Comment 4, Reviewer1, Additional Information:

Here are the rearranged (updated) titles:

2.6 Parameter estimation for deep and wide artificial neural network architectures

2.6.1 Strategies for improving the tractability of parameter estimation

2.6.1.1 Greedy learning

2.6.1.2 Extreme Learning Machine configuration

2.6.1.3 Regularization for robust estimation for hidden node selection

3 Metrics for evaluation of intermittent flow predictions

3.1 How well does the model predict zero-flow states?

3.2 How well does the model predict no-flow persistence?

3.3 How well does the model predict transitions to and from zero-flow states?

3.4 How well does the model predict non-zero flowrates?

4 Input specification for deep and wide ANNs for predicting intermittent streamflows

5 Model evaluation testbeds

6 Results and discussion

6.1 Model calibrations and testing

6.2 Deep and wide topologies for predicting zero-flow states

6.2.1 Comparison of zero flow predictions with shallow model

6.3 Comparison of no-flow state persistence and transitions

6.4 Deep and wide topologies for IRES non-zero flowrate prediction

7 Summary and conclusions

Comment 6, Reviewer1, Additional Information:

Here is the material that are moved to the supplementary material:

Table S1. Contingency Table for Evaluation of Classifier Cell Predictions

Contingency Table		Predicted	
		0 (No-Flow)	+ (Flow)
Observed	0 (No-Flow)	N_{00}	N_{0+}
	+ (Flow)	N_{+0}	N_{++}

Table S2. Transition Matrix for Flow and No-Flow States

Transition Matrix		Next State	
		0 (No-Flow)	1 (Flow)
Current State	0 (No-Flow)	P_{00}	P_{0+}
	1 (Flow)	P_{+0}	P_{++}

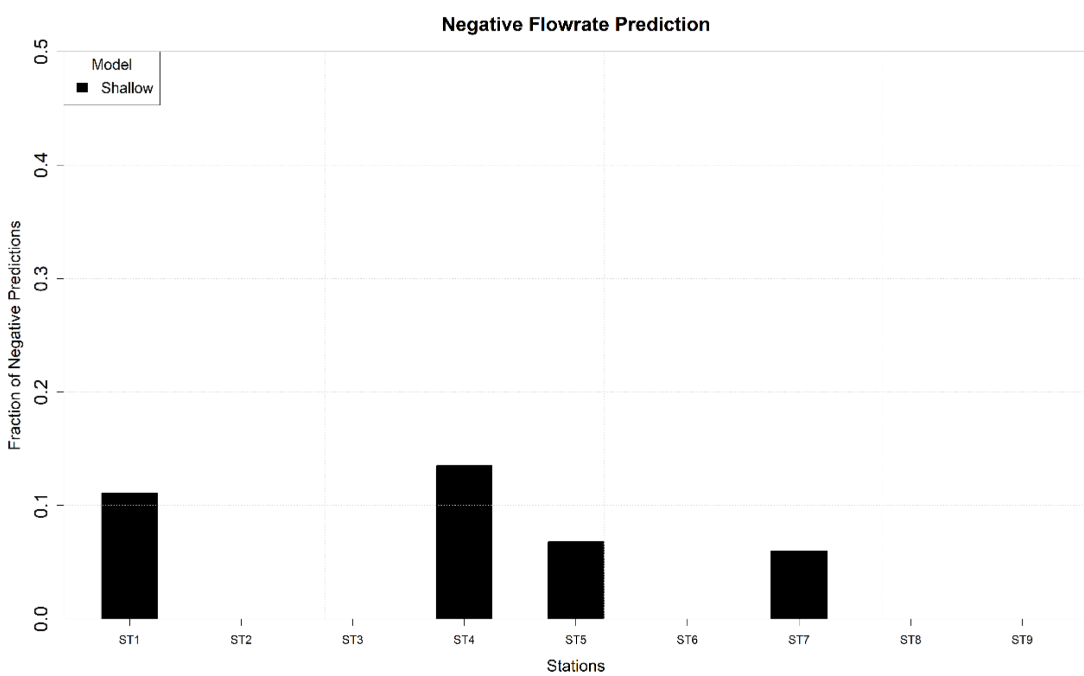

Figure S2. Fraction of negative flow predictions by the shallow, deep and wide models for the nine IRES of study.

Comment 11, Reviewer1, Additional Information:

Here is the TableS3, added to the supplementary material:

Table S3. Statistical features (mean and standard deviation) of the inputs used at the IRES of study.

Station ID	Precipitation (PPT)		Precipitation Lag1 (PPT1)		Soil Moisture Index (SMI)		Soil Moisture Index Lag1 (SMI1)		Potential Evapotranspiration (PET)		Potential Evapotranspiration Lag1 (PET1)		VIC-Estimated Runoff (VICR)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ST1	41.24	42.02	82.74	65.11	745.01	943.12	1486.2	1585.25	68.92	59.81	68.61	58.52	2.76	0.8
ST2	43.23	42.72	86.47	66.47	862.6	1002.63	1716.85	1658.35	77.32	62.08	76.69	60.62	6.21	6.43
ST3	42.56	42.6	84.99	63.48	860.7	1003.85	1714.21	1602.58	80.67	63.21	79.94	61.7	5.73	3.8
ST4	53.85	47.47	107.6	73.36	1061.85	1152.67	2116.2	1905.5	80.74	66.51	80.05	64.84	6.83	2.31
ST5	59.01	50.74	117.98	76.12	1153	1197.29	2302.91	1932.36	81.92	65.54	81.17	63.87	5.3	1.96
ST6	66.32	54.4	132.41	80.48	1299.4	1279.94	2593.6	1986.85	85.39	63.99	84.77	62.68	13.25	16.95
ST7	85.32	67.61	170.65	104.01	1518	1501.72	3049.24	2464.16	79.99	65.19	79.38	63.51	19.76	9.68
ST8	53.01	56.92	105.96	84.48	1224.8	1515.13	2448.8	2311.14	104.88	75.56	104.08	73.72	7.1	3.55
ST9	56.35	64.04	112.79	96.56	1389.2	1759.88	2777	2680.92	113.78	73.04	112.89	71.36	2.49	2.37

Comment 2, Reviewer2, Additional Information:

Here is the updated Figure1:

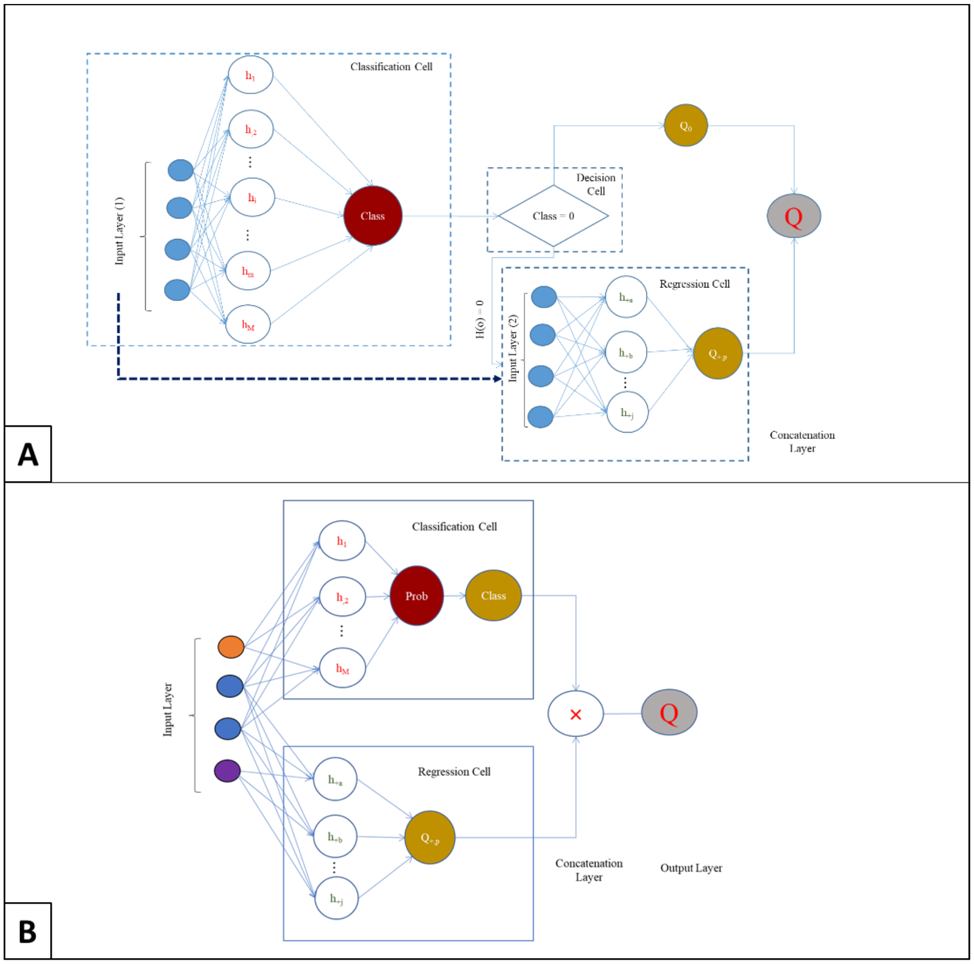

Figure 1. An illustration of the developed architectures for A) the deep model, and B) the wide model for IRES flow prediction.

Comment 5, Reviewer2, Additional Information:

Here is the added introductory section on ANNs:

Artificial Neural Networks (ANN)

ANNs are biologically inspired computational models and parallel-distributed information processing systems (Haykin, 1994; Yaseen et al., 2015). Detailed introductions to ANNs can be found in the literature (Zupan, 1994; Dongare et al., 2012; Adnan et al., 2017). Therefore, and for the sake of brevity, a short overview is provided here. Being known as universal approximators, ANNs has been successfully used in many fields (Wong et al., 1997; Dase and Pawar, 2010; Shrivastava et al., 2012; Qazi et al., 2015; Tealab, 2018), particularly hydrological modeling (Dawson and Wilby, 2001; Joshi and Patel, 2011; Zhang et al., 2018), to model a variety of functions, including streamflow dynamics (Zealand et al., 1999; Kisi, 2007; Gao et al., 2010: Mehr et al., 2015; Li et al., 2019; Malekian and Chitsaz, 2021). A classic single-layer ANN (depicted in figure S1) is a collection of multiple elements, known as units, cells or neurons that are connected in three layers: The input layer, the hidden layer, and the output layer. These layers are connected with links that has weights associated with them. The weights represent the strength (importance) of a connection.

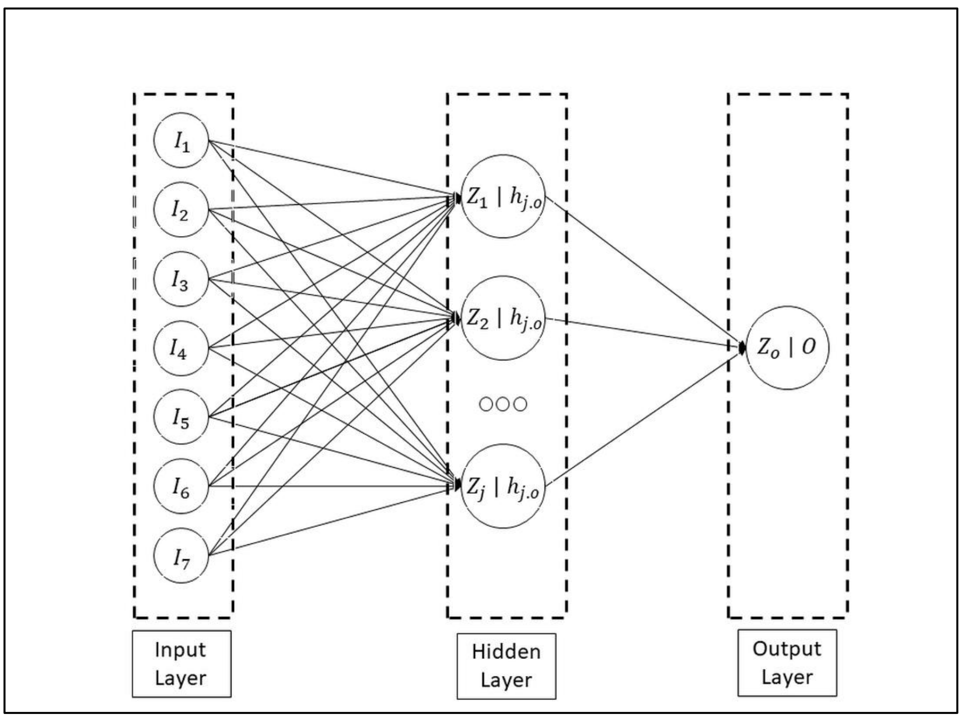

Figure S1. An illustration of a single-layer Artificial Neural Network architecture

The input layer is merely responsible to receive the input information and send it toward the hidden layer. The number of neurons in the input layer is equal to the number of input features (e.g., in the case of having precipitation, potential evapotranspiration, soil moisture index, their first lags, and VIC-estimated runoff, there will be seven input nodes). The nodes in the input layer are fully connected to the nodes in *the hidden layer. Each hidden node receives the product of the input nodes and the weight of the respective connections, sums them up, and then, applies a typically non-linear activation function on it.*

$$Z_j = \sum_{i=1}^{M} w_{i,j} I_i + b_j \quad \forall j = 1, \ldots, H$$

Where Z is the aggregated sum at the j^{th} hidden node, I_i corresponds to the i^{th} input, M is the number of inputs, and H is the total number of hidden nodes within the hidden layer. The bias term (intercept) of the j^{th} node is denoted by, b_j. The output of the hidden node is obtained by passing the aggregated sum from Equation (2) through an activation function:

$$h_{j.o} = g(Z_j)$$

The hidden layer is fully connected to the output layer, meaning that the ultimate product of the hidden layer will be sent to the output layer. The output node also performs a weighted aggregation of the input it receives (i.e., the outputs of the hidden node), and in the case of a binary classifier, it passes the summation value through a sigmoid activation function to obtain a value between 0 and 1.

$$Z_o = \sum_{j=1}^{H} w_{j,o} h_{j,o} + b_o$$

The subscript, o, refers to the output node in the above equation. The output (O) can be computed as:

$$O = \frac{1}{1 + e^{-Z_o}}$$

The value of O is a value between 0 and 1. For a dichotomous variable Q_c, which can take values 0 (no flow) or 1 (flow), the output O corresponds to the probability of obtaining a value of 1 (i.e., flow) or $P(Q_c = 1)$. When the value of O is low (typically below 0.5), there is insufficient evidence of a non-zero flow, and as such Q_c is classified as 0 (i.e., No Flow). When the output O has a value greater than 0.5, Q_c is classified as 1 (i.e., flow). Note that when O is interpreted as $P(Q_c = 1)$, Equation 5 has the same mathematical structure as the Logistic Regression.

In case of a regression problem, as the output of a regression model is continuous and not dichotomous, the activation function for the output cell is taken as linear:

$$Q_p = \left(\sum_{j=1}^{H_r} w_{j,o} h_{j,o} + b_o \right) \times 1$$

Where Q_p is the predicted flowrate within the regression cell, and all other variables have the same meaning as before. Notice that above equation has the same mathematical form as the ordinary linear regression.

Comment 7, Reviewer2, Additional Information:

Here is the updated version of Table 1:

Table 1. Summary of information on the nine streamflow monitoring stations.

Station ID	USGS ID	Stream	Location	Intermittency Ratio	Interquartile Range (cfs)	Maximum Recorded Flowrate (cfs)	Range of Training	Range of Testing
ST1	7233500	Palo Duro Creek	Near Spearman, Hansford County, TX	65%	0.03	197.2	1999/7 – 2015/3	2015/4 – 2020/6
ST2	8079600	Double Mountain Fork Brazos River	At Justiceberg, Garza County, TX	12%	22.15	875.7	1961/12 – 2005/10	2005/11 – 2020/6
ST3	8117995	Colorado River	Near Gail, Borden County, TX	33%	7.59	709.9	1988/3 – 2012/4	2012/5 – 2020/5
ST4	8082700	Millers Creek	Near Munday, TX	41%	1.14	433.8	1963/8 – 2006/3	2006/4 – 2020/6
ST5	8086290	Big Sandy Creek	Above Breckenridge, TX	17%	14.63	1251	1962/2 – 2006/3	2006/4 – 2020/11
ST6	8103900	South Fork Rocky	Near Briggs, Burnet County, TX	24%	10.49	189	1963/4 – 2006/5	2006/6 – 2020/9
ST7	8050840	Range Creek	Near Collinsville, TX	32%	16.65	328.4	1992/10 – 2013/9	2013/10 – 2020/9
ST8	8206700	San Miguel Creek	Near Tilden, TX	32%	12.13	1828	1992/10 – 2013/11	2013/12 – 2020/11
ST9	8212400	Los Olmos Creek	Near Falfurrias, Brooks County, TX	78%	0	137	1999/4 – 2015/6	2015/7 – 2020/11

Comment 8, Reviewer2, Additional Information:

the following captions were updated with more information added:

Updated caption for Table 1. :

"Table 1. Summary of information on the nine streamflow monitoring stations. Intermittency ratio is the ratio of no-flow events to the total number of records."

Updated caption for Figure 3. :

"Figure 3. The observed vs. the predicted streamflow time-series of the shallow, deep, and wide models using SMOTE-balanced and transformed data for stations ST1, ST2, and ST3. The relative location of these stations and their information summary can be found in Figure 2. And Table 1."

Updated caption for Figure 4. :

"Figure 4. The observed vs. the predicted streamflow time-series of the shallow, deep, and wide models using SMOTE-balanced and transformed data for stations ST4, ST5, and ST6. The relative location of these stations and their information summary can be found in Figure 2. And Table 1."

Updated caption for Figure 5. :

"Figure 5. The observed vs. the predicted streamflow time-series of the shallow, deep, and wide models using SMOTE-balanced and transformed data for stations ST7, ST8, and ST9. The relative location of these stations and their information summary can be found in Figure 2. And Table 1."

Updated caption for Table 2. :

"Table 2. Summary of continuous performance evaluation metrics (MAE, RMSE, Pearson's r, and Spearman's Rho) for the shallow, deep, and wide models trained with SMOTE-balanced data during testing with and without applying log transform for the entire flow range. Abbreviations: MAE, mean absolute error; RMSE, root mean squared error."

Updated caption for Figure 7. :

"Figure 7. Comparison of observed No-Flow Persistence (NFP) and the predictions of the shallow, deep and wide models trained with SMOTE-balanced data for the nine IRES of study over the Testing Period."

Updated caption for Figure 8. :

"Figure 8. Comparison of observed Flow Persistence and the predictions of the shallow, deep and wide models trained with SMOTE-balanced data for the nine IRES of study over the Testing Period."

Updated caption for Figure 9. :

"Figure 9. Comparison of observed Flow to No-Flow Transition (F2NFT) and the predictions of the shallow, deep and wide models trained with SMOTE-balanced data for the nine IRES of study over the Testing Period."

Updated caption for Figure 10. :

"Figure 10. Comparison of observed No-Flow to Flow (NF2FT) and the predictions of the shallow, deep and wide models trained with SMOTE-balanced data for the nine IRES of study over the Testing Period."

Comment 9, Reviewer2, Additional Information:

Here is the Table S4:

Table S4. Summary of continuous performance evaluation metrics (MAE, RMSE, Pearson's r, and Spearman's Rho) for the shallow, deep, and wide models during testing without applying log transform for the entire flow range. Abbreviations: MAE, mean absolute error; RMSE, root mean squared error.

Station ID	Model	No Transform			
		MAE	RMSE	Pearson's r	Spearman's Rho
ST1	Con	0.266	0.556	0.32	0.4
ST1	Deep	0.298	0.619	0.56	0.51
ST1	Wide	0.225	0.55	0.33	0.51
ST2	Con	0.766	2.005	0.67	0.72
ST2	Deep	0.72	2.044	0.68	0.8
ST2	Wide	0.717	1.998	0.67	0.76
ST3	Con	0.496	1.92	0.78	0.62
ST3	Deep	0.546	1.918	0.74	0.66
ST3	Wide	0.477	1.919	0.79	0.64
ST4	Con	0.381	0.926	0.74	0.37
ST4	Deep	0.302	0.807	0.78	0.74
ST4	Wide	0.309	0.913	0.74	0.42
ST5	Con	0.894	3.041	0.65	0.65
ST5	Deep	0.938	3.061	0.66	0.67
ST5	Wide	0.868	3.046	0.65	0.64
ST6	Con	0.261	0.42	0.79	0.67
ST6	Deep	0.281	0.433	0.78	0.61
ST6	Wide	0.248	0.418	0.79	0.62
ST7	Con	0.325	0.562	0.95	0.89
ST7	Deep	0.317	0.515	0.95	0.84
ST7	Wide	0.313	0.563	0.95	0.84
ST8	Con	1.398	3.429	0.79	0.25
ST8	Deep	1.201	2.91	0.79	0.49
ST8	Wide	1.258	3.416	0.8	0.24
ST9	Con	0.146	0.559	0.79	0.5
ST9	Deep	0.21	0.588	0.23	0.55
ST9	Wide	0.121	0.558	0.76	0.57