Supplement of Hydrol. Earth Syst. Sci., 26, 1727–1743, 2022
https://doi.org/10.5194/hess-26-1727-2022-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Technical note: Using long short-term memory models to fill data gaps in hydrological monitoring networks

Huiying Ren et al.

Correspondence to: Xingyuan Chen (xingyuan.chen@pnnl.gov)

Supplemental Online Material

1 LSTM Architecture

Detailed information about the LSTM layer architectures are presented in this section.

Each memory unit in the LSTM layer is illustrated in Figure S1. The top panel shows generic representations of an RNN in a looped (left) or chained (right) form, which allows information to be passed to the next successor and persist. While all RNNs have the form of a chain of repeating modules of neural network (i.e., boxes labeled as A in Figure S1), the module being repeated can take different structural design to control the information flow, leading to different variants of RNN. Standard LSTMs use three gates, as shown in the bottom panel of Figure S1, to control the flow of information from one state to another and capture long-term dependencies. Each gate is composed of a linear layer with a sigmoid activation function. A forget gate ($\boldsymbol{f_t}$) decides what information to throw away from the previous memory state by using a sigmoid function that outputs a value between 0 and 1, where 0 represents completely forget the information and 1 represents completely keep the information. An input gate ($\boldsymbol{i_t}$) decides which values from the new input to be used for updating the memory state. The input gate is combined with a vector of new candidate input values out of a tanh layer (generates values between -1 and 1) through pointwise multiplication to yield information to be added to the current state. Finally, an output gate ($\boldsymbol{o_t}$) decides what to output based on the input and the previous memory state. The previous hidden state and the current input are passed to a sigmoid layer of the output gate, while the tanh layer scales the current memory state. Then, pointwise multiplication of the outputs from the tanh and sigmoid layers leads to the output of this repeating module. For a more detailed description of the components of the LSTM unit, the reader is referred to Hochreiter and Schmidhuber (1997).

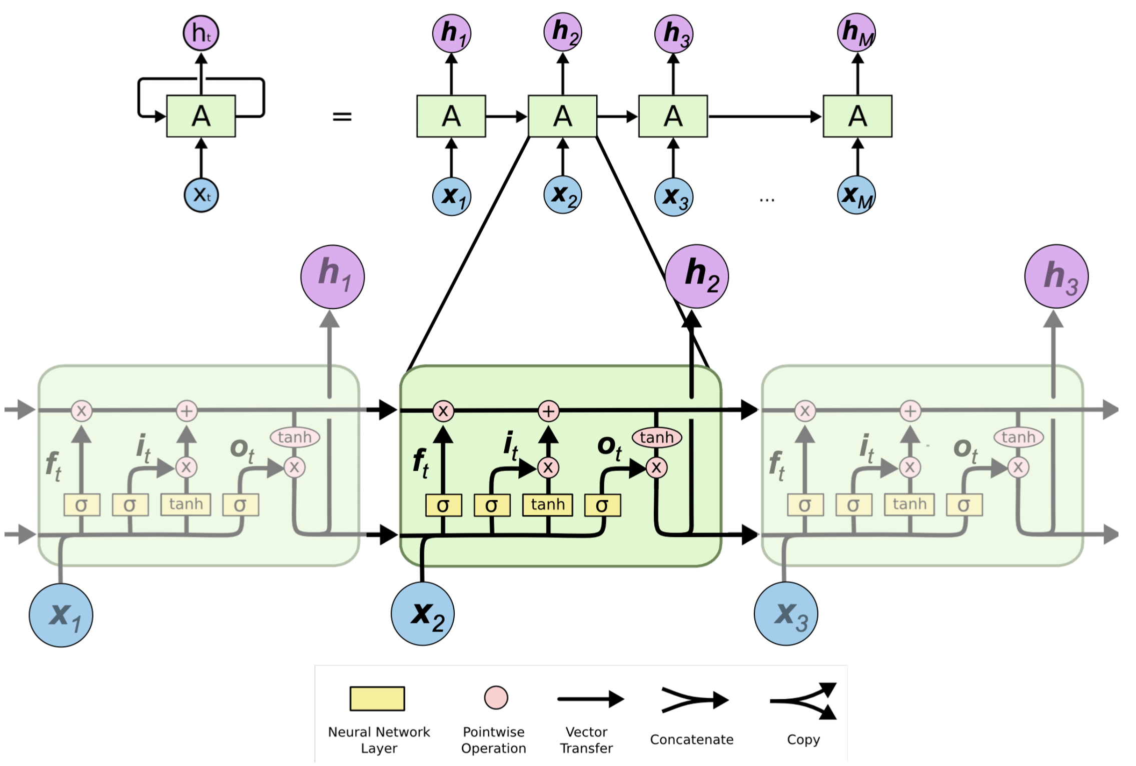

Figure S1. A diagram for network representing an LSTM unit. The top panel shows the looped and chain versions of a generic RNN, where $\boldsymbol{x_t}$ is the input, $\boldsymbol{h_t}$ is the output, and A is the repeating module of the LSTM unit. The bottom panel shows a diagram of the LSTM unit with the three main information gates: a forget gate ($\boldsymbol{f_t}$), an input gate ($\boldsymbol{i_t}$), and an output gate ($\boldsymbol{o_t}$). Images adapted from Olah (2015).

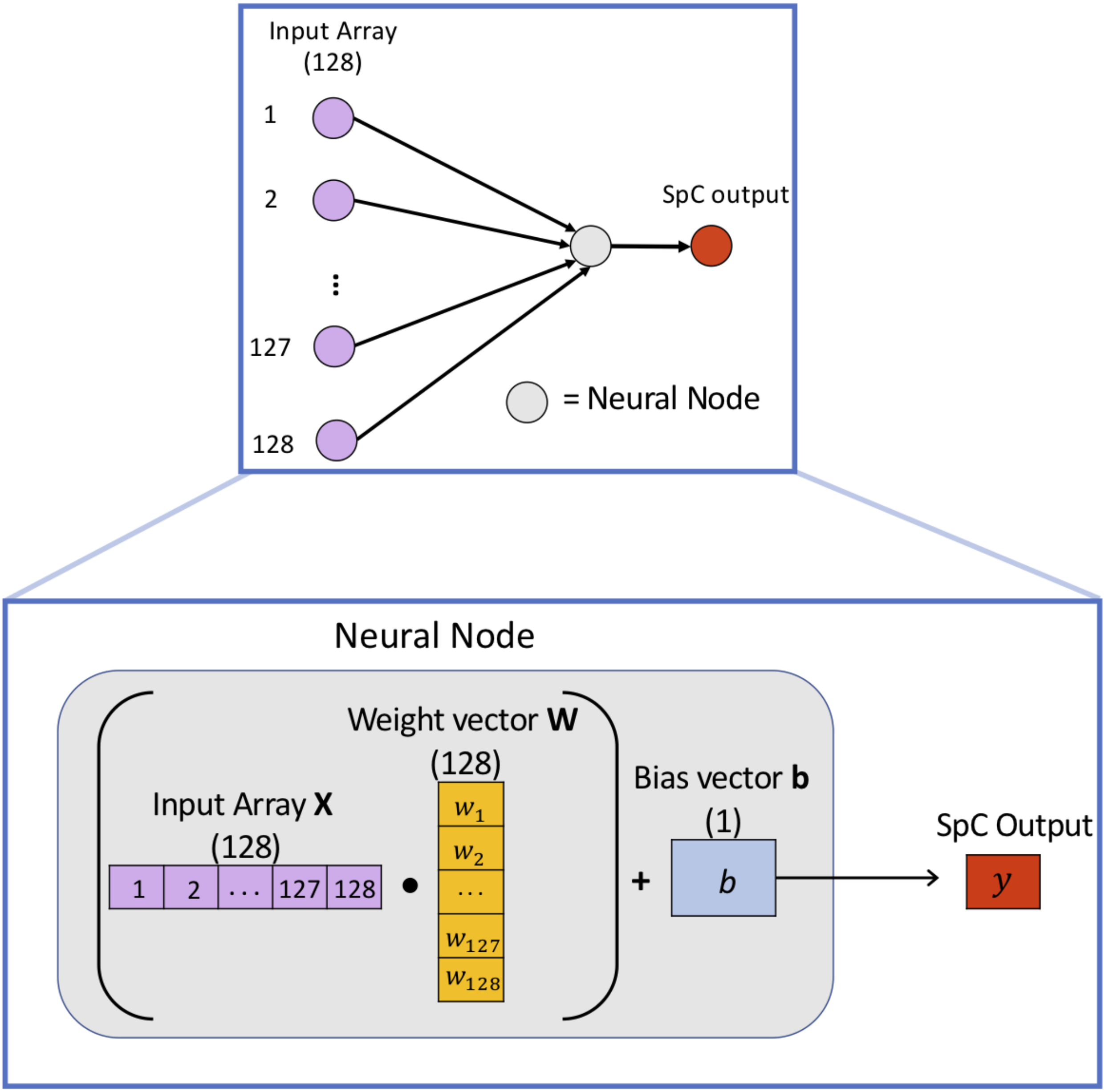

Figure S2. Enhanced detail of the architecture of the dense layer from Figure 4 in section 3.1. Shows the details of how the neural node processes the input data by $W(X) + b = y$, where X is the input array, W is the weight vector of the neural node, w^q is the qth weight of W, b is the bias vector of the neural node, and y is the output.

2 Seasonal ARIMA

The general seasonal ARIMA incorporates both non-seasonal and seasonal factors in multiplicative model. For the time series Y_t, the general seasonal ARIMA can be represented as following equations:

$$\Phi(B^m)\phi(B)\nabla_m^D\nabla^d Y_t = \Theta(B^m)\theta(B)N_t, \tag{1}$$

where N_t is the white noise process.

$$\nabla_m Y_t = Y_t - Y_{t-m}, \tag{2}$$

$$\nabla Y_t = Y_t - Y_{t-m}, \Phi(B^m) = 1 - \Phi_1(B^m) - \Phi_2(B^m) - ... - \Phi_P(B^{Pm}) \tag{3}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B - ... - \phi B \tag{4}$$

$$\Theta(B^m) = 1 - \Theta_1(B^m) - \Theta_2(B^m) - ... - \Theta_Q(B^{Qm}) \tag{5}$$

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B) - ... - \theta_q(B^q) \tag{6}$$

$$B^n Y_t = Y_{t-n} \tag{7}$$

3 Supplemental Figures

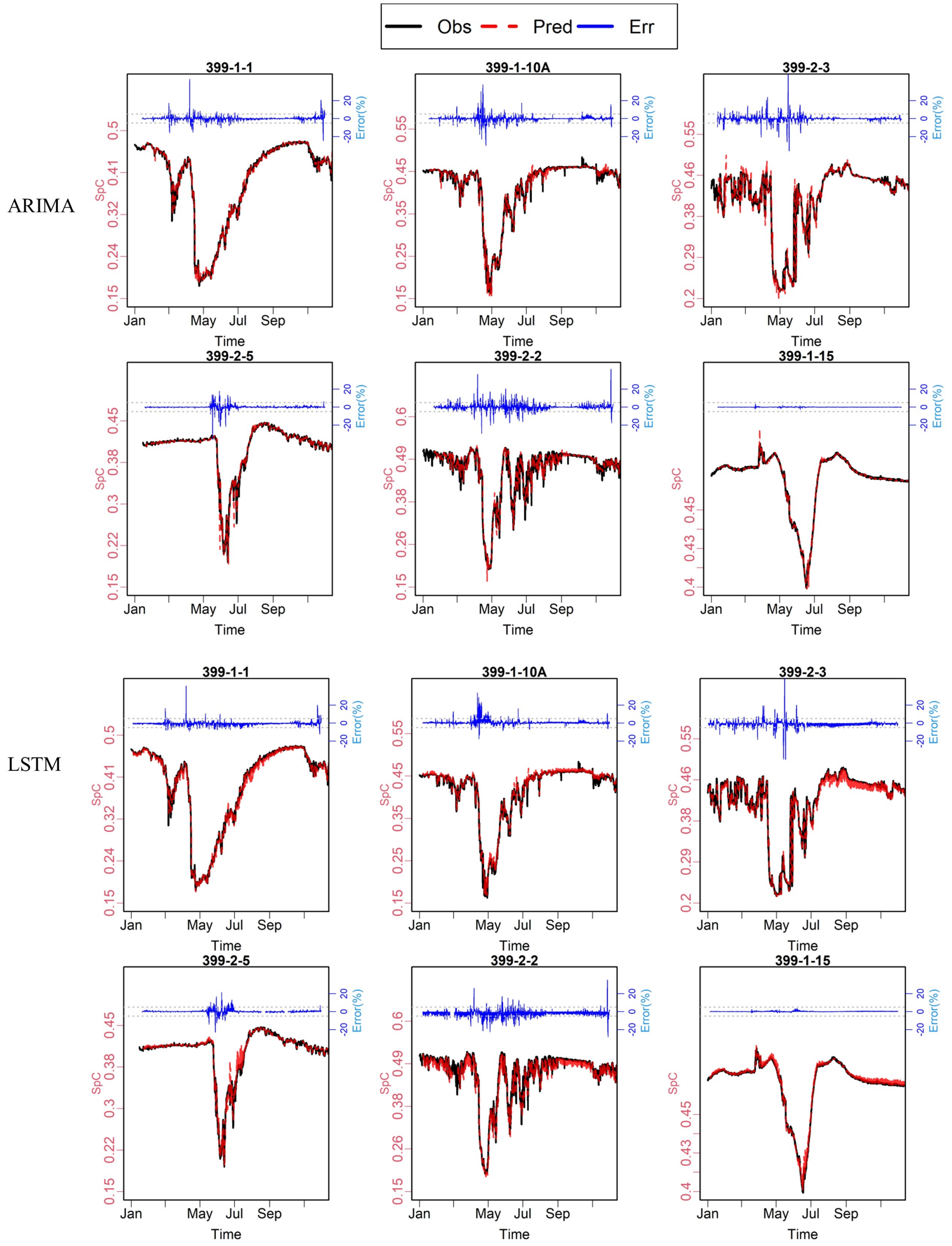

Figure S3. Model prediction comparison between single-well ARIMA and LSTM with observations for 24-hr gap length. The top panel is ARIMA prediction (in red) and relative errors (in blue) and the bottom panel is LSTM prediction (in red) and relative errors (in blue). The observations are in black line for each panel.

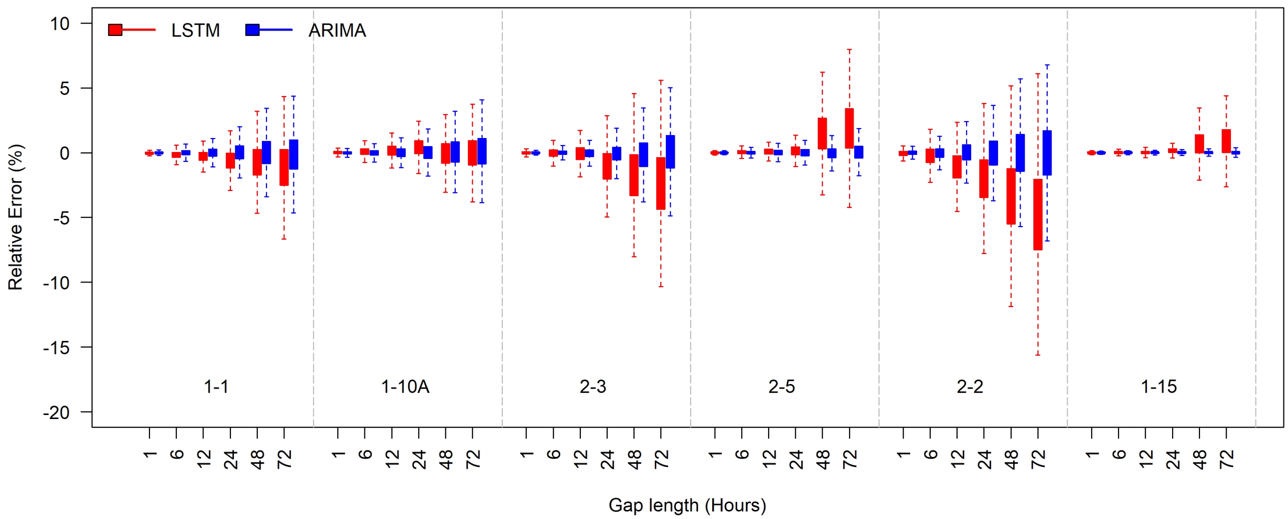

Figure S4. Boxplots of relative errors without outliers for filling SpC gaps of various lengths (1, 6, 12, 24, 48, and 72 hours) at each well during the test periods. The best LSTM and ARIMA models were used for evaluation. The LSTM and ARIMA models are represented by red bars and blue bars, respectively. Outliers are excluded in the boxplot to reveal the distribution of majority data points.

4 Supplemental Tables

Table S1. The best LSTM configurations and performance for a given gap length at each well based on the validation data set (2011): the input time window size (M), the number of units (U) in the LSTM layer, the learning rate (L), the SpC MAPE score, the Akaike Information Criterion (AI) for the model on the validation set, and range of AIC scores for all models for a given gap length and well on the validation set.

Well	Gap Length	M	U	L	MAPE	AIC	AIC Min	AIC Max
1-1	1	96	128	1e-3	0.189	1.31×10^4	-1.13×10^5	2.63×10^4
	6	96	128	1e-3	0.701	3.50×10^4	-9.08×10^4	4.19×10^4
	12	72	32	1e-4	1.24	-8.11×10^4	-8.11×10^4	5.40×10^4
	24	72	32	1e-4	1.95	-7.33×10^4	-7.33×10^4	6.50×10^4
	48	144	32	1e-5	2.85	-6.77×10^4	-6.80×10^4	7.50×10^4
	72	24	64	1e-5	3.67	-3.89×10^4	-6.42×10^4	8.08×10^4
1-10A	1	120	128	1e-3	0.19	4.75×10^4	-7.82×10^4	5.58×10^4
	6	120	128	1e-3	0.685	6.24×10^4	-6.29×10^4	7.39×10^4
	12	120	128	1e-3	1.1	6.69×10^4	-5.82×10^4	8.07×10^4
	24	144	128	1e-3	1.63	7.23×10^4	-5.22×10^4	8.70×10^4
	48	168	64	1e-5	2.16	-2.35×10^4	-4.90×10^4	9.14×10^4
	72	168	64	1e-5	2.39	-2.22×10^4	-4.63×10^4	9.37×10^4
1-15	1	48	32	1e-3	0.0163	-1.31×10^5	-1.32×10^5	1.93×10^4
	6	48	32	1e-3	0.0521	-1.04×10^5	-1.10×10^5	2.68×10^4
	12	48	32	1e-3	0.109	-1.02×10^5	-1.02×10^5	4.18×10^4
	24	96	128	1e-4	0.229	3.26×10^4	-9.25×10^4	5.70×10^4
	48	144	64	1e-3	0.372	-5.27×10^4	-8.50×10^4	7.15×10^4
	72	144	64	1e-3	0.506	-4.98×10^4	-8.08×10^4	8.00×10^4
2-2	1	168	32	1e-3	0.49	-6.06×10^4	-7.54×10^4	6.52×10^4
	6	168	32	1e-3	1.62	-4.86×10^4	-5.66×10^4	7.75×10^4
	12	144	128	1e-5	3.07	8.00×10^4	-4.95×10^4	8.47×10^4
	24	144	128	1e-5	4.42	8.34×10^4	-4.54×10^4	9.15×10^4
	48	144	128	1e-5	6.61	8.69×10^4	-4.19×10^4	9.63×10^4
	72	144	128	1e-5	7.52	8.66×10^4	-4.01×10^4	9.85×10^4
2-3	1	168	64	1e-3	0.142	-8.62×10^4	-1.16×10^5	2.28×10^4
	6	168	64	1e-3	0.369	-6.95×10^4	-9.87×10^4	4.64×10^4
	12	168	64	1e-3	0.663	-5.97×10^4	-8.78×10^4	6.02×10^4
	24	24	64	1e-5	1.09	-5.59×10^4	-7.92×10^4	7.59×10^4
	48	48	64	1e-5	1.7	-4.82×10^4	-7.18×10^4	8.95×10^4
	72	48	64	1e-5	2.28	-4.35×10^4	-6.72×10^4	9.51×10^4
2-5	1	120	32	1e-3	0.109	-1.15×10^5	-1.18×10^5	2.51×10^4
	6	144	64	1e-4	0.332	-6.72×10^4	-9.56×10^4	5.49×10^4
	12	144	64	1e-4	0.586	-5.70×10^4	-8.46×10^4	6.89×10^4
	24	144	64	1e-4	0.999	-4.90×10^4	-7.66×10^4	8.08×10^4
	48	120	64	1e-5	1.39	-4.51×10^4	-7.12×10^4	8.92×10^4
	72	120	64	1e-5	1.77	-4.16×10^4	-6.76×10^4	9.19×10^4

Table S2: Comparison of single-well LSTM and ARIMA models for all synthetic gap lengths in the SpC data. The models are the same ones used in Figure 7. The calculated statistics are: MAPE, Root Mean Squared Error (RMSE), Nash–Sutcliffe model efficiency coefficient (NSE), and Kling-Gupta Efficiency (KGE). The T-Score and P-Value are calculated on the relative errors of the two models for each well and gap length.

Gap Length = 1 hr							
Well	**Model Type**	**MAPE**	**RMSE**	**NSE**	**KGE**	**T-Score**	**P-Value**
1-1	LSTM	0.117	7.76×10^{-4}	1.0	0.999	11.6	6.14×10^{-31}
	ARIMA	0.183	1.23×10^{-3}	1.0	1.0		
1-10A	LSTM	0.185	1.41×10^{-3}	1.0	0.997	-7.52	5.89×10^{-14}
	ARIMA	0.299	2.45×10^{-3}	0.999	0.999		
1-15	LSTM	0.0559	4.00×10^{-4}	1.0	1.0	-4.29	1.79×10^{-5}
	ARIMA	0.0548	4.34×10^{-4}	0.999	1.0		
2-2	LSTM	0.276	2.11×10^{-3}	0.999	0.997	7.13	1.01×10^{-12}
	ARIMA	0.451	3.89×10^{-3}	0.997	0.998		
2-3	LSTM	0.129	1.05×10^{-3}	1.0	0.999	2.88	3.96×10^{-3}
	ARIMA	0.2	1.97×10^{-3}	0.999	0.999		
2-5	LSTM	0.113	1.02×10^{-3}	0.999	0.998	-1.09	2.76×10^{-1}
	ARIMA	0.171	1.62×10^{-3}	0.999	0.999		

Gap Length = 6 hr							
Well	**Model Type**	**MAPE**	**RMSE**	**NSE**	**KGE**	**T-Score**	**P-Value**
1-1	LSTM	0.435	2.92×10^{-3}	0.999	0.995	13.8	4.46×10^{-43}
	ARIMA	0.461	3.29×10^{-3}	0.999	0.999		
1-10A	LSTM	0.589	4.03×10^{-3}	0.996	0.984	-15.4	2.08×10^{-53}
	ARIMA	0.653	4.86×10^{-3}	0.995	0.994		
1-15	LSTM	0.109	9.16×10^{-4}	0.997	0.999	-1.86	6.31×10^{-2}
	ARIMA	0.0747	6.43×10^{-4}	0.999	0.999		
2-2	LSTM	0.981	7.14×10^{-3}	0.989	0.985	4.58	4.79×10^{-6}
	ARIMA	1.01	8.29×10^{-3}	0.984	0.992		
2-3	LSTM	0.517	4.15×10^{-3}	0.997	0.997	4.35	1.37×10^{-5}
	ARIMA	0.521	5.75×10^{-3}	0.993	0.993		
2-5	LSTM	0.314	2.64×10^{-3}	0.996	0.998	-5.03	4.98×10^{-7}
	ARIMA	0.352	3.13×10^{-3}	0.995	0.995		

Gap Length = 12 hr							
Well	**Model Type**	**MAPE**	**RMSE**	**NSE**	**KGE**	**T-Score**	**P-Value**
1-1	LSTM	0.75	5.07×10^{-3}	0.997	0.996	16.0	2.05×10^{-57}
	ARIMA	0.781	5.47×10^{-3}	0.996	0.996		
1-10A	LSTM	0.947	5.96×10^{-3}	0.992	0.969	-19.2	1.88×10^{-81}
	ARIMA	0.947	6.32×10^{-3}	0.991	0.99		
1-15	LSTM	0.166	1.37×10^{-3}	0.994	0.996	-3.41	6.49×10^{-4}
	ARIMA	0.0893	8.16×10^{-4}	0.998	0.998		
2-2	LSTM	1.87	1.22×10^{-2}	0.967	0.981	39.4	0.00
	ARIMA	1.51	1.13×10^{-2}	0.971	0.985		
2-3	LSTM	0.98	7.78×10^{-3}	0.988	0.993	4.42	1.01×10^{-5}
	ARIMA	1.03	1.26×10^{-2}	0.966	0.974		
2-5	LSTM	0.569	4.60×10^{-3}	0.989	0.994	-6.96	3.51×10^{-12}
	ARIMA	0.566	4.86×10^{-3}	0.988	0.989		

Gap Length = 24 hr							
Well	**Model Type**	**MAPE**	**RMSE**	**NSE**	**KGE**	**T-Score**	**P-Value**
1-1	LSTM	1.38	8.33×10^{-3}	0.991	0.988	19.1	1.00×10^{-80}
	ARIMA	1.36	8.98×10^{-3}	0.989	0.994		
1-10A	LSTM	1.37	8.07×10^{-3}	0.986	0.968	-24.6	1.48×10^{-131}
	ARIMA	1.5	9.60×10^{-3}	0.98	0.987		
1-15	LSTM	0.259	1.88×10^{-3}	0.989	0.982	-48.9	0.00
	ARIMA	0.119	1.18×10^{-3}	0.996	0.997		
2-2	LSTM	2.97	1.87×10^{-2}	0.922	0.962	48.1	0.00
	ARIMA	2.23	1.64×10^{-2}	0.939	0.967		
2-3	LSTM	2.15	1.63×10^{-2}	0.945	0.965	21.6	4.69×10^{-102}
	ARIMA	1.72	1.48×10^{-2}	0.954	0.971		
2-5	LSTM	0.929	6.86×10^{-3}	0.976	0.988	-9.6	9.22×10^{-22}
	ARIMA	0.866	7.45×10^{-3}	0.971	0.977		

Gap Length = 48 hr							
Well	**Model Type**	**MAPE**	**RMSE**	**NSE**	**KGE**	**T-Score**	**P-Value**
1-1	LSTM	2.13	1.21×10^{-2}	0.98	0.988	17.8	3.24×10^{-70}
	ARIMA	2.15	1.34×10^{-2}	0.976	0.988		
1-10A	LSTM	2.09	1.09×10^{-2}	0.974	0.911	-16.4	2.68×10^{-60}
	ARIMA	2.17	1.32×10^{-2}	0.962	0.981		
1-15	LSTM	1.0	6.51×10^{-3}	0.869	0.907	-44.2	0.00
	ARIMA	0.168	1.67×10^{-3}	0.991	0.995		
2-2	LSTM	4.64	2.80×10^{-2}	0.825	0.932	60.8	0.00
	ARIMA	2.95	2.04×10^{-2}	0.905	0.952		
2-3	LSTM	3.26	2.04×10^{-2}	0.915	0.919	29.7	4.62×10^{-189}
	ARIMA	2.89	2.38×10^{-2}	0.88	0.91		
2-5	LSTM	2.34	1.22×10^{-2}	0.925	0.928	-33.9	3.18×10^{-243}
	ARIMA	1.25	1.09×10^{-2}	0.937	0.928		

Gap Length = 72 hr							
Well	**Model Type**	**MAPE**	**RMSE**	**NSE**	**KGE**	**T-Score**	**P-Value**
1-1	LSTM	2.56	1.40×10^{-2}	0.974	0.983	26.7	4.91×10^{-154}
	ARIMA	2.57	1.46×10^{-2}	0.971	0.985		
1-10A	LSTM	2.58	1.32×10^{-2}	0.962	0.888	-17.8	5.10×10^{-70}
	ARIMA	2.84	1.75×10^{-2}	0.931	0.957		
1-15	LSTM	1.27	8.18×10^{-3}	0.794	0.845	-42.8	0.00
	ARIMA	0.211	1.79×10^{-3}	0.99	0.988		
2-2	LSTM	5.91	3.44×10^{-2}	0.736	0.914	80.0	0.00
	ARIMA	3.66	2.46×10^{-2}	0.861	0.931		
2-3	LSTM	4.8	2.90×10^{-2}	0.829	0.864	23.5	3.91×10^{-120}
	ARIMA	3.49	2.56×10^{-2}	0.862	0.901		
2-5	LSTM	3.08	1.55×10^{-2}	0.88	0.884	-31.5	9.51×10^{-212}
	ARIMA	1.47	1.06×10^{-2}	0.941	0.96		

Table S3. The best multi-well LSTM configurations and performance for a given gap length based on the validation data set (2011): the input time window size (M), the number of units (U) in the LSTM layer, the learning rate (L), the SpC MAPE score, the Akaike Information Criterion (AI) for the model on the validation set, and range of AIC scores for all models for a given gap length on the validation set.

Gap Length	M	U	L	MAPE	AIC	AIC Min	AIC Max
1	144	128	1e-3	0.154	4.56×10^4	-8.40×10^4	6.11×10^4
6	144	128	1e-3	0.63	6.27×10^4	-6.68×10^4	7.13×10^4
12	24	32	1e-4	1.14	-5.97×10^4	-5.97×10^4	8.08×10^4
24	24	32	1e-4	1.75	-5.43×10^4	-5.43×10^4	8.89×10^4
48	96	128	1e-5	2.69	7.90×10^4	-4.84×10^4	9.46×10^4
72	96	128	1e-5	3.15	8.09×10^4	-4.74×10^4	9.74×10^4

References

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Olah, C.: Understanding LSTM Networks, https://colah.github.io/posts/2015-08-Understanding-LSTMs/, 2015.