Watershed analysis and feature selection based on performance of deep learning streamflow drought models in the Colorado River Basin

Ali Dadkhah, PhD Student, Department of Civil & Environmental Engineering, University of Vermont, Burlington, VT, Ali.Dadkhah@uvm.edu

Donna M. Rizzo, Professor, Department of Civil & Environmental Engineering, University of Vermont, Burlington, VT, Donna.Rizzo@uvm.edu

Kristen Underwood, Research Assistant Professor, Department of Civil & Environmental Engineering, University of Vermont, Burlington, VT, Kristen.Underwood@uvm.edu

Abstract

Hydrological drought is a complex phenomenon and defies a clear consensus on its definition. In earth system and environmental sciences, multiple types of droughts have been identified based on attributes such as duration, drought thresholds, and the impacts and stressors on watersheds. Streamflow is an important and direct indicator of hydrological drought making the accurate prediction of streamflow drought conditions of significant interest for the hydrological forecasting community. As part of a regional drought early warning system project, the U.S. Geological Survey is developing models to predict streamflow drought conditions using a long short-term memory (LSTM) deep learning model using streamflow data from 409 USGS stream gages throughout the Colorado River Basin and the surrounding area. In this study, we leveraged these model outputs, which were characterized using two performance measures; Kling-Gupta efficiency (KGE) on the streamflow percentile and Cohen's Kappa on the classification of 20% (moderate) threshold of hydrologic drought. Model performance ranges from reliable to no predictive skill across watersheds and varies based on the choice of thresholds used to define streamflow drought (i.e., fixed, or daily variable thresholds) over a year.

To support this early warning system project, we used data that comprise model performance measures for predicting streamflow drought at 384 of the 409 gages within the Colorado River Basin region along with an additional 70 attributes compiled at the watershed-scale and associated with each of the 409 gages. We applied an unsupervised machine learning algorithm, a Self-Organizing Map (SOM), to cluster watersheds using their attributes, and then examined the model performance in relation to the clusters to determine if any meaningful patterns existed. We then used a random forest model to identify attributes most predictive of LSTM model performance. The random forest regression models showed that static watershed attributes may not be suitable predictors of LSTM model performance with R^2 values of 0.29 – 0.35 obtained. However, interpretation of SOM results yielded insights into the factors that influence

the predictive ability of the LSTM model and may help identify strategies for further improvements in model performance.

Background

Drought is an extreme event that develops gradually and is often unnoticed until water demands are not met. Given the multitude of direct and indirect detrimental impacts on the environment and various economic sectors, tackling drought at an early stage is of great importance (Yildirim et al., 2022). Streamflow drought is a type of hydrological drought manifesting in below-normal river discharge; accurate predictions of streamflow drought are of great interest for the hydrological forecasting community (Sutanto & van Lanen,2021; van Loon, 2015) and the ability to increase the forecast lead times is a key need in contemporary drought early warning systems. As part of such an early warning system focused on regional drought, the U.S. Geological Survey (USGS) is using what is known as long short-term memory (LSTM) deep learning models to predict streamflow drought conditions at lead times of 0, 7, 14, 30, and 60 days. Methods to forecast streamflow drought are different from streamflow forecasting, in that, the time series of predicted streamflow is post-processed and converted into a time series of drought events by applying drought identification approaches, e.g., the threshold approach or the standardized approach (Sutanto & van Lanen,2022). The threshold approach, or threshold level method, uses daily streamflow time series in order to identify periods of drought, and the onset of a drought event is defined when the daily streamflow drops below a pre-set threshold value; the event continues until the streamflow rises above the threshold. We consider two types of drought threshold approaches in this work, fixed and variable. The fixed approach applies a constant threshold value to the entire time series based on historical streamflow measurements, whereas the variable approach, allows the threshold to vary for each day of the year (Sutanto & van Lanen, 2021; Sarailidis et al., 2019).

Evaluating model performance is an important aspect of any hydrological model, and a variety of evaluation metrics or performance measures are used to calibrate and validate models (Moriasi et al., 2015). Examples such as Nash Sutcliffe Efficiency, NSE (Nash & Sutcliffe, 1970) and Kling-Gupta Efficiency, KGE, (Gupta et al., 2009) are popular for estimating hydrologic model predictive ability. Using these performance measures, model predictive ability can be scored from “reliable” to “no predictive ability” on a watershed-by-watershed basis, and these broad categories of model performance may be correlated to watershed attributes (e.g., mean elevation). Further examination of these model performance patterns may reveal the underlying associations between LSTM model performance and watershed attributes.

This study aims to explore patterns in LSTM model performance for a large-scale study of streamflow drought and the potential linkages of performance metrics to watershed attributes. We used performance measures of USGS LSTM predictive models for streamflow drought from 409 streamflow gages in the Colorado River Basin (CRB) and surrounding area, along with 70 watershed-scale attributes for each gage. We applied an unsupervised machine learning algorithm known as Self-Organizing Map (SOM) to explore linkages between watershed attributes and our target variable (i.e., LSTM model performance) using KGE and Cohen's Kappa statistics as the performance measures. The use of both measures allows investigation of the model performance more generally, using KGE, and for identifying drought event periods specifically, Cohen's Kappa.

Methods

Study Area & Dataset

The USGS has compiled more than 70 watershed attributes for each of the watersheds associated with the 409 USGS streamgages within the Colorado River Basin and surrounding area (Wieczorek et al., 2023) as well as nowcasting (lead time of zero days) LSTM model performance measures (KGE and Cohen's Kappa) in those watersheds (Hamshaw et al., 2023) (Figure 1, Panel a). Watershed attributes include, but are not limited to, information about soil type, topography, land cover and meteorology of watersheds; for a full list of watershed attributes see Table 1 in the Appendix. We used model performance measures corresponding to the LSTM models that predict streamflow drought using fixed and daily variable drought thresholds. The fixed approach applies a constant threshold value to the entire time series based on historical streamflow measurements, whereas the variable approach, allows the threshold to vary for each day of the year (Sutanto & van Lanen, 2021; Sarailidis et al., 2019). In this study, we leveraged these model outputs, which were characterized using two performance measures –Kling-Gupta efficiency (KGE) on the streamflow percentile and Cohen's Kappa on the classification of 20% (moderate) threshold of hydrologic drought; categories of model performance for Cohen's kappa statistics are based on the categories of Landis and Koch (Landis et al., 1977) and for KGE are set based on histogram of KGE values using natural breaks approach (Table 1).

Table 1. The Categories of model performance for Cohen's kappa and KGE.

Performance metric (PM)			
Cohen's kappa		KGE	
Range	Category	Range	Category
PM < 0	No agreement	PM < 0	Poor
0 < PM <0.21	Slight	0 < PM <0.38	Fair
0.21 < PM <0.41	Fair	0.38 < PM <0.55	Moderate
0.41 < PM <0.61	Moderate	0.55 < PM <0.66	Good
0.61 < PM <0.81	Substantial	0.66 < PM <0.78	Very good
0.81 < PM <1	Almost perfect	0.78 < PM <1	Excellent

Data pre-processing

We performed an explanatory data analysis to select a suite of watershed attributes to use as input variables for the clustering algorithm and to remove the redundant attributes. We dropped attributes that were highly correlated to each other (i.e., Pearson correlation>0.8) and reduced the dimensionality of our data set to a more refined set of attributes. We also dropped attributes that had negligible influence on clustering using Principal Component Analysis (PCA).

Self-Organizing Map (SOM), Unified-distance matrix and feature component planes

The SOM is an nonparametric, unsupervised clustering algorithm with advantages over other clustering/classifying methods because of its ability to visualize associations between data. Because of its superior performance over parametric methods for classifying or clustering data of varying types (e.g., continuous, ordinal, nominal), scales, and distributions, the SOM has seen increasing application in the fields of geology (Karymbalis et al, 2010), lake ecology (Pearce et al., 2013), and hydrology (Ley et al., 2011). The SOM iteratively self-organizes the input data (here watershed attributes) to a lower dimensional space, typically a 2-D plane or lattice that can be more easily interpreted by domain experts to explore and visualize their data. For SOM algorithm details, see Kohonen (1990, 2001, 2013). The lattice of a SOM is not a map in the cartographic sense of the word. The x and y axes of these 2-D grids have no specific unit-of-measurement or label, and the absolute distance between any two observations in one part of the map is not the same as in another part of the SOM lattice. Rather, the SOM is designed to map the abstract search space of a given data set. Individual observations are fed into the algorithm and processed to cluster similar observations onto the 2-D lattice in such a way that observations with similar feature patterns (e.g., watershed-scale attributes) will appear closer to one another on the resulting self-organized lattice. Clusters of like observations thus emerge and can be viewed and interpreted by the analyst. Because the algorithm is unsupervised, there are no pre-determined number of clusters; rather relationships among the attribute values drive the clustering.

The "unified distance matrix" (U-matrix), developed by Ultsch and Siemon (1989), is a visualization tool that aids with interpretation of boundaries between clusters. The U-matrix is analogous to topographic shading on a geological map such that darker shading indicates greater differences/boundaries between clusters and lighter shading indicates smaller distances. It is used to display the topological relationships between the nodes in the SOM grid and for identifying patterns in the data. To create the U-matrix, the average distance between each grid node and its neighbors (after training) is computed and assigned to the corresponding grid node. The resulting matrix is visualized as a grayscale image in this work.

The feature component planes are another visualization tool to display and analyze the SOM results. It is a two-dimensional representation of the input data that shows the distribution of each input feature across the trained SOM grid. Feature component planes are created by plotting each input feature as a separate map on the trained SOM lattice, where the value associated with each node represents the average value of that feature for observations that map/cluster to that node. The feature component plane is useful for identifying features (here watershed attributes) that are driving the clustering of the input data. When a target variable (e.g., model performance) is superimposed on the map, it is possible to visualize patterns between the target variable and individual features (i.e., component plane).

In this work, we used a SOM to test whether watersheds that clustered on their physical attributes have identifiable patterns (cluster) with model performance. We used the 'MiniSom' Python package version 2.3.1 (Vettigli, 2018) to train a SOM with a hexagonal lattice topology and lattice size of 10 by 10. The SOM has an initial learning rate of 0.7 and a neighborhood size that decreased linearly from half of the lattice to zero. Training was performed using 1000 iterations. We used the reduced set of watershed attributes (after dropping the redundant attributes) as inputs to the SOM. The LSTM model performance measures (Kappa and KGE metrics) were not used as inputs to the SOM. Instead, once the algorithm had converged, and the observations had self-organized onto the 2-D lattice, the model performance associated with each observation was mapped as a color-coded circle on the lattice. To determine the optimal number of clusters, we then used a subsequent supervised SOM algorithm to classify the trained weights into five groups. To verify watershed clustering, we then visualized the assigned clusters in cartographic space to interpret geographic patterns. Finally, we reviewed the component planes for each of the watershed attributes to explore linkages between the watershed attributes and model performance.

Feature Importance

The set of 50 attributes were then used as inputs to random forest regression models with Cohen's Kappa performance measure for both fixed and variable threshold models as the predictor variable. Random forest is an ensemble learning algorithm that may be employed for identifying feature importance. It uses the Gini importance index or permutation importance index for feature importance measures and ranks the variables based on their contribution to the predictive ability of the model (Breiman, 2001). We used the predictor screening tool in JMP (JMP®, Version 15.0.0. SAS Institute Inc., Cary, NC, 1989–2021) with 1000 decision trees to find the attributes rankings. The predictor screening platform uses random forest to identify and rank potential predictors associated with the target variable, in this work identifying watershed attributes most important for predicting model performance.

Predicting Model Performance from Watershed Attributes

In addition to identifying feature importance (most associated with model performance), we also wanted to investigate using the reduced set of attributes for predicting likely model performance at “ungaged” locations. For this purpose, we trained a random forest regressor in JMP using the bootstrapped forest platform. We held back 30% of the data for testing and used 70% of the data for training. Different number of decision trees, fine-tuned automatically by JMP, were used for different models.

Results & Discussion

Clustering watersheds and identifying meaningful groups of model performance

Based on inputs of 50 watershed attributes, the 409 streamflow gages were grouped into five clusters (Figure 1). We can observe that the clusters obtained through Self-Organizing Maps (SOM) analysis of watersheds showed a clear geographical organization. Specifically, clusters 3, 4, and 5 were mainly concentrated in the upper Colorado basin and Rocky Mountains, which are characterized by higher elevations. On the other hand, clusters 1 and 2 covered the lower Colorado basin and the surrounding lower elevation areas. Cluster 1 was the largest and most widespread in terms of geographic distribution among the clusters.

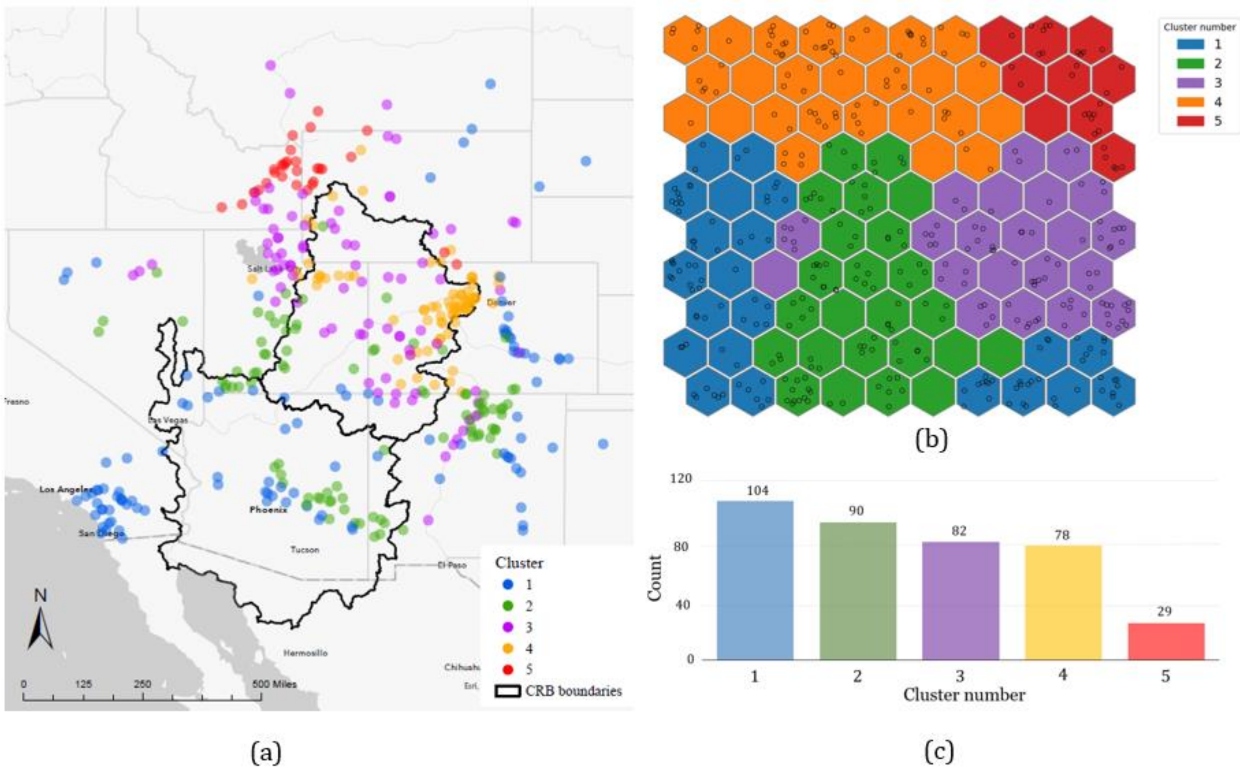

Figure 1. Map of USGS streamgages in the Colorado River Basin and surrounding area; circles are color-coded based on cluster number, (b) The SOM grid, color-coded based on the clusters of trained SOM weights, (c) Histogram of USGS streamgages showing the number of gages in each cluster.

We then used the trained SOM weights to calculate the U-matrix and explore for observable patterns of model performance color-coding clustered observations with respect to model performance on top of the U-matrix. Figures 2 to 5 depict the SOM U-matrix, showcasing the model performance categories superimposed on it across different types of thresholds (fixed and daily variable) and performance measures (Cohen's kappa and KGE). Our analysis of performance metrics did not reveal clear patterns or clusters on the SOM grid, suggesting that static watershed attributes (averaged over 40-years of record) may not be a strong predictor of model performance. However, when examining the four performance measures, we did observe some trends associated with higher and lower performance levels. Furthermore, the performance measures were generally higher for the fixed threshold models compared to the variable threshold models.

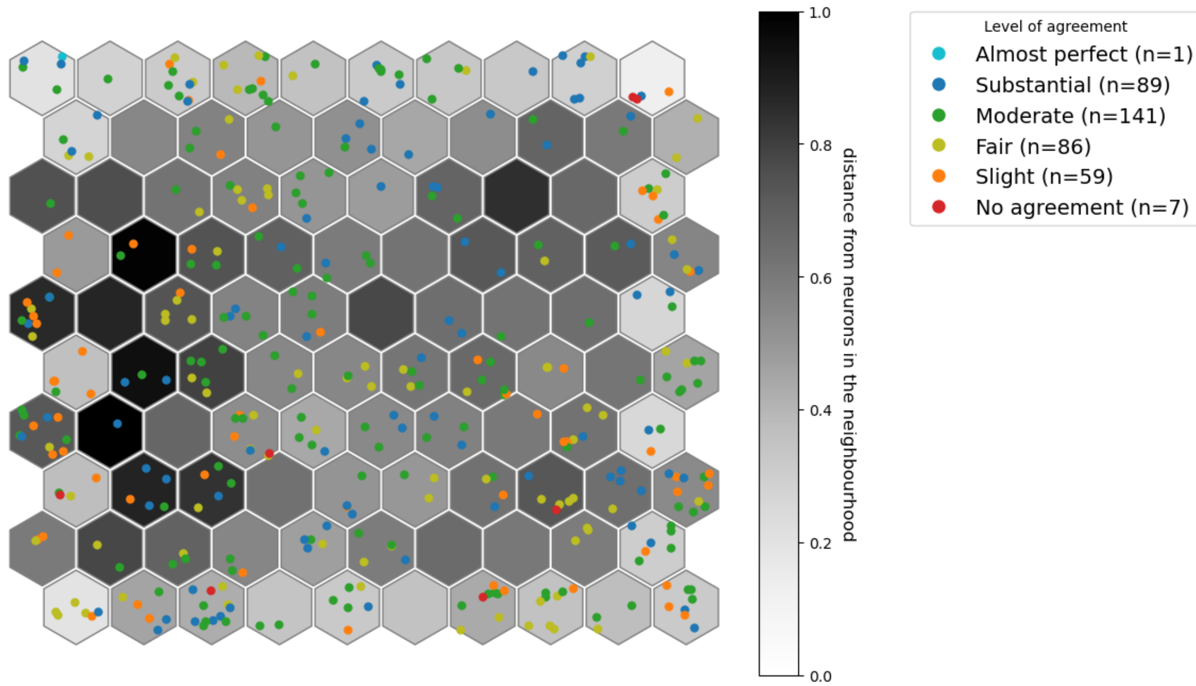

Figure 2. The U-matrix of the SOM, for the model with a fixed threshold, has been overlaid with the performance categories based on Cohen's kappa. Performance metric categories provided in Table 1.

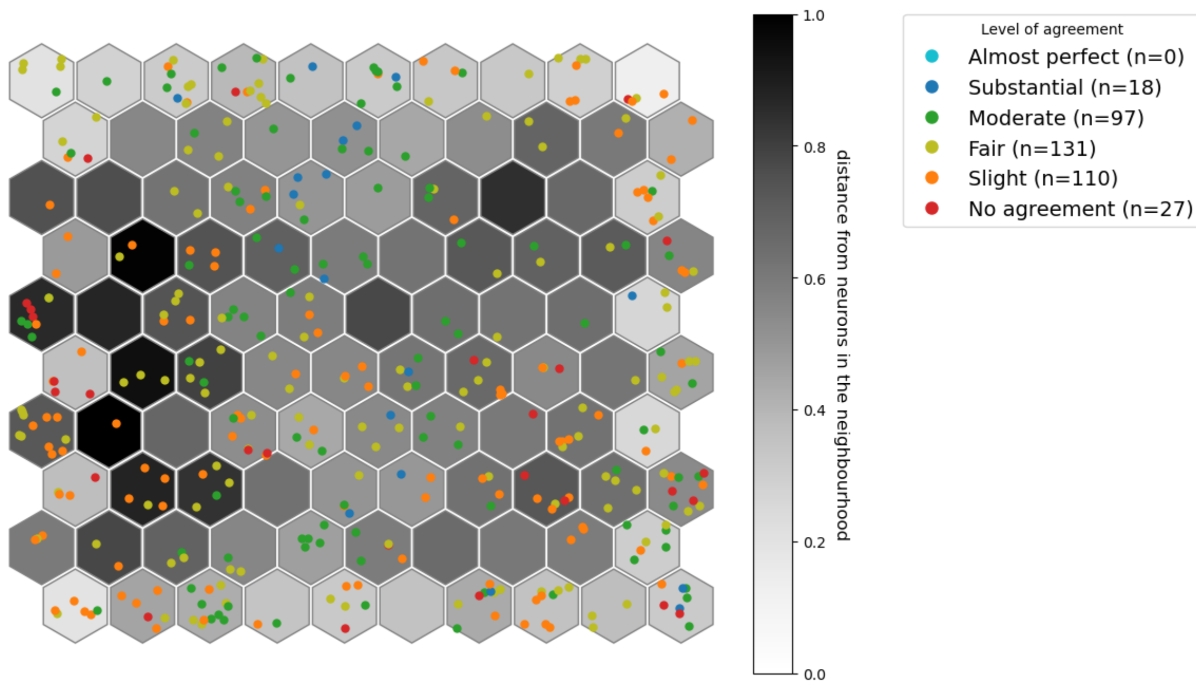

Figure 3. The U-matrix of the SOM, for the model with a daily variable threshold, has been overlaid with the performance categories based on Cohen's kappa

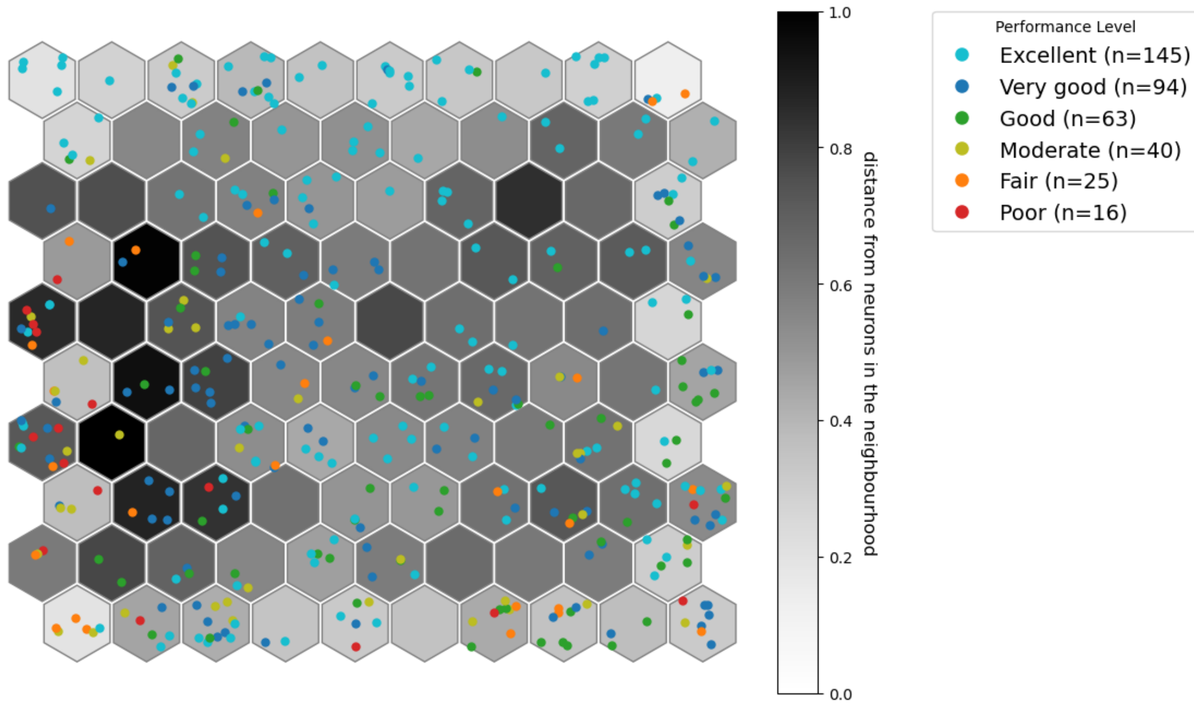

Figure 4. The U-matrix of the SOM, for the model with a fixed threshold, has been overlaid with the performance categories based on KGE

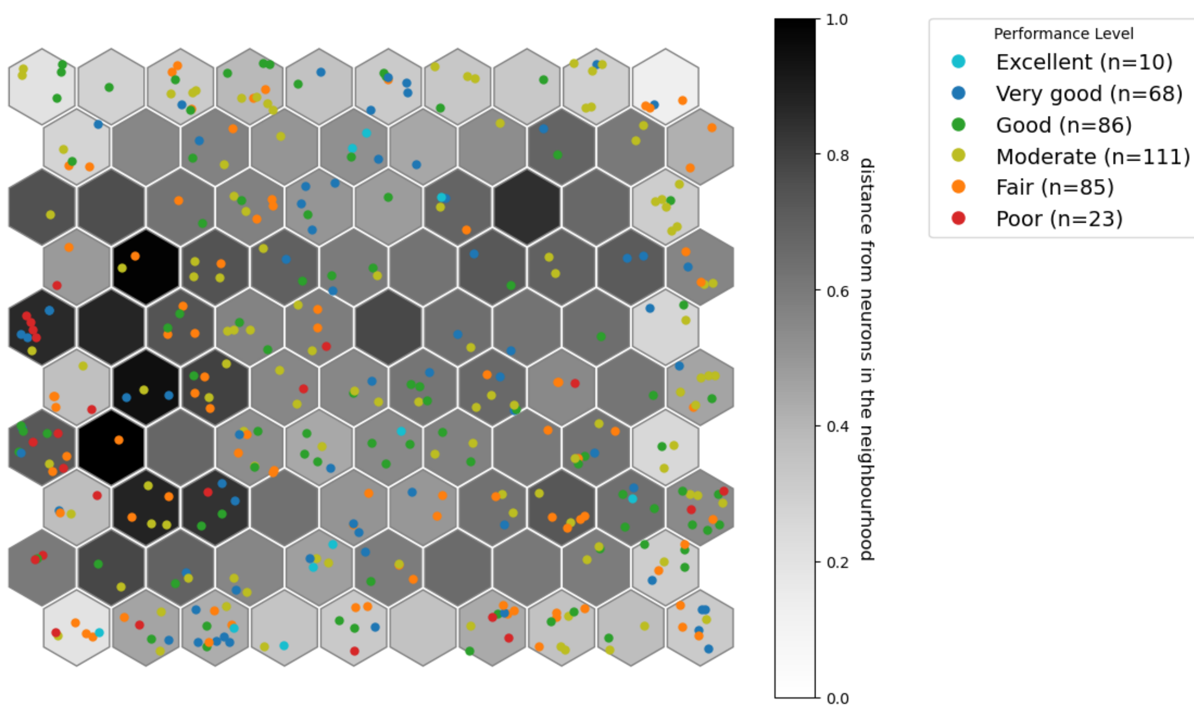

Figure 5. The U-matrix of the SOM, for the model with a daily variable threshold, has been overlaid with the performance categories based on KGE

Identifying linkages between Model Performance & Watershed Attributes

In this section, we present examples of using the feature component planes corresponding to watershed attributes that when overlaid with model performance metrics help interpret relationships between attributes and model performance. The model performance based on Cohen's kappa categories have been superimposed onto these feature component planes, allowing for an examination of the relationships between model performance and individual watershed attributes. Based on results of random forest (next section), we select the reservoir storage intensity (MAXDI_EROM in the list of variables in the Appendix), watershed mean elevation (TOT_ELEV_MEAN) and percentage of forest land cover (TOT_NLCD19_FOREST) to illustrate this process. We focus here specifically on the model performance results from the variable threshold model and Cohen's kappa measure of performance.

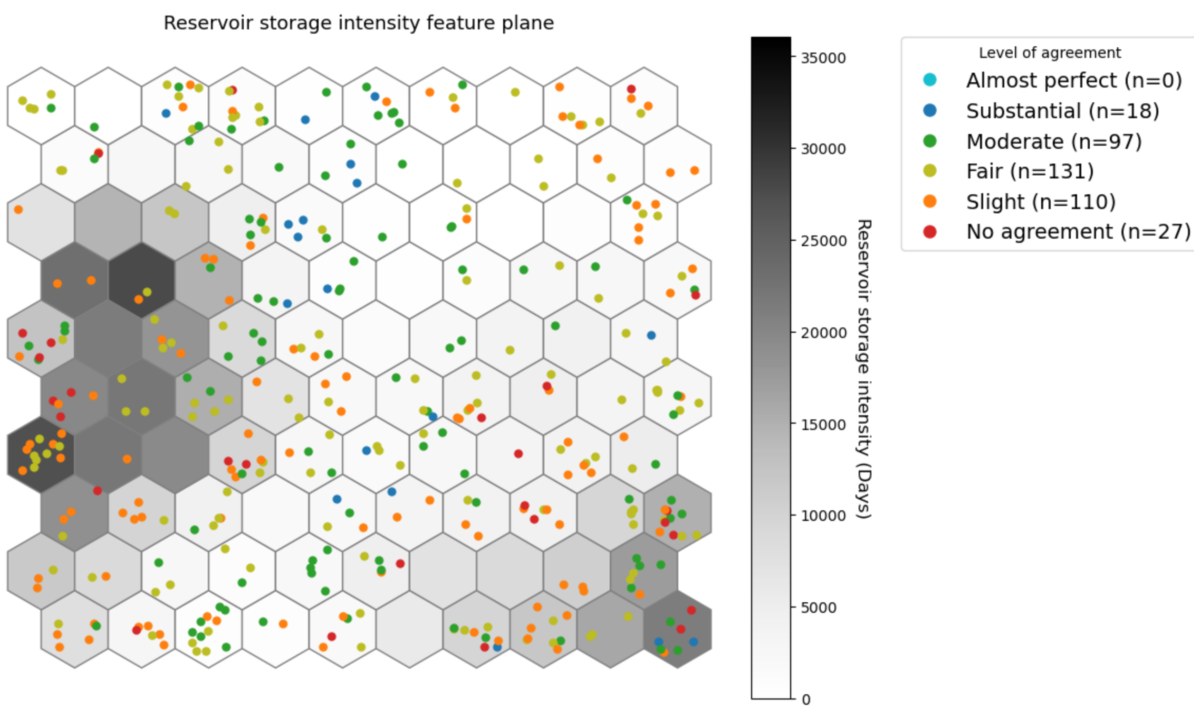

Figure 6. The feature component plane of the reservoir storage intensity, for the model with the daily variable threshold, has been overlaid with the performance categories based on Cohen's kappa

High model performance (substantial category) is mostly corresponding to watersheds with lower reservoir storage intensity (Figure 6) found within Clusters 4 and 2 (Figure 1). This finding may indicate the degree of flow modification in a watershed is influencing model performance. We also see that watersheds with better model performance are mostly contained to higher elevation (Figure 7). However, there are a number of poor performing high elevation watersheds as well, suggesting that factors other than elevation are influencing the accuracy of drought predictions based on the current LSTM models. Based on the component plane of forest cover (Figure 8), high elevation and high forest cover both correspond to higher model performance.

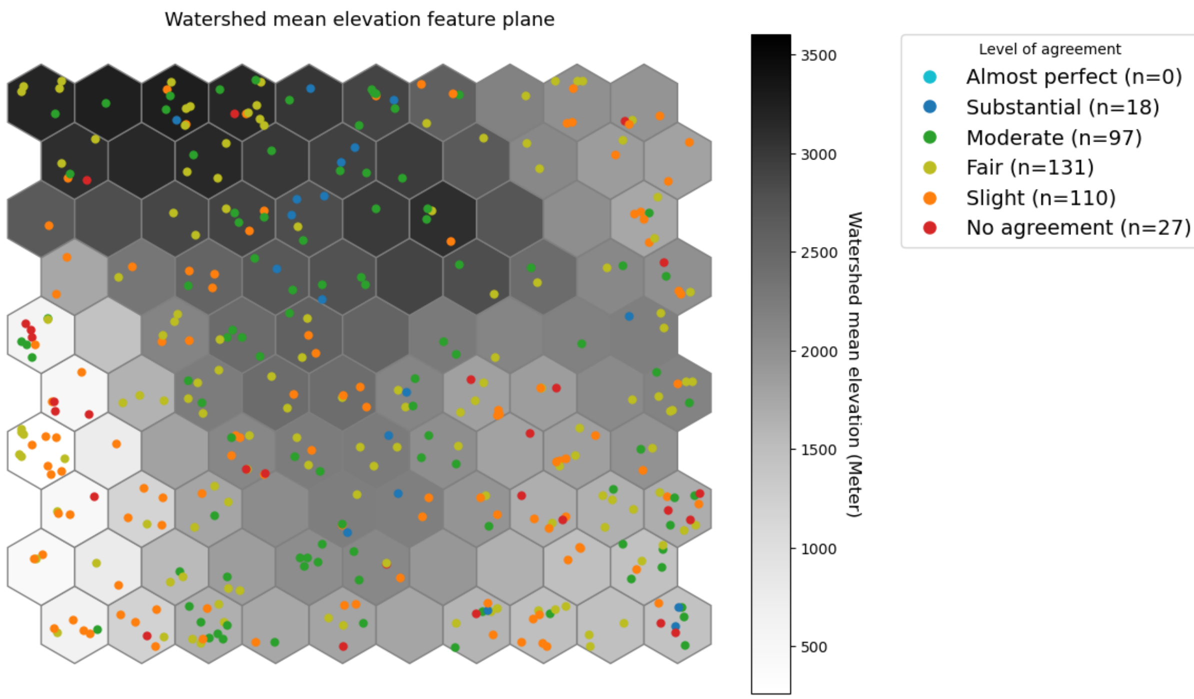

Figure 7. The feature component plane of the watershed mean elevation, for the model with the daily variable threshold, has been overlaid with the performance categories based on Cohen's kappa

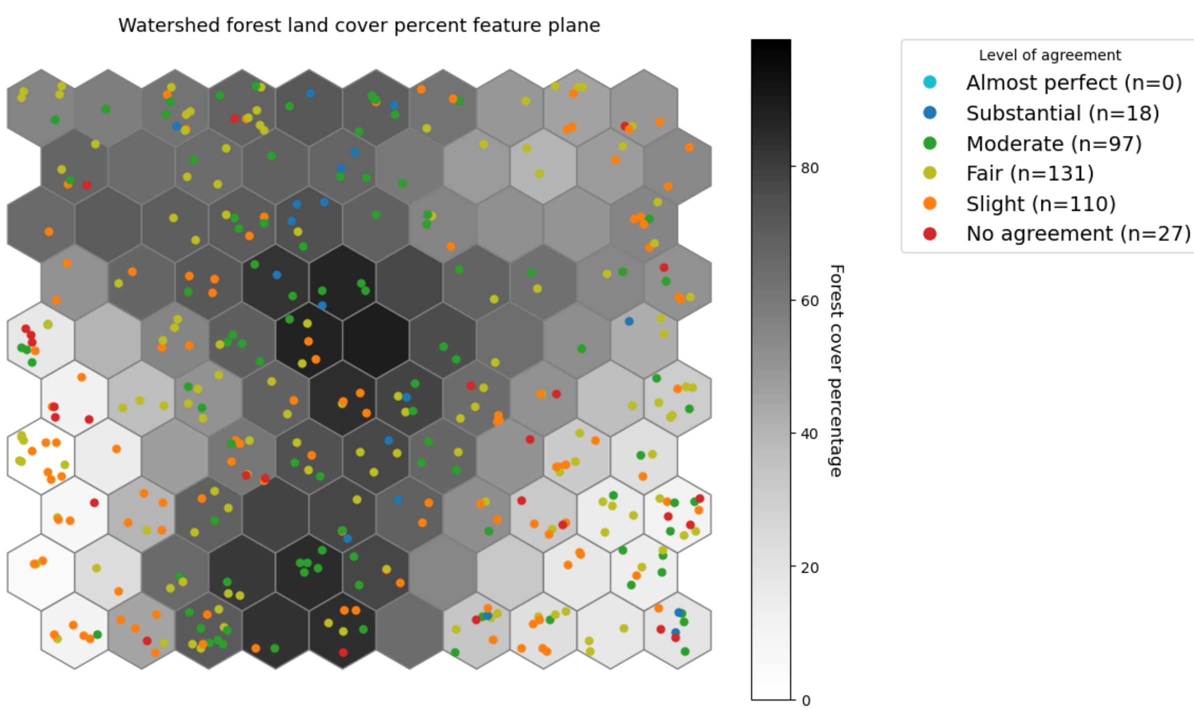

Figure 8. The feature component plane of the watershed forest land cover percent, for the model with the daily variable threshold, has been overlaid with the performance categories based on Cohen's kappa

Assessing Feature Importance for Predicting Model Performance

In this section, we used predictor screening random forest models to find the most important attributes for predicting Kappa performance measures for fixed and variable threshold models. Feature importance rankings are presented in Figure 9. For the fixed threshold model, reservoir storage intensity (MAXDI_EROM) is found to be the strongest indicator of model performance, whereas for the variable threshold mode, mean elevation (TOT_ELEV_MEAN) is the top feature. The feature importance rankings reveal some differences between what drives strong model performance with the variable and fixed models. For the fixed threshold models, features related to geology – topographic wetness index (TOT_TWI), average percent silt in soils (TOT_SILTAVE), average percent clay in soils (TOT_CLAYAVE), depth to bedrock (TOT_EWT) – are in the top 10 ranked features. In contrast, for variable threshold models, features related to land use/cover – percent forest cover (TOT_NLCD19_FOREST), and percent deciduous forest (TOT_NLCD19_41) are in the top ranked features.

Features related to flow modification, baseflow, amount of precipitation, and extensiveness of road network appear in the top features for both fixed and variable threshold models. Of additional note, drainage area (DA_SQKM) does not appear to be an important predictor of performance, indicating the LSTM models are able to scale to different size basins effectively.

Predictor	kappa (fixed threshold) Contribution	Portion	Rank
MAXDI_EROM	0.475965	0.1210	1
TOT_ELEV_MEAN	0.243345	0.0619	2
TOT_TOTAL_ROAD_DENS	0.242740	0.0617	3
Dam_Index	0.213519	0.0543	4
TOT_BFI	0.155556	0.0395	5
TOT_SILTAVE	0.140701	0.0358	6
TOT_TWI	0.117961	0.0300	7
TOT_MINWD6190	0.117735	0.0299	8
TOT_EWT	0.108102	0.0275	9
TOT_SANDAVE	0.106657	0.0271	10

Predictor	kappa (variable threshold) Contribution	Portion	Rank
TOT_ELEV_MEAN	0.469117	0.1181	1
Dam_Index	0.275376	0.0693	2
TOT_ELEV_MAX	0.243149	0.0612	3
TOT_NLCD19_FOREST	0.207094	0.0521	4
TOT_MAXWD6190	0.207026	0.0521	5
TOT_BFI	0.181168	0.0456	6
TOT_NLCD19_41	0.167855	0.0422	7
TOT_TOTAL_ROAD_DENS	0.164180	0.0413	8
MAXDI_EROM	0.161357	0.0406	9
TOT_ARTIFICIAL	0.146142	0.0368	10

Figure 9. Rankings and contributions of top 10 attributes in predicting kappa statistic for fixed threshold model (upper panel) and variable threshold model (lower panel)

Predicting Model Performance from Watershed Attributes

The reduced set of 50 attributes was tested for ability to predict model performance at ungaged sites by withholding a portion of gages (validation set) during training of the random forest models. The R^2 of the models were 0.293 and 0.351 for predicting the Cohen's kappa statistics for fixed and daily variable threshold, respectively (Figure 10). Interestingly, while variable threshold model performance is on average lower than the fixed threshold mode, the predictability of where a fixed threshold model will be more accurate is lower than variable threshold model. While watershed attributes do show some predictive ability of the streamflow drought LSTM model performance, additional factors not captured in the watershed attributes are likely needed to reliably predict model performance.

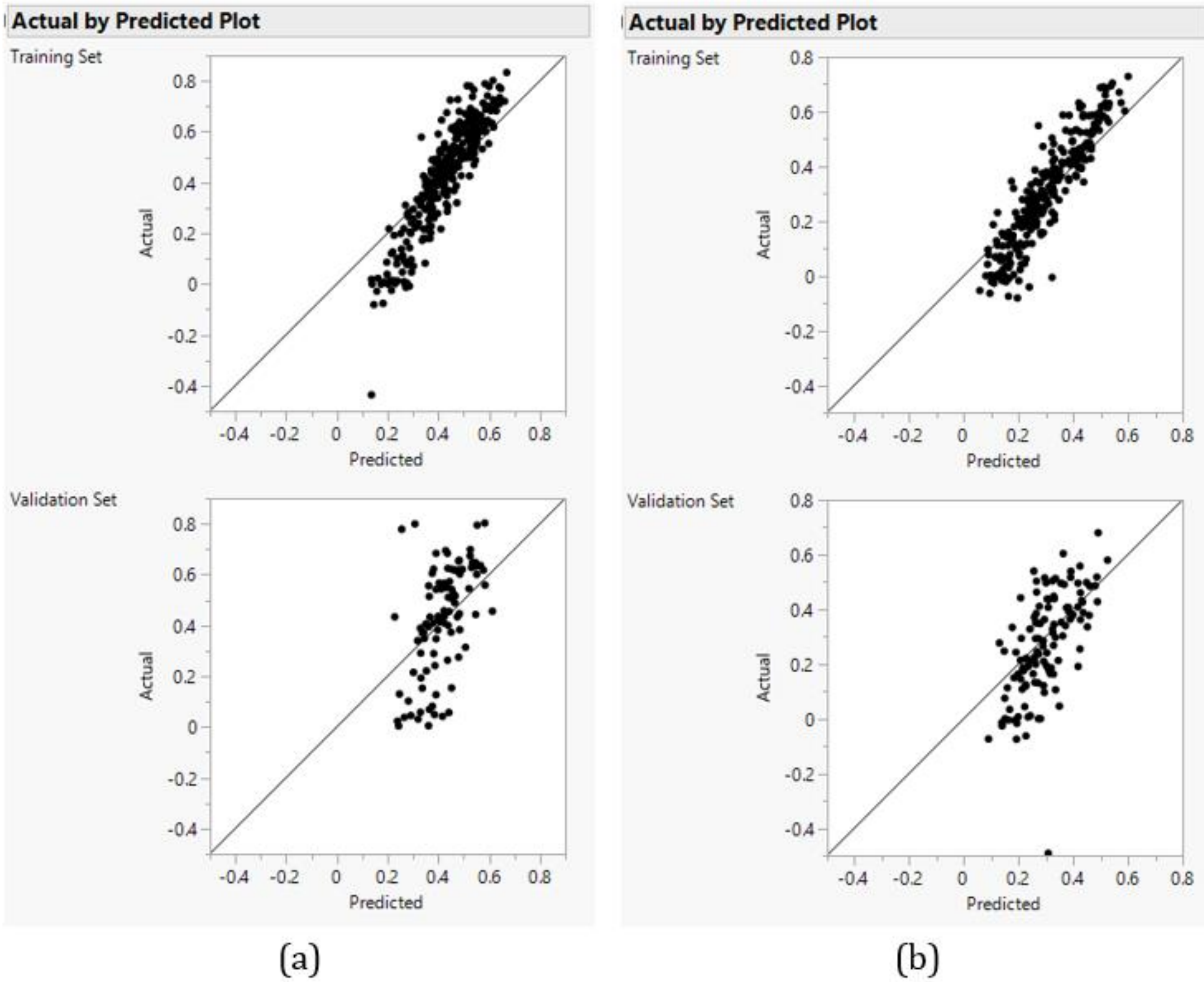

Figure 10. Random Forest actual vs predicted results for Cohen's kappa of (a) fixed threshold models and (b) variable threshold models.

Further analysis could include calculating variables that describe the timeseries and drought event variability more directly. Holistically, this work develops a workflow that the USGS can utilize to prioritize areas for further data collection (e.g., on water use/irrigation) and need for model improvement. Additionally, the further development of a model to predict model performance can aid in estimating areas of likely unreliable model outputs, thereby improving the utility and trustworthiness of a drought early warning system.

Acknowledgements

This research was funded in part by funding from National Science Foundation (NSF) under Vermont EPSCoR Grant No. EPS-1101317, and Grant No. NSF-EAR-2012123, Grant No. 2202706, and through the NOAA cooperative agreement with Alabama (Award NA22NWS4320003)We would like to extend our heartfelt gratitude to Scott

Hamshaw whose valuable insights and assistance were integral to the completion of this paper and thanks to John Hammond and Jared Smith for their suggestions and edits that improved this article.

References

Breiman, L. 2001. “Random forests,” Machine learning, 45(1): 5-32.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. 2009. “Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling,” *Journal of hydrology*, 377(1-2): 80-91.

Hamshaw, S., Watkins, W., and White, E. 2023. “Data-Driven Drought Prediction Project Model Outputs: Daily Streamflow Percentile Predictions for the Colorado River Basin Region,” doi: 10.5066/P97NIH7Y

Karymbalis, E., Gaki-Papanastassiou, K., & Ferentinou, M. 2010. “Fan deltas classification coupling morphometric analysis and artificial neural networks: The case of NW coast of Gulf of Corinth, Greece,” Hellenic Journal of Geosciences, 45: 133-146.

Kohonen, T. 1990. “The self-organizing map,” Proceedings of the IEEE, 78(9): 1464-1480.

Kohonen, T. 2001. “Self-organizing maps of massive databases,” International journal of engineering intelligent systems for electrical engineering and communications, 9(4): 179-186.

Kohonen, T. 2013. “Essentials of the self-organizing map,” Neural networks, 37: 52-65.

Kursa, M. B., & Rudnicki, W. R. 2010. “Feature Selection with the Boruta Package,” Journal of Statistical Software, 36(11): 1–13. https://doi.org/10.18637/jss.v036.i11

Landis, J. Richard, and Gary G. Koch. 1977. “The Measurement of Observer Agreement for Categorical Data,” *Biometrics* 33 (1): 159. https://doi.org/10.2307/2529310

Ley, R., Casper, M. C., Hellebrand, H., & Merz, R. 2011. “Catchment classification by runoff behaviour with self-organizing maps (SOM),” Hydrology and Earth System Sciences, 15(9): 2947-2962.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. 2015. “Hydrologic and water quality models: Performance measures and evaluation criteria,” Transactions of the ASABE, 58(6): 1763-1785.

Nash, J. E., & Sutcliffe, J. V. 1970. “River flow forecasting through conceptual models part I—A discussion of principles,” Journal of hydrology, 10(3): 282-290.

Pearce, A. R., Rizzo, D. M., Watzin, M. C., & Druschel, G. K. 2013. “Unraveling associations between cyanobacteria blooms and in-lake environmental conditions in Missisquoi Bay, Lake Champlain, USA, using a modified self-organizing map,” Environmental science & technology, 47(24): 14267-14274.

Sarailidis, G., Vasiliades, L., & Loukas, A. 2019. “Analysis of streamflow droughts using fixed and variable thresholds,” Hydrological processes, 33(3): 414-431.

Sutanto, S. J., & Van Lanen, H. A. 2021. “Streamflow drought: Implication of drought definitions and its application for drought forecasting,” Hydrology and Earth System Sciences, 25(7): 3991-4023.

Sutanto, S. J., & Van Lanen, H. A. 2022. “Catchment memory explains hydrological drought forecast performance,” Scientific reports, 12(1): 1-11.

Ultsch, A., & Simeon, H. P. 1989. “Exploratory Data Analysis Using Kohonen Networks on Transputers,” Department of Computer Science, University of Dormund, FRG.

Underwood, K. L., Rizzo, D. M., Dewoolkar, M. M., & Kline, M. 2021. “Analysis of reach-scale sediment process domains in glacially-conditioned catchments using self-organizing maps,” Geomorphology, *382*: 107684.

Van Loon, A. F. 2015. “Hydrological drought explained,” Wiley Interdisciplinary Reviews: Water, 2(4): 359-392.

Vettigli, G. (2018). “MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map,” URL: https://github. com/JustGlowing/minisom.

Wieczorek, M. E., Wolock, D. M., & McCarthy, P. M. 2020. “Dam impact/disturbance metrics for the conterminous United States, 1800 to 2015,” USGS Data Release. Reston, VA: USGS.

Yildirim, G., Rahman, A., & Singh, V. 2022. “A Bibliometric Analysis of Drought Indices, Risk, and Forecast as Components of Drought Early Warning Systems,” Water, 14(2): 253. https://doi.org/10.3390/w14020253

Appendix

Table 1. List of watershed attributes and their description

Variable Name	Description
DA_SQKM	Drainage area
MAXDI_EROM	Estimates reservoir storage intensity in units of days based on reservoir storage in a contributing area normalized by the mean annual streamflow. This metric indicates the duration of storage impact upstream from each stream segment relative to the typical flow condition
Dam_Index	Represents the degree of regulation of a river reach based on upstream reservoir storage relative to the 30-year average annual precipitation, as well as the upstream dam and watershed areas
TOT_ELEV_MEAN	Watershed's mean elevation in meters.

TOT_ELEV_MIN	Watershed's minimum elevation in meters.
TOT_ELEV_MAX	Watershed's maximum elevation in meters.
TOT_STREAM_SLOPE	Watershed's average flowline slope in percent.
TOT_STREAM_LENGTH	Watershed's total flowline length in kilometers.
TOT_FSTFZ6190	Watershed average of mean day of the year of first freeze, derived from 30 years of record (1961-1990), 2km PRISM. For example, value of 300 is the 300th day of the year (Oct 27th).
TOT_LSTFZ6190	Watershed average of mean day of the year of last freeze, derived from 30 years of record (1961-1990), 2km PRISM. For example, value of 100 is the 100th day of the year (April 10th).
TOT_MAXP6190	Watershed maximum average annual precipitation in mm
TOT_MAXWD6190	Watershed average of maximum monthly number of days of measurable precipitation, derived from 30 years of record (1961-1990), 2.3km PRISM. This is simply the maximum value of the values WD_JAN_BASIN thru WD_DEC_BASIN.
TOT_MINWD6190	Watershed average of minimum monthly number of days of measurable precipitation, derived from 30 years of record (1961-1990), 2.3km PRISM. This is simply the minimum value of the values WD_JAN_BASIN thru WD_DEC_BASIN.
TOT_PRSNOW	Snow percent of total precipitation estimate, mean for period 1901-2000. From McCabe and Wolock (submitted, 2008), 1km grid.
TOT_RH	Watershed average relative humidity (percent), from 2km PRISM, derived from 30 years of record (1961-1990).
TOT_AET	Watershed mean-annual evapotranspiration, estimated by Senay and others (2013)
TOT_CWD	Watershed's 30-year average number of consecutive days with measurable precipitation per NHDPlus version 2 catchment.
TOT_WDANN	Watershed's value for the 30-year annual average (1961-1990) number of days of measurable precipitation. -9999 denotes NODATA or source data does not cover catchment.
TOT_BFI	Watershed's Base Flow Index (BFI), The BFI is a ratio of base flow to total streamflow, expressed as a percentage and ranging from 0 to 100. Base flow is the sustained, slowly varying component of streamflow, usually attributed to ground-water discharge to a stream.
TOT_CONTACT	Watershed's Subsurface flow contact time index. The subsurface contact time index estimates the number of days that infiltrated water resides in the saturated subsurface zone of the basin before discharging into the stream.
TOT_IEOF	Watershed's Percentage of Horton overland flow as a percent of total flow
TOT_RECHG	Watershed's Mean annual natural ground-water recharge in millimeters per year
TOT_SATOF	Watershed's percentage of Dunne overland flow as a percent of total flow
TOT_TWI	Watershed's average Topographic wetness index, ln(a/S); where ln is the natural log, a is the upslope area per unit contour length and S is the slope at that point. See http://ks.water.usgs.gov/Kansas/pubs/reports/wrir.99-4242.html and Wolock and McCabe, 1995 for more detail
TOT_EWT	Watershed's Average depth to water table relative to the land surface(meters)

TOT_RF7100	Watershed's mean annual average for the Rainfall and Runoff factor ("R factor" of Universal Soil Loss Equation) for the period 1971-2000 in hundreds of foot-ton force-inch/acre-hour per year for the period 1971-2000
TOT_WB5100_ANN	Watershed's Average annual runoff (mm) from McCabe and Wolock's Runoff Model 1951-2000
TOT_MIRAD_2012	Percent of watershed in irrigated agriculture, from USGS 2012 250-m MODIS data
TOT_FRESHWATER_WD	Watershed's freshwater withdrawals from 1995-2000 county-level estimates
TOT_STREAMRIVER	Watershed's percentage of all flowlines reach lengths per NHDPlusV2 catchment that is a stream or river.
TOT_ARTIFICIAL	Watershed's percentage of all flowlines reach lengths per NHDPlusV2 catchment that is an artificial reach. An artificial path is a flowline feature that represents the assumed and generalized flow through a 2-dimensional feature, such as a lake or a wide double-banked stream. The artificial path will carry the GNIS name (Geographical Name Information System from NHDPlusV2 flowine shapefile's item GNIS of the major stream it represents and not the waterbody feature.
TOT_CANALDITCH	Watershed's percentage of all flowlines reach lengths per NHDPlusV2 catchment that is a canal.
TOT_CONNECTOR	Watershed's percentage of all flowlines reach lengths per NHDPlusV2 catchment that is a connector. The NHDFlowline feature type connector establishes a known, but non-specific (unseen) connection between two non-adjacent geometric network (flowline) segments that have flow. These features are usually associated with the results of a dye-trace injection that ties the two surface water features together through some groundwater connection.
TOT_PIPELINE	Watershed's percentage of all flowlines reach lengths per NHDPlusV2 catchment that is a pipe line. Pipelines are specifically man-made structures of steel, concrete, or polymers that direct surface water flows from one area to another.
TOT_STRM_DENS	Watershed's flowline catchment stream density calculated as stream length (meters) divided by catchment(s) area (square kilometers).
TOT_RDX	Number of roads to stream crossings per watershed.
TOT_TOTAL_ROAD_DENS	Density of all road types per watershed. Density is defined as the length of road divided by the catchment area.
TOT_HGA	Percentage of Hydrologic Group A soil. -9999 denotes NODATA, usually water. Hydrologic group A soils have high infiltration rates. Soils are deep and well drained and, typically, have high sand and gravel content.
TOT_HGB	Percentage of Hydrologic Group B soil. -9999 denotes NODATA, usually water. Hydrologic group B soils have moderate infiltration rates. Soils are moderately deep, moderately well drained, and moderately coarse in texture.
TOT_HGC	Percentage of Hydrologic Group C soil. -9999 denotes NODATA, usually water. Hydrologic group C soils have slow soil infiltration rates. The soil profiles include layers impeding downward movement of water and, typically, have moderately fine or fine texture.
TOT_HGD	Percentage of Hydrologic Group D soil. -9999 denotes NODATA, usually water. Hydrologic group D soils have very slow infiltration rates. Soils are clayey, have a high water table, or have a shallow impervious layer.
TOT_SILTAVE	Average percent of silt in soil per watershed

TOT_CLAYAVE	Average percent of clay in soil per watershed.
TOT_SANDAVE	Average percent of sand in soil per watershed.
TOT_KFACT	Average value for the K factor per watershed.
TOT_KFACT_UP	Average value for the K factor in the upper soil horizon per watershed.
TOT_NO10AVE	Average percent by weight of soil material less than 3 inches in size that passes through a No. 10 sieve (2 millimeters) per watershed.
TOT_NO200AVE	Average percent by weight of soil material less than 3 inches in size that passes through a No. 200 sieve (.074 millimeters) per watershed.
TOT_NO4AVE	Average percent by weight of soil material less than 3 inches in size that passes through a No. 4 sieve (5 millimeters) per watershed.
TOT_OM	Average value for the range in organic matter content (percent by weight) per watershed.
TOT_PERMAVE	Average value for the range in permeability (inches per hour) per watershed.
TOT_RFACT	Average Rainfall Runoff Factor from Universal Soil Loss Equation per watershed.
TOT_ROCKDEP	Average value for the range in the total soil thickness examined (inches) per watershed.
TOT_BDAVE	Average value for the value of bulk density (grams per cubic centimeter) per watershed.
TOT_AWCAVE	Average value for the range in available water capacity (fraction) per watershed.
TOT_WTDEP	Average value for the range in depth to the seasonally high water table (feet) per watershed.
TOT_SRL25AG	Estimated percent presence of the soil restrictive layer in the upper 25 centimeters of agricultural land per watershed.
TOT_NLCD19_11	2019 watershed percentage of land-use and land-cover type Open Water: All areas of open water, generally with less than 25 percent cover of vegetation or soil.
TOT_NLCD19_12	2019 watershed percentage of land-use and land-cover type Perennial Ice/Snow: All areas characterized by a perennial cover of ice and/or snow, generally greater than 25 percent of total cover.
TOT_NLCD19_52	2019 watershed percentage of land-use and land-cover type Shrub/Scrub: Areas dominated by shrubs less than 5 meters tall. Shrub canopy is typically greater than 20 percent of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions.
TOT_NLCD19_71	2019 watershed percentage of land-use and land-cover type Grassland/Herbaceous: Areas dominated by graminoid or herbaceous vegetation, generally greater than 80 percent of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
TOT_NLCD19_81	2019 watershed percentage of land-use and land-cover type Pasture/Hay: Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.
TOT_NLCD19_82	2019 watershed percentage of land-use and land-cover type Cultivated Crops: Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for

	greater than 20 percent of total vegetation. This class also includes all land being actively tilled.
TOT_NLCD19_DEVELOPED	2019 watershed percentage of land-use and land-cover type Developed (either Open, Low Intensity, Medium Intensity, or High Intensity)
TOT_NLCD19_FOREST	2019 watershed percentage of land-use and land-cover type Forest (either Mixed, Evergreen, or Deciduous)
TOT_NLCD19_WETLAND	2019 watershed percentage of land-use and land-cover type Wetland (either Herbaceous or Woody)