Comparing Large-Scale Hydrological Model Predictions with Observed Streamflow in the Pacific Northwest: Effects of Climate and Groundwater*

MOHAMMAD SAFEEQ,[*] GUILLAUME S. MAUGER,[+] GORDON E. GRANT,[#] IVAN ARISMENDI,[@]
ALAN F. HAMLET,[&] AND SE-YEUN LEE[+]

[*] *College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon*
[+] *Climate Impacts Group, University of Washington, Seattle, Washington*
[#] *Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon*
[@] *Department of Fisheries and Wildlife, Oregon State University, Corvallis, Oregon*
[&] *Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, Indiana*

(Manuscript received 22 November 2013, in final form 28 May 2014)

ABSTRACT

Assessing uncertainties in hydrologic models can improve accuracy in predicting future streamflow. Here, simulated streamflows using the Variable Infiltration Capacity (VIC) model at coarse ($^1/_{16}°$) and fine ($^1/_{120}°$) spatial resolutions were evaluated against observed streamflows from 217 watersheds. In particular, the adequacy of VIC simulations in groundwater- versus runoff-dominated watersheds using a range of flow metrics relevant for water supply and aquatic habitat was examined. These flow metrics were 1) total annual streamflow; 2) total fall, winter, spring, and summer season streamflows; and 3) 5th, 25th, 50th, 75th, and 95th flow percentiles. The effect of climate on model performance was also evaluated by comparing the observed and simulated streamflow sensitivities to temperature and precipitation. Model performance was evaluated using four quantitative statistics: nonparametric rank correlation ρ, normalized Nash–Sutcliffe efficiency NNSE, root-mean-square error RMSE, and percent bias PBIAS. The VIC model captured the sensitivity of streamflow for temperature better than for precipitation and was in poor agreement with the corresponding temperature and precipitation sensitivities derived from observed streamflow. The model was able to capture the hydrologic behavior of the study watersheds with reasonable accuracy. Both total streamflow and flow percentiles, however, are subject to strong systematic model bias. For example, summer streamflows were underpredicted (PBIAS = −13%) in groundwater-dominated watersheds and overpredicted (PBIAS = 48%) in runoff-dominated watersheds. Similarly, the 5th flow percentile was underpredicted (PBIAS = −51%) in groundwater-dominated watersheds and overpredicted (PBIAS = 19%) in runoff-dominated watersheds. These results provide a foundation for improving model parameterization and calibration in ungauged basins.

1. Introduction

Climate changes anticipated over the next few decades pose challenges to resource managers seeking the most effective strategies to adapt, maintain, and restore rivers, watersheds, and aquatic ecosystems. Because water resources are particularly sensitive to changes in climate, managers benefit from accurate analyses of historical streamflows and predictions of future hydrologic behavior. Accurate estimation of runoff, especially during dry seasons, is extremely critical to plan for hydroelectric power generation (Hamlet et al. 2010), agriculture and municipal water supply (Roy et al. 2012), aquatic habitat (Battin et al. 2007), and water-based recreation (Farley et al. 2011). Both empirical and numerical models have been routinely used for predicting future streamflows and improving understanding of hydrological functioning at varying spatial and temporal scales. In large watershed– and regional-scale studies, land surface models (LSMs) such as the catchment model (Koster et al. 2000), Community Land Model

[*] Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/JHM-D-13-0198.s1.

Corresponding author address: Mohammad Safeeq, College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 3200 SW Jefferson Way, Corvallis, OR 97331.
E-mail: mohammad.safeeq@oregonstate.edu

DOI: 10.1175/JHM-D-13-0198.1

(Oleson et al. 2010), Noah model (Ek et al. 2003), Sacramento Soil Moisture Accounting Model (Burnash et al. 1973), Unified Land Model (Livneh et al. 2011), and Variable Infiltration Capacity (VIC) model (Liang et al. 1994) are commonly used (Koster et al. 2010; Nijssen et al. 2014; Vano et al. 2012; Wang et al. 2009; Xia et al. 2012, 2014). In the U.S. Pacific Northwest (PNW), the large-scale VIC model has been widely employed to study regional-scale changes in snowpack (Hamlet et al. 2005), water resources (Hamlet et al. 2007; Liu et al. 2013), droughts (Shukla and Wood 2008), and energy (Hamlet et al. 2010).

The VIC model, typically implemented at a spatial resolution of 1/8° and 1/16°, is calibrated and validated using observed or naturalized streamflow from large rivers (Hamlet et al. 2013; Matheussen et al. 2000). The number of watersheds used for calibration and validation has been quite variable but is generally limited to available gauged watersheds. One of the assumptions in calibrating the model against observed streamflows in large watersheds is that calibrated parameters are applicable to subwatersheds within the larger basins. This assumption may not be valid, however, in regions and basins with strong hydrogeological differences, thereby introducing some degree of uncertainty into model predictions at finer spatial scales. Conversely, improving the topographic representation by increasing the model spatial resolution or model calibration over small watersheds may be adversely affected by errors in the meteorological driving data, resulting in a calibrated model with compensating errors, that is, getting the right answer for the wrong reasons. Examining the sources and magnitudes of uncertainty at the small watershed scale can help interpret and constrain predictions of direction, magnitude, and timing of future streamflow changes and thereby improve decision making.

Sources of hydrologic modeling uncertainty can be classified as parametric (Beven and Binley 1992; Duan et al. 2006) or structural (Butts et al. 2004; Refsgaard and Knudsen 1996). Parametric uncertainties are associated with the model input data and parameter values, whereas structural uncertainties are associated with the model formulation. Both parametric and structural uncertainties can be minimized, but this may require different strategies for each type in terms of model selection, forcing, calibration, and parameterization. Identifying the major sources of uncertainties and distinguishing which of these are due to model forcing, parameter estimation, and/or model structure is fundamental to minimizing uncertainties (Beven and Freer 2001; McMichael et al. 2006). Various techniques have been developed [e.g., Generalized Likelihood Uncertainty Estimation (GLUE), bootstrapping, Monte Carlo based, Bayesian method, and machine learning] and utilized for model uncertainty analysis (Beven 2011; Shrestha 2010). These techniques can be implemented within many different parameter spaces and model structures (Butts et al. 2004; Clark and Vrugt 2006; Gupta et al. 1998; Jin et al. 2010; Shen et al. 2012). Despite the scientific merits of exploring parameter space and tradeoffs between various model structures, such an approach will be computationally intensive at a regional scale such as the PNW. In fact, for an LSM such as VIC, full hydrologic calibration at a small scale is extremely resource intensive (Oubeidillah et al. 2014), and model calibration is often restricted to a subset of large basins (Hamlet et al. 2013) or grid cells (Troy et al. 2008). Even when high-performance supercomputing is available, exhaustive calibrations and validations of LSMs have to rely on assimilated (Oubeidillah et al. 2014) or naturalized (Hamlet et al. 2013; Vano et al. 2012) streamflow time series because of the lack of unregulated stream gauges. Given all of these limitations, it becomes important to evaluate and assess whether any model inherits systematic biases, whether these are more prevalent in some landscapes than others, and whether these biases can be reduced to improve model performance. Any evaluation of bias should also address how the choice of model (Vano et al. 2012), meteorological data (Elsner et al. 2014), or even parameterization scheme (Tague et al. 2013) affects model behavior.

Here, we examine the source of a range of parametric and structural uncertainties associated with the VIC model. We focus on parametric uncertainties associated with the scale of model resolution, potential biases in meteorological forcing variables, and structural uncertainties introduced by how the model handles watersheds that are dominated by either runoff or groundwater flow paths. We emphasize the latter because runoff- and groundwater-dominated watersheds have been shown to respond quite differently to climate change, and ensuring adequate representation of watersheds with different runoff dynamics is vital for accurate streamflow forecasting (Safeeq et al. 2013; Tague and Grant 2009; Tague et al. 2008, 2013; Waibel et al. 2013). Thus, we have two overarching questions: 1) Can we improve model accuracy by increasing topographic representation and hence theoretically better capturing hillslope-scale processes? and 2) Are model uncertainties consistent across watersheds in a geologically heterogeneous landscape such as the PNW? In this study, we focus on the VIC model because of its increasing use in water resource assessment and planning in the Pacific Northwest. However, the issue of deep-groundwater representation is not limited to VIC alone. Explicit representation of deep groundwater is not a part of any LSM and is approximated instead by extended soil profiles (Vano et al. 2012).

The VIC model conceptualizes infiltration, surface, and subsurface flow processes as occurring within a soil layer that can be made up of two or more (typically three) sublayers. The top soil layer relates to soil infiltration and surface runoff via the variable infiltration curve whereas base flow processes are controlled primarily by the lowest soil layer. The VIC model does not explicitly mimic the movement of water into and out of deep groundwater. Rather, the formulation of base flow in VIC follows the conceptual ARNO rainfall–runoff model (Todini 1996) that relates base flow as a function of soil moisture in the lowest soil layer. The base flow curve based on the ARNO model is linear under low soil moisture and becomes nonlinear toward saturation. This results in a rapid base flow response in wet conditions and a relatively slower response under dry conditions. Movement of water into and out of shallow groundwater can be fairly represented by increasing the bottom layer soil depth, hence increasing the residence time, or alternatively coupling with a groundwater flow model. The first approach can reproduce groundwater dynamics but requires site-specific calibration of bottom soil layer depth and related drainage parameters—a difficult task at the regional scale due to data limitations. On the other hand, the use of coupled surface water groundwater models is computationally inefficient and has shown limited success (Jin and Sridhar 2010; Rosenberg et al. 2013). As a result, capturing groundwater dynamics in areas with deep-groundwater-fed streams, such as the Oregon High Cascades, remains a challenge. Indeed, simulating climate change scenarios without accounting for deep-groundwater influences may lead to predictions of greater relative decline in summer streamflow (Tague et al. 2008). Such biased predictions of streamflow can potentially affect decisions and adaptation plans for future water scenarios.

The role of streamflow contributions from deep groundwater and the sensitivity of streamflow to precipitation and temperature under different geological regimes have not yet been tested for the VIC model. Some of the limitations of a regional-scale application of the VIC model surfaced in an earlier study using 55 streamflow gauges across the PNW (Wenger et al. 2010). However, the focus of this previous study was to evaluate model performance in terms of ecologically relevant flow metrics. In the present study, the objective is to quantify modeling uncertainties in hydrologic predictions due to both geological and climatic factors, with the goal of improving predictions of streamflow in the PNW and elsewhere. Specifically, we examined hydrological predictions using the VIC model in 217 watersheds located across Oregon (OR) and Washington (WA) in the PNW region of the United States. We explored uncertainties in 1) predicted total streamflow at annual and seasonal time scales, 2) five percentiles calculated based on predicted daily streamflows, and 3) predicted annual and seasonal streamflow sensitivities to a change in temperature and precipitation. Model performance evaluations at annual and seasonal time scales are useful for water resource assessment under climate change, whereas model evaluations using daily flow metrics (Olden and Poff 2003; Wenger et al. 2010) are useful for characterizing the entire hydrograph and assessing uncertainties in future ecological and in-stream flow requirements. Additionally, temperature- and precipitation-based hydrologic sensitivity metrics are useful for forecasting water resource vulnerability under climate change (Vano and Lettenmaier 2014; Vano et al. 2012). Our findings over a range of temporal scales help demonstrate under which circumstances the VIC model can be applied with confidence and point to future improvements for model predictions at the local/regional scale.

2. Methods

a. *Data*

1) OBSERVATIONS

We obtained daily streamflow time series from 217 unregulated watersheds from the U.S. Geologic Survey (UGSG; U.S. Geological Survey 2013) and the Oregon Water Resources Department (Water Resources Department 2013; Fig. 1). These watersheds are part of the USGS Hydro-Climatic Data Network (HCDN; Slack et al. 1993) and recently updated Geospatial Attributes of Gages for Evaluating Streamflow (GAGES) network (Falcone et al. 2010). Mean watershed elevation ranged from 106 to 2273 m MSL. Drainage areas for most (~75%) of the 217 watersheds were less than 500 km^2 (Fig. 2a). All selected watersheds had a minimum record length of 20 years of complete daily streamflow during the span of water years (wy) from 1950 to 2006. Among the 217 watersheds used to evaluate the model performance, 21% ($n = 45$) of the stream gauges have daily streamflow that spanned the entire 57-yr period (1950–2006) and 68% ($n = 148$) of the stream gauges have more than 30 years of streamflow record (Fig. 2b). The total number of stream gauges during any given year varied between 121 and 189; their spatial distribution was defined by data availability, with most of them located on the western side of the Cascade Mountains. A majority (~70%) of the watersheds are located between 500 and 1500 m mean elevation (Fig. 2c). The average precipitation ranges

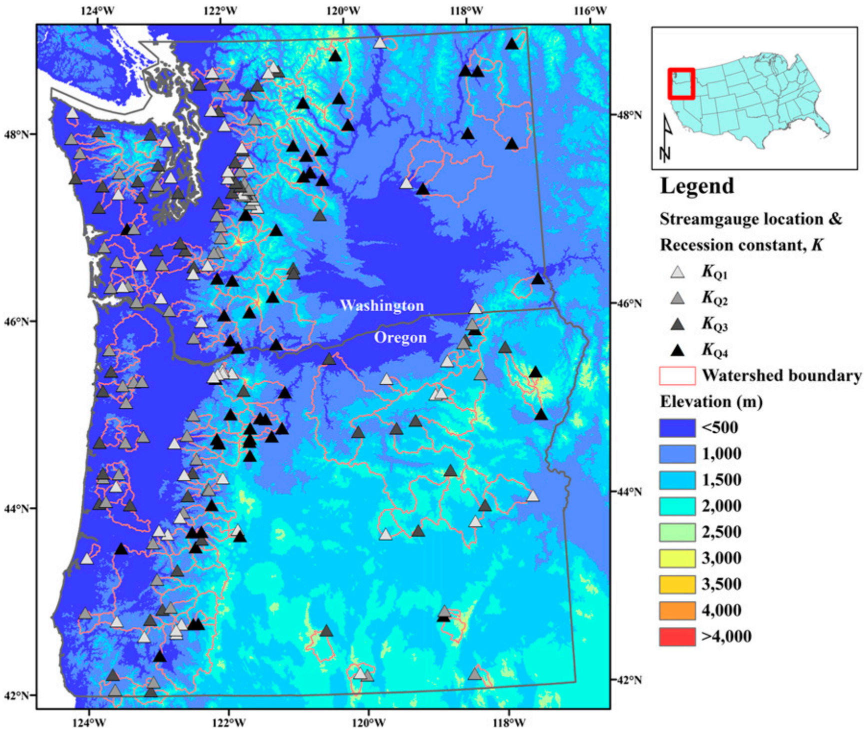

FIG. 1. Stream gauge locations (triangles) for 217 study watersheds with drainage boundaries (red lines) and shaded elevation map. Watersheds were divided into groups based on quartiles and shaded from light gray (K_{Q_1}) to black (K_{Q_4}): $0.810 \leq K_{Q_1} < 0.941$, $0.941 \leq K_{Q_2} < 0.950$, $0.950 \leq K_{Q_3} < 0.963$, and $0.963 \leq K_{Q_4} < 0.989$.

from <1 to as much as 4.5 m (Fig. 2d). The number of watersheds classified as an early (flow timing <150 or before 17 February), intermediate (150 ≤ flow timing ≤ 200), and late (flow timing >200 or after 18 April) hydrologic regime were 92, 100, and 25, respectively (Fig. 2e). Watersheds classified as an early, intermediate, and late hydrologic regime represent rain, mixture of rain and snow, and snow-dominated streams, respectively (Wenger et al. 2010).

2) VIC modeling

We used simulated streamflow from the VIC model applied at 1/16° (~6 km × 6 km) spatial resolution over the period 1950–2006 from Hamlet et al. (2013). Daily minimum and maximum temperatures, precipitation data from the National Climatic Data Center and Environment Canada, and wind speed data from reanalysis products (Hamlet and Lettenmaier 2005; Kalnay et al. 1996) were gridded at 1/16° spatial resolution from Elsner et al. (2010), based on the techniques developed by Maurer et al. (2002) and Hamlet and Lettenmaier (2005). The model was calibrated and validated on a monthly time step utilizing streamflow data from 11 major watersheds within the Columbia River basin, following the approach of Yapo et al. (1998). The Nash–Sutcliffe efficiency NSE for 11 major watersheds ranged between 0.74 and 0.89 during calibration periods and 0.68 and 0.93 during validation periods. The calibrated model parameters were further validated at 80 streamflow gauging stations in the Columbia River basin where NSE ranged from <0 to over 0.9. Further description of the model calibration and validation procedure can be found in Hamlet et al. (2013). In addition to VIC modeling at 1/16°, we also utilized simulated streamflow from the most recent VIC implementation in the Columbia River basin at a fine spatial resolution (1/120°, or about 800 m × 800 m) over the period 1950–2006. This 1/120°-resolution model was built primarily to better capture finescale snow dynamics by providing more realistic radiative forcing at local scales. However, climate, soil, and vegetation forcing variables for 1/120° model implementation were resampled from those developed at 1/16°.

The VIC model simulates infiltration, runoff, and base flow processes based on empirically derived relationships that characterize the average gridcell condition (Liang et al. 1994). To contrast simulated streamflow with observed values, we estimated simulated watershed streamflow by adding the daily runoff and base flow values from the entire VIC grid cells, both whole and

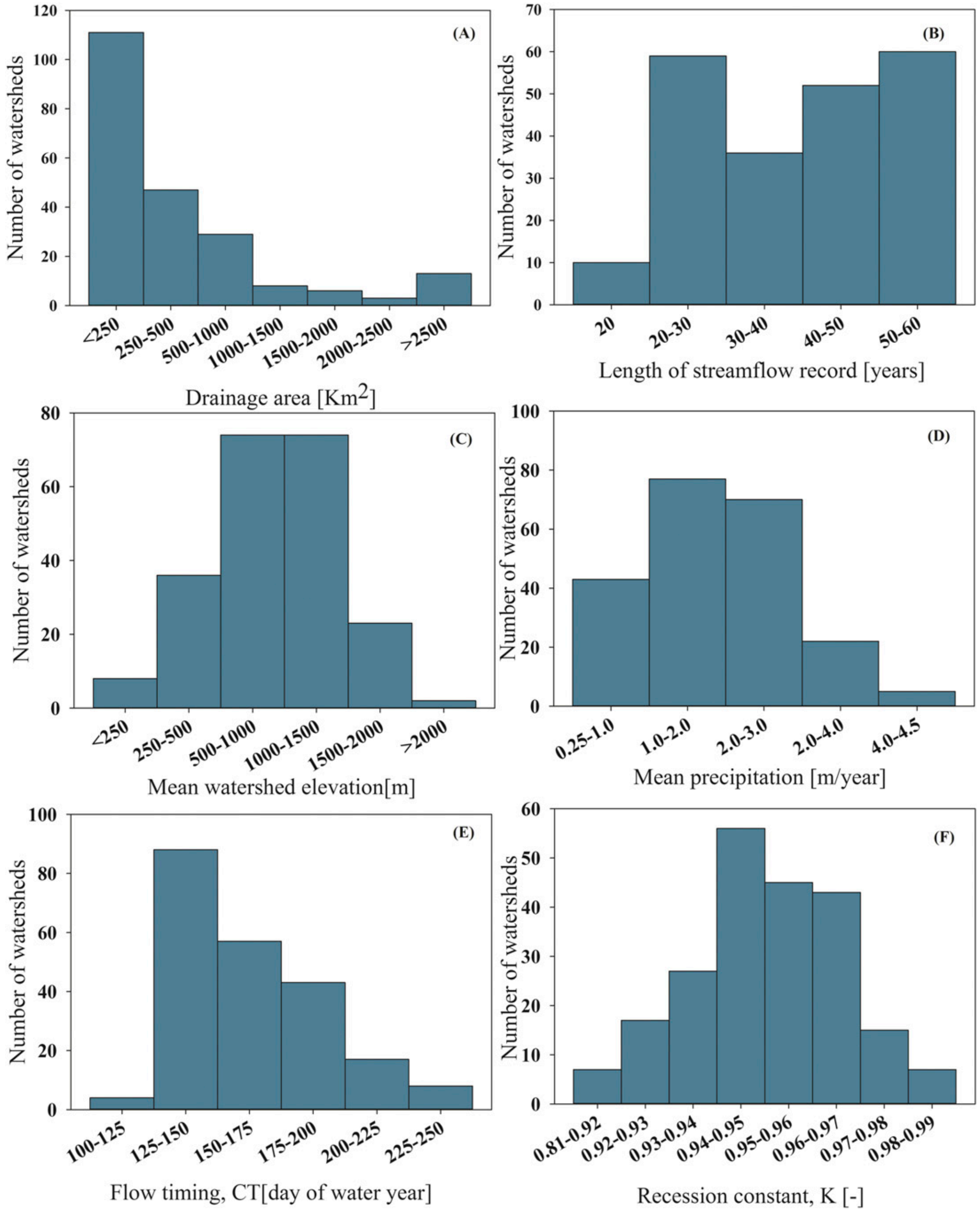

FIG. 2. Distribution of watersheds by (a) drainage area, (b) length of streamflow record, (c) mean elevation, (d) mean precipitation, (e) flow timing, and (f) recession constant.

partial cells based on area weighting, within each watershed boundary. No channel routing algorithm was employed in this analysis, and we assumed that all the runoff exits the watershed on the same day. This assumption is not an issue for comparisons of annual and seasonal streamflow but could be problematic for flow percentiles based on daily flows. However, since the majority (n = 158) of the selected watersheds are small (<500 km^2; Fig. 2a), the influence of channel routing on model performance is likely to be small compared with other modeling uncertainties.

b. Model evaluation metrics

To explore seasonal and annual biases in model performance, observed and simulated daily streamflow data were converted into time series of seasonal and annual time scales on a water-year (October–September) basis. Seasons were defined as fall [October–December

(OND)], winter [January–March (JFM)], spring [April–June (AMJ)], and summer [July–September (JAS)]. Hereafter, the total streamflows are referred to as Q_{wy} for water year and Q_{OND}, Q_{JFM}, Q_{AMJ}, and Q_{JAS} for the fall, winter, spring, and summer seasons, respectively. In addition, five streamflow percentiles were used to characterize the modeling uncertainty in matching the overall hydrologic regime of the watersheds. Both low and moderately low streamflows were characterized by the annual 5th (Q_5) and 25th (Q_{25}) percentiles, respectively. Similarly, the high and moderately high streamflows were characterized by the 75th (Q_{75}) and 95th (Q_{95}) percentiles, respectively. Annual 50th percentile values were used to characterize mean streamflow.

Uncertainties associated with the VIC-simulated streamflow were assessed by comparing the concurrent observed and simulated streamflows using four quantitative statistics for model performance: the nonparametric rank correlation coefficient ρ, the NSE, the root-mean-square error RMSE, and the percent bias PBIAS. The rank correlation is a nonparametric measure that shows the model's ability to reproduce the observed temporal patterns of interannual variability in streamflow. We used rank correlation instead of the Pearson product-moment correlation to specifically focus on evaluating the model performance in capturing interannual variability rather than the strength of the linear relationship between observed and simulated streamflow. Typically, values of Pearson correlation greater than 0.7 or a coefficient of determination greater than 0.5 are considered acceptable (Moriasi et al. 2007). Following this, we used a threshold of ρ greater than 0.7 as an acceptable model performance. The NSE is a measure of overall goodness of fit between observed and simulated data, with NSE = 1 being the optimal value (1:1 relationship). Since the value of NSE ranges between $-\infty$ and 1.0, we rescaled it between 0 and 1 and refer to it hereafter as the normalized NSE or NNSE (Nossent and Bauwens 2012). While the optimal value of NNSE remains 1, a value of 0.5 corresponds with a 0 value for the NSE. Model performance is considered satisfactory when NSE (NNSE) is greater than 0.5 (0.67) (Moriasi et al. 2007). We also used RMSE and PBIAS to quantify the magnitude of model error. RMSE provides the overall error, and PBIAS measures the average tendency of the simulated data to be larger (positive PBIAS) or smaller (negative PBIAS) than their observed counterparts (Gupta et al. 1999). The RMSE can be decomposed into its systematic component RMSE_s and unsystematic component RMSE_u using a linear regression (Willmott et al. 1985). Also known as the linear bias, RMSE_s is a measure of discrepancy between simulated and observed data caused by poor calibration, forcing errors, and/or unaccounted for processes in the model. The discrepancy between simulated and observed data caused by random processes is measured by RMSE_u. When the ratio of RMSE_s to RMSE_u is greater than one, the RMSE is largely composed of systematic bias, which can potentially be removed through calibration. However, a ratio of RMSE_s to RMSE_u less than one indicates that the RMSE is largely composed of unsystematic or random bias, and further improvement in model performance will require model and forcing refinement. Optimal values of RMSE and PBIAS are zero, indicating accurate model prediction. Model performance is considered satisfactory when PBIAS is within ±25% (Moriasi et al. 2007).

c. *Impact of climate variability on model performance*

We assess the impact of climate variability on model performance by comparing the simulated and observed streamflow sensitivities to precipitation and temperature. Following Sankarasubramanian et al. (2001), we defined the precipitation sensitivity of streamflow S_P as the percent change in total streamflow Q_t over annual and seasonal time t divided by the percentage change in annual precipitation P:

$$S_P = \mathrm{median}\left[\frac{Q_t - \overline{Q}}{P - \overline{P}}\left(\frac{\overline{P}}{\overline{Q}}\right)\right], \qquad (1)$$

where $\overline{Q}$ is the long-term sample mean of streamflow total over a period t and $\overline{P}$ is the long-term sample mean of average annual precipitation. Similarly, temperature sensitivity of streamflow S_T (% °C^{-1}) was defined as the percent change in total Q_t over a period t per unit change in average mean daily temperature T_{avg} between October and June (°C). We omit the summer months from the temperature sensitivity analysis to specifically focus on evaluating the model performance during the snow accumulation and melt period, a particularly critical set of processes influencing model performance in this region. Much of the snowfall (>90%) in this region occurs between October and May (Knowles et al. 2006). Although over 50% of the snowpack melts by the end of April, the total snowmelt period can extend until late spring at some locations (results not shown). The temperature sensitivity of streamflow S_T is given by

$$S_T = \mathrm{median}\left[\frac{Q_t - \overline{Q}}{T_{\mathrm{avg}} - \overline{T}}\left(\frac{100}{\overline{Q}}\right)\right], \qquad (2)$$

where $\overline{T}$ is the long-term sample mean of T_{avg}. Both S_P and S_T were calculated using observed and simulated

$Q_{\rm wy}$, $Q_{\rm OND}$, $Q_{\rm JFM}$, $Q_{\rm AMJ}$, and $Q_{\rm JAS}$. In the past, this sensitivity approach has been used at the annual time scale (Patil and Stieglitz 2012; Safeeq and Fares 2012; Vano et al. 2012). Comparing the additional seasonal-scale sensitivities with respect to change in annual precipitation and October–June temperature provides insight into how accurately the model represents the seasonal carryover of above (snow) and below (groundwater) land surface storage. Additionally, the metrics S_P and S_T can be used as a measure of model performance, irrespective of structural and parametric uncertainties, in characterizing streamflow under climate change.

d. Impact of deep groundwater on model performance

We used the hydrograph recession constant K as a metric for evaluating the relative contribution of deep groundwater to streamflow. This metric effectively distinguishes the relative contribution of shallow versus deep groundwater (Safeeq et al. 2013). Following Vogel and Kroll (1992), an automated recession algorithm was employed to search all 10-days-or-longer recession segments from the historical record of daily streamflow. The peak and end of each recession segment was defined as the point when the 3-day moving average of streamflow began to recede and rise, respectively. The beginning of recession (inflection point) was identified following the method of Arnold et al. (1995). To minimize the effects of snowmelt, recession segments were excluded between the onset of the snowmelt-derived streamflow pulse and 15 August. Days of snowmelt pulse onset were determined following the method of Cayan et al. (2001), with mean flow calculated for calendar days 9–248. Similar to Vogel and Kroll (1992), spurious observations were avoided by only accepting the pairs of receding streamflow (Q_t, Q_{t-1}) when $Q_t > 0.7Q_{t-1}$. The recession constant K can be given by

$$K = \exp\left[-\exp\left(\frac{1}{m}\sum_{t=1}^{m}\{\ln(Q_{t-1} - Q_t) - \ln[0.5(Q_t + Q_{t+1})]\}\right)\right], \qquad (3)$$

where m is the total number of pairs of consecutive daily streamflow Q_{t-1} and Q_t at each site. The K parameter ranges between 0 and 1, representing the lowest and highest possible groundwater contribution, respectively. Watersheds were divided into four groups based on quartiles: $0.810 \leq K_{Q_1} < 0.941$, $0.941 \leq K_{Q_2} < 0.950$, $0.950 \leq K_{Q_3} < 0.963$, and $0.963 \leq K_{Q_4} < 0.989$. Quartile technique was used for grouping over other commonly used techniques (i.e., k-means clustering) for simplicity and in an effort to keep the sample size consistent between K groups. The distribution of watersheds by K between the 217 watersheds is shown in Fig. 2f.

A summary of watershed characteristics and hydrologic conditions across all study watersheds and under different K regimes is presented in Table 1. The K_{Q_4} watersheds have higher drainage areas, base flow indices, and mean watershed elevations; lower annual precipitation; and colder temperatures ($T_{\rm avg}$) as compared to K_{Q_1}, K_{Q_2}, and K_{Q_3}. As a result of slower hydrograph recession (i.e., higher groundwater contribution), the low flow (Q_5 and Q_{25}) increases and high flow (Q_{75} and Q_{95}) diminishes between watershed groups K_{Q_1} and K_{Q_4}. The centroid of timing is nearly one month earlier in K_{Q_1} and K_{Q_2} watersheds as compared to K_{Q_4}. The model evaluation metrics under different K regimes were compared using Kruskal–Wallis one-way analysis of variance (ANOVA) on ranks. If the ANOVA revealed statistically significant differences ($p < 0.05$), a post hoc Dunn's multiple comparison test was used to determine which K regimes were different at a significance level of 0.05.

e. Impact of meteorological forcing

Since meteorological data for the VIC model are generated through spatial interpolation of irregularly spaced point measurements, this interpolation adds a potential source of uncertainty, especially at higher elevations where point measurements are scarce. To attempt to quantify this uncertainty, we compared the VIC meteorological forcing data (i.e., precipitation and temperature) and simulated snow water equivalent (SWE) against independent observations from the Natural Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) sites. Although comparing gridded data with point measurements can be somewhat misleading, it nonetheless provides useful information about the potential errors in meteorological driving data (and particularly precipitation data) at small scales. Daily precipitation, maximum and minimum temperatures, and SWE data were downloaded from 148 sites (70 in OR and 78 in WA). We also added data from three [Climatic Station at Watershed 2 (CS2MET), Primary Meteorological Station (PRIMET), and Hi-15 Meteorological Station (H15MET)] additional meteorological sites at the H.J. Andrews Experimental Forest. Many of these sites only extend from 1978 to present and do not have concurrent meteorological records. Hence, we have only included the sites with at least 10 years of data. This criterion resulted in 109 stations with daily precipitation, 34 stations with maximum temperature, 31 stations with minimum temperature, and 106 stations with daily SWE data.

TABLE 1. Summary of watershed characteristics and hydrologic conditions across all study watersheds and under different K regimes. The values reported here are the median from respective n number of watersheds. Categories: K is recession constant; P is annual precipitation; BFI is base flow index; CT is centroid of flow timing; μ_o and σ_o are observed mean and std dev during 1950–2006; μ_s and σ_s are simulated mean and std dev during 1950–2006; and S_T and S_P are temperature and precipitation sensitivities of streamflow, respectively. Notice K, BFI, drainage area, elevation, S_T, and S_P are single data points and not time series; hence, std dev (σ) are not reported.

	K_{Q_1} (n = 54)				K_{Q_2} (n = 54)				K_{Q_3} (n = 55)				K_{Q_4} (n = 54)				Regional (n = 217)			
	μ_o	σ_o	μ_s	σ_s	μ_o	σ_o	μ_s	σ_s	μ_o	σ_o	μ_s	σ_s	μ_o	σ_o	μ_s	σ_s	μ_o	σ_o	μ_s	σ_s
K	0.932	—	—	—	0.946	—	—	—	0.955	—	—	—	0.969	—	—	—	0.950	—	—	—
BFI	0.457	—	—	—	0.514	—	—	—	0.589	—	—	—	0.734	—	—	—	0.555	—	—	—
CT (day)	144	15	143	12	140	14	139	12	155	13	149	12	181	12	176	13	157	14	151	12
P (mm)	2015	333	—	—	2274	375	—	—	1819	312	—	—	1444	256	—	—	1851	319	—	—
T_{avg} (Oct–Jun; °C)	5.3	0.6	—	—	6.4	0.6	—	—	5.1	0.6	—	—	3.2	0.7	—	—	4.7	0.6	—	—
Drainage area (km^2)	59	—	—	—	215	—	—	—	405	—	—	—	428	—	—	—	227	—	—	—
Elev (m)	854	—	—	—	778	—	—	—	943	—	—	—	1270	—	—	—	948	—	—	—
Streamflow total (mm)																				
Q_{OND}	487	202	386	148	486	220	495	180	345	162	316	122	173	70	164	70	345	156	326	134
Q_{JFM}	549	179	536	166	695	208	721	199	507	183	415	151	173	71	195	78	462	157	476	153
Q_{AMJ}	247	87	279	76	272	87	295	79	323	95	275	83	371	95	388	111	318	93	314	88
Q_{JAS}	39	20	70	22	53	18	68	23	70	28	67	24	123	34	79	40	71	25	72	27
Q_{wy}	1514	349	1355	296	1631	383	1689	361	1453	345	1345	300	956	228	927	227	1311	320	1335	294
Flow percentile (mm)																				
Q_5	0.19	0.07	0.34	0.09	0.32	0.09	0.33	0.10	0.42	0.10	0.35	0.08	0.60	0.13	0.27	0.05	0.39	0.10	0.30	0.08
Q_{25}	0.56	0.23	0.75	0.24	0.72	0.25	0.83	0.23	0.78	0.23	0.73	0.15	1.01	0.23	0.56	0.22	0.78	0.23	0.76	0.22
Q_{50}	4.14	0.96	3.71	0.81	4.46	1.05	4.62	0.99	3.98	0.94	3.68	0.82	2.62	0.62	2.54	0.62	3.59	0.88	3.66	0.80
Q_{75}	5.19	1.36	5.09	1.24	5.69	1.63	6.32	1.48	4.97	1.33	4.90	1.24	3.19	0.83	3.60	0.97	4.48	1.23	4.90	1.24
Q_{95}	14.69	3.45	10.75	2.62	14.44	4.02	13.84	2.85	11.97	3.22	10.90	2.53	7.01	2.08	7.97	1.99	11.22	2.96	10.58	2.43
S_T (% °C^{-1})																				
Q_{OND}	−0.98	—	−4.36	—	−3.09	—	−4.06	—	−3.01	—	−3.38	—	1.84	—	−3.89	—	−1.07	—	−4.11	—
Q_{JFM}	−8.43	—	−4.69	—	−11.73	—	−9.60	—	−8.92	—	−7.42	—	−3.60	—	−4.82	—	−8.19	—	−6.62	—
Q_{AMJ}	−1.40	—	−0.91	—	−0.17	—	0.36	—	−0.54	—	−2.17	—	1.26	—	0.95	—	−0.34	—	−0.46	—
Q_{JAS}	0.96	—	1.12	—	3.92	—	1.01	—	−1.43	—	0.36	—	−3.44	—	−7.64	—	−0.76	—	−0.66	—
Q_{wy}	−2.51	—	−1.34	—	−3.41	—	−4.41	—	−2.93	—	−2.51	—	−1.13	—	−0.92	—	−2.60	—	−2.22	—
S_P (% %$^{-1}$)																				
Q_{OND}	1.42	—	1.38	—	1.61	—	1.54	—	1.55	—	1.38	—	1.13	—	1.44	—	1.42	—	1.44	—
Q_{JFM}	1.38	—	1.40	—	1.29	—	1.31	—	1.42	—	1.30	—	1.39	—	1.17	—	1.38	—	1.31	—
Q_{AMJ}	0.81	—	0.87	—	0.67	—	0.81	—	0.89	—	0.95	—	1.01	—	1.02	—	0.87	—	0.90	—
Q_{JAS}	1.41	—	0.88	—	1.00	—	1.11	—	1.18	—	0.93	—	1.25	—	2.06	—	1.19	—	1.19	—
Q_{wy}	1.30	—	1.25	—	1.32	—	1.23	—	1.34	—	1.30	—	1.25	—	1.31	—	1.30	—	1.27	—

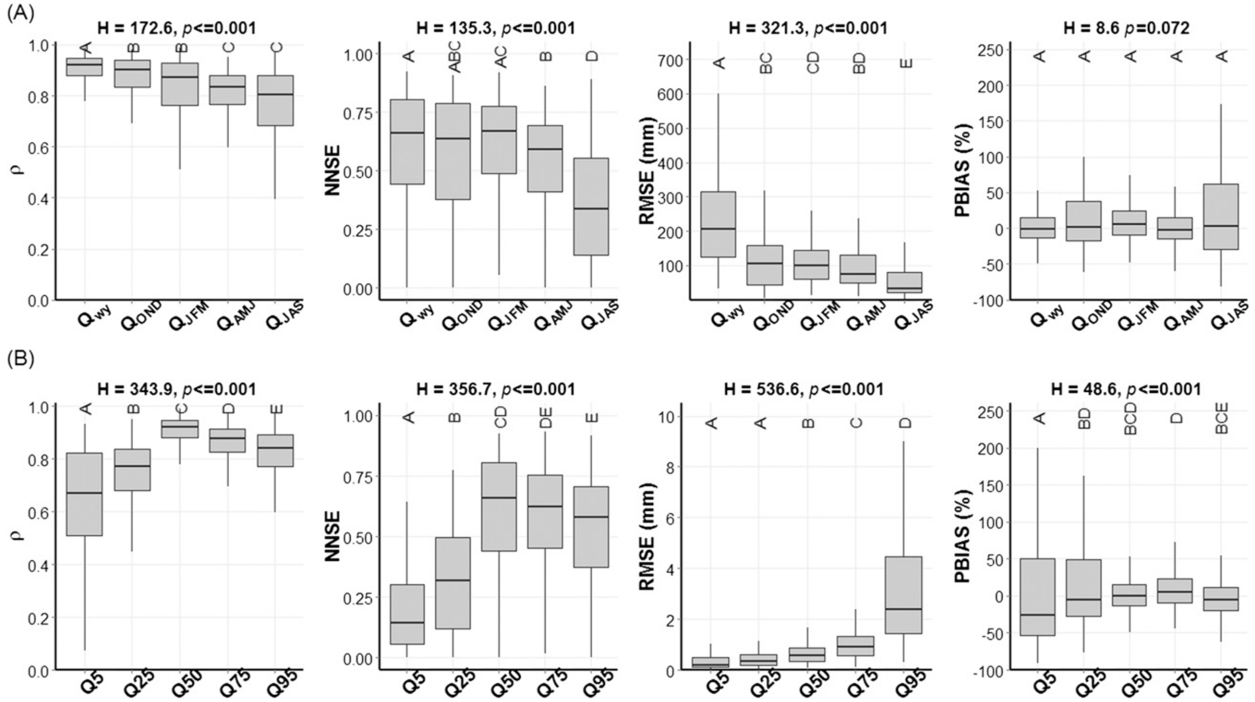

FIG. 3. Model performance metrics (ρ, NNSE, RMSE, and PBIAS) for predicting (a) total streamflow and (b) flow percentiles at 1/16° spatial resolution. The line inside the box represents the median value, the box itself represents the interquartile range (IQR; 25th–75th percentile range), and the whiskers are the lowest and highest values that are within 1.5(IQR) of the 25th and 75th percentiles. The Kruskal–Wallis rank sum statistic H and corresponding probability value of the test p are shown at the top. Model performances denoted with same letters are not significantly different (Kruskal–Wallis and post hoc Dunn's test, $p < 0.05$).

3. Results

a. *Total streamflow and flow percentiles*

The VIC model performed well in capturing the interannual variability in observed annual and seasonal total streamflow, as measured by rank correlation coefficients (i.e., ρ; Fig. 3a). The median value of ρ from 217 individual watersheds was the highest for the annual time scale and diminished as seasons progressed from fall to summer (Table 2). Additionally, the interquartile range of ρ from individual watersheds was smallest for Q_{wy} and relatively large for Q_{JFM} and $Q_{\mathrm{JAS},}$ showing greater variability in model performance between watersheds in the latter two cases. The percentage of watersheds with $\rho > 0.7$ decreased from 96% to 73% for Q_{wy} and Q_{JAS}, respectively. This indicates that in 27% of the watersheds, interannual variability in Q_{JAS} was not satisfactorily captured by the model. The model performed better for high-flow percentiles as compared to low flows (Fig. 3b): the median value of ρ ranged from 0.92 for Q_{50} to 0.67 for Q_5 (Table 2). The interquartile range was small for Q_{50} and increased toward the more extreme flows, increasing more for low- than high-flow percentiles (Fig. 3b). For example, the model performed poorly ($\rho \leq 0.7$) in 29% of watersheds for Q_{25} and in 54% of watersheds for Q_5, whereas only 4% of watersheds had similarly low correlations for Q_{50}, 7% for Q_{75}, and 14% for Q_{95}. This indicates that the model consistently performed better in predicting the interannual variability in mean and high flows across all selected watersheds but underperformed in predicting low flows.

Rank correlations are useful for evaluating model sensitivities in terms of interannual variability but do not provide information on absolute model biases. In contrast, NNSE along with RMSE and PBIAS provide overall goodness of fit between observed and simulated hydrographs. The percentage of total watersheds with NNSE below 0.67 ranged from 50% for Q_{wy} to 90% for Q_{JAS}, indicating strong disagreement between modeled and observed flows in a large number of watersheds. Similarly, strong disagreements between modeled and observed low-flow percentiles were also found (Table 2). The NNSE was below the 0.67 threshold in 93% of watersheds for Q_{25} and in 95% of watersheds for Q_5, whereas only 51% of watersheds had similarly low NNSE for Q_{50}, 55% for Q_{75}, and 65% for Q_{95}. As compared to ρ, the NNSE values were lower for both total flow and flow percentiles, indicating a systematic absolute model bias. This was confirmed from the

TABLE 2. Median values of the model performance statistics at 1/16° and 1/120° spatial resolution in predicting total streamflow and flow percentiles from 217 watersheds across OR and WA. The significant differences (Wilcoxon rank-sum test, $p < 0.05$) between 1/16° and 1/120° are marked with an asterisk.

Streamflow	ρ	NNSE	RMSE (mm)	RMSE$_s$ (mm)	RMSE$_u$ (mm)	RMSE$_s$/RMSE$_u$	PBIAS (%)
1/16°							
Q_{wy}	0.92	0.66	206.67	142.99	105.77	1.70	−0.55*
Q_{OND}	0.90	0.64	105.30	72.29	51.73	1.70	1.38
Q_{JFM}	0.87	0.67	99.18	60.32	67.30	1.16	5.17
Q_{AMJ}	0.84	0.59	75.74	55.58	45.13	1.38	−2.98*
Q_{JAS}	0.80	0.34	32.79	26.08	15.12	2.47	3.01
Q_5	0.67	0.14	0.20	0.19	0.05	4.91	−25.92
Q_{25}	0.77	0.32	0.33	0.25	0.12	2.54	−5.26
Q_{50}	0.92	0.66	0.57	0.39	0.29	1.71	−0.55*
Q_{75}	0.88	0.62	0.92	0.65	0.54	1.52	5.14*
Q_{95}	0.84	0.58	2.38	1.69	1.24	1.64	−4.85
1/20°							
Q_{wy}	0.92	0.65	221.51	157.62	105.19	1.83	3.01*
Q_{OND}	0.90	0.62	109.81	77.09	49.99	1.80	−4.03
Q_{JFM}	0.86	0.65	106.54	62.28	70.31	1.16	1.83
Q_{AMJ}	0.85	0.56	82.25	56.59	48.89	1.46	11.14*
Q_{JAS}	0.82	0.32	35.44	29.04	14.76	2.39	11.71
Q_5	0.70	0.18	0.20	0.20	0.06	3.70	−13.89
Q_{25}	0.79	0.28	0.36	0.30	0.12	2.42	3.16
Q_{50}	0.92	0.65	0.61	0.43	0.29	1.83	3.01*
Q_{75}	0.88	0.59	1.01	0.76	0.57	1.64	12.88*
Q_{95}	0.84	0.55	2.45	1.75	1.30	1.72	−2.32

proportionally higher RMSE$_s$ values as compared to RMSE$_u$ (Table 2). Although, the median RMSE was large for Q_{wy} (206.7 mm) and Q_{95} (2.5 mm), the corresponding PBIAS was small. The median absolute PBIAS was very good (<6%) for all total streamflows and flow percentiles except for Q_5 (PBIAS = −26%). Although median PBIAS was satisfactory in the majority of cases, the range was quite variable and, in some cases, absolute PBIAS was larger than 25% (Fig. 3b). The percentage of watersheds with absolute PBIAS > 25% in predicting total flow increased from 23% for Q_{wy} to 66% for Q_{JAS}. Similarly, the percentage of watersheds with absolute PBIAS > 25% in predicting flow percentiles increased from 23% for Q_{50} to 82% for Q_5.

Increasing model spatial resolution from 1/16° to 1/120° alone resulted in no improvement in model performance (Fig. S1 in the supplemental material). This can be primarily attributed to the fact that both simulations were driven by the same vegetation and soil parameterization. Additionally, effects of improved small-scale snow dynamic and evaporation as a result of more realistic radiative forcing at 1/120° scale as compared to 1/16° may not be apparent at the watershed scale. Based on the Wilcoxon rank-sum test, which is a nonparametric method to test if the two population distributions are the same, the effect of model resolution was only significant ($p < 0.05$) for PBIAS in Q_{wy}, Q_{AMJ}, Q_5, Q_{50}, and Q_{75}. As compared to the 1/16° representation, the 1/120°-resolution interquartile range in flow percentiles shifted toward more positive PBIAS. In the case of Q_5, there was a 50% reduction in median PBIAS under 1/120° simulations (median PBIAS = −13%) as compared to 1/16° (median PBIAS = −25%). A similar change was also observed for Q_{wy} and Q_{AMJ}. For example, the median PBIAS in Q_{AMJ} increased from an underestimation (−3%) to an overestimation (11%) under 1/16° and 1/120° spatial resolution, respectively (Table 2). Although some differences in the model performance were statistically significant, there was no clear or substantial improvement from the finer-resolution simulation. Therefore, we only present the 1/16° VIC modeling compared to the observed values unless otherwise noted.

b. Precipitation and temperature sensitivities

The comparisons of average observed and simulated precipitation sensitivity (i.e., S_P) across the 217 watersheds showed consistently low correspondence, with only 18% of the variance in S_P explained by the model (Fig. 4a). The median sensitivity of observed streamflow to precipitation (i.e., S_P) across all 217 watersheds ranged from 0.87 in spring to 1.42 in fall (Table 1). In other words, an increase of annual precipitation by 1% resulted in a 1.42% increase in Q_{OND}. In comparison to observed S_P, the median S_P derived from simulated streamflow ranged from 0.90 during spring to 1.44 in fall. On a regional scale, the simulated median S_P across the

217 watersheds were comparable to the observed S_P in all seasons and water years (Table 1). However, at the individual watershed level, the model largely underpredicts S_P, particularly in watersheds with observed $S_P > 2.0$ (Fig. 4a). This disagreement between observed and simulated S_P was within ±25% in the majority (50%–76%) of watersheds for annual, fall, and winter streamflows. In the spring and summer, only 43% and 26% of the watersheds had results within a ±25% error between simulated and observed S_P. These results indicate that model performance in capturing S_P was even lower for summer as compared to other seasons. Although the model performed less well for precipitation (Fig. 4), uncertainties about the magnitude and direction of future precipitation changes (Mote and Salathé 2010) make this aspect less critical in the PNW, but it may be a factor when modeling other regions.

Future changes in temperature are more certain in this region, and model performance in capturing S_T was improved as compared to S_P across all time scales as inferred by the higher coefficient of regression between observed and simulated S_T (Fig. 4b). However, overall model performance in simulating S_T was unsatisfactory, with only 33%–45% of the variance in observed S_T explained by the VIC model. The model largely underpredicted S_T in 50%–60% watersheds and overpredicted in 25%–28% watersheds by at least 25% across all seasons. The median sensitivity of observed streamflow to temperature S_T across all watersheds ranged from -8.19 (% °C^{-1}) during winter to -0.34 (% °C^{-1}) in spring (Table 1). The negative S_T value indicates a decline in streamflow with increasing T_{avg}. An increase in T_{avg} by 1°C will result in as much as nearly an 8% decline in Q_{JFM} and 0.34% decline in Q_{AMJ}. On the annual time scale, an increase in T_{avg} by 1°C will result in a 2.6% decline in observed Q_{wy} and a 2.2% decline in simulated Q_{wy}. Although there was no major difference in model performance across the seasons and water year in simulating S_T based on NNSE and RMSE values (Fig. 4b), the median simulated S_T for Q_{OND} was significantly higher (Table 1).

c. *Impact of deep groundwater on model performance*

1) TOTAL STREAMFLOW AND FLOW PERCENTILES

Performance metrics among K_{Q_1}, K_{Q_2}, K_{Q_3}, and K_{Q_4} watersheds revealed differences in model performance in different geological terrains (Fig. 5). In 80% of possible cases, the Kruskal–Wallis and post hoc Dunn's multiple comparison test showed statistically significant differences in model performance metrics for total streamflow and flow percentiles between two or more watershed groups (Table S1 in the supplemental material). The model performance in capturing interannual variability in Q_{OND} and Q_{JFM} based on ρ was significantly lower ($p < 0.05$) in K_{Q_4} as compared to K_{Q_1} watersheds. However, the opposite was true for Q_{AMJ}, where the model performed significantly ($p < 0.05$) better in K_{Q_4} as compared to K_{Q_1} watersheds. In terms of flow percentiles, the difference in ρ was only significant between K_{Q_1} and K_{Q_2} watersheds for Q_{25}, Q_{50}, and Q_{75}, where the model performed better in the latter case. For Q_5 the model performed significantly better in K_{Q_2} as compared to K_{Q_1}, K_{Q_3}, and K_{Q_4}. The NNSE in K_{Q_1} was significantly lower ($p < 0.05$) for Q_{OND} and Q_{wy} as compared to K_{Q_2} and K_{Q_3}, respectively. Similarly, the NNSE in K_{Q_4} was significantly lower for Q_{OND} and Q_{JFM} as compared to K_{Q_2} and K_{Q_3}, respectively. The model performed poorly in K_{Q_1} as compared K_{Q_3} watersheds for Q_{25}, Q_{50}, and Q_{95}. The RMSE was significantly lower for K_{Q_4} as compared to K_{Q_1} or K_{Q_2} during Q_{wy}, Q_{OND}, and Q_{JFM}, which is not surprising given the overall lower flow during these seasons in K_{Q_4} watersheds (Table 1). Similarly, the RMSE values for Q_{50}, Q_{75}, and Q_{95} in K_{Q_1} were significantly higher as compared to K_{Q_4} watersheds. However, despite overall higher Q_{AMJ} and Q_{JAS} in K_{Q_4} watersheds (Table 1), there was no statistical difference in RMSE values during these seasons among the different groups of watersheds. This was unexpected given that deep groundwater, which was not explicitly modeled by VIC, exerts a greater influence on Q_{AMJ} and Q_{JAS} as compared to streamflow in other seasons. Considering the model limitation, the RMSE values in Q_{AMJ} and Q_{JAS} for K_{Q_4} were expected to be higher than the K_{Q_1} watersheds. However, although not statistically significant, the corresponding RMSE values in K_{Q_1} and K_{Q_4} are slightly higher as compared to K_{Q_2} and K_{Q_3} watersheds. This pattern seems to be consistent in terms of ρ and NNSE as well with the model performing better overall in K_{Q_2} and K_{Q_3} as compared to K_{Q_1} and K_{Q_4} watersheds.

The effect of deep groundwater in simulating Q_{JAS} and extreme flow percentiles (i.e., Q_5, Q_{25}, and Q_{95}) was most evident in terms of PBIAS. The model significantly ($p < 0.05$) overpredicted (median PBIAS = 48%) Q_{JAS} in K_{Q_1} and underpredicted (median PBIAS = -13%) in K_{Q_4} watersheds. Similarly, Q_5 was significantly overpredicted (median PBIAS = 19%) in K_{Q_1} and underpredicted (median PBIAS = -51%) in K_{Q_4} watersheds (Fig. 5). In contrast, Q_{95} was significantly underpredicted (median PBIAS = -17%) in K_{Q_1} and overpredicted (median PBIAS 11%) in K_{Q_4} watersheds. These results indicate that the model performance was significantly influenced by the absence/presence of groundwater and that base flow recedes quickly in groundwater-dominated

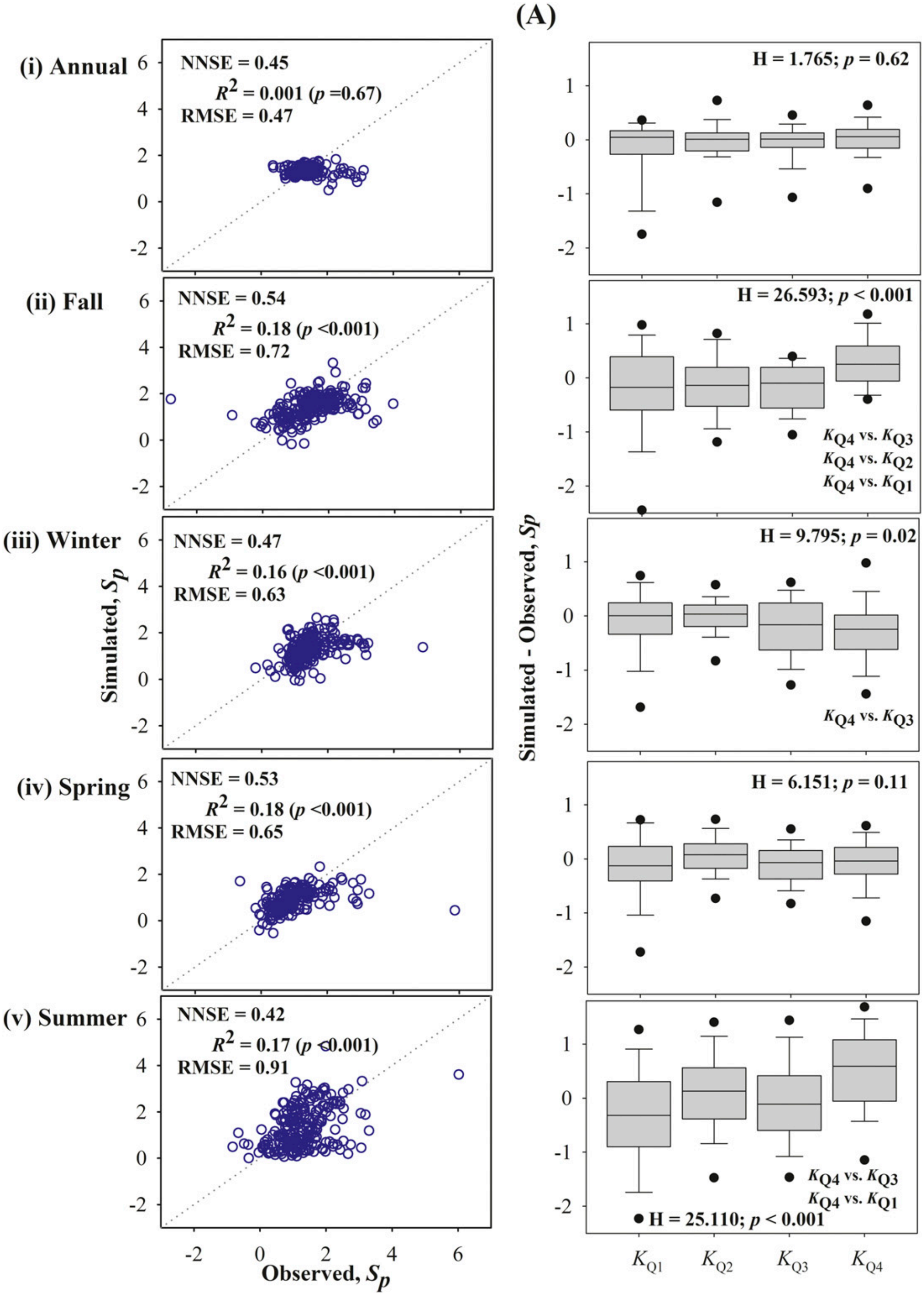

FIG. 4. Model performance at the 1/16° spatial resolution in predicting annual and seasonal streamflow sensitivity to (a) annual precipitation S_P and (b) October–June temperature S_T. The corresponding box plots show the difference between simulated and observed S_p in (a) and S_T in (b) for different watersheds grouped based on K values. The line inside the box represents the median value, the box itself represents the IQR, the whiskers are the lowest and highest values that are within 1.5(IQR) of the 25th and 75th percentiles, and the dots represent 5th and 95th percentiles. The Kruskal–Wallis rank sum statistic (i.e., H) and corresponding probability value of the test (p) are shown with the box plots along with the significant (Kruskal–Wallis and post hoc Dunn's test, $p < 0.05$) K groups.

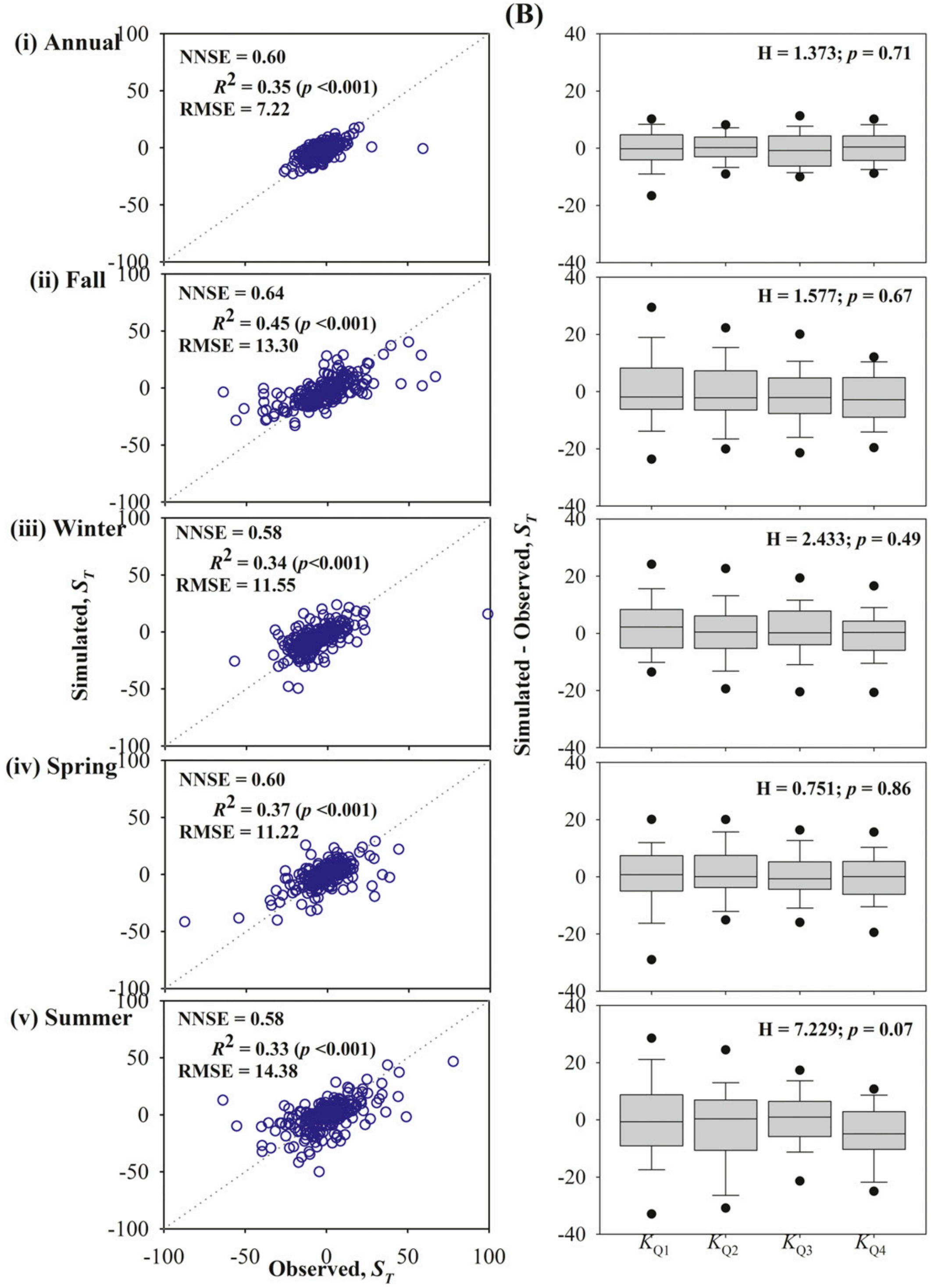

FIG. 4. (*Continued*)

watersheds and slowly in runoff-dominated watersheds. Although the magnitude of error (both RMSE and PBIAS) was comparable among the different groups of watersheds, a systematic shift in the direction of error (over- or underestimation) based on groundwater influence could be problematic.

The spatial pattern among the watersheds with PBIAS $< -25\%$ in Q_{JAS} and Q_5 was consistent with

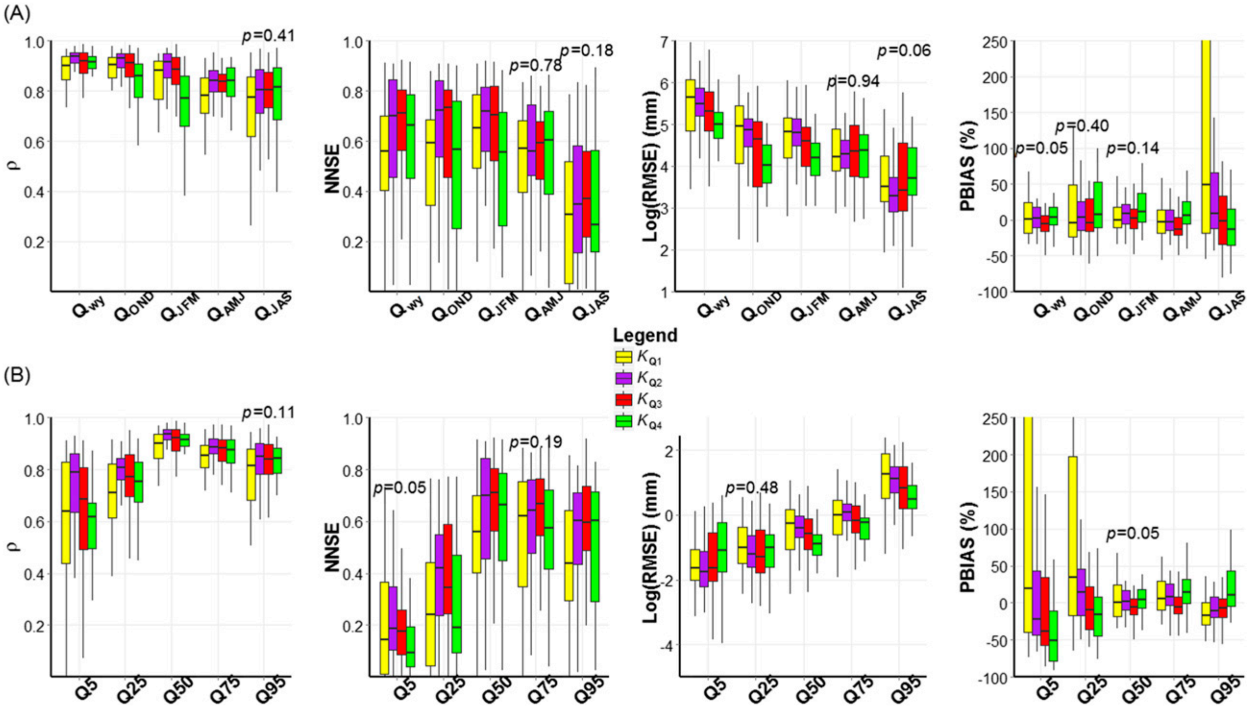

FIG. 5. Effect of deep groundwater (K) on model performance at the 1/16° spatial resolution in predicting (a) total streamflow and (b) flow percentiles. The line inside the box represents the median value, the box itself represents the IQR, and the whiskers are the lowest and highest values that are within 1.5(IQR) of the 25th and 75th percentiles. Model performances are significantly different between different groups of K (Kruskal–Wallis and post hoc Dunn's test, $p < 0.05$) unless denoted with p values (note that complete ranges of PBIAS values are shown in Fig. S2 in the supplemental material).

geological terrains in the PNW (Fig. 6). Most of the watersheds with PBIAS $< -25\%$ had high K values and were located along the Cascades. This was not surprising given the fact that most of these watersheds are sourced from the High Cascades. However, there were watersheds on the Olympic Peninsula in WA and around the Wallowa Mountains in OR with PBIAS $< -25\%$. These places did not have deep-groundwater systems as in the Cascades, but they sustained relatively higher summer base flows, presumably because of late-melting snowpacks.

2) Precipitation and temperature sensitivities

Comparisons of observed and simulated S_P showed sharp differences in model performance between different groups of watersheds based on K (Fig. 4). The error in S_P, calculated as the difference between simulated and observed S_P, was significantly different ($p < 0.05$) for Q_{OND} and Q_{JAS} in K_{Q_4} (groundwater dominated) as compared to K_{Q_1} (runoff dominated) watersheds. The model significantly overpredicted S_P for Q_{OND} and Q_{JAS} in K_{Q_4} as compared to K_{Q_1}. We did not see any significant influence of deep groundwater on model performance in terms of S_T (Fig. 4b). This indicates that model performance was not influenced by the presence/absence of groundwater in terms of temperature sensitivity of streamflow. However, S_P during fall and summer season streamflow was strongly influence by the presence/absence of groundwater.

d. *Uncertainty in model meteorological forcing*

Comparing gridded (1/16° resolution) meteorological forcing and simulated SWE to point measurements at SNOTEL sites revealed that gridded precipitation compared reasonably well with measurements at seasonal and annual time scales, with average ρ and NNSE values larger than 0.67, except during spring (Table 3). Although average RMSE values ranged between 39 mm during summer and 209 mm on the annual time scale, average PBIAS remained <2%. The gridded precipitation values were slightly higher than measured values during fall, spring, and summer and lower during winter. The average RMSE for annual and seasonal precipitation (Table 3) was comparable to those values for total streamflow (Table 2). However, the PBIAS in precipitation and total streamflow at the seasonal time scale did not agree. For example, the model overpredicted winter flows despite negative PBIAS in winter

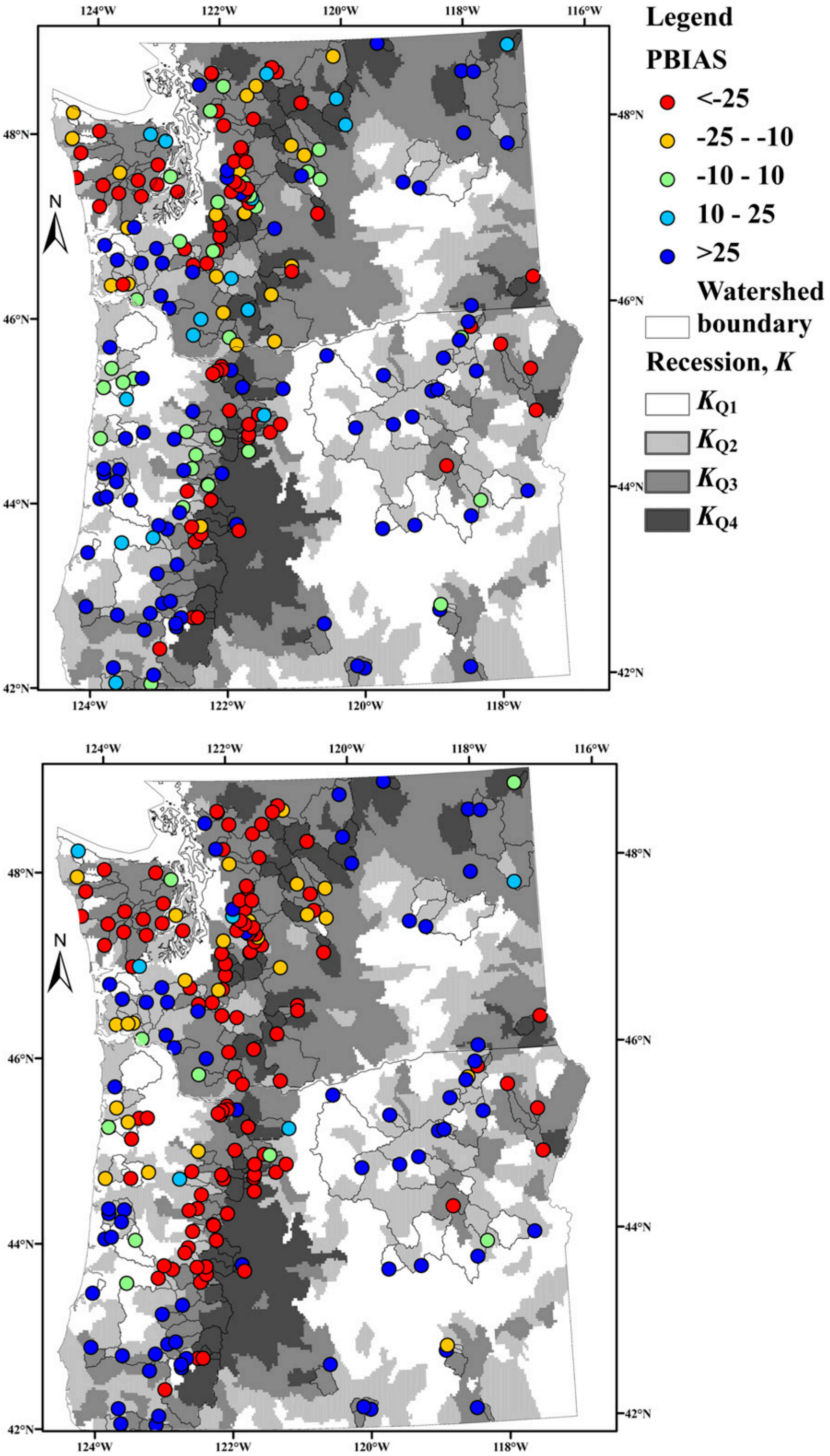

FIG. 6. Spatial variability in predicting (top) summer streamflow (i.e., Q_{JAS}) and (bottom) 5th flow percentile (i.e., Q_5) at $^1/_{16}$° spatial resolution. Recession constant K is indicated by gray shading at 5th field hydrologic unit code scale; K data are from Safeeq et al. (2014).

TABLE 3. Summary of the performance statistics showing the agreement between 1/16° spatial resolution gridded precipitations, max and min temperatures, and model-simulated SWE with those measured from SNOTEL sites across OR and WA.

Performance statistics	wy	Fall	Winter	Spring	Summer
Precipitation (n = 109)					
ρ	0.86	0.88	0.83	0.73	0.77
NNSE	0.68	0.75	0.69	0.61	0.70
RMSE (mm)	208.51	100.80	106.43	63.07	38.61
RMSE_s (mm)	151.44	65.48	68.84	38.45	26.99
RMSE_u (mm)	130.59	71.73	75.41	47.25	26.67
$\mathrm{RMSE}_s/\mathrm{RMSE}_u$	1.22	0.96	1.00	0.88	1.05
PBIAS (%)	−0.33	0.70	−1.57	0.78	0.23
Max temp (n = 34)					
ρ	0.63	0.66	0.70	0.81	0.77
NNSE	0.12	0.26	0.32	0.21	0.16
RMSE (°C)	1.87	2.36	2.04	2.50	2.67
RMSE_s (°C)	1.81	2.19	1.75	2.41	2.58
RMSE_u (°C)	0.39	0.73	0.83	0.55	0.56
$\mathrm{RMSE}_s/\mathrm{RMSE}_u$	4.78	3.13	2.36	4.84	5.42
PBIAS (%)	14.57	38.11	40.56	15.35	10.55
Min temp (n = 31)					
ρ	0.62	0.56	0.64	0.75	0.56
NNSE	0.10	0.33	0.35	0.22	0.13
RMSE (°C)	1.66	1.86	1.76	2.14	2.49
RMSE_s (°C)	1.61	1.71	1.55	2.08	2.40
RMSE_u (°C)	0.34	0.65	0.73	0.42	0.56
$\mathrm{RMSE}_s/\mathrm{RMSE}_u$	4.91	2.68	2.26	4.73	4.35
PBIAS (%)	13.94	−57.71	40.59	−89.07	−23.79
SWE (n = 106)					
ρ	0.71	0.72	0.71	0.71	—
NNSE	0.21	0.22	0.25	0.04	—
RMSE (mm)	142.25	70.40	267.15	233.40	—
RMSE_s (mm)	131.75	66.53	246.55	216.14	—
RMSE_u (mm)	38.16	17.37	79.71	60.14	—
$\mathrm{RMSE}_s/\mathrm{RMSE}_u$	6.06	4.83	5.10	121.13	—
PBIAS (%)	−35.18	−43.30	−37.95	36.92	—

precipitation, while the opposite (underprediction) was true for the spring season. This was not surprising given the seasonal carryover of precipitation in the form of groundwater and SWE. As opposed to RMSE in streamflow (Table 2), the RMSE in precipitation was equally composed of both systematic and unsystematic components (Table 3). The comparisons of gridded and measured maximum and minimum temperatures showed strong bias with NNSE values less than 0.67 and RMSE ranging between 1.7° and 2.7°C (Table 3). As opposed to precipitation, RMSE in temperatures were mostly systematic.

The average ρ between observed and simulated SWE was higher than the acceptable threshold of 0.7 for good model performance in terms of ρ (Table 3). However, average NNSE values were all below the 0.6 threshold for good model performance in terms of NNSE at all four time scales, indicating large absolute bias. The model underpredicted the average SWE during fall and winter and overpredicted average SWE during spring. Since most of the SNOTEL sites used in this study are located at elevations above the average elevation of K_{Q_3} (90%) and K_{Q_4} (66%), it was not possible to disentangle the role of bias in SWE to streamflow between runoff- and groundwater-dominated watersheds. However, large underpredictions of Q_{JAS} in groundwater-dominated watersheds (Fig. 5), despite an overprediction of spring SWE and higher gridded spring and summer precipitation as compared to those measured at SNOTEL sites (Table 3), provide confirmation of the importance of considering groundwater contributions when simulating streamflow.

4. Discussion

Our analysis of VIC model performance, comparing time series of total streamflow and flow percentiles of observed and simulated streamflow, revealed both strengths and weaknesses in the model that are important to understand for successful model applications. The model performed reasonably well in capturing interannual variability in observed streamflow (as reflected by relatively high ρ), which gives confidence in using the model to estimate and simulate hydrologic trends for climate change assessments (Hamlet et al. 2007; Hidalgo et al. 2009). However, performance was poorer in predicting the magnitude and interannual variability in observed low flows (i.e., Q_5 and Q_{25}) and total summer streamflow (i.e., Q_{JAS}). This poorer performance could be problematic, given the importance of summer flows for aquatic organisms (Arismendi et al. 2013; Beer and Anderson 2013) and municipal water supply (Barnett et al. 2005). Although routing was not explicitly included in streamflow simulation, we explored the possible effect of routing on model performance (Fig. 3). We found that in large watersheds (drainage area >500 km^2), where lack of routing was expected to affect model performance the most, both low flow (Q_5) and peak flow (Q_{95}) were better predicted than in small watersheds (drainage area <500 km^2). The effect of routing was evident on Q_{25}, where the model performance (median PBIAS = 12%) deteriorated in large watersheds as compared to small watersheds (median PBIAS = −7%). However, since the majority of our studied watersheds were <500 km^2 (Fig. 2a), the effect of routing on model performance for our study sites is expected to be minimal. We also found no improvement in model performance for low-flow metrics (Q_5 and Q_{25}) as a result of site-specific calibration and validation. However, the peak flows (Q_{95}) in calibrated watersheds were better predicted. This indicates that reducing the model uncertainty from low flows will require a different strategy for model calibration. Not only was VIC calibrated at the monthly time scale, the NSE, which was used to calibrate the VIC model, is biased toward the peak-flow portion of the hydrograph rather than the low-flow portion.

The strong disagreement between observed and simulated sensitivities to changes in both precipitation and temperature does not necessarily imply poor model performance. As shown by Elsner et al. (2014), the choice of gridded meteorological dataset can influence the streamflow sensitivity to changes in climate. In this study, observed sensitivities were calculated using gridded precipitation and temperature data. However, small-scale variability in precipitation and temperature may not be accurately captured in these datasets, leading to biased estimates of observed sensitivities (S_P and S_T). This alone could cause substantial disagreement between observed and simulated sensitivities, more so for precipitation than temperature. Thus, higher uncertainties in precipitation as compared to temperature sensitivities are not surprising (Fig. 4). Although quantifying the error associated with observed sensitivities is challenging in the absence of a landscape-level-independent record of precipitation and temperature variability, it would be interesting to look at how the choice of gridded meteorological data affects observed S_P and S_T.

The model showed strong systematic bias in both runoff- and groundwater-dominated watersheds, especially total summer streamflow and low-flow percentile. Although the magnitudes of RMSE in runoff- and groundwater-dominated watersheds are comparable for both Q_5 and Q_{JAS}, the difference in PBIAS was statistically significant. Most importantly, the contrasting (positive and negative) PBIAS error in the prediction of Q_5 and Q_{JAS} highlights the varying level of uncertainty in model predictions between runoff- and groundwater-dominated watersheds. In groundwater-dominated watersheds, PBIAS in Q_5 and Q_{JAS} shows an overall underestimation as opposed to an overestimation in runoff-dominated watersheds. In general, the model tends to perform better in intermediate (i.e., K_{Q_2} and K_{Q_3}) across all model performance metrics. This is somewhat contrary to the findings of Wenger et al. (2010), who showed diminishing model performance with increasing groundwater contribution. However, considering the fact that the model neither accurately captures the groundwater dynamics nor is capable of producing zero (or near zero) flows, corresponding under- and overestimation of low-flow percentiles and total summer streamflow are not surprising. In either extreme case where streams either gain water from inflow of groundwater (e.g., K_{Q_4} watersheds) or lose water by outflow to groundwater (e.g., K_{Q_1} watersheds), base flow recession was not accurately captured by the model. Under these circumstances, increasing the model resolution alone may theoretically provide better topographic representation of the landscape but not necessarily improve model performance. On the other hand, coupling a groundwater model to VIC is not a feasible option because of both huge computational resources and time requirements and because of the extensive amount of data needed for groundwater parameterization. Most of these data are typically not available at a regional scale. Also, as pointed out by Wenger et al. (2010), these finescale biases become less important at the larger scale when extreme hydrogeological systems (i.e., K_{Q_1} and K_{Q_4}) mix with intermediate systems (i.e., K_{Q_2} and K_{Q_3}).

Although comparisons of meteorological data forcing and simulated SWE at SNOTEL sites apparently show large discrepancies, these results should be interpreted with caution. This comparison relies on point measurements of precipitation, temperature, and SWE from SNOTEL sites against corresponding mean values generated by the model over a ~6-km grid. Point measurements are typically made under open tree canopies, whereas average SWE from a ~6-km grid not only ignores local topographic effects but also includes the effect of existing vegetation over the entire grid cell. SWE values derived from SNOTEL sites have been shown to be as much as 200% higher when compared with mean SWE over a 1-, 4-, and 16-km^2 grid (Molotch and Bales 2005). Hence, a large underestimation of SWE by the model was not surprising and could be entirely due to the sampling nature of the snow datasets. Comparing VIC-simulated SWE with those derived from remote sensing [e.g., spatially distributed SWE derived from Moderate Resolution Imaging Spectroradiometer (MODIS) imagery] could provide a better measure of model performance in capturing snow dynamics.

These findings on model performance have broad implications for using large-scale LSMs in landscape-level planning as well as future model improvement. As mentioned earlier, geological differences among the watershed as inferred by K along with errors in meteorological forcing can significantly affect model performance. However, the relative contribution of structural (lack of groundwater) and parametric (meteorological forcing) error is still unclear. Because of the nature of the landscape where geology, snow, and elevation all are geographically correlated, it was difficult to disentangle their individual effects on model performance. For example, summer-flow and low-flow percentiles were underpredicted in groundwater-dominated watersheds and overpredicted in runoff-dominated watersheds. But in this region, groundwater-dominated watersheds are typically located at high elevations, where meteorological forcings are also uncertain and presumably underreported. The meteorological forcing data for VIC were interpolated based on Cooperative Observer Program (COOP) stations that are predominantly located at lower elevations. Under these circumstances, it would be unfair to call the error in groundwater-dominated watersheds entirely structural. Most likely it was a result of both structural and parametric, and more research is needed to quantify the individual impact.

Finally, since most of the model error is systematic ($\mathrm{RMSE}_s/\mathrm{RMSE}_u > 1$), a more rigorous site-specific calibration may help improve model performance. At a regional scale where climate and geology varies significantly, a geologically (Tague et al. 2013) or landscape-based (Patil et al. 2013) model parameterization, as opposed to transferring the calibrated model parameters based on spatial proximity and physical similarity (Oudin et al. 2008), could help reduce the uncertainty due to presence/absence of groundwater. Additionally, relying on only streamflow for model calibration in regions such as PNW, where shapes of the hydrographs largely depend on the water stored in the form of SWE and groundwater, may be problematic. Availability of remotely sensed datasets such as MODIS-based evapotranspiration and snow cover and Gravity Recovery and Climate Experiment (GRACE)-based terrestrial water storage provide opportunities for multicriteria parameter estimation (Livneh and Lettenmaier 2012). Site-specific calibration can also compensate for the uncertainty in meteorological forcing (Elsner et al. 2014). Although, as pointed out by Elsner et al. (2014), uncertainties in meteorological forcing could still be challenging in forecasting the effects of climate change. Uncertainties in gridded meteorological data (Elsner et al. 2014) can be minimized by utilizing the measurements from SNOTEL sites. The SNOTEL sites are typically located at higher elevations than the currently used COOP stations for gridded meteorological VIC forcing data (Hamlet and Lettenmaier 2005).

5. Summary and conclusions

This study provides an assessment of the large-scale Variable Infiltration Capacity (VIC) model for predicting hydrologic regimes of small watersheds in the Pacific Northwest. Since large-scale hydrologic models of this type are typically not calibrated for small watersheds, knowing the uncertainties and their relationship to topographic, geologic, and climatic controls is quite valuable. Model performance and associated uncertainties were assessed by comparing VIC-simulated and observed streamflows from 217 watersheds in terms of total flow at annual and season time scales and flow percentiles. In addition to streamflow, we also compared the model meteorological forcing with independent observations from 109 stations with daily and monthly precipitation, maximum and minimum temperatures, and SWE data. The effect of deep groundwater on model performance was assessed by grouping watersheds based on the streamflow recession constant K, following Safeeq et al. (2013).

Overall, the model was able to capture the hydrologic behavior of these watersheds with reasonable accuracy as measured by the Spearman rank correlation. Both total streamflow and flow percentiles, however, are subject to strong systematic model error. Although the magnitude of relative bias (i.e., PBIAS) between groundwater- and runoff-dominated watersheds are comparable, summer streamflow and lower-percentile flows in runoff-dominated watersheds are predominantly overestimated and consistently underestimated in groundwater-dominated watersheds. The model performed poorly in capturing the sensitivity of streamflow to changes in both temperature and precipitation across all seasons. Our findings also suggest strong disagreements between gridded and observed meteorological forcing and simulated and observed SWE, which could be contributing to model bias. Since groundwater- and snow-dominated watersheds overlap geographically, disentangling the individual impact on model bias was challenging. However, since most of the model bias was systematic, a careful site-specific or geologically driven model calibration using not only streamflow but also SWE would be expected to improve model performance.

Predicting changes in future streamflow under climate change at the regional scale is essential for planning and developing mitigation strategies. The VIC and other LSMs help scientists and resource managers answer "what if" questions in a quantitative manner based on future climate and land use changes as projected by global climate models. This study highlights some of the uncertainties in model-simulated streamflow and how it may vary under different hydrogeological terrains and time scales. Our results also provide a basis for developing model calibration and parameterization strategies for future modeling work in this region that might better account for landscape differences in terms of groundwater contribution.

Acknowledgments. The authors gratefully acknowledge funding support from the Oregon Watershed Enhancement Board, Bureau of Land Management (Oregon), and the U.S. Forest Service Region 6 and Pacific Northwest

Research Station. The manuscript benefitted from the thoughtful comments of Sarah Lewis and three anonymous reviewers.

REFERENCES

Arismendi, I., M. Safeeq, S. L. Johnson, J. B. Dunham, and R. Haggerty, 2013: Increasing synchrony of high temperature and low flow in western North American streams: Double trouble for coldwater biota? *Hydrobiologia,* **712,** 61–70, doi:10.1007/s10750-012-1327-2.

Arnold, J. G., P. M. Allen, R. Muttiah, and G. Bernhardt, 1995: Automated base flow separation and recession analysis techniques. *Ground Water,* **33,** 1010–1018, doi:10.1111/j.1745-6584.1995.tb00046.x.

Barnett, T. P., J. C. Adam, and D. Lettenmaier, 2005: Potential impacts of a warming climate on water availability in snow dominated regions. *Nature,* **438,** 303–309, doi:10.1038/nature04141.

Battin, J., M. W. Wiley, M. H. Ruckelshaus, R. N. Palmer, E. Korb, K. K. Bartz, and H. Imaki, 2007: Projected impacts of climate change on salmon habitat restoration. *Proc. Natl. Acad. Sci. USA,* **104,** 6720–6725, doi:10.1073/pnas.0701685104.

Beer, W. N., and J. J. Anderson, 2013: Sensitivity of salmonid freshwater life history in western US streams to future climate conditions. *Global Change Biol.,* **19,** 2547–2556, doi:10.1111/gcb.1224.

Beven, K. J., 2011: *Rainfall–Runoff Modelling: The Primer.* 2nd ed. John Wiley, 457 pp.

——, and A. Binley, 1992: The future of distributed models: Model calibration and predictive uncertainty. *Hydrol. Processes,* **6,** 279–298, doi:10.1002/hyp.3360060305.

——, 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, doi:10.1016/S0022-1694(01)00421-8.

Burnash, R. J., R. L. Ferral, and R. A. McGuire, 1973: A generalized streamflow simulation system: Conceptual modeling for digital computers. Dept. of Commerce/NWS/CDWR Rep., 204 pp.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen, 2004: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *J. Hydrol.,* **298,** 242–266, doi:10.1016/j.jhydrol.2004.03.042.

Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D. H. Peterson, 2001: Changes in the onset of spring in the western United States. *Bull. Amer. Meteor. Soc.,* **82,** 399–415, doi:10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2.

Clark, M. P., and J. A. Vrugt, 2006: Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters. *Geophys. Res. Lett.,* **33,** L06406, doi:10.1029/2005GL025604.

Duan, Q., and Coauthors, 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.,* **320,** 3–17, doi:10.1016/j.jhydrol.2005.07.031.

Ek, M., and Coauthors, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Elsner, M. M., and Coauthors, 2010: Implications of 21st century climate change for the hydrology of Washington State. *Climatic Change,* **102,** 225–260, doi:10.1007/s10584-010-9855-0.

——, S. Gangopadhyay, T. Pruitt, L. Brekke, N. Mizukami, and M. Clark, 2014: How does the choice of distributed meteorological data affect hydrologic model calibration and streamflow simulations? *J. Hydrometeor.,* **15,** 1384–1403, doi:10.1175/JHM-D-13-083.1.

Falcone, J. A., D. M. Carlisle, D. M. Wolock, and M. R. Meador, 2010: GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology,* **91,** 621–621, doi:10.1890/09-0889.1.

Farley, K. A., C. Tague, and G. E. Grant, 2011: Vulnerability of water supply from the Oregon Cascades to changing climate: Linking science to users and policy. *Global Environ. Change,* **21,** 110–122, doi:10.1016/j.gloenvcha.2010.09.011.

Gupta, H. V., S. Sorooshian, and P. O. Yapo, 1998: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resour. Res.,* **34,** 751–763, doi:10.1029/97WR03495.

——, ——, and ——, 1999: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng.,* **4,** 135–143, doi:10.1061/(ASCE)1084-0699(1999)4:2(135).

Hamlet, A. F., and D. P. Lettenmaier, 2005: Production of temporally consistent gridded precipitation and temperature fields for the continental United States. *J. Hydrometeor.,* **6,** 330–336, doi:10.1175/JHM420.1.

——, P. W. Mote, M. P. Clark, and D. P. Lettenmaier, 2005: Effects of temperature and precipitation variability on snowpack trends in the western United States. *J. Climate,* **18,** 4545–4561, doi:10.1175/JCLI3538.1.

——, ——, ——, and ——, 2007: Twentieth-century trends in runoff, evapotranspiration, and soil moisture in the western United States. *J. Climate,* **20,** 1468–1486, doi:10.1175/JCLI4051.1.

——, S. Y. Lee, K. E. B. Mickelson, and M. M. Elsner, 2010: Effects of projected climate change on energy supply and demand in the Pacific Northwest and Washington State. *Climatic Change,* **102,** 103–128, doi:10.1007/s10584-010-9857-y.

——, M. M. Elsner, G. S. Mauger, S.-Y. Lee, I. Tohver, and R. A. Norheim, 2013: An overview of the Columbia Basin climate change scenarios project: Approach, methods, and summary of key results. *Atmos.–Ocean,* **51,** 392–415, doi:10.1080/07055900.2013.819555.

Hidalgo, H. G., and Coauthors, 2009: Detection and attribution of streamflow timing changes to climate change in the western United States. *J. Climate,* **22,** 3838–3855, doi:10.1175/2009JCLI2470.1.

Jin, X., and V. Sridhar, 2010: An integrated surface water–groundwater modeling in the Upper Snake River basin, Idaho. *2010 Fall Meeting,* San Francisco, CA, Amer. Geophys. Union, Abstract H21B-1026.

——, C.-Y. Xu, Q. Zhang, and V. P. Singh, 2010: Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *J. Hydrol.,* **383,** 147–155, doi:10.1016/j.jhydrol.2009.12.028.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Knowles, N., M. D. Dettinger, and D. R. Cayan, 2006: Trends in snowfall versus rainfall in the western United States. *J. Climate,* **19,** 4545–4559, doi:10.1175/JCLI3850.1.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *J. Geophys. Res.,* **105,** 24 809–24 822, doi:10.1029/2000JD900327.

——, S. P. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle, 2010: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nat. Geosci.,* **3,** 613–616, doi:10.1038/ngeo944.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.,* **99,** 14 415–14 428, doi:10.1029/94JD00483.

Liu, M., J. C. Adam, and A. F. Hamlet, 2013: Spatial–temporal variations of evapotranspiration and runoff/precipitation ratios responding to the changing climate in the Pacific Northwest during 1921–2006. *J. Geophys. Res. Atmos.,* **118,** 380–394, doi:10.1029/2012JD018400.

Livneh, B., and D. P. Lettenmaier, 2012: Multi-criteria parameter estimation for the Unified Land Model. *Hydrol. Earth Syst. Sci.,* **16,** 3029–3048, doi:10.5194/hess-16-3029-2012.

——, P. J. Restrepo, and D. P. Lettenmaier, 2011: Development of a unified land model for prediction of surface hydrology and land–atmosphere interactions. *J. Hydrometeor.,* **12,** 1299–1320, doi:10.1175/2011JHM1361.1.

Matheussen, B., R. L. Kirschbaum, I. A. Goodman, G. M. O'Donnell, and D. P. Lettenmaier, 2000: Effects of land cover change on streamflow in the interior Columbia River basin (USA and Canada). *Hydrol. Processes,* **14,** 867–885, doi:10.1002/(SICI)1099-1085(20000415)14:5<867::AID-HYP975>3.0.CO;2-5.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate,* **15,** 3237–3251, doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

McMichael, C. E., A. S. Hope, and H. A. Loaiciga, 2006: Distributed hydrological modelling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation. *J. Hydrol.,* **317,** 307–324, doi:10.1016/j.jhydrol.2005.05.023.

Molotch, N. P., and R. C. Bales, 2005: Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resour. Res.,* **41,** W11421, doi:10.1029/2005WR004229.

Moriasi, D., J. Arnold, M. Van Liew, R. Bingner, R. Harmel, and T. Veith, 2007: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE,* **50,** 885–900, doi:10.13031/2013.23153.

Mote, P. W., and E. P. Salathé, 2010: Future climate in the Pacific Northwest. *Climatic Change,* **102,** 29–50, doi:10.1007/s10584-010-9848-z.

Nijssen, B., and Coauthors, 2014: A prototype global drought information system based on multiple land surface models. *J. Hydrometeor.,* **15,** 1661–1676, doi:10.1175/JHM-D-13-090.1.

Nossent, J., and W. Bauwens, 2012: Application of a normalized Nash–Sutcliffe efficiency to improve the accuracy of the Sobol' sensitivity analysis of a hydrological model. *Geophysical Research Abstracts,* Vol. 14, Abstract EGU2012-237. [Available online at http://meetingorganizer.copernicus.org/EGU2012/EGU2012-237.pdf.]

Olden, J. D., and N. L. Poff, 2003: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Res. Appl.,* **19,** 101–121, doi:10.1002/rra.700.

Oleson, K. W., and Coauthors, 2010: Technical description of version 4.0 of the Community Land Model (CLM). NCAR Tech. Note NCAR/TN-478+STR, 257 pp., doi:10.5065/D6FB50WZ.

Oubeidillah, A., S.-C. Kao, M. Ashfaq, B. Naz, and G. Tootle, 2014: A large-scale, high-resolution hydrological model parameter data set for climate change impact assessment for the conterminous US. *Hydrol. Earth Syst. Sci.,* **18,** 67–84, doi:10.5194/hess-18-67-2014.

Oudin, L., V. Andréassian, C. Perrin, C. Michel, and N. Le Moine, 2008: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resour. Res.,* **44,** W03413, doi:10.1029/2007WR006240.

Patil, S., and M. Stieglitz, 2012: Controls on hydrologic similarity: Role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrol. Earth Syst. Sci.,* **16,** 551–562, doi:10.5194/hess-16-551-2012.

Patil, S. D., P. J. Wigington, S. G. Leibowitz, and R. L. Comeleo, 2013: Use of hydrologic landscape classification to diagnose streamflow predictability in Oregon. *J. Amer. Water Resour. Assoc.,* **50,** 762–776, doi:10.1111/jawr.12143.

Refsgaard, J. C., and J. Knudsen, 1996: Operational validation and intercomparison of different types of hydrological models. *Water Resour. Res.,* **32,** 2189–2202, doi:10.1029/96WR00896.

Rosenberg, E. A., E. A. Clark, A. C. Steinemann, and D. P. Lettenmaier, 2013: On the contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin. *Hydrol. Earth Syst. Sci.,* **17,** 1475–1491, doi:10.5194/hess-17-1475-2013.

Roy, S. B., L. Chen, E. H. Girvetz, E. P. Maurer, W. B. Mills, and T. M. Grieb, 2012: Projecting water withdrawal and supply for future decades in the U.S. under climate change scenarios. *Environ. Sci. Technol.,* **46,** 2545–2556, doi:10.1021/es2030774.

Safeeq, M., and A. Fares, 2012: Hydrologic response of a Hawaiian watershed to future climate change scenarios. *Hydrol. Processes,* **26,** 2745–2764, doi:10.1002/hyp.8328.

——, G. Grant, S. Lewis, and C. Tague, 2013: Coupling snowpack and groundwater dynamics to interpret historical streamflow trends in the western United States. *Hydrol. Processes,* **27,** 655–668, doi:10.1002/hyp.9628.

——, ——, ——, M. Kramer, and B. Staab, 2014: A geohydrologic framework for characterizing summer streamflow sensitivity to climate warming in the Pacific Northwest, USA. *Hydrol. Earth Syst. Sci. Discuss.,* **11,** 3315–3357, doi:10.5194/hessd-11-3315-2014.

Sankarasubramanian, A., R. M. Vogel, and J. F. Limbrunner, 2001: Climate elasticity of streamflow in the United States. *Water Resour. Res.,* **37,** 1771–1781, doi:10.1029/2000WR900330.

Shen, Z. Y., L. Chen, and T. Chen, 2012: Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: A case study of SWAT model applied to Three Gorges Reservoir region, China. *Hydrol. Earth Syst. Sci.,* **16,** 121–132, doi:10.5194/hess-16-121-2012.

Shrestha, D. L., 2010: *Uncertainty Analysis in Rainfall–Runoff Modelling: Application of Machine Learning Techniques.* CRC Press, 224 pp.

Shukla, S., and A. W. Wood, 2008: Use of a standardized runoff index for characterizing hydrologic drought. *Geophys. Res. Lett.,* **35,** L02405, doi:10.1029/2007GL032487.

Slack, J., A. Lumb, and J. Landwehr, 1993: Hydro-Climate Data Network (HCDN): Steamflow data set, 1874–1988: USGS Water-Resources Investigations Rep. 93-4076, CD-ROM. [Available online at http://pubs.usgs.gov/wri/wri934076/1st_page.html.]

Tague, C., and G. E. Grant, 2009: Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions. *Water Resour. Res.,* **45,** W07421, doi:10.1029/2008WR007179.

——, G. Grant, M. Farrell, J. Choate, and A. Jefferson, 2008: Deep groundwater mediates streamflow response to climate warming in the Oregon Cascades. *Climatic Change,* **86,** 189–210, doi:10.1007/s10584-007-9294-8.

——, J. Choate, and G. Grant, 2013: Parameterizing sub-surface drainage with geology to improve modeling streamflow responses to climate in data limited environments. *Hydrol. Earth Syst. Sci.,* **17,** 341–354, doi:10.5194/hess-17-341-2013.

Todini, E., 1996: The ARNO rainfall–runoff model. *J. Hydrol.,* **175,** 339–382, doi:10.1016/S0022-1694(96)80016-3.

Troy, T. J., E. F. Wood, and J. Sheffield, 2008: An efficient calibration method for continental-scale land surface modeling. *Water Resour. Res.,* **44,** W09411, doi:10.1029/2007WR006513.

U.S. Geological Survey, cited 2013: USGS water data for the nation. [Available online at http://waterdata.usgs.gov/nwis/.]

Vano, J. A., and D. P. Lettenmaier, 2014: A sensitivity-based approach to evaluating future changes in Colorado River discharge. *Climatic Change,* **122,** 621–634, doi:10.1007/s10584-013-1023-x.

——, T. Das, and D. P. Lettenmaier, 2012: Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *J. Hydrometeor.,* **13,** 932–949, doi:10.1175/JHM-D-11-069.1.

Vogel, R. M., and C. N. Kroll, 1992: Regional geohydrologic–geomorphic relationships for the estimation of low-flow statistics. *Water Resour. Res.,* **28,** 2451–2458, doi:10.1029/92WR01007.

Waibel, M. S., M. W. Gannett, H. Chang, and C. L. Hulbe, 2013: Spatial variability of the response to climate change in regional groundwater systems—Examples from simulations in the Deschutes basin, Oregon. *J. Hydrol.,* **486,** 187–201, doi:10.1016/j.jhydrol.2013.01.019.

Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier, 2009: Multimodel ensemble reconstruction of drought over the continental United States. *J. Climate,* **22,** 2694–2712, doi:10.1175/2008JCLI2586.1.

Water Resources Department, cited 2013: Tools and data. [Available online at http://www.oregon.gov/owrd/pages/pubs/toolsdata.aspx/.]

Wenger, S. J., C. H. Luce, A. F. Hamlet, D. J. Isaak, and H. M. Neville, 2010: Macroscale hydrologic modeling of ecologically relevant flow metrics. *Water Resour. Res.,* **46,** W09513, doi:10.1029/2009WR008839.

Willmott, C. J., and Coauthors, 1985: Statistics for the evaluation and comparison of models. *J. Geophys. Res.,* **90,** 8995–9005, doi:10.1029/JC090iC05p08995.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.,* **117,** D03110, doi:10.1029/2011JD016051.

——, J. Sheffield, M. B. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, and E. F. Wood, 2014: Evaluation of multi-model simulated soil moisture in NLDAS-2. *J. Hydrol.,* **512,** 107–125, doi:10.1016/j.jhydrol.2014.02.027.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1998: Multi-objective global optimization for hydrologic models. *J. Hydrol.,* **204,** 83–97, doi:10.1016/S0022-1694(97)00107-8.