Journal of Hydrology 597 (2021) 126327

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Augmenting geophysical interpretation of data-driven operational water supply forecast modeling for a western US river using a hybrid machine learning approach

Sean W. Fleming [a,b,c,*], Velimir V. Vesselinov [d], Angus G. Goodbody [a]

[a] National Water and Climate Center, Natural Resources Conservation Service, US Department of Agriculture, 1201 NE Lloyd Blvd., Suite 802, Portland, OR 97232-1274, USA
[b] College of Earth, Ocean, and Atmospheric Sciences and Water Resources Graduate Program, Oregon State University, 104 CEOAS Admin. Bldg., Corvallis, OR 97331-5503, USA
[c] Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, 2020-2207 Main Mall, Vancouver BC V6T 1Z4 Canada
[d] Computational Earth Sciences Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87544, USA

ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Andrea Castelletti, Associate Editor

Keywords:
Explainable machine learning
Water resources management
Hydropower
River forecasting
Regression
Probabilistic prediction

ABSTRACT

In the largely dry and increasingly heavily populated western US, operational modeling systems for seasonal river runoff volume forecasting are key elements of the practical water and hydropower management infrastructure. Explainability of model results in terms of known hydroclimatic processes and conditions is a core requirement for these systems. To improve geophysical interpretability of a standard statistical modeling approach to operational water supply forecasting (WSF), we introduce a hybrid statistical-artificial intelligence method. The procedure involves using a recently developed unsupervised machine learning algorithm designed for improved explainability (non-negative matrix factorization with k-means clustering, NMFk) to extract a compact basin-scale hydroclimatic index from available precipitation and snowpack data; that index is then used as the predictor variate in a largely conventional probabilistic regression on seasonal water supply. The resulting method, dominant-signal NMFk regression, is applied to a challenging forecast site, the Owyhee River, drawn from the US Department of Agriculture Natural Resources Conservation Service WSF system. Outcomes demonstrate that improved interpretability and plausibility relative to conventional statistical methods are achieved through physical consistency of NMFk results with nonnegativity of the environmental data being analyzed. In particular, the nonnegativity property facilitates identifying potential geophysical relationships to input variable type (snow water equivalent vs. accumulated precipitation), location, and underlying hydrologic processes; and it encourages nonnegative river runoff predictions, improving physical realism of WSFs over conventional statistical approaches in certain cases. The method also offers straightforward interpretation of relationships to known forms of climate variability. However, testing suggests that with these capabilities come limitations. Its primary anticipated role, at present, is to augment geophysical interpretation when needed, by serving as a complement alongside other methods in a next-generation US West-wide operational forecasting system.

1. Introduction

As margins between water supply and demand continue to narrow in the mostly dry American West (loosely defined as the US westward of the Great Plains), the importance of maximizing efficiency of the region's massive river management infrastructure around drinking, irrigation, and industrial water supplies and hydropower grows. In this region, water supply forecasts (WSFs) refer to quantitative model predictions of boreal spring-summer total river flow volume at a given location of interest, typically issued once per month or for some locations more often, beginning the preceding autumn or winter and continuing through late spring or early summer. Such forecasts of

* Corresponding author at: National Water and Climate Center, Natural Resources Conservation Service, US Department of Agriculture, 1201 NE Lloyd Blvd., Suite 802, Portland, OR 97232-1274, USA.
E-mail address: sean.fleming@usda.gov (S.W. Fleming).

https://doi.org/10.1016/j.jhydrol.2021.126327
Received 25 November 2020; Received in revised form 6 April 2021; Accepted 11 April 2021
Available online 20 April 2021
0022-1694/Published by Elsevier B.V.

forthcoming seasonal water supply availability, based largely on measurements of the winter-spring mountain snowpack that provides much of the region's river flow volumes, have long been a crucial requirement for the water resource optimization process.

Methods for generating these hydrologic forecasts fall into two general categories: process-simulation models that aim to explicitly represent the underlying physics of watershed-scale runoff generation, and data-driven phenomenological models that account for the physics implicitly using empirical input–output mappings of predictors to predictands. A tremendous variety of specific models, with a wide range of complexity and applicability, fall under these broad umbrellas (see reviews and syntheses by, for example, Singh and Woolhiser, 2002; Perkins et al., 2009; Gelfan and Motovilov, 2009; Bourdin et al., 2012; Weber et al., 2012; Cunderlik et al., 2013; Hrachowitz and Clark, 2017; Fleming and Gupta, 2020). Even modest incremental improvements in WSF skill can reap tens to hundreds of millions of dollars of public benefit per year for a single river basin in the western US (Yao and Georgakakos, 2001; Hamlet et al., 2002). Consequently, there has been intense interest in improving the prediction accuracy of WSF models in this and adjacent regions (e.g., Garen, 1998; Mahabir et al., 2003; Hsieh et al., 2003; Wood and Lettenmaier, 2006; Kennedy et al., 2009; Gobena and Gan, 2009; 2010; Rosenberg et al., 2011; Gobena et al., 2013; Robertson et al., 2013; Fleming and Dahlke, 2014; Demargne et al., 2014; Pagano et al., 2009; Trubilowicz et al., 2015; Harpold et al., 2016; Najafi and Moradkhani, 2016; Mendoza et al., 2017; Lehner et al., 2017; Fleming and Goodbody, 2019).

Another crucial aspect of WSF beyond improving predictive accuracy, however, is geophysical interpretation of modeling methods and results. This issue may be particularly significant in the operational hydrologic forecasting community, by which we mean large institutions, particularly but not exclusively government agencies, tasked with routine generation and distribution of WSFs used for responsible, and sometimes high-stakes, decision-making in the public interest, with corresponding institutional accountabilities for the reliability and timeliness of that information. The ability to readily determine how model behaviors relate to physical hydrologic processes is necessary for meeting professional responsibilities around assessing and communicating the reliability of forecasts used for these high-impact decisions, and for verification diagnostics. More broadly, physical interpretability is also critically important for explaining key aspects of the forecasts to clients, who often include the general public, such as why a river volume prediction increased or decreased and by how much since the last forecast date. Such readily communicated hydroclimatological 'storylines' go beyond simply improving client relations; runoff volume forecasts at some locations in the western US are requirements specified in legislation, legal decisions, or international treaties, like various biological opinions (BiOps) and the Columbia River Treaty, which govern water management in certain high-stakes basins and are therefore subject to intense public and even political scrutiny. Physical interpretability is therefore a key design criterion for all operational WSF systems, including those based on statistical and machine learning methods (e.g., Garen, 1992; Weber et al., 2012; Fleming and Goodbody, 2019).

Achieving such geophysical explainability in a practical WSF system can be challenging. Even process-simulation models, which have the advantage of explicitly capturing geophysical processes, can suffer from interpretability issues; a well-known example is equifinality of predictions produced by different representations or parameterizations of the underlying physics (e.g., Beven and Binley, 1992; Beven, 2006). Data-driven methods can be still more susceptible. Statistical regression models are the most widely used WSF technique operationally in western North America due to intrinsically much lower modeling system development and operation costs, similar or better forecast skill, easier incorporation of emerging new data types, greater operational simplicity and robustness, and easier and more accurate estimates of forecast uncertainty, relative to process-based methods (e.g., Gobena and Gan, 2010; Gobena et al., 2013; Risley et al., 2005; Fleming and Dahlke, 2014; Hsieh et al., 2003; Grantz et al., 2005; Harpold et al., 2016; Regonda et al., 2006; Rosenberg et al., 2011; Pagano et al., 2014; Minxue et al., 2016; Moradkhani and Meier, 2010; Mendoza et al., 2017; Robertson et al., 2013). However, these logistical and prediction performance advantages can be partially offset by less explicit representation of underlying physical processes. Additionally, WSF interpretability considerations for modern data-driven methods based on machine learning intersect with more acute, and much wider, questions around the ostensibly black-box nature of these techniques. Exploration of machine learning for river flow modeling began 25 years ago (Hsu et al., 1995; Minns and Hall, 1996), but in spite of significant ongoing advances (e.g., Lima et al., 2017; Kratzert et al., 2018), machine learning has largely failed to transition into production systems (e.g., Abrahart et al., 2012; Fleming and Gupta, 2020). In particular, river forecasting applications of AI to date have consisted almost exclusively of research studies with a narrow emphasis on improved simulation accuracy as expressed in terms of some conventional goodness-of-fit metric, with little attention to improving explainability; this has been a key reason for lack of uptake of machine learning by the operational community (e.g., Abrahart et al., 2012; Fleming et al., 2015; Fleming and Goodbody, 2019).

Such obstacles are gradually proving surmountable. For example, the artificial intelligence (AI) community is responding to explainability concerns with a major ongoing drive to develop glass-box machine learning. In environmental and geophysical applications, there is a limited but well-established track record of using AI to discover and explain governing physical processes (e.g., Cannon and McKendry, 2002; Fleming, 2007; Kratzert et al., 2019; Ellenson et al., 2020). Additionally, in hydrologic prediction, new machine learning-based systems are being developed and implemented in such a way that they explicitly incorporate and obey domain-specific expert hydrometeorological knowledge (e.g., Fleming et al., 2015; Cannon, 2018; Fleming and Goodbody, 2019; Oh and Orth, 2019; Xu et al., 2019; Wang et al., 2019), amounting to a form of theory-guided machine learning (Karpatne et al., 2017).

We contribute to this process of improving the geophysical trustworthiness and explainability of data-driven operational WSF models by introducing a new hybrid method. The technique pairs an established statistical regression modeling approach with a new unsupervised machine learning technique, nonnegative matrix factorization with k-means clustering (NMFk; Vesselinov et al., 2018), for hydroclimatic signal extraction. As discussed below, NMFk was specifically created around the need for physical plausibility and interpretability. We then apply the resulting WSF method, dominant-signal NMFk regression, to a particularly complex and difficult forecast site within the WSF system operated by the National Water and Climate Center of the US Department of Agriculture's Natural Resources Conservation Service (NRCS). This is the largest stand-alone WSF system in the American West, with approximately 600 forecast locations in the Colorado, Columbia, Missouri, Rio Grande, and other basins, and to our knowledge it is the world's largest statistically based operational WSF system. For relevance to the operational WSF problems that are our primary focus, the implementation here approximately follows standard practices for setting up such systems in western North America, including predictor and predictand choices.

We stress that the goal of our study is the relatively new topic of exploring novel techniques to improve hydroclimatic interpretation of machine learning-based WSF models, rather than the incremental accuracy improvements that have often been the focus of machine learning applications in hydrology; and further, that it is framed within the practical context of augmenting a well-established production WSF system.

2. Experimental design and data

2.1. General problem setup

To explore dominant-signal NMFk regression in a realistic WSF context, we set up the overall prediction problem in a manner similar to the existing NRCS forecast system, which is in turn broadly similar to most other regression-based operational WSF models in western North America. This general structure is as follows. The predictand is spring-summer runoff volume, which is usually measured at a US Geological Survey streamgage with adjustments as needed for upstream diversions, or at some other hydrometric monitoring site. Predictors consist of snow water equivalent (SWE) and wintertime-to-date accumulated precipitation measurements at mountain climate monitoring stations, predominantly NRCS SNOTEL or similar sites. Various other datasets, like antecedent streamflow, are occasionally used as supplemental predictors operationally but were not employed in this particular implementation, consistent with the existing operational NRCS model for the study basin described below. Note that research on data-driven WSF systems has extensively tested additional predictor types, like remotely sensed SWE, gridded precipitation datasets, seasonal-scale numerical climate model forecasts, and other products, but so far, these experimental predictors have not experienced significant operational adoption in statistical WSF models. Similarly, to the limited extent that machine learning-based WSF models have transitioned into genuinely operational WSF settings, they use predictors similar to those in current-generation statistical regression-based WSF models.

To illustrate, a typical regression-based WSF model might predict, on March 1, the upcoming April-September cumulative flow volume at a given point on a given river, using as predictors March 1 SWE and October-February total precipitation measured at SNOTEL sites within or near the watershed upstream of the streamgage. The number of such sites varies widely depending on the basin, but about a half-dozen to two dozen is roughly typical. How these predictors are used in the regression procedure varies depending on the technique as described in Section 3.5 below, but as part of the modeling process, the input datasets are usually amalgamated in some way into an index that serves as the actual regression predictor, or to use the machine learning nomenclature, a feature that is presented to the supervised learning algorithm. We use the same standard approach here.

2.2. Owyhee River test case

In the interest of conciseness, and because the goal of dominant-signal NMFk regression is to improve ability to identify and interpret relevant geophysical processes and relationships, we mainly focus on a reasonably thorough exploration of results and interpretations for one watershed (though generality is briefly examined in our assessment of capabilities and limitations in Section 4.5). The Owyhee River (Fig. 1) is known from operational NRCS experience to be one of the more challenging forecast points in the western US, with relatively low accuracy, tendency of statistical models to occasionally generate physically unrealistic negative-valued runoff predictions in dry years, and nonstationary and non-Gaussian prediction residuals that complicate prediction interval estimation; it is also relatively sparsely monitored compared to some other basins. Its headwaters lie in the remote mountains of northwestern Nevada and southwestern Idaho. It flows northward through eastern Oregon to empty into the Snake River, a major tributary of the international Columbia River. The region is semi-arid, and the annual hydrograph is dominated by spring runoff generated mainly by melting of wintertime mountain snowpack. The US Bureau of Reclamation operates Owyhee Dam and its approximately 80 km long reservoir, primarily to provide agricultural irrigation water.

We consider data from the US Geological Survey streamgage at Rome, Oregon and precipitation and snowpack data across the upstream drainage area from the SNOTEL network of remotely operated and telemetered snow and climate stations. As an illustrative example that captures the main features of established operational WSF practices, and therefore meaningfully tests the characteristics of dominant-signal NMFk regression in an operational WSF context, we take the forecast issue date to be April 1, our predictand is April-July total accumulated river runoff volume, and our predictors are wintertime-to-forecast-date accumulated precipitation and forecast-date SWE at several SNOTEL sites within and near the basin. This gives a total of one predictand and 18 predictors (Table 1). For model training and testing, a standard 30-year hydroclimatic normal period (1986–2015) is used. This is also typical practice in operational WSF and reflects, among other considerations, trade-offs between better model development using longer records vs. modest available record lengths at many observation locations in the region (see, e.g., Fleming and Goodbody, 2019). As such, the input dataset is a matrix containing 30 samples of 18 variables, and the output dataset is a vector of length 30. Further details of this test case are provided by Fleming and Goodbody (2019), and the data are freely available at wcc.sc.egov.usda.gov/reportGenerator.

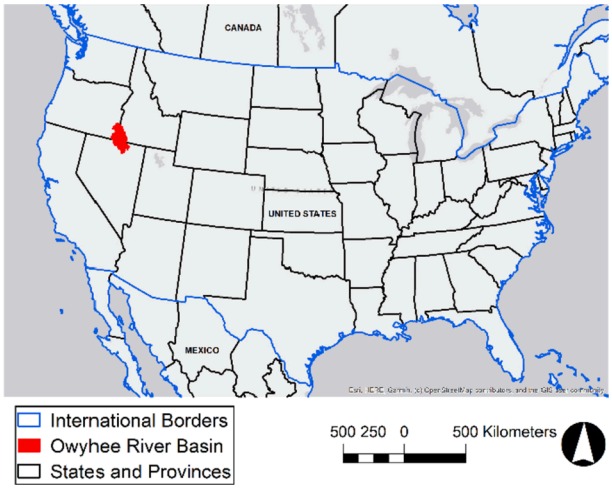

Fig. 1. Location of study basin.

3. Method

3.1. General

The overall procedure for forming a prediction involves using NMFk

Table 1
Predictor variables for April 1 forecast of April-July flow volume, Owyhee River near Rome, Oregon. P: October-March accumulated precipitation; SWE: April 1 snow water equivalent.

Variable ID	SNOTEL site ID	SNOTEL site name	Variable type
X1	336	Big Bend, Nevada	SWE
X2	373	Buckskin Lower, Nevada	SWE
X3	476	Fawn Creek, Nevada	SWE
X4	498	Granite Peak, Nevada	SWE
X5	548	Jack Creek Upper, Nevada	SWE
X6	573	Laurel Draw, Nevada	SWE
X7	654	Mud Flat, Idaho	SWE
X8	774	South Mountain, Idaho	SWE
X9	811	Taylor Canyon, Nevada	SWE
X10	336	Big Bend, Nevada	P
X11	373	Buckskin Lower, Nevada	P
X12	476	Fawn Creek, Nevada	P
X13	498	Granite Peak, Nevada	P
X14	548	Jack Creek Upper, Nevada	P
X15	549	Jacks Peak, Nevada	P
X16	573	Laurel Draw, Nevada	P
X17	654	Mud Flat, Idaho	P
X18	774	South Mountain, Idaho	P

to extract a compact basin-scale hydroclimatic index from available precipitation and snowpack data, and then using that index as the predictor variate in a probabilistic regression on seasonal water supply. This process, and other information from NMFk, contribute to formation of geophysical interpretations. Details of the method and additional context, including its relationships to data-driven operational WSF methods in current widespread operational use, are described below.

3.2. Non-negative matrix factorization with k-means clustering

NMFk is an unsupervised learning method used for feature extraction from datasets that are inherently nonnegative, and for which the identified features should in turn be nonnegative to be physically realistic and readily interpretable. Details on the NMFk algorithm and its implementation are discussed in Alexandrov and Vesselinov (2014), Vesselinov et al. (2018), and Vesselinov et al. (2019a, 2019b). The following is a brief summary.

Let us define a nonnegative-valued observational data matrix $\mathbf{X}$ of size (n,m), where m is the number of observable variables and n is the number of samples of each variable over time. The first step in NMFk analysis is to decompose the data matrix $\mathbf{X}$ into a nonnegative signal matrix $\mathbf{W}$ of size (n,k) and nonnegative mixing matrix $\mathbf{H}$ of size (k,m):

$$\mathbf{X} \cong \mathbf{W} \times \mathbf{H} \tag{1}$$

where k is an unknown number of features present in the data, that is, hidden variables. In other words, a feature represents a time series of an unmeasurable "master variable" hidden in the data. Specifically, the mixing matrix $\mathbf{H}$ describes sets of specific patterns across the variables; and the signal matrix $\mathbf{W}$ combines all extracted features over time, that is, it represents time variability in how strongly the various patterns in $\mathbf{H}$ are expressed. Thus, for example, the i^{th} $(i \in [1, k])$ extracted feature (hidden master variable) is represented by two vectors in the $\mathbf{W}$ and $\mathbf{H}$ matrices, respectively: the i^{th} signal (column) with length n and the i^{th} length-m mixing (row) vector. NMFk is a form of blind source separation: through multiplication of estimated $\mathbf{W}$ and $\mathbf{H}$ matrices, we can obtain an estimate of $\mathbf{X}$ which is a reconstruction of the original data matrix; the reconstruction would account for how the extracted signals in $\mathbf{W}$ are mixed (as defined by $\mathbf{H}$) to obtain the original observable variables in $\mathbf{X}$.

The optimal number of hidden signals k_{opt} is unknown a priori and is estimated by performing a series of nonnegative matrix factorizations for different values of k; $k = 1, 2, \cdots, d$. The maximum value d cannot exceed n or m. This is achieved by minimizing the following objective function O based on the Frobenius norm for all possible values of k:

$$O = \|\mathbf{X} - \mathbf{W} \times \mathbf{H}\|^F \tag{2}$$

For each k value in the range $1, 2, \cdots, d$, nonnegative matrix factorization is performed multiple times (typically, on the order of 1000 times) based on random initial guesses for $\mathbf{W}$ and $\mathbf{H}$ matrices. The best estimate of O for a given k from all these runs is applied to define the reconstruction error for each k value: $\varepsilon(k)$. The resulting multiple solutions of $\mathbf{H}$ (or alternatively $\mathbf{W}$; typically, it is preferred to cluster the smaller matrix) are clustered into k clusters using a customized k-means clustering. During clustering, we enforce the condition that each of the k clusters contain equal number of members which is equal to the number of performed multiple random runs (e.g., 1000 solutions). After clustering, the average silhouette width $S(k)$ is computed. This metric (Vesselinov et al. 2018; see also Rousseeuw, 1987) measures how well the random NMF solutions are clustered for given value of k. The values of $S(k)$ theoretically can vary from -1 to 1. Typically, $S(k)$ declines sharply after an optimal number, k_{opt}, is reached. The k_{opt} value is selected to be equal to the maximum number of signals that accurately reconstructs the observational data matrix $\mathbf{X}$ as estimated by $\varepsilon(k_{opt})$ and the average silhouette width $S(k_{opt})$ is close to 1.

Similar to principal component analysis (PCA), and in contrast to classical NMF, NMFk allows identification of the optimal number of features (which are somewhat equivalent to basis vectors or eigenvectors). NMFk analyses also lead to strictly additive features that are parts of the data (Paatero and Tapper, 1994). Similar to classical NMF (and in contrast to PCA), NMFk's ability to identify readily understandable features enables the discovery of new causal structures and unknown mechanisms hidden in the data (Cichocki et al., 2009). NMFk and its multi-dimensional version, nonnegative tensor factorization (NTFk), have recently been used for various types of analyses on observational data and model outputs (see above references). Note also that the NMFk/NTFk algorithms allow for missing entries in the data matrix or tensor $\mathbf{X}$, and that the algorithms are capable of reconstructing such data gaps based on the signal/mixing matrices extracted from the available data. NMFk/NTFk are also capable of estimating uncertainties associated with the number of features, $\mathbf{W}$ and $\mathbf{H}$ estimates, and reconstructions of $\mathbf{X}$. Though these additional functionalities around missing data estimation and matrix uncertainty estimation are potentially useful in some contexts, in this study we focus on leveraging the inherent interpretability advantages of NMFk by integrating it with the probabilistic regression methods more typically used in in WSF.

The NMFk algorithm is written in the Julia language. Though patented by Los Alamos National Laboratory, open source code and documentation, examples, and tests are available online at http://github.com/TensorDecompositions/NMFk.jl. We emphasize that the foregoing is only a brief conceptual summary of NMFk; theoretical and numerical implementation details are non-trivial and can be found in the references cited above.

3.3. Dominant-signal NMFk regression

The NMFk signal for $i = 1$, which we refer to here as the dominant signal, for $\mathbf{X}$ consisting of snow water equivalent (SWE) and precipitation time series at various locations within a basin, is taken to be an index of wintertime hydroclimatic conditions (see discussion in Section 2) and used as the predictor in a linear regression on spring-summer runoff volume:

$$< V(t) >= \beta_0 + \beta_1 I(t) \tag{3}$$

where $<V>$ is the expectation value of water supply volume in year t = [1, …, n]; I is the NMFk-derived basin-scale hydroclimatic index (i.e., the first column of $\mathbf{W}$); and β_0, β_1 are coefficients estimated by ordinary least squares (OLS), with β_1 providing a measure of the unit sensitivity of flow to hydroclimatic variability to the extent it is captured by the available input datasets and the signal extraction algorithm. Prediction intervals and various goodness-of-fit metrics (see below and Section 4.5) were assessed using cross-validated model predictions calculated largely following the widely used method of Garen (1992). Motivations for cross-validation in data-driven WSF modeling are that out-of-sample performance metrics better reflect true predictive skill than in-sample measures, and that record lengths are rarely sufficient in western North American WSF applications (see Section 2) to simply partition data into fully disjoint training and testing subsets of adequate lengths (e.g., Garen, 1992; see also Bergmeir and Benitez, 2012; Syed, 2011; Koul et al., 2018). The statistics and machine learning communities have developed many variants, with the two data science branches sometimes adopting different philosophies (e.g., Bergmeir and Benitez, 2012). We followed standard practice for data-driven operational WSF systems in the western US by using leave-one-out cross-validation, which experience has shown to provide a reliable estimate of prediction error (e.g., Garen, 1992; Pagano et al., 2004; Rosenberg et al., 2012; Lehner et al., 2017; Fleming and Goodbody, 2019). Note also that empirical partial autocorrelation functions of regression residuals from $<V>$ do not exhibit statistically significant serial dependence, again typical of WSF problems in the western US (e.g., Garen, 1992; Fleming and Goodbody, 2019).

Most operational WSF systems are probabilistic, by which is meant

that the best estimate provided by regression (or other methods) is accompanied by prediction bounds. In NRCS practice, these are the 0.1, 0.3, 0.7, and 0.9 quantiles of a probability distribution centered at $<V>$, which correspond, respectively, to the 90, 70, 30, and 10% exceedance probability flows. A common heuristic method, which is also widely applied in WSF, assumes a stationary normal distribution with a standard deviation equal to the cross-validated regression standard error or RMSE (e.g., Garen, 1992; Hyndman and Athanasopoulos, 2013).

Ideally, however, these quantiles should be estimated with a probability model that can flexibly accommodate non-Gaussian and heteroscedastic regression residuals, as these more complex error structures are seen for many NRCS basins, including the Owyhee River test case (Fleming and Goodbody, 2019). To do this, we employ a modified version of the aforementioned heuristic that uses the Box-Cox transform:

$$y(t)^{(\psi)} = \begin{cases} \dfrac{y(t)^{\psi} - 1}{\psi}, & \psi \neq 0 \\ \ln[y(t)], & \psi = 0 \end{cases} \quad (4)$$

where ψ is a parameter and $y(t)^{(\psi)}$ denotes the Box-Cox transform of variable y at time t. Both the observed and model-predicted cross-validated hindcast time series of water supply volume are transformed into homoscedastic normally distributed data in this way. The transformed datasets are then differenced to obtain the residual time series, which is normally distributed in transform space; the corresponding root mean square error (RMSE) can therefore be used as a convenient approximate metric of the standard deviation of the transform-space residuals. The transform-space α^{th} quantile forecast is estimated as:

$$Q_{\alpha}[V(t)]^{(\psi)} = < V(t) >^{(\psi)} + z\left(RMSE_{CV}{}^{(\psi)}\right) \quad (5)$$

where z are corresponding z-scores under the normal distribution and $<V(t)>^{(\psi)}$ is the transform of the best estimate of runoff. An inverse Box-Cox transform is applied to the result to obtain final estimates of the α^{th} quantile prediction bounds. This method generates, around a deterministic best-estimate model prediction, confidence intervals that (if needed) are asymmetric about $<V>$, show excessive kurtosis, and have time-variant widths or (otherwise) are Gaussian and stationary. This post-hoc approach, which appears to have been introduced to conceptual process simulation-based hydrologic modeling by Feyen et al. (2008) and was adapted to data-driven runoff volume forecasting with cross-validation by Fleming and Goodbody (2019), is superficially similar to but fundamentally different from the use of predictand transforms (including the Box-Cox transform) prior to model-building (e.g., Garen, 1992; Wang et al., 2012).

Regression modeling was performed in the R scientific computing environment. Forward and inverse Box-Cox transforms were performed using the `forecast` R package (Hyndman, 2017), which also estimates optimal ψ.

3.4. Comparison to standard statistical WSF methods

Additional regression models were developed using the same dataset described above but different methods: composite index regression, and principal component regression (PCR). These are the two most common approaches for statistically based operational WSF in western North America and serve as meaningful points of comparison.

PCR uses principal component analysis (PCA) to address dimensionality and multicollinearity. It was adapted to WSF by NRCS (Garen, 1992) and forms the primary basis for its official WSFs. PCA creates a new set of variables, obtained by projecting the original data onto a new orthogonal coordinate system defined by eigenvectors of the data correlation matrix. The resulting principal component scores are mutually uncorrelated and are used as possible predictors in an otherwise conventional regression. In NRCS experience, the leading PCA mode typically contains at least roughly 70% (or often much more) of the input dataset total variance and captures overall year-to-year variability in winter-spring snowpack available for spring-summer runoff generation. Higher modes are occasionally retained for some basins. For most rivers, however, the leading mode is the only PCA mode retained in the operational WSF regression model on the basis of t-tests of regression significance. For convenience we refer to this configuration as leading-mode PCR.

In composite index regression, dimensionality reduction and multicollinearity handling are accomplished by averaging together the input datasets for a given watershed. This amounts to calculating the mean areal cumulative precipitation and snowpack across the basin on the basis of available station data. The resulting index captures wintertime climatic inputs to the basin and is used as the sole predictor in a linear regression (see Garen, 1992). The method is a de facto application of stacking, which improves signal-to-noise ratio by the square root of the number of input variables, where signal in this context refers to the regionally coherent component of year-to-year climatic variability across all the input variable locations and types (e.g., Telford et al., 1990; Monteleoni et al., 2011; Fleming and Barton, 2015). More complex variants exist but can have disadvantages (Perkins et al., 2009; Garen, 1992). For convenience, we refer to regression of water supply upon an average of the input variables as simple-index regression.

All three modeling frameworks considered invoke a similar overall architecture: an index of wintertime hydroclimatic conditions is formed from SNOTEL or other similar data, which then serves as the predictor variate in a linear regression model. They differ mainly in how the predictive index is created, and of particular interest to us here, how thoroughly or easily it can be interpreted.

In simple-index regression, no information about relationships between input variables is generated during index creation. This restricts possibilities for interpretation without taking additional steps like correlation analyses. In PCR, the eigenvectors contain information about relationships between variates in the input data matrix that can be used to understand the final predictive model. However, such interpretation is hampered by two characteristics. First, there is normally no non-negativity constraint, and in general both the eigenvector and scores contain a mixture of positive and negative values. Second, the signs on the eigenvectors and scores are mathematically arbitrary in the sense that one can multiply both by -1 and obtain the same net contribution to the total dataset variance and dynamics (that is, a positive-valued eigenvector entry corresponding to a certain input variable may represent either an above- or below-average value of that variable, depending on the arbitrary polarity of the scores time series). In contrast, typical operational WSF predictor variables (SWE and precipitation; Section 2) are strictly nonnegative, so that negative-valued PCA results are physically unrealistic. A consequence is that, in WSF practice, PCA is normally treated only as a data pre-processing trick to facilitate application of classical regression to multicollinear problems, and detailed PCA results (e.g., eigenvectors) are typically ignored during interpretation of PCR-based WSF models. In dominant-signal NMFk regression, the mixing matrix and signals provide information loosely akin to PCA eigenvectors and scores. In contrast, however, the NMFk mixing matrix and signals are nonnegative, like the SWE and precipitation datasets they represent; they are therefore more physically realistic, facilitating geophysical interpretation.

These potential improvements in physical explainability motivated our exploration of NMFk as an alternative feature extraction approach prior to regression modeling. Other advantages and disadvantages of dominant-signal NMFk regression relative to standard statistical models grew clearer after application to the WSF test case. These are discussed in Section 4.5.

3.5. Post-hoc analysis steps

Two additional analytical steps were performed during assessment and interpretation of results from dominant-signal NMFk regression. The first was a simple application of information theory to support

preliminary geophysical interpretation of the mixing matrix. (For general background on information theory, see Shannon (1948) and Pierce (1980); hydrologic applications include, e.g., Amorocho and Espildora (1972), Caselton and Husain (1980), Krasovskaia (1995), Weijs et al. (2010), Fleming and Dahlke (2014), and Nearing and Gupta (2015)). The 18 mixing matrix entries for the leading signal are divided into three bins (low, middle, high) with category cutoffs equally spaced between the minimum and maximum values. The Shannon entropy is then calculated as $-\Sigma p_i \log_2(p_i)$, where p_i is the estimated probability that any given mixing matrix entry falls within the i^{th} bin, found by counting how many times entries fall within each bin. This gives the information content, in bits, of the dominant-signal entries in the mixing matrix. The same category cutoffs are then used to calculate the probabilities and corresponding entropy only for those mixing matrix entries corresponding to snow data, giving an estimate of the quantity of information provided by SWE measurements as captured in the NMFk dominant signal. The procedure is repeated again to find the information content of mixing matrix entries for the precipitation data alone. Implications of these results and caveats to their interpretation are discussed in Section 4.3.

Additionally, to test interpretability of the NMFk-derived basin-scale hydroclimatic index in terms of well-established large-scale climate patterns, we examined relationships between it and publicly available indices for nine modes of climate variability known to be at least partially related to North American precipitation and snowpack. Spearman rank (nonparametric) correlation was used to robustly estimate statistical significance of linear and monotonically nonlinear associations without making distributional assumptions. We consider El Niño-Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), Pacific-North America pattern (PNA), trans-Niño index (TNI), North Atlantic Oscillation (NAO), Interdecadal Pacific Oscillation (IPO), Arctic Oscillation (AO), Atlantic Multidecadal Oscillation (AMO), and North Pacific Gyre Oscillation (NPGO). Note that these patterns are not all mutually independent; for instance, statistical and physical relationships exist between the IPO and PDO, and the NAO and AO. Monthly indices were obtained from the National Oceanic and Atmospheric Administration Earth System Research Laboratory's Physical Sciences Division (e.g., www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.data for ENSO indices) except for the NPGO index, which was obtained from the Ocean Climate & Ecosystem Science group at Georgia Tech (www.o3d.org/npgo/). This analysis is intended only to confirm climatic interpretability of the dominant NMFk signal by comparing it to indices of atmosphere–ocean oscillations for which hydrologic teleconnections in some regions of North America are already reasonably well-known (e.g., Mantua et al., 1997; Garen, 1998; Hamlet and Lettenmaier, 1999; Kennedy et al., 2009; Gobena et al., 2013; Fleming and Dahlke, 2014; Enfield et al., 2001; Kingston et al., 2006; McCabe et al., 2004; Pascolini-Campbell et al., 2015; Vincent et al., 2015; Wu et al., 2005). As such, the analysis is not exhaustive. For instance, though exploring other averaging intervals could be valuable, we focus on wintertime-to-forecast-date mean values of these large-scale climate indices, which corresponds to the seasonal timeframe of our basin-scale NMFk index (see Section 2). Similarly, while ENSO has several indices which are in general closely correlated, we use only Niño 3.4, a common choice for ENSO teleconnection identification in water resource studies. Likewise, certain climate oscillations may demonstrate parabolic teleconnections in some locations, like ENSO in northern California (e.g., Wu et al., 2005; Fleming and Dahlke, 2014), but only linear and monotonically nonlinear relationships are considered here, as again is common practice in hydroclimatic analysis.

4. Results and discussion

4.1. Raw NMFk results for Owyhee River test case

NMFk identified three viable solutions for the data matrix described in Section 2, containing $k_{opt} = 2$, 3, and 6 signals. Cursory scoping regressions on water supply volume suggested that within each of these solutions, signal $i = 1$ ($i \in [1, k_{opt}]$), which we refer to as the dominant signal (see above), provided the best in-sample WSF predictive ability. The 2-signal NMFk solution was selected over the 3- and 6-signal solutions for the remainder of the analysis, primarily because it provides a more compact and parsimonious representation of predictor dataset dynamics, and secondarily because its dominant signal provided slightly better flow volume regression results.

Raw results from the 2-signal NMFk solution are illustrated in Fig. 2. The mixing matrix entries for the dominant signal exhibit moderate heterogeneity across predictor variates, and its signal correlates reasonably well ($R = 0.74$) with observed flow volumes. The mixing matrix values for the second signal are considerably more heterogeneous, and the corresponding signal correlates poorly ($R = -0.13$) with runoff volume. As per Section 3.3, the dominant signal was used as the predictor variate in the regression; results are described in the remainder of Section 4 below. Geophysical interpretations specifically of the signal time series and mixing matrix entries corresponding to the dominant NMFk signal are provided in Sections 4.3 and 4.4.

4.2. Physical acceptability of runoff volume predictions

Observed and predicted runoff and associated prediction bounds are shown in Fig. 3. A notable feature is that the dominant-signal NMFk regression-based best estimate and its prediction bounds are without exception nonnegative. Application of regression methods to WSF in some arid or semi-arid basins in the US West can lead to 90% exceedance flows and, in some cases, best estimates, that are negative-valued during dry years (for discussion of these effects and why they occur, see Fleming and Goodbody, 2019). Physically, of course, river runoff volume cannot take on negative values. This is a known issue for the Owyhee River forecast point in particular, and the leading-mode PCR and simple-index regression models both produce negative-valued best estimates at one or more sample times when applied to the identical dataset as dominant-signal NMFk regression. This makes the predictions from conventional statistical WSF methods unacceptable, and in operational WSF practice using such techniques, subjective choices and manual applications of predictand transforms are required to sidestep the problem when it occurs. That NMFk by construction gives strictly nonnegative regression predictors does not guarantee nonnegative best-estimate regression model predictions, but it does seem to promote this required characteristic of a physically plausible data-driven water supply forecast model.

4.3. Preliminary geophysical interpretation of NMFk mixing matrix maps

Fig. 4 presents the dominant-signal NMFk mixing matrix from Fig. 2 in map format to facilitate interpretation. Many of the precipitation and SWE input variables are co-located because they both come from the same SNOTEL site (blue dots on Fig. 4), so for convenience of visualization and interpretation, the dominant-signal mixing matrix entries are separated into maps corresponding to SWE (Fig. 4a) and precipitation (Fig. 4b) input variables. We emphasize, however, that all the dominant-signal NMFk mixing matrix entries are a single vector, which can be viewed as coefficients that weight the contributions of each input variable (SWE and precipitation measurements at various locations) to the net dominant signal in the NMFk solution; separation of the entries into those corresponding to SWE versus precipitation is done only to facilitate viewing and interpretation. Long-term average SWE and precipitation fields across the watershed are also shown for reference on Fig. 4a and 4b, respectively,

as are maps of topography (Fig. 4c) and mean air temperature (Fig. 4d). The sizes of the blue dots in Fig. 4a and 4b provide the magnitude of the corresponding mixing matrix entries.

These mixing matrix maps provide opportunities for physical

Fig. 2. Two-signal NMFk solution (dimensionless). Top: mixing matrix entries corresponding to the 18 input variables listed in Table 1. Bottom: associated signal time series. See Figs. 4 and 5 for more detailed representations of the dominant signal (signal 1) and interpretive information, including juxtapositions of its mixing matrix entries against maps of watershed characteristics, and of the signal time series against indices for several modes of large-scale climatic variability.

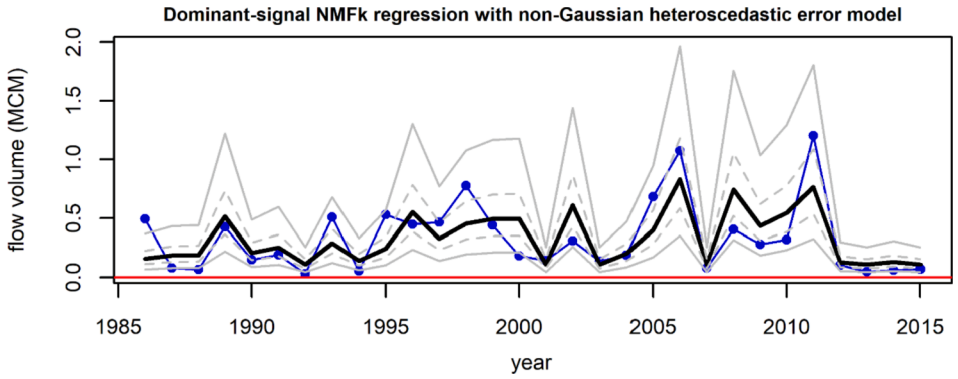

Fig. 3. Dominant-signal NMFk regression forecasts of spring-summer water supply volume for the Owyhee River issued on April 1 in millions of cubic meters (MCM). Blue dots and connecting line: observations; solid black line: best-estimate predictions; dashed gray lines: cross-validated prediction intervals corresponding to 30% and 70% exceedance probability flows; solid gray lines: 10% and 90% exceedance probability flows; red horizontal line: zero-flow marker.

interpretation of patterns detected by the NMFk algorithm. The largest entries in the mixing matrix correspond to SWE, rather than precipitation. This suggests that SWE contributes more than precipitation to the dominant pattern (primary hidden variable; see Section 3.2) that underlies the mixed SWE-precipitation spatiotemporal dataset, which in turn seems intuitively reasonable. SWE data in some sense tell more about overall hydroclimatic conditions than total precipitation data alone, as snowpack reflects the cumulative impact of several geophysical and biophysical controls, including precipitation, temperature, wind speed, forest canopy, and so forth. That the mixing matrix entries for precipitation are nearly uniform across sites whereas those for SWE show significant variation further reinforces this notion that the SWE data contain more information about underlying patterns of hydroclimatic variability in this watershed.

We can provide a back-of-the-envelope quantitative representation of this effect using information theory, following the methodology described in Section 3.5. The Shannon entropy of all 18 dominant-signal mixing matrix entries is 1.1 bits. The value rises to 1.4 bits if we use the same category cutoffs but calculate the probabilities and corresponding entropy only for mixing matrix entries corresponding to snow data. Repeating the procedure again for dominant-signal mixing matrix entries corresponding to the precipitation data alone gives 0 bits, as they all fall within the same bin. This result should not be overinterpreted to mean precipitation data offer no additional value to a basin-scale index of hydroclimatic variability beyond what snowpack data provide, and more complex analysis steps could provide a more refined view of comparative information contents. With these caveats in mind, however, the outcome does seem to support the physical interpretation above around snow data capturing more information about a wider variety of processes than precipitation.

As described in Sections 2 and 3, as used here NMFk is applied to all the hydrometeorological input data to create an index of wintertime hydroclimatic conditions. As such, relationships to runoff volume are, by construction, not part of this first step in the dominant-signal NMFk regression process. It is nevertheless interesting to note that, especially for an April 1 forecast date, by which point seasonal snow accumulation has typically peaked, SWE is widely known by operational hydrologists to be a better predictor of runoff than wintertime-to-date precipitation. That is, the NMFk result that snow data captures more information than precipitation data is additionally consistent with the generally

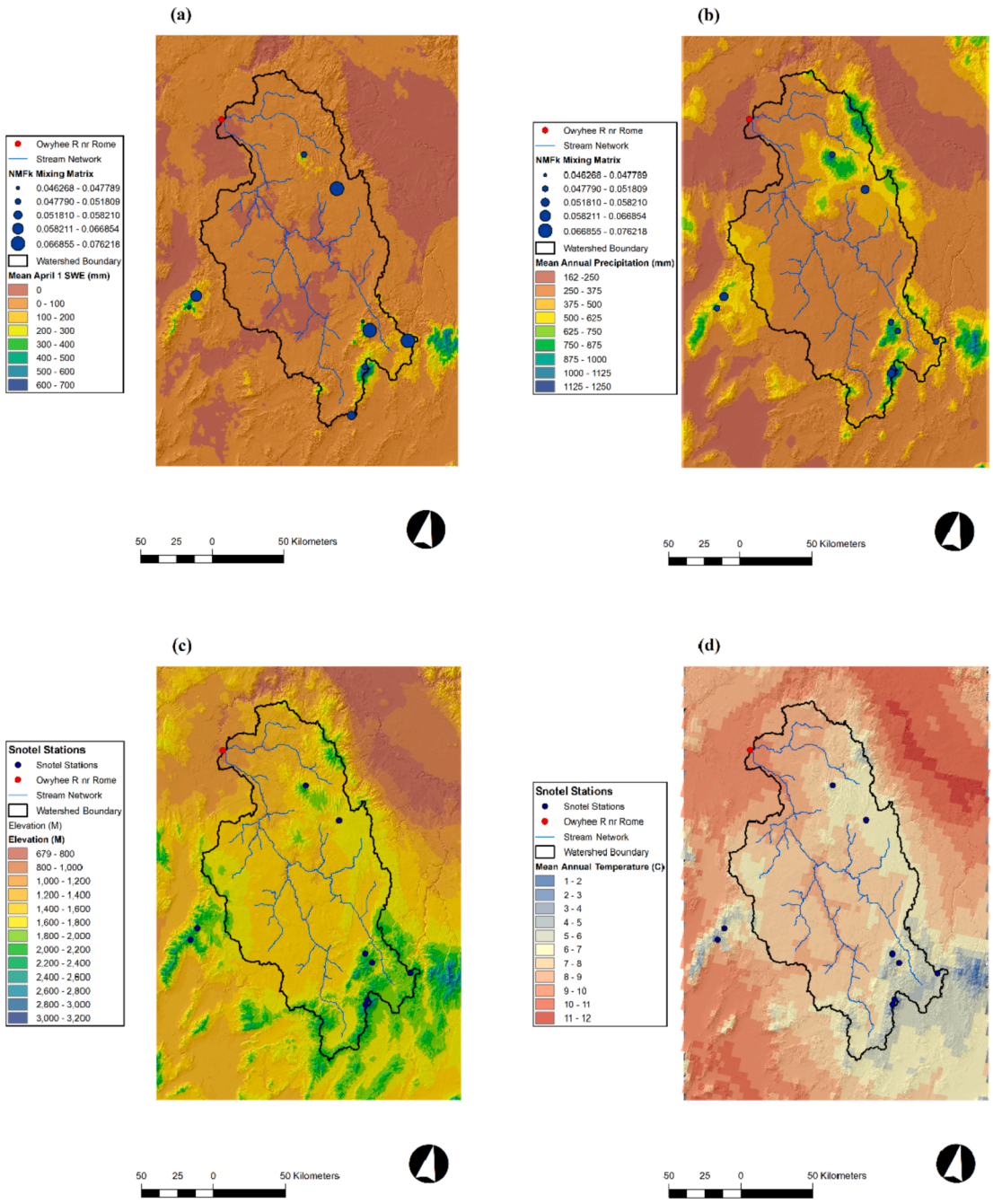

Fig. 4. Dominant-signal NMFk mixing matrix entries at each SNOTEL station (4a: SWE variables, 4b: precipitation variables). Size of blue dots in 4a and 4b indicate magnitude of corresponding matrix entry; note nonnegativity. Blue dots in 4c and 4d only denote locations of SNOTEL stations. For interpretative context, mean SNODAS April 1 SWE (4a), mean PRISM precipitation (4b), elevation (4c), and mean temperature (4d) are provided. SNODAS and PRISM background fields are available at https://www.nohrsc.noaa.gov/nsa and https://prism.oregonstate.edu and are shown solely for context here.

established fact that springtime snow data captures more information relevant to WSF than precipitation does.

The NMFk mixing matrix maps may also suggest hypotheses to help explain spatial structure in this underlying variability pattern. For instance, many of the larger mixing matrix entries correspond to higher-elevation SNOTEL sites. These generally experience greater spring snowpack than lower-elevation sites and therefore dominate overall basin-scale water availability and thus its year-to-year variability. Conversely, one of the largest mixing matrix entries corresponds to snowpack at Mud Flat, near the northeastern boundary of the watershed, which is a relatively warm, low-elevation location that, on average, experiences complete melting of the seasonal snowpack in early April, several weeks before most of the other SNOTEL sites considered here. Thus, SWE data at this site might amount to a de facto snow presence/absence indicator on the April 1 measurement date used here and, in turn, as a powerful index of high- versus low-snow years, potentially making a large contribution to the overall basin-scale hydroclimatic signal. These NMFk-implied hypotheses could form a basis for additional research on how different input data types and locations contribute to overall basin-wide hydroclimatic signal indexing and predictive value in water supply forecasting.

Note that simple-index regression cannot suggest any such interpretations around the underlying patterns in the dataset and hypotheses for their physical origins; similarly, the eigenvector in leading-

mode PCA can be physically interpreted, but in practice it presents more barriers to doing so than the NMFk mixing matrix because both magnitude and polarity need to be taken account by the interpreter (Section 3.4). In contrast, the NMFk mixing matrix streamlines interpretation because it is strictly nonnegative like the environmental data types it represents, and only magnitude of the entries needs to be taken into account by the hydrologic scientist or engineer. This relative simplicity is particularly valuable if the model development and interpretation process may have to be repeated dozens or, in the NRCS system, hundreds of times across sites.

4.4. *Preliminary geophysical interpretation of NMFk dominant signal*

The precipitation- and SWE-based NMFk dominant signal represents year-to-year watershed-scale cryospheric and meteorological variability. Such expressions of interannual to interdecadal climate variation, in turn, generally tend to reflect net superposition of several organized modes of large-scale coupled ocean–atmosphere circulation patterns. These modes provide a well-established basis for framing investigations of water resource variations in both explanatory and predictive contexts (e.g., Redmond and Koch, 1991; Mantua et al., 1997; Garen, 1998; Hamlet and Lettenmaier, 1999; Werner et al., 2004; Hsieh et al., 2006; Fleming et al., 2006; Moradkhani and Meier, 2010; Gobena et al., 2013; Fleming and Dahlke, 2014; Beckers et al., 2016). Interpretability of the NMFk-derived basin-scale hydroclimatic index in terms of these circulation patterns was therefore examined using the methods and data discussed in Section 3.5. Fig. 5 illustrates the NMFk dominant signal from Fig. 2 alongside the large-scale climate indices,

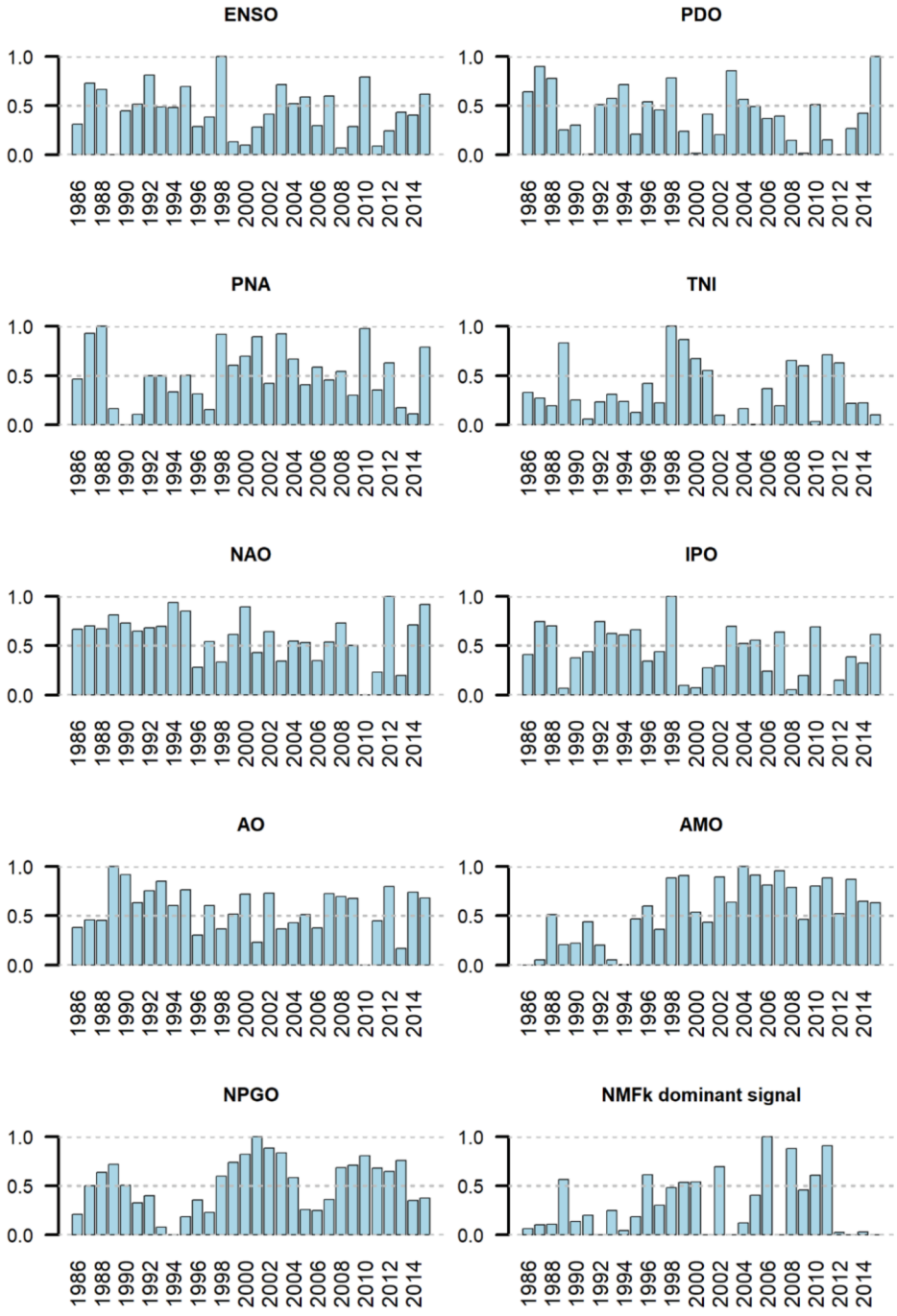

Fig. 5. Rescaled indices for several climate oscillations, and dominant NMFk signal.

which for purposes of visual comparison have been rescaled to the interval [0,1] in this graphic.

We found clear statistical evidence ($p < 0.05$ and occasionally $p \sim 0.01$) for monotonic correlations between the NMFk basin-scale hydroclimatic index and ENSO, PDO, TNI, and IPO. Conclusive ($p < 0.05$) statistical evidence was not found for associations with NAO, PNA, and AMO, though this does not preclude the presence of subtle teleconnections to one or more of these patterns that might be unambiguously identified using longer observational records or alternative averaging intervals, for example. There was no statistical support in this case for AO and NPGO teleconnections. Overall, the set of teleconnection analysis results is broadly consistent with prior work in western North America, and in particular the southern Columbia Basin (e.g., Mantua et al., 1997; Garen, 1998; Hamlet and Lettenmaier, 1999; Kennedy et al., 2009; Gobena et al., 2013; Fleming and Dahlke, 2014; Enfield et al., 2001; Kingston et al., 2006; McCabe et al., 2004; Pascolini-Campbell et al., 2015; Vincent et al., 2015; Wu et al., 2005). Note that for each of those climate patterns demonstrating statistically significant relationships to the NMFk index (ENSO, PDO, IPO, and TNI), the corresponding linear correlation coefficient is as high as (for ENSO) or higher than (for PDO, IPO, and TNI) the average of the correlation coefficients between the respective climate pattern and each of the 18 input variables. Overall, the results demonstrate that, like indices of watershed-scale wintertime precipitation and snowpack conditions derived using other methods such as PCA, explanations of interannual variability in the dominant-signal NMFk index can be easily and clearly framed in reference to specific known hemispheric- to global-scale climate processes.

4.5. Capabilities and limitations

Considered collectively, the results show that dominant-signal NMFk regression generates intermediate analytical products (refer again to Figs. 4 and 5, and Sections 4.3 and 4.4) and final WSF predictions (Fig. 3 and Section 4.2) that meet or beat the physical interpretability, and for runoff predictions also the physical plausibility, of statistical WSF models of the general types conventionally used in operational WSF for the western US. The outcomes are broadly consistent with expectations based on how these various modeling approaches are structured (Section 3.4). Note that we have focused on an apples-to-apples comparison of several index-based linear regression modeling methods based on exactly the same input data, with the only free experimental parameter being the approach used for feature extraction. Other regression and regression-like methods are available for the supervised learning component of data-driven WSF, including far more complex machine learning approaches, as are optimization methods for input variable selection, such as tree-based search methods and genetic algorithms (e.g., Garen, 1992; Fleming and Goodbody, 2019). Some of these techniques can significantly improve predictive performance in terms of both summary goodness-of-fit measures and physical reliability and, in some cases, explainability. These could in principle be combined with NMFk-based feature extraction. As a result, the outcomes presented here provide a pessimistic assessment of the values of the approaches used, including dominant-signal NMFk regression.

However, testing also revealed limitations and drawbacks. Interestingly, improvements in a priori physical basis, geophysical interpretability, and geophysical plausibility come at the price of poorer cross-validated (Section 3.3) statistical goodness-of-fit metrics (Table 2). Measures considered here include root mean square error (RMSE), which is similar to regression standard error and provides a measure of the typical prediction error that might be expected from the model; correlation coefficient (R) and coefficient of determination (R^2), where the former describes how faithfully the predictions reproduce patterns of interannual variability in observed flow volumes, and the latter gives the proportion of variance explained by the model; and ranked probability skill score (RPSS), a measure of the probabilistic skill of the model, framed in terms of its ability, relative to a naïve climatology forecast, to predict the probability of dry, normal, or wet years as defined by terciles of the observed flow volumes (e.g., Wiegel et al., 2007; Guihan, 2014; Fleming and Goodbody, 2019). However, these conventional accuracy metrics do not penalize the non-physical predictions made by the standard statistical models or reward the physical acceptability of predictions made by dominant-signal NMFk regression. This limits the ability of such measures to meaningfully describe and diagnose model quality in this application (see again Section 4.2). Broad-based quantitative fitness-for-purpose rankings of hydrologic models are complex, include subjective components, and are best formed in competitive team evaluation processes beyond the scope of the current paper (see Cunderlik et al., 2013). Nevertheless, a more comprehensive and balanced portrayal of model performance is provided in Table 2 by additionally noting whether the model-predicted best estimates, and the lowest of the associated prediction intervals considered in standard WSF applications (Section 3.3), meet the physicality requirement of being nonnegative for all available sample times. In this broader view, the dominant-signal NMFk regression performance results are more mixed compared to conventional methods.

Table 2
Model performance. Root mean square error, correlation coefficient, coefficient of determination, and ranked probability skill score quantify out-of-sample deterministic and probabilistic forecast accuracy. Bottom two metrics are binary indices of physical plausibility of model predictions, in particular, whether the best-estimate volume (V) or the lowermost operationally used prediction uncertainty interval on V, the 0.10 quantile flow estimate ($Q_{0.10}$), are nonnegative at all available sample times. See text for details.

Metric	Dominant-signal NMFk regression	Leading-mode PCR	Simple-index regression
RMSE	0.22	0.19	0.19
R	0.68	0.78	0.77
R^2	0.46	0.61	0.59
RPSS	0.26	0.32	0.47
$\{<V(t)>\} \geq 0 \; \forall \; t = [1, \ldots, N]$?	Y	N	N
$\{\langle Q_{0.10}[V(t)]\rangle\} \geq 0 \; \forall \; t = [1, \cdots, N]$?	Y	N	N

Additionally, though nonnegativity of NMFk outcomes matches the physical characteristics of typical WSF predictors and therefore simplifies (and thus encourages) deeper interpretations relative to the more conventional alternative of PCA (see Sections 3.4 and 4.3), it bears noting that nonnegativity is not a strict requirement for such interpretation. There is a long history of geophysically interpreting eigenvectors derived from PCA of fundamentally nonnegative hydrologic quantities. Examples include watershed regime classification or detecting the impacts of climatic variability and change on streamflow (e.g., Bartlein, 1982; Guetter and Georgakakos, 1993; Lins, 1997; Fleming et al., 2006). NMFk only makes this interpretive process more intuitive and accessible.

The physical interpretability advantages of NMFk are also premised on the exclusive use of nonnegative WSF predictors. This is consistent with most operational WSF practices in the US West, as standard statistical WSF predictors are SWE, precipitation, and in some cases antecedent streamflow, all of which are strictly nonnegative. That said, some additional predictor types that have seen extensive experimentation and occasional operational implementation may not be nonnegative. For instance, though not currently used in its production systems, ENSO indices were introduced into long lead-time US West operational forecasting by NRCS as a de facto early-season surrogate for winter precipitation and snow data (Garen, 1998) and have also been operationally adopted elsewhere (e.g., Gobena et al., 2013). Dominant-signal NMFk regression is not directly applicable to mixed-sign predictors, although of course such predictors could be rescaled to render them nonnegative as in Fig. 5.

Another applicability question is associated with a distinct technical characteristic of NMFk: unlike PCA, the suite of k_{opt} signals in a NMFk solution are not mutually uncorrelated. From a rigorous statistical modeling perspective, using more than one of these multicollinear signals as candidate features for a multiple linear regression is therefore potentially problematic (e.g., Garen, 1992). In the context of conventional linear statistical regression modeling, then, the WSF role of NMFk may be limited to creating a univariate watershed-scale hydroclimatic index.

Finally, scoping applications to other rivers in the western US suggest that dominant-signal NMFk regression is a generally viable WSF method, but also that the strengths of its advantages and limitations, as compared to standard statistical models, vary between basins. To illustrate, consider another existing NRCS forecast point, the Yellowstone River at Corwin Springs, located in the Rocky Mountain headwaters of the Missouri Basin. Similar to the Owyhee River, dominant-signal NMFk regression provided serviceable but decreased performance on conventional goodness-of-fit metrics compared to simple-index regression and PCR. However, conventional statistical regression models for Yellowstone WSF do not produce negative-valued predictions, so the tendency of dominant-signal NMFk regression to encourage physically realistic nonnegative volume predictions is irrelevant here, unlike the Owyhee River. The NMFk mixing matrix entries for the dominant signal offered clear physical interpretations for the Yellowstone River, as it did for the Owyhee River. That said, we should expect the resulting hydroclimatic 'storyline' to differ from basin to basin, depending on details of precipitation and SWE monitoring sites, predominant watershed-scale terrestrial hydrologic processes, overall climatic characteristics, and potentially other factors in each basin. This was seen to be the case. For the Yellowstone River, the primary pattern in the dominant-signal NMFk mixing matrix entries involved both SWE and precipitation data from a single SNOTEL site that stood apart from results for both SWE and precipitation at all other sites. This pattern presumably reflects local-scale characteristics at the anomalous SNOTEL station, potentially including but not necessarily limited to its instrumentation, local land cover/land use, or microclimate. Similar to the Owyhee River, the dominant NMFk signal for Yellowstone showed correlations with various large-scale climate indices consistent with established understanding of these teleconnections in western North America. All things considered, dominant-signal NMFk regression provided useful interpretive insights into the Yellowstone River but might be considered a somewhat less attractive candidate for operational WSF here than in the Owyhee basin.

In the following section, we discuss how the various capabilities and limitations described above may guide potential next steps for exploring the ways that dominant-signal NMFk regression might be further developed, and ultimately deployed in production systems for operational WSF.

5. Conclusions

Operational water supply forecasts are a cornerstone of water management in the largely arid US West. Ability to tell a hydrologically meaningful 'story' around what the forecast models are telling us is a requirement for understanding and evaluating the models and communicating their outcomes to clients. While AI has conventionally had a reputation as an uninterpretable black box, recent advances in physics-aware AI are changing that perception. Here, we leverage one of these advances, NMFk, to improve geophysical plausibility and interpretability compared to traditional statistical models. The result is a hybrid that pairs this new theory-guided, glass-box unsupervised learning algorithm with a largely conventional statistical prediction model. Application demonstrates that it facilitates both easier geophysical interpretability and better geophysical plausibility than established data-driven WSF methods. This demonstration, in combination with increasing interest in improving the physical interpretability of data-driven WSF systems, in turn suggests that continued research and development is warranted on using NMFk for feature extraction in WSF.

One possibility follows on the existence of multiple (k_{opt}) signals. As noted in Section 4.5, these are not mutually independent, possibly limiting their role as predictors in conventional multiple linear regression modeling. However, NMFk appears to reduce dimensionality very effectively, and the compressed signals it generates may therefore prove useful as candidate features to a supervised machine learning-based prediction system relating NMFk signals to seasonal flow volume. Such supervised AI methods, like neural networks for instance, do not strictly require feature independence but do typically benefit from reduced input data dimensionality, as this in turn reduces the required size and complexity of the network topology and attendant training and interpretation complications.

We currently anticipate that the primary use of dominant-signal NMFk regression is as a complement to, rather than a replacement for, established approaches within WSF systems. The goal in doing so would be to provide additional geophysical interpretive information as needed for specific forecasting problems that require special attention, such as particularly difficult forecast locations like the Owyhee River we focused on here. Our intention is to experiment with integration of NMFk as an additional feature extraction technique alongside PCA in a multi-method machine learning metasystem that has been developed as the basis of the next generation of the US West-wide NRCS operational WSF model (Fleming and Goodbody, 2019).

More broadly, our results seem to reinforce the value of practically minded and selective integrations of existing knowledge, methods, and processes with certain new machine learning techniques as they emerge, to the extent that such techniques may fill certain specific known gaps (in this particular case, improving the physical interpretability of watershed-scale hydroclimatic signal extraction). This philosophy may suggest a template for investigating emergent AI technologies in the context of applied hydrometeorological prediction systems, where a broad spectrum of quantitative and qualitative considerations, including but extending far beyond prediction skill, determine operational desirability of available modeling methods and guide design criteria (e.g., Weber et al., 2012; Cunderlik et al., 2013; Fleming and Goodbody, 2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Velimir V. Vesselinov is supported by LANL LDRD grants 20180060DR and 20190020DR. Processed SNODAS data products for Fig. 4 were provided by Jiunn-Der (Geoffrey) Duh, Geography Department, Portland State University. We thank two anonymous reviewers and the associate editor for helpful comments on an earlier version of this manuscript.

References

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Mount, N.J., See, L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L., 2012. Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Progress in Physical Geography 36, 480–513.

Alexandrov, B.S., Vesselinov, V.V., 2014. Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. Water Resources Research 50, 7332–7347.

Amorocho, J., Espildora, B., 1972. Entropy in the assessment of uncertainty in hydrologic systems and models. Water Resources Research 9, 1511–1522.

Bartlein, P.J., 1982. Streamflow anomaly patterns in the USA and southern Canada – 1951–1970. Journal of Hydrology 57, 49–63.

Beckers, J.V.L., Weerts, A.H., Tijdeman, E., Welles, E., 2016. ENSO-conditioned weather resampling method for seasonal streamflow prediction. Hydrology and Earth System Sciences 20, 3277–3287.
Bergmeir, C., Benitez, J.M., 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences 191, 192–213.
Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320, 18–36.
Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 279–298.
Bourdin, D.R., Fleming, S.W., Stull, R.B., 2012. Streamflow modelling: a primer on applications, approaches, and challenges. Atmosphere-Ocean 50, 507–536.
Cannon, A.J., 2018. Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. Stochastic Environmental Research and Risk Assessment 32, 3207–3225.
Cannon, A.J., McKendry, I.G., 2002. A graphical sensitivity analysis for statistical climate models: application to Indian Monsoon rainfall prediction by artificial neural networks and multiple linear regression models. International Journal of Climatology 22, 1687–1708.
Caselton, W.F., Husain, T., 1980. Hydrologic networks: information transmission. ASCE Journal of the Water Resources Planning Division 106, 503–520.
Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.-I., 2009. Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. John Wiley and Sons, Chichester, UK.
Cunderlik, J.M., Fleming, S.W., Jenkinson, R.W., Thiemann, M., Kouwen, N., Quick, M., 2013. Integrating logistical and technical criteria into a multiteam, competitive watershed model ranking procedure. ASCE Journal of Hydrologic Engineering 18, 641–654.
Demargne, J., Wu, L., Regonda, S.K., Brown, J.D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H.D., Fresch, M., Schaake, J., Zhu, Y., 2014. The science of NOAA's operational hydrologic ensemble forecast service. Bulletin of the American Meteorological Society January, 80–98.
Ellenson, A., Pei, Y., Wilson, G., Ozkan-Haller, H.T., Fern, X., 2020. An application of a machine learning algorithm to determine and describe error patterns within wave model output. Coastal Engineering 157.
Enfield, D.B., Mestas-Nuñez, A.M., Trimble, P.J., 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophysical Research Letters 28, 2077–2080.
Feyen, L., Kalas, M., Vrugt, J.A., 2008. Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization. Hydrological Sciences Journal 53, 293–308.
Fleming, S.W., 2007. Artificial neural network forecasting of nonlinear Markov processes. Canadian Journal of Physics 85, 279–294.
Fleming, S.W., Barton, M., 2015. Climate trends but little net water supply shifts in one of Canada's most water-stressed regions over the last century. Journal of the American Water Resources Association 51, 833–841.
Fleming, S.W., Bourdin, D.R., Campbell, D., Stull, R.B., Gardner, T., 2015. Development and operational testing of a super-ensemble artificial intelligence flood-forecast model for a Pacific Northwest river. Journal of the American Water Resources Association 51, 502–512.
Fleming, S.W., Dahlke, H.E., 2014. Parabolic northern-hemisphere river flow teleconnections to El Niño-Southern Oscillation and the Arctic Oscillation. Environmental Research Letters 9. https://doi.org/10.1088/1748-9326/9/10/104007.
Fleming, S.W., Goodbody, A.G., 2019. A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US West. IEEE Access 7, 119943–119964.
Fleming, S.W., Gupta, H.V., 2020. The physics of river prediction. Physics Today 73, 46–52.
Fleming, S.W., Moore, R.D., Clarke, G.K.C., 2006. Glacier-mediated streamflow teleconnections to the Arctic Oscillation. International Journal of Climatology 26, 619–636.
Garen, D.C., 1992. Improved techniques in regression-based streamflow volume forecasting. Journal of Water Resources Planning and Management 118, 654–669.
Garen DC. 1998. ENSO indicators and long-range climate forecasts: usage in seasonal streamflow volume forecasting in the western United States, American Geophysical Union Fall Conference, San Francisco, CA.
Gelfan, A.N., Motovilov, Y.G., 2009. Long-term hydrological forecasting in cold regions: retrospect, current status, and prospect. Geography Compass 3 (5), 1841–1864.
Gobena, A.K., Gan, T.Y., 2009. Statistical ensemble seasonal streamflow forecasting in the South Saskatchewan River basin by a modified nearest neighbors resampling. Journal of Hydrologic Engineering 14, 628–639.
Gobena, A.K., Gan, T.Y., 2010. Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. Journal of Hydrology 385, 336–352.
Gobena, A.K., Weber, F.A., Fleming, S.W., 2013. The role of large-scale climate modes in regional streamflow variability and implications for water supply forecasting: a case study of the Canadian Columbia Basin. Atmosphere-Ocean 51, 380–391.
Grantz, K., Rajagopalan, B., Clark, M., Zagona, E., 2005. A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. Water Resource Research 41, W10410. https://doi.org/10.1029/2004WR003467.
Guetter, A.K., Georgakakos, K.P., 1993. River outflow of the conterminous United States, 1939–1988. Bulletin of the American Meteorological Society 74, 1873–1891.
Guihan, R., 2014. Integrating Emerging River Forecast Center Streamflow Products into the Salt Lake City Parley's Drinking Water System. University of Massachusetts-Amherst, Masters Degree Project.
Hamlet, A.F., Lettenmaier, D.L., 1999. Columbia River streamflow forecasting based on ENSO and PDO climate signals. Journal of Water Resource Planning and Management 125, 333–341.
Hamlet, A.F., Huppert, D., Lettenmaier, D.P., 2002. Economic value of long-lead streamflow forecasts for Columbia River hydropower. ASCE Journal of Water Resources Planning and Management 128, 91–101.
Harpold, A.A., Sutcliffe, K., Clayton, J., Goodbody, A., Vazquez, S., 2016. Does including soil moisture observations improve operational streamflow forecasts in snow-dominated watersheds? Journal of the American Water Resources Association 53, 179–196.
Hrachowitz, M., Clark, M.P., 2017. The complementary merits of competing modelling philosophies in hydrology. Hydrology and Earth System Sciences 21, 3953–3973.
Hsieh, W.W., Wu, A., Shabbar, A., 2006. Nonlinear atmospheric teleconnections. Geophysical Research Letters 33, L07714.
Hsieh WW, Yuval, Li J; Shabbar A, Smith S. 2003. Seasonal prediction with error estimation of Columbia River streamflow in British Columbia. Journal of Water Resource Planning and Management, 129, 146-149.
Hsu, K.-L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resources Research 31, 2517–2530.
Hyndman, R.J., 2017. forecast: Forecasting functions for time series and linear models. R package version 8, 1. http://github.com/robjhyndman/forecast.
Hyndman RJ, Athanasopoulos G. 2013. Forecasting: principles and practice. OTexts, Melbourne, Australia. http://otexts.org/fpp/. Accessed on 22 September 2017.
Karpatne, A., Atluri, G., Faghmous, J.H., Steinback, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: a new paradigm for scientific discover from data. IEEE Transactions on Knowledge and Data Engineering 29, 2318–2331.
Kennedy, A.M., Garen, D.C., Koch, R.W., 2009. The association between climate teleconnection indices and Upper Klamath seasonal streamflow: Trans-Niño index. Hydrological Processes 23, 973–984.
Kingston, D.G., Lawler, D.M., McGregor, G.R., 2006. Linkages between atmospheric circulation, climate, and streamflow in the northern North Atlantic: research prospects. Progress in Physical Geography 30, 143–174.
Koul, A., Becchio, C., Cavallo, A., 2018. Cross-validation approaches for replicability in psychology. Frontiers in Psychology 9, 1117.
Krasovskaia, I., 1995. Quantification of the stability of river flow regimes. Hydrological Sciences Journal 40, 587–598.
Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences 22, 6005–6022.
Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences 23, 5089–5110.
Lehner, F., Wood, A.W., Llewellyn, D., Blatchford, D.B., Goodbody, A.G., Pappenberger, F., 2017. Mitigating the impacts of climate nonstationarity on seasonal streamflow predictability in the US southwest. Geophysical Research Letters 44, 12208–12217.
Lima, A.R., Hsieh, W.W., Cannon, A.J., 2017. Variable complexity online sequential extreme learning machine, with applications to streamflow prediction. Journal of Hydrology 555, 983–994.
Lins, H.F., 1997. Regional streamflow regimes and hydroclimatology of the United States. Water Resources Research 33, 1655–1667.
Mahabir, C., Hicks, F.E., Fayek, A.R., 2003. Application of fuzzy logic to forecast seasonal runoff. Hydrological Processes 17, 3749–3762.
Mantua NJ, Hare Steven R, Zhang Y, Wallace JM, Francis RC. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society, 78, June, 1069-1079.
McCabe, G.J., Palecki, M.A., Betancourt, J.L., 2004. Pacific and Atlantic Ocean influences on multidecadal drought frequency in the United States. Proceedings of the National Academy of Sciences 101, 4136–4141.
Mendoza, P.A., Wood, A.W., Clark, E., Rothwell, E., Clark, M.P., Nijssen, B., Brekke, L.D., Arnold, J.R., 2017. An intercomparison of approaches for improving operational seasonal streamflow forecasts. Hydrology and Earth System Sciences 21, 3915–3935.
Minns, A.W., Hall, M.J., 1996. Artificial neural networks as rainfall runoff models. Hydrological Sciences Journal 41, 399–417.
Minxue, H., Whitin, B., Hartman, R., Henkel, A., Fickenschers, P., Staggs, S., Morin, A., Imgarten, M., Haynes, A., Russo, M., 2016. Verification of ensemble water supply forecasts for Sierra Nevada watersheds. Hydrology 3, 35. https://doi.org/10.3390/hydrology3040035.
Monteleoni C, Schmidt GA, Saroha S, Asplund. 2011. Tracking climate models. Journal of Statistical Analysis and Data Mining, 4, 372-392.
Moradkhani, H., Meier, M., 2010. Long-lead water supply forecast using large-scale climate predictors and independent component analysis. Journal of Hydrologic Engineering 15, 744–762.
Najafi, M.R., Moradkhani, H., 2016. Ensemble combination of seasonal streamflow forecasts. Journal of Hydrologic Engineering 21. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001250.
Nearing, G.S., Gupta, H.V., 2015. The quantity and quality of information in hydrologic models. Water Resources Research 51, 524–538.
Oh, S., Orth, R., 2019. Getting the best of both worlds: physics-guided machine learning for hydrologic modeling. Presentation at the American Geophysical Union Fall Meeting, San Francisco, CA.
Paatero, P., Tapper, U., 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111–126.

Pagano, T.C., Garen, D.C., Perkins, T.R., Pasteris, P.A., 2009. Daily updating of operational statistical seasonal water supply forecasts for the western US. Journal of the American Water Resources Association 45, 767–778.
Pagano, T.C., Garen, D.C., Sorooshian, S., 2004. Evaluation of official western US seasonal water supply outlooks, 1922–2002. Journal of Hydrometeorology 5, 896–909.
Pagano, T.C., Wood, A.W., Werner, K., Tama-Sweet, R., 2014. Western US water supply forecasting: a tradition evolve. Eos, Transactions, AGU 95, 28–29.
Pascolini-Campbell, M.A., Seager, R., Gutzler, D.S., Cook, B.I., Griffin, D., 2015. Causes of interannual to decadal variability of Gila River streamflow over the past century. Journal of Hydrology: Regional Studies 3, 494–508.
Perkins, T.R., Pagano, T.C., Garen, D.C., 2009. Innovative operational seasonal water supply forecasting technologies. Journal of Soil and Water Conservation 64, 15–17.
Pierce, J.R., 1980. An Introduction to Information Theory: Symbols, Signals, and Noise, 2nd Ed. Dover, New York.
Redmond, K.T., Koch, R.W., 1991. Surface climate and streamflow variability in the western United States and their relationship to large scale circulation indices. Water Resources Research 27, 2381–2399.
Regonda, S.K., Rajagopalan, B., Clark, M., Zagon, E., 2006. A multimodel ensemble forecast framework: application to spring seasonal flows in the Gunnison River Basin. Water Resources Research 42. https://doi.org/10.1029/2005WR004653.
Robertson, D.E., Pokhrel, P., Wang, Q.J., 2013. Improving statistical forecasts of seasonal streamflows using hydrological model output. Hydrology and Earth System Sciences 17, 579–593.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.
Rosenberg, E.A., Wood, A.W., Steinemann, A.C., 2011. Statistical applications of physically based hydrologic models to seasonal streamflow forecasts. Water Resources Research 47. https://doi.org/10.1029/2010WR010101.
Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27 (379–423), 623–656.
Singh, V.P., Woolhiser, D.A., 2002. Mathematical modeling of watershed hydrology. ASCE Journal of Hydrologic Engineering 7, 270–292.
Syed, A.R., 2011. A Review of Cross Validation and Adaptive Model Selection. Georgia State University, Thesis.
Telford, W.M., Geldart, L.P., Sheriff, R.E., 1990. Applied Geophysics, 2nd Ed. Cambridge University Press, Cambridge, UK.
Trubilowicz JW, Chorlton E, Dery SJ, Fleming SW. 2015. Satellite remote sensing for water resource applications in British Columbia. Innovation, 19, March/April, 18-20.
Vesselinov, V.V., Alexandrov, B.S., O'Malley, D., 2018. Contaminant source identification using semi-supervised machine learning. Journal of Contaminant Hydrology 212, 134–142.
Vesselinov, V.V., Alexandrov, B.S., O'Malley, D., 2019a. Nonnegative tensor factorization for contaminant source identification. Journal of Contaminant Hydrology 220, 66–97.
Vesselinov, V.V., Mudunuru, M.K., Karra, S., O'Malley, D., Alexandrov, B.S., 2019b. Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing. Journal of Computational Physics 15, 85–104.
Vincent, L.A., Zhang, X., Brown, R.D., Feng, Y., Mekis, E., Milewska, E.J., Wan, H., Wang, X.L., 2015. Observed trends in Canada's climate and influence of low-frequency variability modes. Journal of Climate 28, 4545–4560.
Wang, Q.J., Shrestha, D.L., Robertson, D.E., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. Water Resources Research 48, W05514. https://doi.org/10.1029/2011WR010973.
Wang, N., Zhang, D., Chang, H., Li, H., 2019. Deep learning of subsurface flow via theory-guided neural network. Presentation at the American Geophysical Union Fall Meeting, San Francisco, CA.
Weber, F., Garen, D., Gobena, A., 2012. Invited commentary: themes and issues from the workshop "Operational River Flow and Water Supply Forecasting'.'. Canadian Water Resources Journal/Revue canadienne des ressources hydriques 37, 151–161.
Wiegel, A.P., Liniger, M.A., Appenzeller, C., 2007. The discrete Brier and ranked probability skill scores. Monthly Weather Review 135, 118–124.
Weijs, S.V., Schoups, G., Van De Giesen, N., 2010. Why hydrological predictions should be evaluated using information theory. Hydrology and Earth System Sciences 14, 2545–2558.
Werner, K., Brandon, D., Clark, M., Gangopadhyay, S., 2004. Climate index weighting schemes for NWS ESP-based seasonal volume forecasts. Journal of Hydrometeorology 5, 1076–1090.
Wood AW, Lettenmaier DP. 2006. A test bed for new seasonal hydrologic forecasting approaches in the western United States. Bulletin of the American Meteorological Society, December, 1699-1712.
Wu, A., Hsieh, W.W., Shabbar, A., 2005. The nonlinear pattern of North American winter temperature and precipitation associated with ENSO. Journal of Climate 18, 1736–1752.
Xu, T., Longyang, Q., Tyson, C., Zeng, R., Neilson, B.T., Tarboton, D.G., 2019. Hybrid physically-based and deep learning modeling of a snow dominated mountainous karst watershed. Presentation at the American Geophysical Union Fall Meeting, San Francisco, CA.
Yao, H., Georgakakos, A., 2001. Assessment of Folsom Lake response to historical and potential future climate scenarios, 2, reservoir management. Journal of Hydrology 249, 176–196.