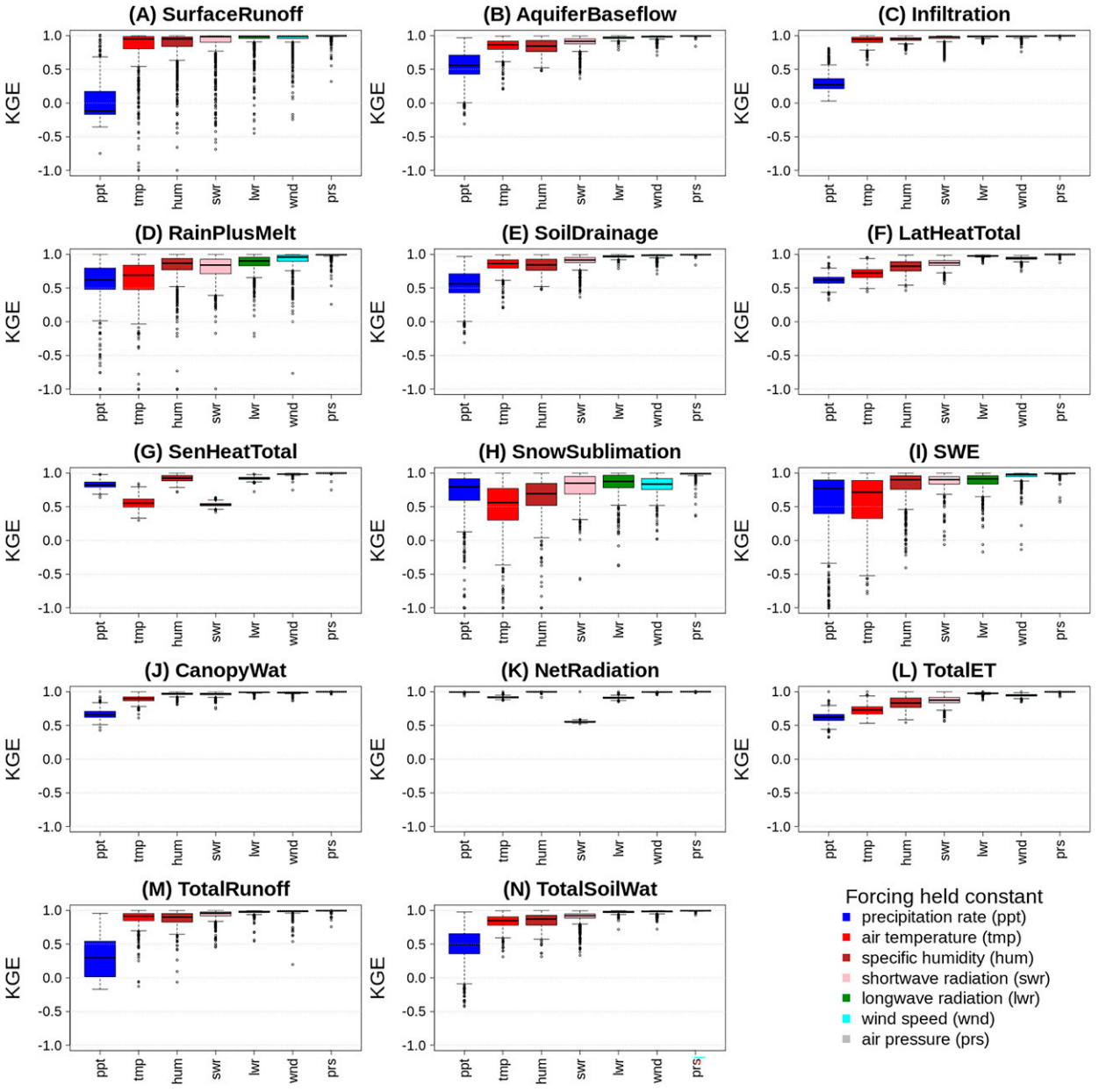

FIG. 3. (a)–(n) Range in CAMELS basins DEFAULT-KGE values for each forcing dataset grouped by SUMMA output variable.

values across all CAMELS basins (671 DEFAULT-KGE values/boxplot) by forcing, providing an overall perspective on the influence of temporally aggregating the hourly forcings on an output variable. In general, the water fluxes and stores are more strongly affected by constant-hourly *ppt*, while energy fluxes (other than LatHeatTotal; Fig. 3f) are more strongly affected by constant-hourly *tmp* and *swr*. The outliers in the boxplots indicate that analysis of individual basins could provide very different results than the overall perspective described below.

Based on the median DEFAULT-KGE values, forcing datasets using hourly-constant *ppt* had the largest overall influence on the SUMMA output variables out of all the hourly-constant forcings (lowest DEFAULT-KGE values). The lowest median DEFAULT-KGE values by output variable are seen for SurfaceRunoff in Fig. 3a. When compared to the DEFAULT-KGE values from the other forcing datasets, constant *ppt* produces some of the lowest median DEFAULT-KGE values for all output variables with the exception of SenHeatTotal (Fig. 3g, which was more strongly affected by hourly-constant *tmp* and *swr*), SnowSublimation (Fig. 3h; *tmp* and *hum* are lower), SWE (Fig. 3i; *tmp* is lower), and NetRadiation (Fig. 3k; *swr*, *lwr*, and *tmp* are lower). Forcing datasets using constant *tmp* rank an overall second in

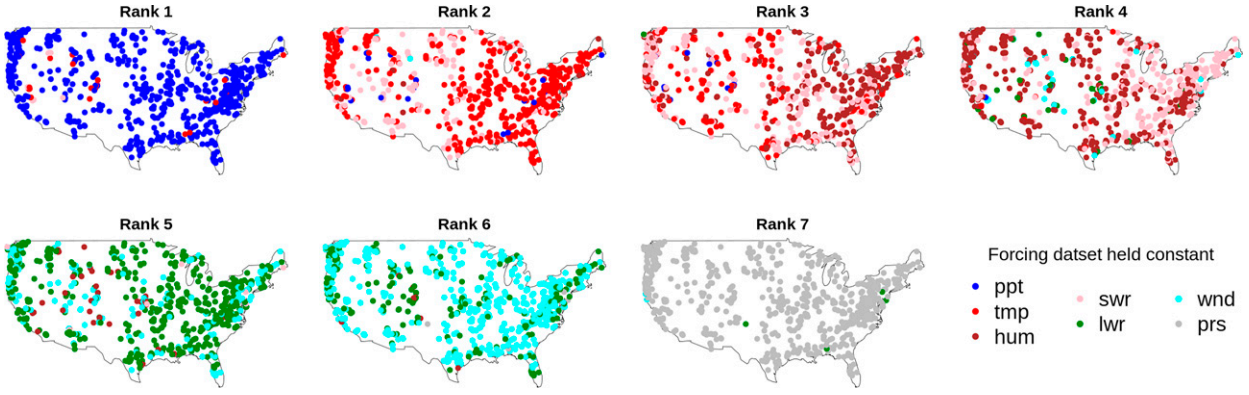

FIG. 4. Spatial distribution of basin DEFAULT-KGE values summed over SUMMA output variables and ranked from low (rank 1; most sensitive) to high (rank 7; least sensitive) summed KGEs.

influence on output variables. When compared to the median DEFAULT-KGE values from the other forcing datasets (Fig. 3), constant *tmp* produces the lowest median DEFAULT-KGE values for SnowSublimation (Fig. 3h) and SWE (Fig. 3i). Forcing datasets using constant *hum* rank an overall third in influence on output variables and based on median DEFAULT-KGE values for constant *hum*, show the greatest influence on SnowSublimation values (Fig. 3h). The lowest median DEFAULT-KGE values using constant *swr* are produced for SenHeatTotal (Fig. 3g) and NetRadiation (Fig. 3k). Note, NetRadiation is essentially only affected by constant values of *swr* (Fig. 3k). Forcing datasets using constant *lwr* and constant *wnd* have some influence on RainPlusMelt (Fig. 3d), SnowSublimation (Fig. 3h), and SWE (Fig. 3i) values, with outliers showing greater influence in individual basins. Forcing datasets using constant *prs* produced minimal effect on any of the outputs indicating that diurnal variations in *prs* have minimal influence on any of the output variables.

Next, for each CAMELS basin, the DEFAULT-KGE values produced from the seven forcing datasets 2–8 (FORCINGS box in Fig. 2) were summed over the 14 model output variables and the resulting seven sums were then ranked for each basin. Figure 4 shows this ranking of the seven forcing datasets spatially by basin, while Fig. 5 shows, by rank, both the associated range in the basin mean forcing value and the basin counts (671 total) for each forcing dataset held constant. Given the large variability of the ranked DEFAULT-KGE values across basins and forcings, the broader-scale patterns may not apply to an individual basin.

In Figs. 4 and 5, *ppt* is most commonly ranked having the largest differences between the hourly varying and hourly constant simulations (rank 1 for lowest summed KGE; 633 basin counts), with a small number of basins being most sensitive to hourly constant *tmp* (red; 26 counts) and *swr* (pink; 12 counts). Note the rank 1 *ppt* basins are associated with higher basin mean *ppt* values (Fig. 5a), rank 1 *tmp* basins are associated with lower basin mean *tmp* values (Fig. 5b), and rank 1 *swr* basins are associated with higher basin mean *swr* values (Fig. 5d). The second most important source of differences (rank 2) is predominantly *tmp* (red) with smaller counts of *swr* (pink) that are concentrated in the central CONUS. Rank 3 consists predominantly of *hum* (brown) concentrated in the eastern CONUS, followed by *swr* (pink) and *tmp* (red). Rank 4 are predominantly *hum* (brown) followed by *swr* (pink) that are concentrated in the eastern CONUS. Note the basins with the higher basin mean *hum* values are associated with the

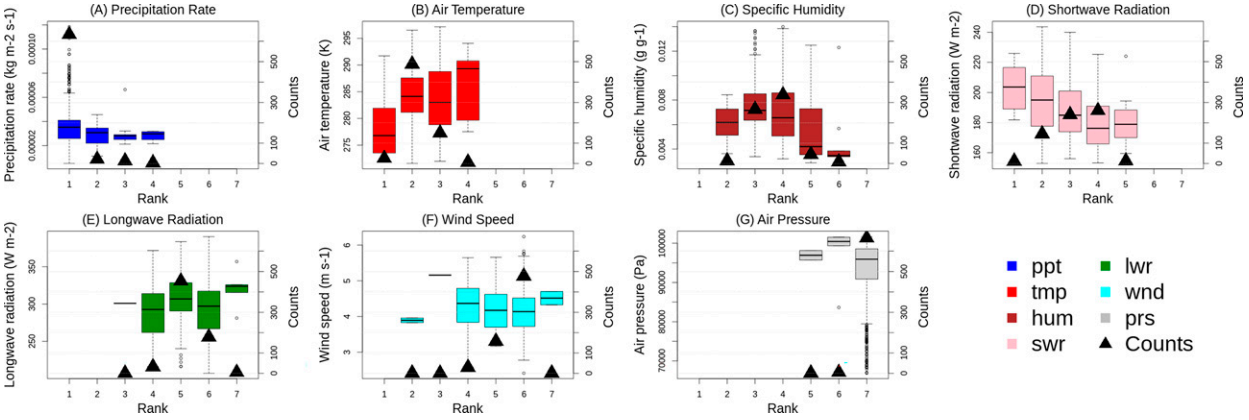

FIG. 5. Range in basin mean hourly forcing values and basin counts associated with DEFAULT-KGE ranked sums from Fig. 4.

lower *hum* ranks (Fig. 4c; ranks 2–4). Rank 5 are predominantly *lwr* (green), followed by *wnd* (cyan) and *hum* (brown) that are concentrated in the central CONUS. Rank 6 are predominantly *wnd* (cyan) with some *lwr* (green). Rank 7 (highest summed KGEs; least sensitive) are almost all *prs* (gray).

b. *Parameter space (LHS)*

Not surprisingly, the parameter sets from the LHS resulted in SUMMA simulations with different KGE values (not shown) but we analyzed the results to determine whether these different parameter sets affected the relative order of model output sensitivities to constant forcing in a given basin. We repeated the analysis from the previous section for each of the 10 LHS parameter sets with the results summarized in Fig. 6. Each row in Fig. 6 shows the relative counts for each forcing dataset for a different SUMMA output variable by rank 1–7 (columns). The *x* axis for each individual plot indicates each of the 10 LHS parameter sets. The *y* axis for each individual plot indicates the relative forcing dataset counts by LHS parameter set. The counts in Fig. 6 are relatively consistent across the parameter sets for each rank and output variable. For example, the counts for LatHeatTot rank 1 (Fig. 6, row 7) are divided between *ppt*, *tmp*, and *swr* (highest to lowest) and show similar counts across all 10 LHS parameter sets. In other words, the relative influence of these three forcing variables on LatHeatTot is relatively insensitive to the value of the LHS parameters. The similarity in output sensitivity over the LHS parameter sets (each *x*-axis bar) indicates that forcing importance may not be dependent on the parameter space allowing for model output sensitivity to forcing aggregation to be tested prior to parameter calibration.

c. *Model configuration (CONFIG)*

We repeated the analysis again, but now varied the SUMMA configuration (with the default parameter values). As before, the CONFIG-KGE values were first examined by summing, for each basin, the seven KGE values from forcing datasets 2–8 (FORCINGS box in Fig. 2) for each of the eight SUMMA configurations and each output variable. For each output variable, the range in the summed-KGE values over the CAMELS basins are shown as boxplots for each SUMMA configuration (1–8) in Fig. 7. Note that the *y*-axis range varies for the plots in Fig. 7 to better demonstrate the boxplot patterns for each output variable (total possible range from −7 to 7). NetRadiation (Fig. 7n) is the only output variable that shows similar summed-KGE ranges over the configurations, indicating minimal configuration influence on NetRadiation values. For the other output variables, the size of the boxes reflects the influence of the model decision options (CONFIGURATIONS box in Fig. 2) on the output variables. Most evident in Fig. 7 are the similar sensitivity patterns according to each windPrfile decision option (SUMMA configurations 1, 3, 5, 7 "Exponential" versus 2, 4, 6, 8 "LogBelowCanopy"; CONFIGURATIONS box in Fig. 2) on every output variable with the exception of CanopyWat (Fig. 7f), which shows the effects of the snowIncept decisions (SUMMA configurations 1, 2, 5, 6 "lightSnow" versus 3, 4, 7, 8 "stickySnow"; CONFIGURATIONS box in Fig. 2).

Next, the CONFIG-KGE values were summed over the SUMMA output variable and the sums were ranked for each of the eight SUMMA configurations. Figure 8 summarizes the basin counts by configuration for each rank (columns) and constant forcing dataset (rows). The count summaries in Fig. 8 show that the SUMMA configuration can make a difference in the ranking of forcing influence. Note the higher counts in *tmp* and the lower counts in *ppt* associated with the Exponential model decision for the wind profile (SUMMA configurations 1, 3, 5, 7 (BlE, BsE JlE, JsE) through the canopy (windPrfile) in the rank 1 counts. In general, the ranking for a given forcing changes only by one rank as a result of a change in SUMMA configuration (i.e., ranks 5 and 6 for *lwr* and *wnd*, ranks 3 and 4 for *hum* and *swr*), indicating that configuration may have a small effect on forcing influence.

d. *Parameter space and model configuration (LHS+CONFIG)*

A final set of 640 SUMMA runs was made for each of the 671 CAMELS basins (LHS+CONFIG; RUNS box in Fig. 2) by combining the eight input forcing datasets (FORCINGS box in Fig. 2) with the 10 LHS parameter sets (PARAMETERS box in Fig. 2) and the eight SUMMA configurations (CONFIGURATIONS box in Fig. 2). For each CAMELS basin, the LHS+CONFIG-KGE values were calculated and summed and ranked for each individual SUMMA output variable. Then, for each rank (1–7), the counts for each CAMELS basin were summed (similar to Fig. 6). Figure 9 shows the relative basin LHS+CONFIG-KGE rank counts (plots in columns, seven ranks) for each SUMMA output variable (plots in rows, 14 output variables) for each of the eight SUMMA configurations run with each of the 10 LHS parameter sets (80 SUMMA runs on the *x* axis for each plot). Each plot row in Fig. 9 shows the relative counts for each forcing dataset for a different SUMMA output variable. Note that the SUMMA output variables were ordered to match those in Fig. 6 (which grouped similar count patterns over the ranks). The SUMMA runs on the *x*-axis groups the results by SUMMA configuration (i.e., the first 10 SUMMA runs in Fig. 9 display results for the first SUMMA configuration (BlE) with LHS 1–10, etc.). The blocky nature in the boxplots depicts the influence of the SUMMA model decision on the SUMMA output variables, which can change with rank, indicating the varying influence of forcing dataset and SUMMA configuration on outputs. For example, when examining the rank 1 barplots in column 1, the influence of the stomResist decision is somewhat evident in the 2-block pattern of *ppt* and *tmp* in the Total SoilWat bars [SUMMA configurations 1–4 (BlE, BlL, BsE, BsL) versus 5–8 (JlE, JlL, JsE, JsL)]. In contrast, the influence of the snowIncept decision is very evident in the 4-block pattern of *ppt* and *tmp* in the CanopyWat bars [SUMMA configurations 1, 2, 5, 6 (BlE, BlL, JlE, JlL) versus 3, 4, 7, 8 (BsE, BsL, JsE, JsL)]; and the influence of the windPrfile decision is very evident in the 4-block pattern of *ppt* and *tmp* in the SWE bars [SUMMA configurations 1, 3, 5, 7 (BlE, BsE JlE, JsE) versus 2, 4, 6, 8 (BlL, BsL JlL, JsL)].

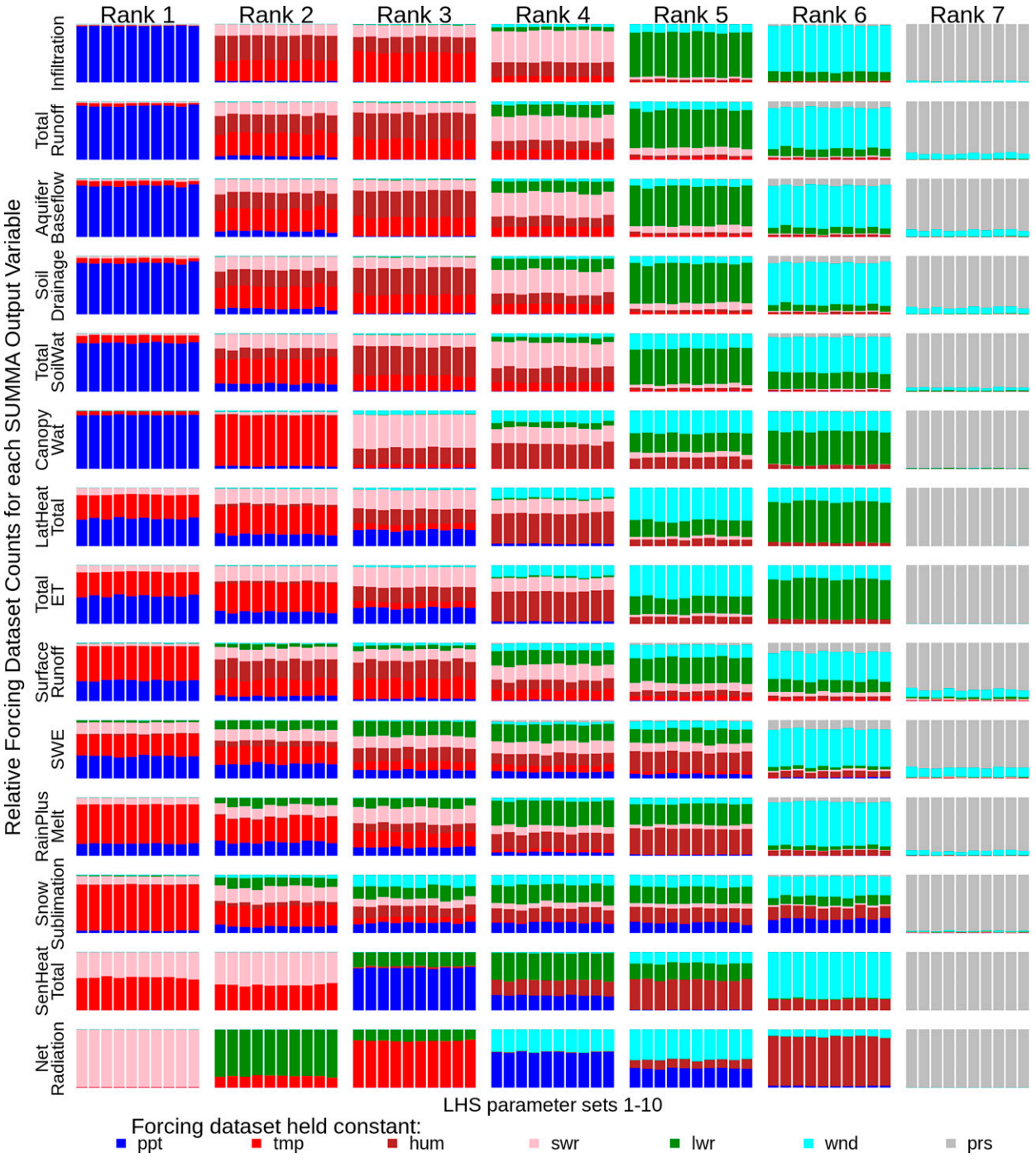

FIG. 6. Relative CAMELS basin KGE rank counts (y axis, total counts = 671) by LHS parameter set (x axis, 10 parameter sets) for ranks 1–7 (columns) for the 14 SUMMA output variables (rows).

5. Summary and discussion

SUMMA applied across the 671 CAMELS basins distributed across the CONUS (Fig. 1) provides insights into the influence and regional variability of different input model forcings. The methodology introduced in this paper was used to examine hourly model output sensitivities to constant daily forcings in combination with the influence of parameter values and model configuration. To synthesize the above KGE analysis across SUMMA model runs (the sets outlined in the RUNS box in Fig. 2; DEFAULT, LHS, CONFIG, LHS+CONFIG), a summary is shown in Figs. 10a–c.

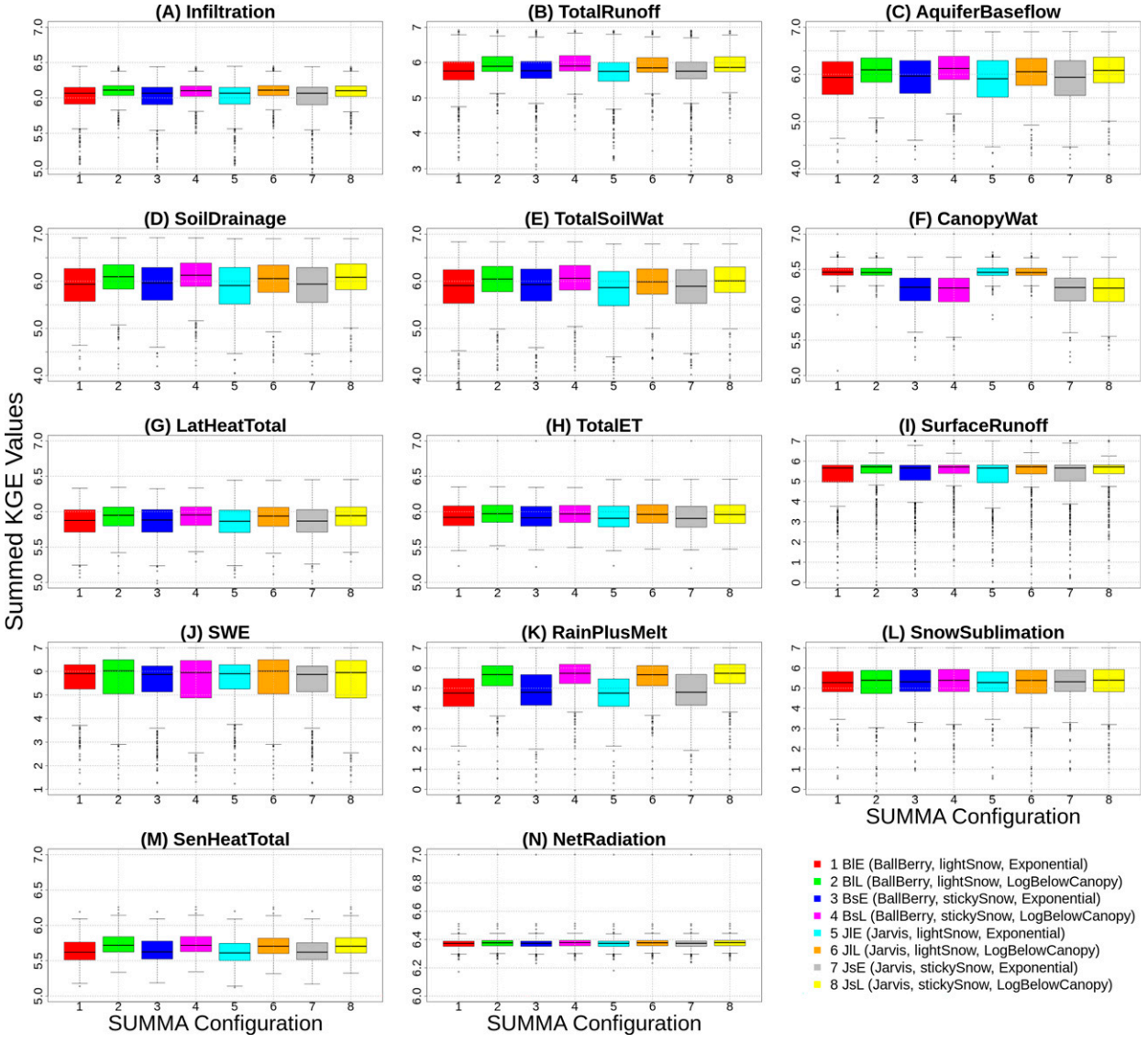

FIG. 7. Range in the summed seven KGE values from forcing datasets 2–8 over the CAMELS basins for eight SUMMA configurations. Note *y*-axis range changes with output variable, with a total possible range from −7 to 7.

Figure 10a summarizes the range in KGE values over the forcing datasets from the DEFAULT set of SUMMA model runs for each output variable (Table 2). The forcing influence varies by model output [*x* axis, compare Infiltration (*ppt*), SnowSublimation (*tmp*), and NetRadiation (*swr*)] and by model location (as evidenced by the large range in the boxplots).

Figures 10b and 10c display the KGE ranges for each SUMMA output variable and each forcing dataset on the *x* axis, respectively. Four boxplots are shown for each *x*-axis output variable: tan, purple, cyan, and magenta boxplots display the KGE ranges from the DEFAULT, LHS, CONFIG, and LHS+CONFIG runs, respectively. In Fig. 10b, the KGE range by SUMMA output variable shows which output variables are sensitive to a change in forcings, with Infiltration, CanopyWat, and NetRadiation showing the least and SnowSublimation showing the greatest overall sensitivity (based on the median values). The similar KGE ranges in the four boxplots indicate that the model parameterization and configuration do not have a significant influence on output sensitivity to forcings on the basins overall, even though the SUMMA configuration can make a difference at the individual basin level (Fig. 8). This means that choice of model configuration for a particular basin should also consider the accuracy of the available input forcings (because different configurations are affected more heavily by different forcings).

In Fig. 10c, the KGE by forcing dataset provides information on which forcing variables are most influential on model outputs. Overall, the forcing dataset with *prs* held constant shows the least influence and the forcing dataset with *ppt*

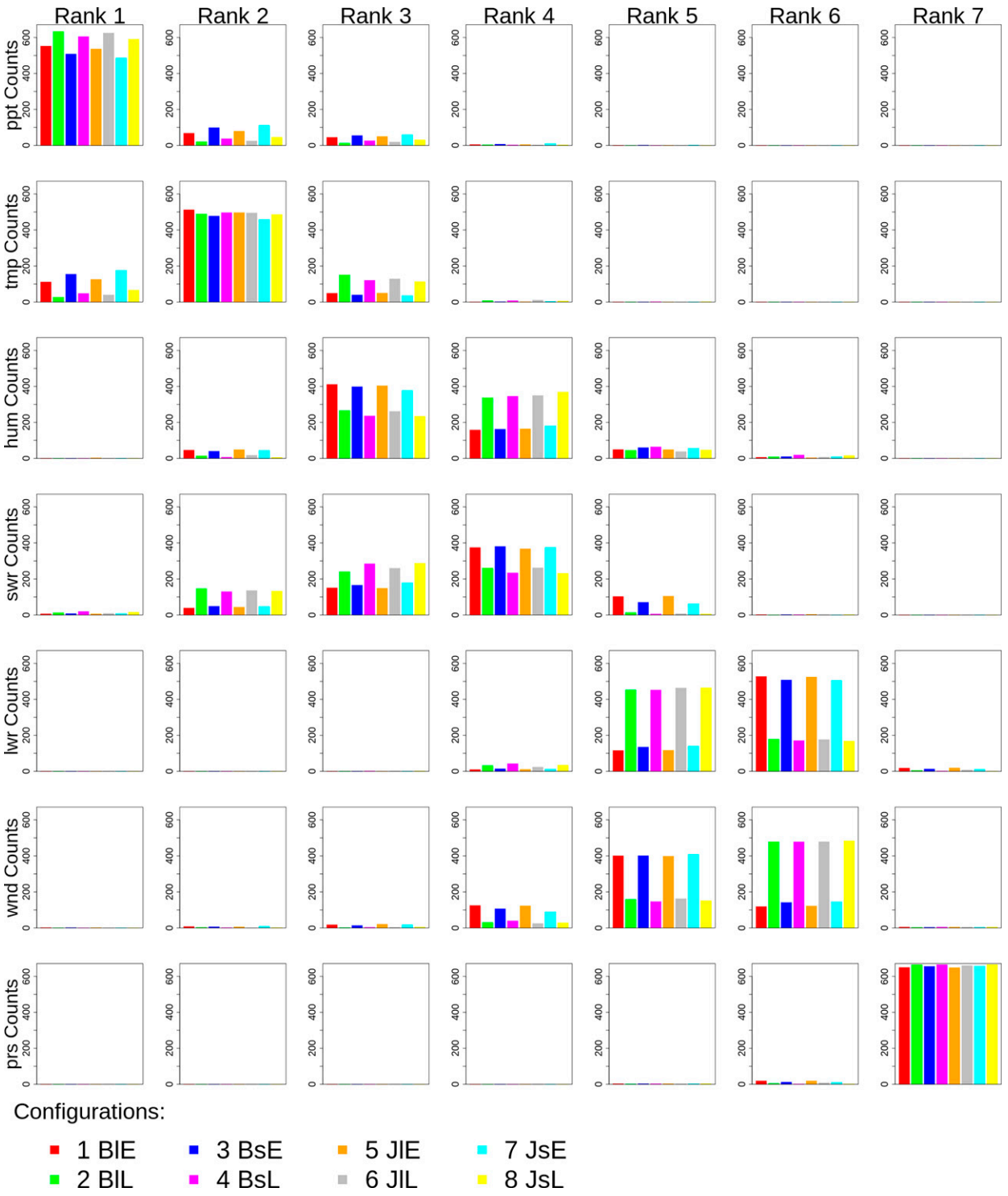

FIG. 8. Summary of counts for each SUMMA configuration for ranks 1–7 (columns) and forcing dataset (rows).

held constant shows the greatest influence (based on the median values). While these results may not be surprising, note that the individual basin results are highly variable; in most cases the scaled KGE range in the boxplots is from −1 to 1. KGE values close to 1 indicate that using that forcing as a constant daily value will have no impact on outputs and therefore forcing does not need to be temporally distributed over the 24-h period. This applies to *prs* in most cases

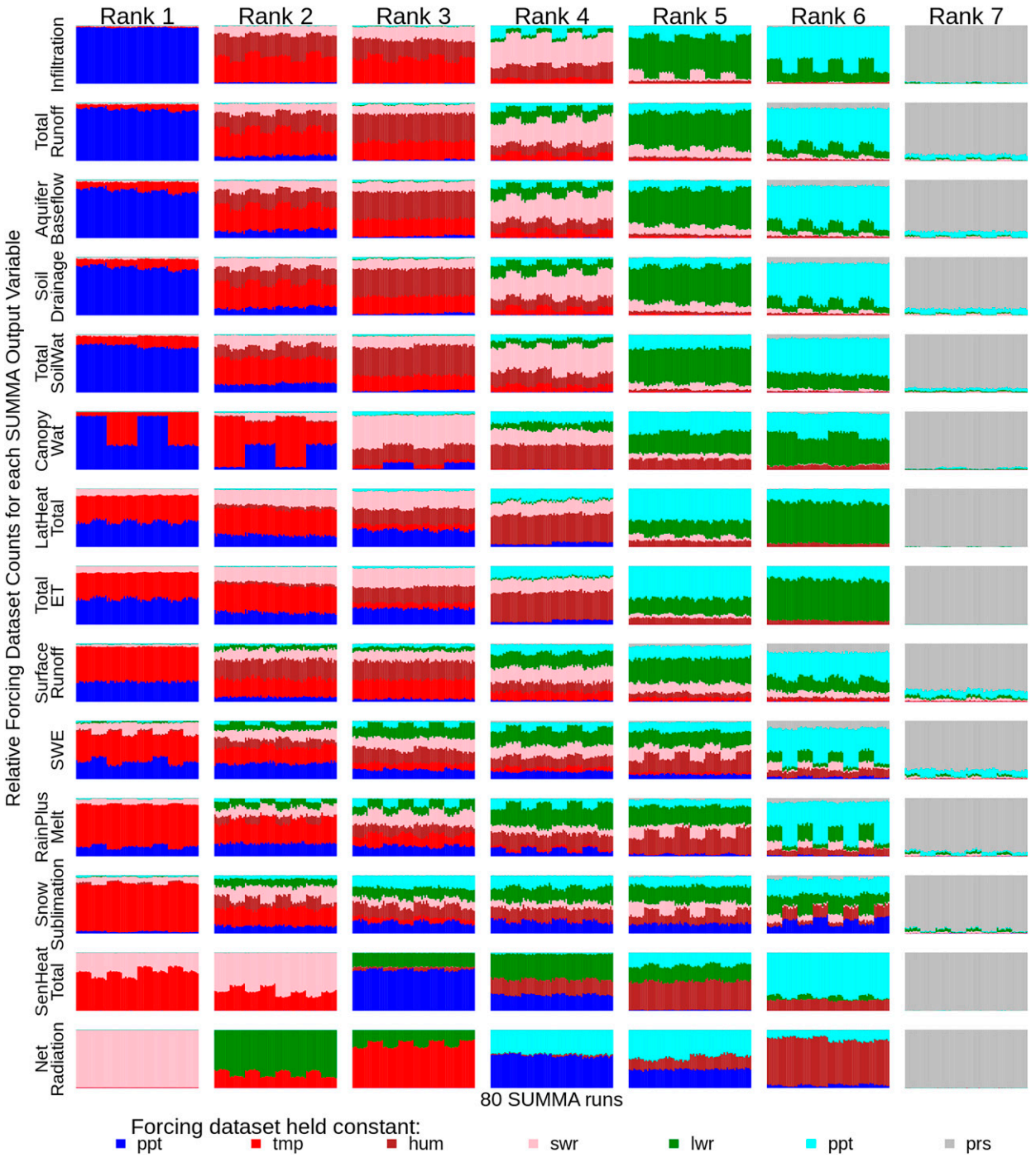

FIG. 9. Relative CAMELS basin KGE rank counts (*y* axis, total counts = 671) by LHS+CONFIG SUMMA runs (*x* axis, 80 runs) for ranks 1–7 (columns) for the 14 SUMMA output variables (rows). The 80 SUMMA runs are in blocks of 10 with the following model configuration order: BlE, BlL, BsE, BsL, JlE, JlL, JsE, JsL.

and *wnd* and *lwr* in some cases. Conversely, the lower KGE values associated with hourly constant *ppt*, *tmp*, *hum*, and *swr* values indicate greater sensitivity and suggest that model simulations require subdaily time series for these forcings in most cases. The basins with model outputs most influenced by constant *ppt*, *hum*, and *swr* tended to have higher basin mean *ppt* (Fig. 5a), *hum* (Fig. 5c), and *swr* (Fig. 5d), respectively. The basins with model outputs most influenced by constant *tmp* tended to have relatively lower basin mean *tmp* (Fig. 5b).

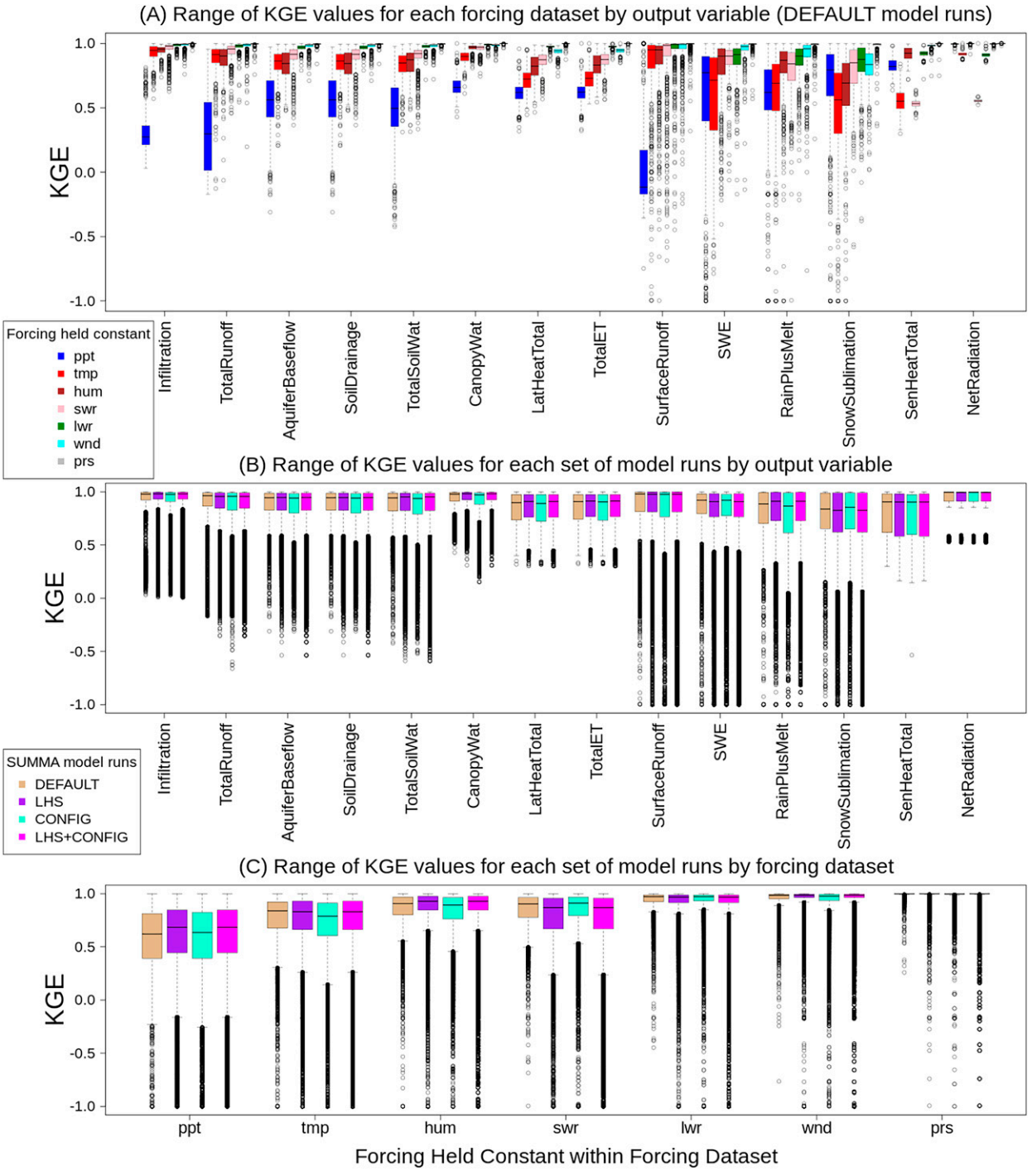

FIG. 10. Summary of KGE range for (a) each forcing dataset by SUMMA output variable using only the DEFAULT SUMMA model runs and each set of SUMMA model runs by (b) SUMMA output variable and (c) forcing dataset.

The 671 CAMELS basins provide a unique dataset to test relations between forcing sensitivity and basin characteristics across the CONUS and potentially determine linkages between these sensitivities and site characteristics. For a more thorough analysis, a finer-resolution parameter space could have been tested, more model configurations could have been added, and specific periods of hydrologic events, extremes, or seasons could have been examined. We chose to expand the computational expense of the problem by using the full set of CAMELs basins with their heterogeneous geocharacteristics,

instead of using fewer basins and a finer resolution parameter set with more model configurations. Larger differences were found between basins with dissimilar geocharacteristics, than within individual basins with alternative modeling decisions. While we did look at seasonal results (not shown), here we presented a detailed evaluation over the entire 5-yr period, as seasonal results did not exhibit differences meriting intense exploration at the shorter time scales. However, it is possible that a seasonal analysis may provide interesting results at the individual basin level.

6. Conclusions

This study used the 671 CAMELS basins to improve understanding of how uncertainties in meteorological forcing data affect hydrologic model simulations, shedding light on the influence of seven forcing variables (precipitation rate, air temperature, specific humidity, shortwave radiation, longwave radiation, wind speed, and air pressure) across the contiguous United States (CONUS). SUMMA was used to study the impact of parameter space and model configuration on the influence and regional variability of input model forcings.

Results show that the impacts of temporal forcing aggregation errors affect model outputs differently across the CONUS. Aggregation error in this context covered a range of timing errors in the forcing data, but also simultaneously covered a broader scope of magnitude errors. Overall, changing from an hourly varying to an hourly precipitation rate was found to have the most influence on the hydrologic outputs, and air pressure was found to have the least influence, but results vary at the individual basin level. In general, basins where hydrologic outputs are most sensitive to a constant precipitation rate, shortwave radiation, or specific humidity had relatively higher mean values of these forcings compared to other basins. Conversely, basins with hydrologic outputs most sensitive to constant air temperature had relatively lower mean values of air temperature, compared to other basins.

Comparing the results over the parameter space indicated that the use of a particular parameter set is not critical in determining the most and least influential forcing variables for a given basin. This means that forcing sensitivity can be tested prior to parameter calibration. Comparing the results over the choice of model configuration indicated that the choice of model physics changes the relative influence of a forcing in a given basin (choice of model configuration should consider the accuracy of the input forcings). Future research to improve meteorological forcing data should be directed at improving the forcing variables that are critical to the important simulated outputs, given the desired model physics.

Acknowledgments. This work was supported by the National Science Foundation under collaborative Grants 1664061, 1664119, and 1664018 for the development of HydroShare (http://www.hydroshare.org). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data availability statement. The input forcings can be obtained as a NetCDF file in the Hydroshare resource (Mizukami and Wood 2021). The SUMMA setup for the CAMELS basins can be obtained from the summa_camels folder of the HydroShare resources (Choi et al. 2021). The SUMMA code can be obtained from https://github.com/NCAR/summa.

REFERENCES

Addor, N., A. Newman, N. Mizukami, and M. P. Clark, 2017a: The CAMELS data set: Catchment attributes for large-sample studies, version 2.0. UCAR/NCAR, accessed 1 January 2021, https://doi.org/10.5065/D6G73C3Q.

——, ——, ——, and ——, 2017b: The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst. Sci.*, **21**, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017.

Andreadis, K. M., P. Storck, and D. P. Lettenmaier, 2009: Modeling snow accumulation and ablation processes in forested environments. *Water Resour. Res.*, **45**, W05429, https://doi.org/10.1029/2008WR007042.

Ball, J. T., I. E. Woodrow, and J. A. Berry, 1987: A model predicting stomatal conductance and its contribution to the control of photosynthesis under ifferent environmental conditions. *Progress in Photosynthesis Research*, J. Biggins, Ed., Springer, 221–224, https://doi.org/10.1007/978-94-017-0519-6_48.

Beck, H., M. Pan, P. Lin, J. Seibert, A. Dijk, and E. Wood, 2020: Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *J. Geophys. Res. Atmos.*, **125**, https://doi.org/10.1029/2019JD031485.

Bennett, A. R., J. J. Hamman, and B. Nijssen, 2020: MetSim: A Python package for estimation and disaggregation of meteorological data. *J. Open Source Software*, **5**, 2042, https://doi.org/10.21105/joss.02042.

Beven, K., 2006: A manifesto for the equifinality thesis. *J. Hydrol.*, **320**, 18–36, https://doi.org/10.1016/j.jhydrol.2005.07.007.

——, 2019: How to make advances in hydrological modelling. *Hydrol. Res.*, **50**, 1481–1494, https://doi.org/10.2166/nh.2019.134.

——, and J. Freer, 2001: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.*, **249**, 11–29, https://doi.org/10.1016/S0022-1694(01)00421-8.

Choi, Y., and Coauthors, 2021: SUMMA Simulations using CAMELS Datasets on CyberGIS-Jupyter for Water, HydroShare, accessed 1 January 2021, https://www.hydroshare.org/resource/50e9716922dc487981b71e2e11f3bb5d/.

Choudhury, B. J., and J. L. Monteith, 1988: A four-layer model for the heat budget of homogenous land surfaces. *Quart. J. Roy. Meteor. Soc.*, **114**, 373–398, https://doi.org/10.1002/qj.49711448006.

Clark, M. P., and A. G. Slater, 2006: Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeor.*, **7**, 3–22, https://doi.org/10.1175/JHM474.1.

——, and Coauthors, 2015a: A unified approach for process-based hydrologic modeling: Part 1. Modeling concept. *Water Resour. Res.*, **51**, 2498–2514, https://doi.org/10.1002/2015WR017198.

——, and Coauthors, 2015b: A unified approach for process-based hydrologic modeling: Part 2. Model implementation and case studies. *Water Resour. Res.*, **51**, 2515–2542, https://doi.org/10.1002/2015WR017200.

——, and Coauthors, 2015c: The structure for unifying multiple modeling alternatives (SUMMA), version 1.0: Technical description. NCAR Tech. Note NCAR/TN-514+STR, 50 pp., https://doi.org/10.5065/D6WQ01TD.

——, and Coauthors, 2021: The abuse of popular performance metrics in hydrologic modeling. *Water Resour. Res.*, **57**, https://doi.org/10.1029/2020WR029001.

Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.*, **108**, 8842, https://doi.org/10.1029/2002JD003118.

Downer, C. W., and F. L. Ogden, 2004: GSSHA: A model for simulating diverse streamflow generating processes. *J. Hydrol. Eng.*, **9**, 161–174, https://doi.org/10.1061/(ASCE)1084-0699(2004)9:3(161).

Gelati, E., B. Decharme, J.-C. Calvet, M. Minvielle, J. Polcher, D. Fairbairn, and G. P. Weedon, 2018: Hydrological assessment of atmospheric forcing uncertainty in the Euro-Mediterranean area using a land surface model. *Hydrol. Earth Syst. Sci.*, **22**, 2091–2115, https://doi.org/10.5194/hess-22-2091-2018.

Ghatak, D., B. Zaitchik, S. Kumar, M. A. Matin, B. Bajracharya, C. Hain, and M. Anderson, 2018: Influence of precipitation forcing uncertainty on hydrological simulations with the NASA South Asia land data assimilation system. *Hydrology*, **5**, 57, https://doi.org/10.3390/hydrology5040057.

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

Hedstrom, N. R., and J. W. Pomeroy, 1998: Measurements and modelling of snow interception in the boreal forest. *Hydrol. Processes*, **12**, 1611–1625, https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1611::AID-HYP684>3.0.CO;2-4.

Hrachowitz, M., and M. P. Clark, 2017: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrol. Earth Syst. Sci.*, **21**, 3953–3973, https://doi.org/10.5194/hess-21-3953-2017.

Jarvis, P., 1976: The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philos. Trans. Roy. Soc. London*, **B273**, 593–610, https://doi.org/10.1098/rstb.1976.0035.

Kato, H., M. Rodell, F. Beyrich, H. Cleugh, E. van Gorsel, H. Liu, and T. P. Meyers, 2007: Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP Sites. *J. Meteor. Sci.*, **87A**, 187–204, https://doi.org/10.2151/jmsj.85A.187.

Magnusson, J., N. Wever, R. Essery, N. Helbig, A. Winstral, and T. Jonas, 2015: Evaluating snow models with varying process representations for hydrological applications. *Water Resour. Res.*, **51**, 2707–2723, https://doi.org/10.1002/2014WR016498.

Mahat, V., D. G. Tarboton, and N. P. Molotch, 2013: Testing above- and below-canopy representations of turbulent fluxes in an energy balance snowmelt model. *Water Resour. Res.*, **49**, 1107–1122, https://doi.org/10.1002/wrcr.20073.

Mathevet, T., C. M. V. Andréassian, and C. Perrin, 2006: A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. *IAHS Publ.*, **307**, 211–219, https://iahs.info/uploads/dms/13614.21–211-219-41-MATHEVET.pdf.

McCabe, G. J., and S. L. Markstrom, 2007: A monthly water-balance model driven by a graphical user interface. USGS Open-File Rep. 2007-1088, 6 pp., https://doi.org/10.3133/ofr20071088.

Mizukami, N., and A. Wood, 2021: NLDAS Forcing NetCDF using CAMELS datasets from 1980 to 2018. HydroShare, accessed 1 January 2021, http://www.hydroshare.org/resource/a28685d2dd584fe5885fc368cb76ff2a.

Mockler, E. M., K. P. Chun, G. Sapriza-Azuri, M. Bruen, and H. S. Wheater, 2016: Assessing the relative importance of parameter and forcing uncertainty and their interactions in conceptual hydrological model simulations. *Adv. Water Resour.*, **97**, 299–313, https://doi.org/10.1016/j.advwatres.2016.10.008.

Morris, M. D., 1991: Factorial sampling plans for preliminary computational experiments. *Technometrics*, **33**, 161–174, https://doi.org/10.1080/00401706.1991.10484804.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models. Part I. A discussion of principles. *J. Hydrol.*, **10**, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

Newman, A. J., and Coauthors, 2015a: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeor.*, **16**, 2481–2500, https://doi.org/10.1175/JHM-D-15-0026.1.

——, and Coauthors, 2015b: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci.*, **19**, 209–223, https://doi.org/10.5194/hess-19-209-2015.

Niu, G. Y., and Z. L. Yang, 2004: Effects of vegetation canopy processes on snow surface energy and mass balances. *J. Geophys. Res.*, **109**, D23111, https://doi.org/10.1029/2004JD004884.

Poulin, A., F. Brissette, R. Leconte, R. Arsenault, and J. S. Malo, 2011: Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *J. Hydrol.*, **409**, 626–636, https://doi.org/10.1016/j.jhydrol.2011.08.057.

Raleigh, M. S., J. D. Lundquist, and M. P. Clark, 2015: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.*, **19**, 3153–3179, https://doi.org/10.5194/hess-19-3153-2015.

Santos, L., G. Thirel, and C. Perrin, 2018: Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrol. Earth Syst. Sci.*, **22**, 4583–4591, https://doi.org/10.5194/hess-22-4583-2018.

Seiller, G., F. Anctil, and R. Roy, 2017: Design and experimentation of an empirical multistructure framework for accurate, sharp and reliable hydrological ensembles. *J. Hydrol.*, **552**, 313–340, https://doi.org/10.1016/j.jhydrol.2017.07.002.

Sobol', I. M., 1993: Sensitivity analysis for nonlinear mathematical models. *Math. Modell. Comput. Exp.*, **1**, 407–414.

Sperna Weiland, F. C., J. A. Vrugt, R. L. P. H. van Beek, A. H. Weerts, and M. F. P. Bierkens, 2015: Significant uncertainty in global scale hydrological modeling from precipitation data errors. *J. Hydrol.*, **529**, 1095–1115, https://doi.org/10.1016/j.jhydrol.2015.08.061.

Tang, G., M. P. Clark, A. J. Newman, A. W. Wood, S. M. Papalexiou, V. Vionnet, and P. H. Whitfield, 2020: SCDNA: A serially complete precipitation and temperature dataset for North America from 1979 to 2018. *Earth Syst.*

Sci. Data, **12**, 2381–2409, https://doi.org/10.5194/essd-12-2381-2020.

Wagener, T., and F. Pianosi, 2019: What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. *Earth-Sci. Rev.*, **194**, 1–18, https://doi.org/10.1016/j.earscirev.2019.04.006.

Wood, A. W., and N. Mizukami, 2022: SUMMA CAMELS Dataset. HydroShare, accessed 1 January 2021, http://www.hydroshare.org/resource/0513cf5e792a4dc4acd0ca77a8146036.

Zaitchik, B. F., M. Rodell, and F. Olivera, 2010: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resour. Res.*, **46**, W06507, https://doi.org/10.1029/2009WR007811.