H12C-06 - Practical Glass-Box Machine Learning for Seasonal Water Supply Forecasting, with Applications to the Owyhee and Yellowstone Rivers: Regression Using Climate Indices Derived from SNOTEL Data Using Nonnegative Matrix Factorization with k-Means Clustering

Monday, 12 December 2022

08:50 - 09:00

McCormick Place - E451b

Abstract

Hydroclimatic explainability – a straightforward, concise, intuitive, and defensible 'storyline' framed around commonly accepted atmospheric and terrestrial hydrologic processes and conditions – is a key requirement for many forms of hydrologic modeling. This includes seasonal water supply forecasts (WSFs), which are the information backbone of the massive water management infrastructure in the mostly dry and increasingly thirsty American West. WSF information is consumed by a wide variety of forecast product users, including the general public, and clear explanations of how and why a certain prediction was arrived at are necessary for establishing geophysical plausibility and effective client engagement. The interpretability of WSF techniques used operationally by service delivery organizations (SDOs) could bear improvement. For example, process-based simulation modeling suffers from equifinality of model parameterization and physics, which deeply complicates geophysical inference. In addition, principal component regression, the most widespread WSF technique in western North America, generates eigenvectors and scores that are inconsistent with the inherent nonnegativity of predictor and predictand variables (e.g., snow water equivalent, accumulated precipitation, and streamflow volume) typically considered in operational WSF systems. Broadly speaking, these issues are only compounded in artificial intelligence (AI)-based approaches, which in the past were often viewed as black-box prediction methods. However, nonnegative matrix factorization with k-means clustering (NMFk) is a new unsupervised machine learning technique developed specifically to improve, relative even to conventional statistical pattern recognition methods, the explainability of spatiotemporal patterns extracted from geophysical datasets by ensuring that a nonnegativity condition is satisfied. Here we integrate NMFk with statistical regression in an otherwise mostly conventional and proven WSF framework. The resulting dominant-signal NMFk regression method was tested for WSF in the Owyhee and Yellowstone Rivers and was found, subject to some caveats, to offer substantial improvement in geophysical explainability relative to two conventional statistical modeling techniques.

First Author

F

Sean William Fleming

US Department of Agriculture

Oregon State University

Authors

V

Velimir monty V Vesselinov

Los Alamos National Laboratory

G

Angus Graham Goodbody

Natural Resource Conservation Service

View Related