Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions

Using Deep Learning to Fill Spatio-Temporal Data Gaps in Hydrological Monitoring Networks

Huiying Ren[1], Erol Cromwell[2], Ben Kravitz[3,4], and Xingyuan Chen[4]

[1]Earth Systems Science Division, Pacific Northwest National Laboratory, WA, USA
[2]Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory, WA, USA
[3]Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, USA
[4]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, WA, USA

Correspondence: Xingyuan Chen (Xingyuan.Chen@pnnl.gov)

Abstract. Long-term spatio-temporal changes in subsurface hydrological flow are usually quantified through a network of wells; however, such observations often are spatially sparse and temporal gaps exist due to poor quality or instrument failure. In this study, we explore the ability of deep neural networks to fill in gaps in spatially distributed time-series data. We selected a location at the U.S. Department of Energy's Hanford site to demonstrate and evaluate the new method, using a 10-year spatio-temporal hydrological dataset of temperature, specific conductance, and groundwater table elevation from 42 wells that monitor the dynamic and heterogeneous hydrologic exchanges between the Columbia River and its adjacent groundwater aquifer. We employ a long short-term memory (LSTM)-based architecture, which is specially designed to address both spatial and temporal variations in the property fields.

The performance of gap filling using an LSTM framework is evaluated using test datasets with synthetic data gaps created by assuming the observations were missing for a given time window (i.e., gap length), such that the mean absolute percentage error can be calculated against true observations. Such test datasets also allow us to examine how well the original nonlinear dynamics are captured in gap-filled time series beyond the error statistics. The performance of the LSTM-based gap-filling method is compared to that of a traditional, popular gap-filling method: autoregressive integrated moving average (ARIMA). Although ARIMA appears to perform slightly better than LSTM on average error statistics, LSTM is better able to capture nonlinear dynamics that are present in time series. Thus, LSTMs show promising potential to outperform ARIMA for gap filling in highly dynamic time-series observations characterized by multiple dominant modes of variability. Capturing such dynamics is essential to generate the most valuable observations to advance our understanding of dynamic complex systems.

1 Introduction

Long-term hydrological monitoring using distributed well networks is of critical importance for many areas, including understanding how ecosystems respond to chronic or extreme perturbations, as well as informing policies and decisions related to

natural resources and environmental issues (Wett et al., 2002; Taylor and Alley, 2002; Grant and Dietrich, 2017). One of the most common methods of collecting hydrological data in groundwater is through wells (Güler and Thyne, 2004; Strobl and Robillard, 2008; Lin et al., 2012); however, wells are necessarily sparse, leaving spatial gaps in the dataset. Moreover, most well data will also have temporal gaps due to instrument failure or poor quality of measurements for numerous reasons. These data gaps degrade the quality of the dataset and increase the uncertainty in the spatial and temporal patterns that are derived from them. Gap filling is necessary for developing understanding of the underlying system as well as for use in creating continuous, internally consistent boundary conditions for numerical models. Many natural systems exhibit nonlinear or/and nonstationary behaviors due to evolving nonlinear dynamics, which makes it challenging to reproduce those complex patterns while filling in data gaps.

Various statistical methods have been developed to fill gaps in spatio-temporal datasets, with the most commonly used being the autoregressive integrated moving average (ARIMA) method (Han et al., 2010; Zhang, 2003). For any given spatial location, ARIMA uses temporal autocorrelation to predict unobserved data points in a time series. Spatio-temporal autocorrelations can be considered by using multivariate ARIMA and space-time autoregressive models (Kamarianakis and Prastacos, 2003; Wikle et al., 1998; Kamarianakis and Prastacos, 2005); however, ARIMA cannot capture nonlinear trends because it assumes a linear dependence between adjacent observations (Faruk, 2010; Valenzuela et al., 2008; Ho et al., 2002). In addition, all existing space-time ARIMA models assume fixed global autoregressive and moving average terms, which would fail to capture evolving dynamics in highly dynamic systems (Pfeifer and Deutrch, 1980; Griffith, 2010; Cheng et al., 2012, 2014). Spectral-based methods, such as singular spectrum analysis, maximum entropy method, and Lomb-Scargle periodogram, have been used to account for nonlinear trends while filling in gaps in spatio-temporal datasets (Ghil et al., 2002; Hocke and Kämpfer, 2008; Kondrashov and Ghil, 2006). However, these methods use a few optimal spatial or temporal modes occurring at low frequencies to predict the missing values, with the other higher frequency components discarded as noise, which may lead to reduced accuracy of the statistical models in fitting the observations and in predicting missing values (Kondrashov et al., 2010; Wang et al., 2012). Kriging and maximum likelihood estimation used in spatial and spatio-temporal gap filling often face computational challenges as they require computing the covariance matrix of the data vector, which can be quite large (Katzfuss and Cressie, 2012; Eidsvik et al., 2014). Other nonlinear methods have been explored with some success, including expectation-maximization or Bayesian probabilistic inference including hierarchical models, Gaussian process, and Markov chain Monte Carlo; the spatial and temporal correlations are most effectively captured by using models that build dependencies in different stages or hierarchies (Calculli et al., 2015; Banerjee et al., 2014; Datta et al., 2016; Finley et al., 2013; Stroud et al., 2017). In general, the expectation-maximization algorithm and Bayesian-based methods are sensitive to the choice of initial values and prior distributions in parameter space (Katzfuss and Cressie, 2011, 2012). Moreover, the prior distributions with all the associated parameters in both the spatial and temporal domains need to be specified, which becomes increasingly difficult in more complex systems.

Deep neural networks (DNNs) (Schmidhuber, 2015) are data-driven tools that, in principle, could provide a powerful way of extracting the nonlinear spatio-temporal patterns hidden in the distributed time-series data without knowing their explicit forms (Längkvist et al., 2014). They are increasingly been used in geoscience domains to extract patterns and insights from

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions

the streams of geospatial data and to transform the understanding of complex systems (Reichstein et al., 2019; Shen, 2018; Sun, 2018; Sun et al., 2019; Gentine et al., 2018). The umbrella term of DNN contains numerous categories of architectures, depending on the problem at hand. For the analyses in this paper, which are focused on filling gaps in time-series data, a natural choice of architecture is recurrent neural networks (RNNs) (Connor et al., 1994; Olah, 2015). These networks take sequences (e.g., time series) as input and output single values or sequences that follow. They are designed to use information about previous events to make predictions about future events, essentially by letting the model "remember." However, for longer sequences of data, RNNs have been shown to lose memory from previously trained data, i.e., they "forget" (Hochreiter et al., 2001). This affects the performance of RNNs, particularly for data where the beginning of a sequence impacts the prediction, since this information becomes exponentially less impactful for the prediction as the size of the sequence increases. Long short-term memory (LSTM) networks are variations of RNNs that are explicitly designed to avoid this problem by using memory cells to retain information about relevant past events (Ma et al., 2015). RNNs and LSTMs have been successfully applied to text prediction (Graves, 2013), text translation (Wu et al., 2016), speech recognition (Graves et al., 2013), and image captioning (You et al., 2016). (Specifics on LSTM architectures are described in Section 3.1.) This makes LSTMs well suited for the problem at hand, particularly for data where multiple timescales of variability can affect responses (Liu et al., 2016; Song et al., 2017).

This study aims to evaluate the potential of using a DNN architecture that utilizes LSTMs for filling gaps in a spatio-temporal environmental dataset, using a test case that focuses on understanding the interactions between a regulated river and contaminated groundwater aquifer. The use of a DNN is compared with traditional approaches (e.g., ARIMA) to identify situations in which a DNN outperforms more commonly used methods as well as what the optimal configurations might be for this particular application.

2 Study Site and Data Description

A 10-year (2008–2018) hourly spatio-temporal dataset was collected from a network of groundwater wells that monitor temperature, specific conductance (SpC), and water-table elevation at the 300 Area of the U.S. Department of Energy Hanford site, located in southeastern Washington State. The well network was built to monitor the attenuation of legacy contaminants. The time series of river stage from 2008 to 2018 (Figure 1) shows large and dynamic fluctuations in river stage. These fluctuations are due not only to natural processes but also to regulation of the river by the upstream hydroelectric dam operations (Song et al., 2018), which on average vary ∼0.5 m diurnally and up to ∼2-3 m annually. The water elevation dynamics in each groundwater well is driven by river stage fluctuations, which in turn influence contaminant recharge to groundwater and lead to highly complex transport behaviors of the contaminants at the site (Arntzen et al., 2006; Zachara et al., 2016).

The intrusion of river water into the adjacent groundwater aquifer causes mixing of two water bodies with distinct geochemistry and stimulates biogeochemical reactions at the interface. The river water has lower SpC (0.1–0.2 mS/cm) than groundwater (averaging $\sim$ 0.4 mS/cm). Groundwater has a nearly constant temperature (16–17$^{\circ}C$) as opposed to seasonally varying river temperature (3–22$^{\circ}C$). The highly heterogeneous coarse-textured aquifer (Zachara et al., 2013) interacts with

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions

dynamic river stages to create complex river intrusion and retreat pathways and dynamics (Chen et al., 2013, 2012). The time series of multi-year SpC and temperature observations at the select set of wells in the network have demonstrated these complicated processes of river water intrusion into our study site (Figure 1). For wells that are farther inland (e.g., well 1-15), temperatures remain consistently within the groundwater temperature range and SpC has three noticeable dips (dropping from 0.5 to 0.4 mS/cm range), coinciding with the high river stages in years 2011, 2012, and 2017, which are featured with higher peak river stages than other years so the river water could intrude further into the groundwater aquifer. Wells close to the shoreline (e.g., wells 1-1, 1-10A, 2-2, and 2-3) tend to be strongly affected by river water intrusion in spring and summer. As such, the dynamic patterns of SpC and temperature correspond well with river stage fluctuations, specifically that SpC decreases and temperature increases with increasing river stage. Fluctuations of SpC in well 2-2 appear to be stronger and at higher frequency than in other wells, likely indicating its higher connectivity with the river. Well 2-5 is located at an intermediate distance from the river compared to other wells shown in Figure 1, so the intrusion of river water is evident in most of the years except in low-flow years such as 2009 and 2015, during which both SpC and temperature remain nearly unchanged.

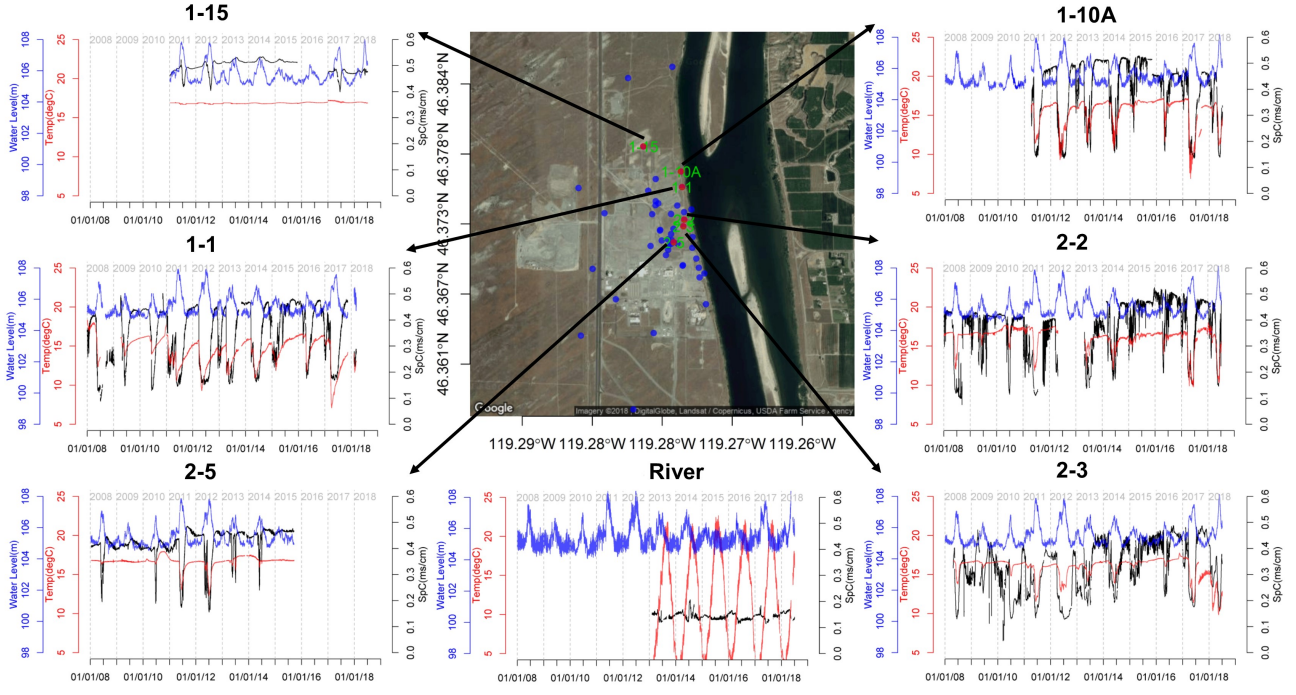

Figure 1. Groundwater monitoring well network at the 300 Area of the Hanford site and the monitoring data at select wells. Each well represented by a dot is instrumented to measure groundwater elevation, temperature, and SpC. The wells selected for this study are marked with red dots with well names. The three variables monitored in wells and in the river are shown in time-series plots with blue (water elevation), black (SpC), and red (temperature) lines.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

To understand the dynamic behaviors of all the variables in each well at the study site, we perform spectral analysis on multi-year SpC observations at each selected well using a continuous wavelet transform. The continuous wavelet transform is widely used for time–frequency analysis of time series and relies on a "mother wavelet," which is chosen to be the Morlet wavelet (Grossmann and Morlet, 1984) to deal with the time-varying frequency and amplitude in time-series data at this site (Stockwell et al., 1996; Grinsted et al., 2004). The Wavelet Power Spectrum (WPS) for the multiyear SpC time series and the normalized global power spectrum (average WPS over the time domain for each well) are displayed in the first two columns in Figure 2. Data gaps are indicated as blank regions; examples include early year 2009 at well 1-1, the beginning of year 2011 at well 1-10A, and the later part of year 2012 at well 2-2. At wells 1-1, 1-10A, 2-3, 2-5, and 2-2, the strong intensities of SpC signals appear at the half-year and yearly frequencies; however, well 1-15 has a different pattern in that most of its high intensities are below the 256-hour frequency. The averaged WPS more clearly shows the contrast in behaviors: wells 1-1, 1-10A, 2-3, and 2-5 have a dominant frequency at half a year; well 2-2 has multiple dominant frequencies at daily, monthly, and seasonal scales; while well 1-15 has similar intensities at half-year and hourly scales. Applying this information to the task at hand, we hypothesize that gap filling at well 2-2 could be more challenging due to the multiple sources of its dynamic behaviors, manifested as significant powers at multiple frequencies.

Since the dynamics of the system are driven by the river stage, we perform magnitude-squared wavelet coherence analysis via the Morlet wavelet to reveal dynamic correlations between the SpC and river stage time series (Grinsted et al., 2004; Vacha and Barunik, 2012). Wavelet coherence in the time-frequency domain is plotted in the third column in Figure 2 and the average coherence is plotted in the fourth column; statistically significant values at the 95th percent confidence interval are indicated with red points. Red regions in the third column indicate that the two signals are highly correlated, while blue regions indicate lower correlations. SpC correlates well with river stage in every well at multiple time scales, with the exception of well 1-15. High correlations are found in wells close to the river (e.g., 1-1, 1-10A, 2-2, and 2-3) at half-year and yearly frequencies. High correlations in wells farther from the river (e.g., 1-15 and 2-5) are shifted towards longer periods and less persistent in time.

As can be clearly seen in Figure 2, many of the wells have long data gaps, which have unknown effects on our ability to estimate dynamics from the wavelet spectra. As such, gap filling is needed to infer observations and guide modeling of the underlying system. Figure 3 provides a summary of gap lengths for the overall network of monitoring wells. The majority of the gap lengths of all the three monitored variables are less than 50 hours. Therefore, in our investigations we explore the ability of the methods in filling gaps at 24-, 48-, and 72-hour lengths.

3 Gap-Filling Methods

In this section, we describe two methods we use to fill gaps in SpC measurements at select wells: a DNN model using several LSTM network layers and the traditional ARIMA model for comparison and assessing the strengths and limitations of the LSTM-based model. In both models, an input with M data points (input window length) is provided to predict outputs of N time steps that follow the input window (output window length).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

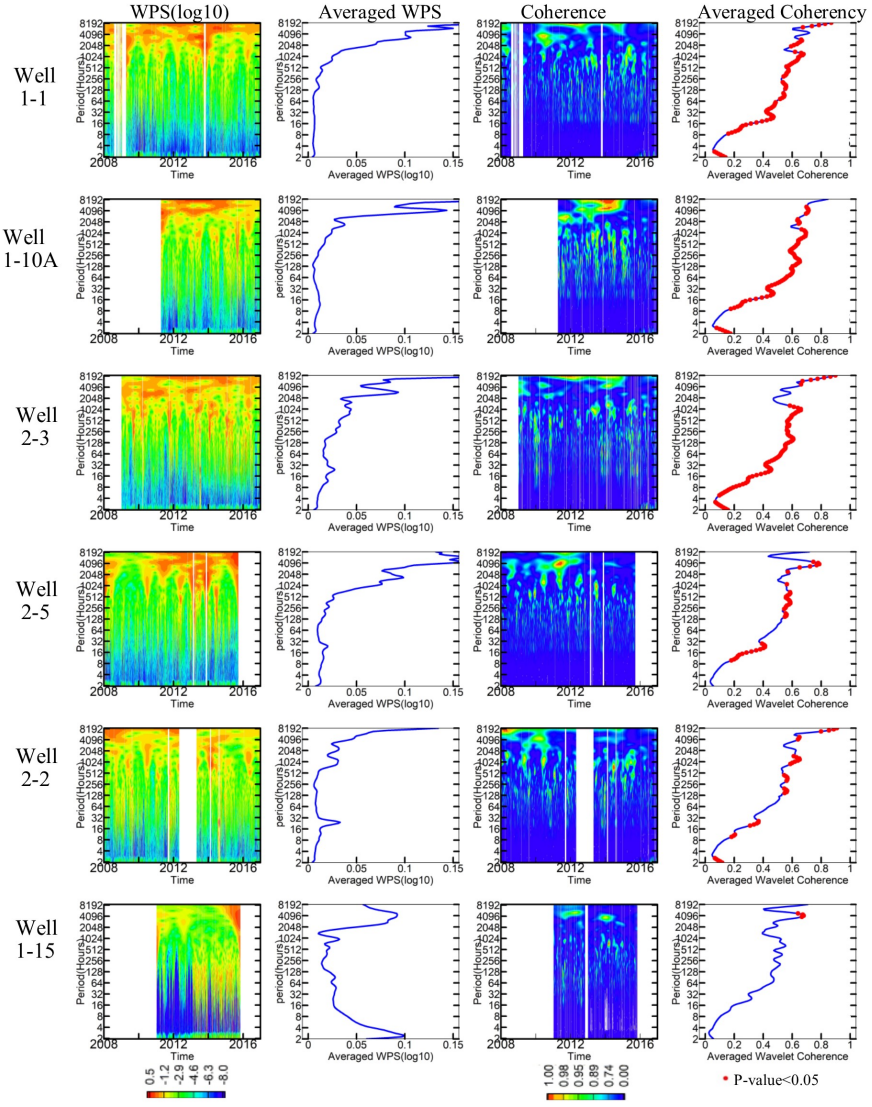

Figure 2. WPS analysis of SpC at each well. The first column is the spectrogram of SpC in each well; the second column is the averaged WPS; the third column is the coherence between SpC in each well and the river stage; and the fourth column is the averaged coherence with $p < 0.05$ values indicated in red.

3.1 Stacked LSTM Architecture

We designed the DNN architecture to train models of input size M and output size N to fill gaps of various lengths in groundwater well measurements. The DNN architecture is shown in Figure 4, which contains three LSTM layers, followed by two consecutive dropout layers, a convolutional layer, and a final output dense layer. This model architecture is generally

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

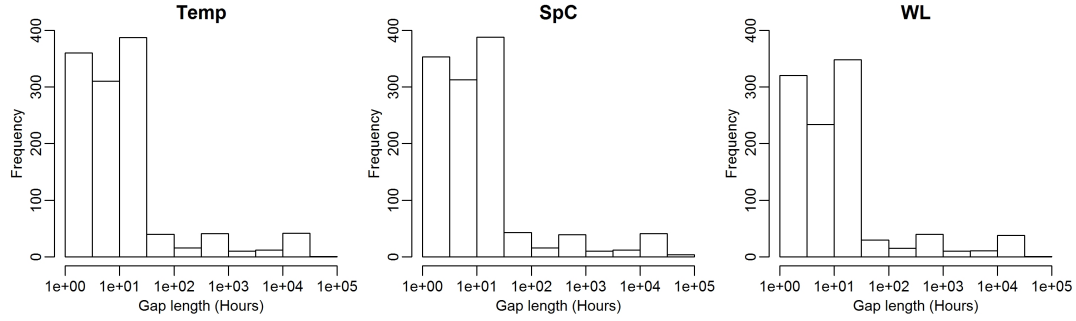

Figure 3. Histograms of gap lengths for each monitored variable, aggregated across all wells in the monitoring network during 2008-2018.

described as a stacked LSTM model, given that the LSTM layers are "stacked" on top of each other. Each input and predicted output contain the following three well measurements: water level (m), temperature ($°C$), and SpC (mS/cm), leaving the model to generalize nonlinear connections among them. Stacked LSTM layers take advantage of the temporal correlations of the measurements to improve model performance. Each of the three LSTM layers has 128 units, with the last LSTM layer returning an M x 128 output, which is fed into two consecutive dense layers with dropout of 0.3, reducing the output from M x 128 to M x 64. A dense layer is a neural network where every input neuron is connected to every output neuron with a weight matrix and bias vector. Dropout is a regularization technique that randomly disables a select fraction of neurons during training to enhance robust model performance and prevent overfitting (Hinton et al., 2012). The output from the second dense layer is fed into a convolutional layer with 24 filters of size M-N+1, reducing the output size to N x 24. Finally, a dense layer is applied to yield a model output of our desired size, N x 3. The detailed structures of the LSTM layers, dense layer, and convolutional layer are provided in the supplemental material.

Each memory unit in the LSTM layer is further illustrated in Figure 5. The top panel shows generic representations of an RNN (Olah, 2015) in a looped (left) or chain (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 5), 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 5, to control the flow of information from one state to another and capture long-term dependencies. Each gate is composed of a sigmoid neural net layer and a pointwise multiplication operation. A forget gate (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 (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 (generate values between -1 and 1) through pointwise multiplication to generate information to be added to the current state. Finally, an output gate (O_t) decides what to output based on the input and previous memory state. The sigmoid layer of the output gate decides what parts of the memory state will be output, while the tanh layer scales the current memory state. The pointwise

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

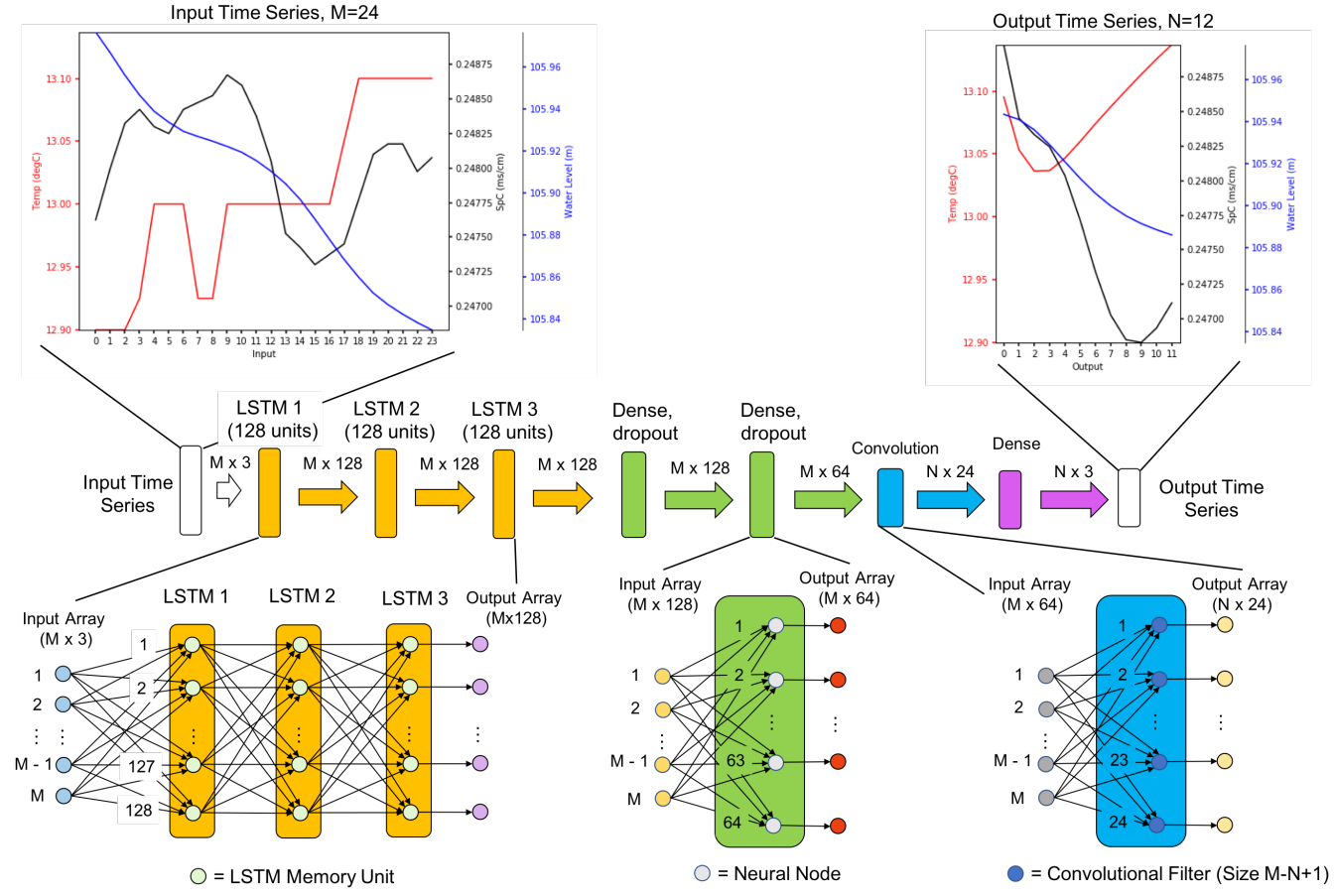

Figure 4. Architecture of the stacked LSTM models, where M is the input window size and N is the output window size. Includes example input and output data with $M = 24$ and $N = 12$. For a more detailed diagram of the LSTM layers, dense layer, and convolutional layer, see Figures S1, S2, and S3, respectively, in the Supplemental Online Material.

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 Olah (2015) and Ma et al. (2015).

3.2 Training the Stacked LSTM Models

Training data for the stacked LSTM models are created by finding data segments of $M + N$ hours that have no missing values, i.e., no gaps in the data, for all three measurements over a specified monitoring window. The well data are then preprocessed by normalizing all measurements to fall between 0 and 1 using different scaling factors for each variable, as temperature measurements are on a scale of 10^1, SpC is on a scale of 10^{-1}, and water level is on a scale of 10^2. The model is trained for 30 iterations (i.e., epochs) over the training data. We use an Adam optimizer (Kingma and Ba, 2014) for training and the mean-

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

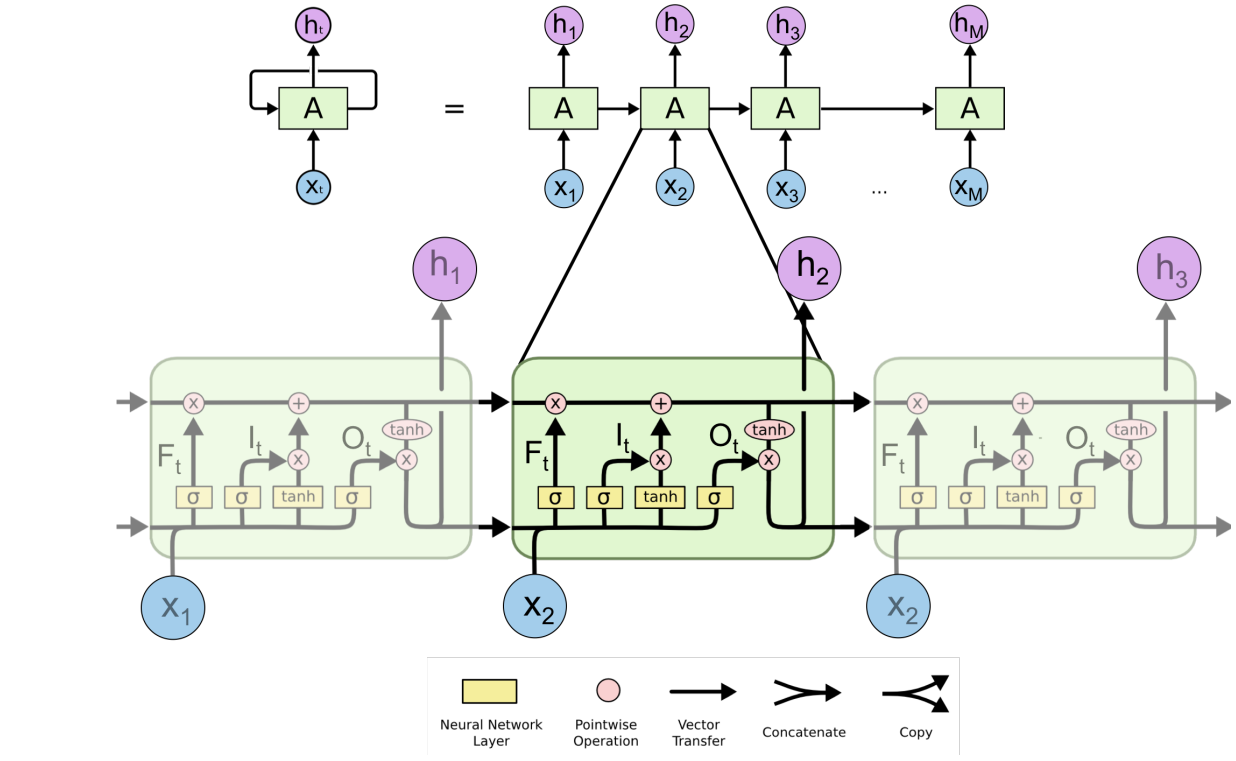

Figure 5. A diagram for network representing an LSTM unit. The top panel shows the looped and chain versions of a generic RNN, where x_t is the input, h_t is the output, and A is the repeating module of the LSTM unit. The bottom panel shows a diagram of the LSTM unit with the three main information gates: a forget gate (F_t), an input gate (I_t), and an output gate (O_t). Images from Olah (2015).

squared error for the loss function. We train a stacked LSTM model for each desired combination of M, N at each well given a certain amount of training data. The set of alternatives we consider for each model configuration parameter is shown in Table 1. Training wells were chosen based on adequate data availability and their distance from the river. Excluding combinations with $M < N$, a total of 810 unique models (135 model configurations for each of six wells) are trained.

Table 1. Parameters varied during LSTM model training. No models are trained where $M < N$.

Input Window (M)	**Output Window** (N)	**Training Wells**	**Training Period**	**Testing Period**
24, 48, 72, 96,	1, 6, 12, 24, 48, 72,	1-1, 1-10A, 2-2	2 years (2012-2013), 4 years (2012-2015)	1 year (2011)
120, 144, 168 (hours)	120, 144, 168 (hours)	2-3, 2-5, 1-15	6 year (2010, 2012-2016)	

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

3.3 Testing the Stacked LSTM Models for Gap Filling

To evaluate the accuracy of the trained stacked LSTM models in filling in measurement gaps, we assume synthetic gaps of various lengths (e.g., 1, 24, 48, and 72 hours, referred to as gap scenarios hereafter) exist in a dataset containing all three variables monitored in year 2011 at each testing well (same as the training wells shown in Table 1). Given a model configuration (M, N, and training period), an LSTM model is trained for the six training wells, which is then tested on filling in synthetic data gaps of various lengths in all the wells. There are 36 training-testing well pairs in total. During testing, each model is given the first M hours of data from the time series preceding the occurrence of a gap (assuming no missing values in these M hours). Then the model is used to fill in the first missing value in the gap by taking the first value of the predicted N hours from the model. This gap-filled value is then treated as if it was observed when repeating this procedure to fill in the gap of the next hour. The input data window keeps sliding in this way, hour by hour, until the model has filled the entire gap. The accuracy of the gap-filling model is evaluated by calculating the mean absolute percentage error (MAPE; %) between the values that are filled in (i.e., predicted) and the true values:

$$MAPE = 100 \times \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\text{Prediction} - \text{Observation}}{\text{Observation}}\right|, \tag{1}$$

where n is the number of data points being missing.

3.4 ARIMA model

ARIMA is one of the most general classes of models for extrapolating time series to produce forecasts and we used it as a baseline to compare and assess the LSTM-based gap-filling method. ARIMA is applicable to nonstationary processes in that the dataset can be made stationary by differencing if necessary. Differencing, autogressive, and moving average components make up a nonseasonal ARIMA(p,d,q) model given by:

$$Y_t = c + \phi_1 Y_{t-1}^d + \phi_p Y_{t-p}^d + ... + \theta_1 e_{t-1} + \theta_q e_{t-q} + e_t, \tag{2}$$

where ϕs and θs are polynomials of orders p and q, respectively, each containing no roots inside the unit circle. es are the error terms, $Y_\cdot^d$ is $Y_\cdot$ differenced d times, and c is a constant. Seasonal structure can be added with parameters $(P,D,Q)_m$ to the base ARIMA model to become ARIMA$(p,d,q)(P,D,Q)_m$, including a periodic component containing m periods. $c \neq 0$ implies a polynomial of order $d+D$ in the forecast function.

The main task in ARIMA-based forecasting is to select appropriate model orders, i.e., the values of p,q,d,P,Q,D. If d and D are known, we can select the orders p,q,P,Q via an information criterion such as the Akaike Information Criterion (AIC):

$$AIC = -2\log(L) + 2(p+q+P+Q+k), \tag{3}$$

where $k=1$ if $c \neq 0$ and 0 otherwise, and L is the maximized likelihood of the model fitted to the differenced data. The best fitted parameters of the ARIMA model can be determined by minimizing the AIC.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Similar to the LSTM-based gap filling, an ARIMA model is built for each combination of input and output window sizes for each well using the `auto.arima` function from the R package called `forecast` (Hyndman et al., 2007). The length of output N in ARIMA corresponds to the gap lengths. Each trained ARIMA model is only tested on the well that is used for training the model, not tested on any other wells (as is done in the LSTM approach). Accuracy of ARIMA-based gap filling can also be evaluated on the same synthetic gaps as in the LSTM approach using the same MAPE metric shown in Eq. (1).

4 Results and Discussion

4.1 Performance of LSTM in filling gaps

We evaluate the accuracy of LSTM in filling gaps of various lengths following the steps described in section 3.3. MAPEs are summarized for different LSTM model configuration parameters and for testing and training wells, as shown in Figure 6 for SpC. Other similar analyses for groundwater table and temperature are done but not shown here because SpC is our primary interest for this study. Each MAPE shown in the plots represents an average of a group of models with one parameter (corresponding to each x-axis) fixed at the given value while all the other parameters, including training-testing well pairs and gap scenarios, cycle through their possible combinations. As can be seen in Figure 6 (a), using more training data improves the model performance on average, consistent with a standard observation in machine learning applications that model performance is highly dependent on the amount of training data available. When the amount of training data increases from 2 to 4 and 6 years, the MAPE drops slightly with consistent variability across all combinations. We therefore conclude that 2 years of data is sufficient to train the LSTM models we need for gap filling, although 4 years of training data would be better.

In Figure 6 (b), model performance deteriorates as the gap length increases. This is because the performance of LSTMs tends to degrade as it loses ground truth information from its input to predict, i.e., the model begins to transition from interpolation to extrapolation. We also observe that models with a daily 24-hour input window outperform other models with longer input windows as shown in Figure 6 (c). This likely results from an optimal number of memory units for capturing daily and subdaily memories. The output window lengths are shown in log scale in Figure 6 (d) to allow sufficient separation between the smaller time windows (i.e., 1, 6, and 12 hours). Daily and subdaily output windows yield comparable performances in gap filling the SpC time series, with the 12-hour output window slightly outperforming its 1-, 6-, and 24-hour counterparts. There is a significant performance deterioration when the output window increases from 24 hours to 48 hours and beyond. Overall, an input window of 24 hours and an output windows of 12 hours appear to be a robust model configuration for all wells considered.

We also tested how models trained on one well perform in filling gaps in other wells (defined as testing wells), given their vast differences in dynamic signatures. The performance of models trained on each well is shown in Figure 6 (e), from which we observe that models trained on wells 1-15, 2-3,1-10A, and 2-5 performed comparably in filling in gaps in the other wells, with the models of well 1-15 leading with a small margin. Models trained on well 2-2 yielded the largest error when tested to fill gaps in the 2011 testing data of all wells including itself. When the performance is grouped by wells being tested, as shown in Figure 6 (f), all models can perform very well in filling in gaps for well 1-15 and reasonably well for wells 1-1, 1-10A, 2-3, and 2-5, while all models appear to have difficulty in filling gaps for well 2-2. When selecting the optimal LSTM configuration

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

of 24-hour input window and 12-hour output window using 4 years of training data, the models trained on well 2-3 perform the best in filling gaps of various lengths for all the 6 wells.

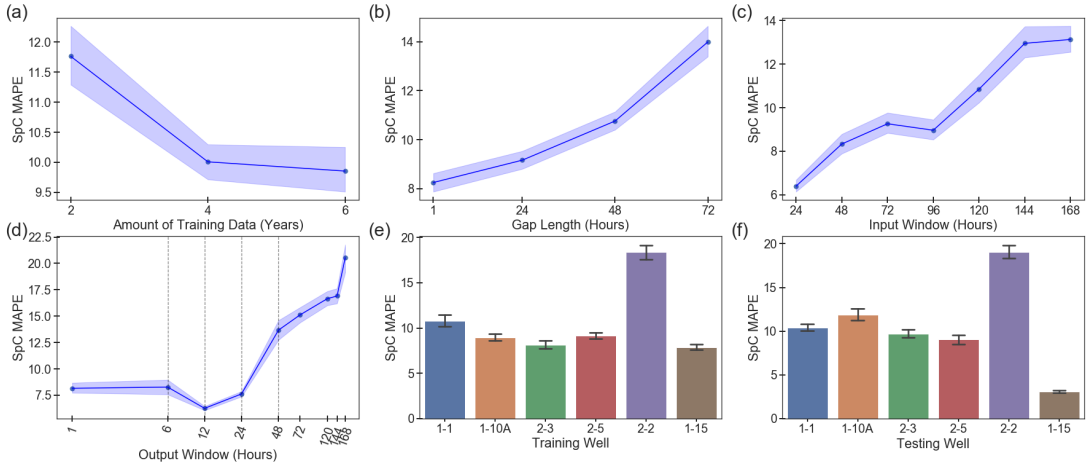

Figure 6. Gap filling performance for SpC evaluated against multiple model configuration parameters (a-d) or grouped by training and testing wells (e and f). (a) average MAPE vs. number of years of training data, (b) average MAPE vs. gap lengths, (c) average MAPE vs. input window size M, (d) average MAPE vs. output window size N, (e) average MAPE aggregated by wells used to train the models, and (f) average MAPE aggregated by wells being tested on. 95% confidence intervals of the averaged MAPE value are shown in shaded area in plots (a) -(d) and as the error bars in (e) and (f).

4.2 ARIMA and LSTM comparisons

Both ARIMA and LSTM approaches are tested in filling gaps of various lengths. Figure 7 shows the interquartile ranges of relative errors calculated for each data point that is assumed to be missing in the testing data by setting n=1 in Eq. (1), each bounded by its 25th to 75th percentiles under different gap lengths. The relative errors shown in Figure 7 are the results using the best model configurations trained for each well by the two approaches respectively. The gray lines representing the $\pm 5\%$ relative error range for each model correspond to typical measurement errors of the SpC sensors deployed at the site. Most of the relative errors yielded from both gap-filling methods are within $\pm 5\%$ measurement error except for well 2-2 with 48- and 72-hour gaps. The ARIMA models tend to perform better than the LSTM models in terms of error statistics. For both approaches, the relative errors increase as the gap length increases as expected. The relative errors in the ARIMA models tend to distribute symmetrically on both sides of 0, whereas errors in the LSTM models skew toward the negative side for wells 1-1 and 2-2 and towards the positive side for wells 1-10A and 2-3. Also for both approaches, the smallest and largest errors occur at wells 1-15 and 2-2, respectively. For well 1-15, the relative errors for all three gap lengths are very close to 0. Well 2-2 has the largest relative errors over the testing window. It is noted that the optimal input window size M for the LSTM models is

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions

smaller than that required by the ARIMA method for all the wells tested, indicating that LSTM models can rely on less input information than the ARIMA models to produce predictions of comparable accuracy. We also note that the optimal LSTM model configuration selected for each well based on its testing performance is different from that selected based on testing performance averaged across a range of training-testing well pairs.

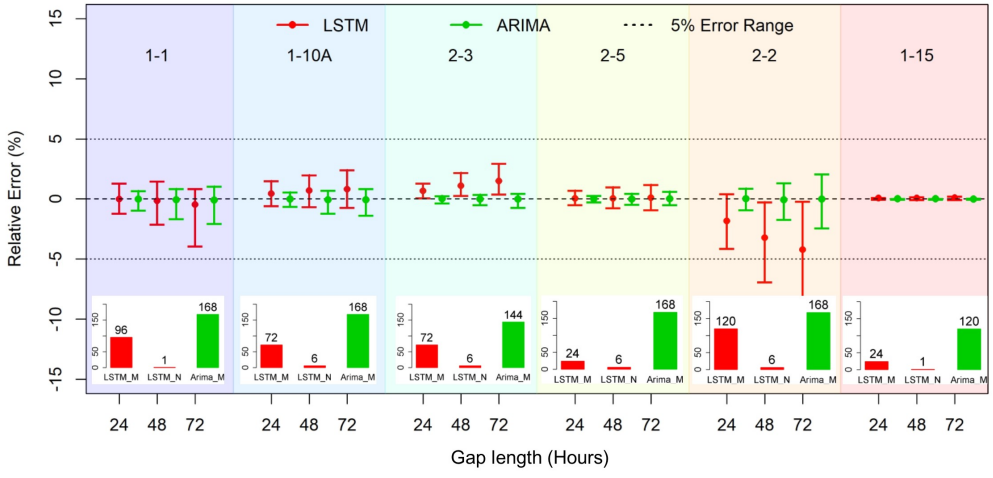

Figure 7. Summary of relative errors for filling gaps of various lengths (i.e., 24, 48, and 72 hours) for best LSTM and ARIMA models tested for each well. Their corresponding model input and output configuration (M and N for LSTM and and M for ARIMA) are shown for each well along the horizontal axis.

In addition to the error statistics, it is important to examine how well a gap-filling method can capture the desired dynamic patterns in the gap-filled time series. Therefore, the SpC time series reproduced by the gap-filling methods for the testing dataset with 24-hour synthetic gaps are evaluated against the real time series. Model configurations are the same as those used in error statistics comparison. As shown in the first and second columns of Figure 8, the ARIMA approach (column 1) can capture the smooth changes in the observations but not abrupt changes that occur over a short time window (i.e., at higher frequency); these occur in all wells except 1-15. This is an indication that ARIMA fails to capture higher frequency dynamics and nonlinear trends despite having smaller errors on average. The LSTM approach, on the other hand, is able to better resolve nonlinearity, nonstationary, and highly dynamic temporal patterns in time series, despite not having as small relative errors as the ARIMA approach. This holds for nearly all wells, including well 1-15, which has less dynamic behavior. Both gap filling methods exhibit difficulties in filling gaps for well 2-2 in terms of both relative errors and capturing real dynamic patterns, especially during January, October, and November when the SpC appeared to be highly dynamic.

To investigate how the relative performance of both gap-filling methods depends on the inherent dynamics in each time series, wavelet analyses results for the testing SpC dataset are extracted from the multi-year analyses (shown earlier in Figure 2). As shown in Figure 8 for the testing window of year 2011, the time windows of high relative errors are found to approximately co-

locate with the time when high-frequency signals are gaining more power. The LSTM models tend to outperform the ARIMA models during those time windows. Wells 1-1, 1-10A, 2-3, and 2-5 share similar seasonal patterns in WPS, with the highest intensity bin above 1024 hours. Among these four wells, well 2-5 has its greatest intensity above 2048 hours across the entire year. For well 1-15, the strongest intensities tend to group into three bands: one at 2048 hours across the entire year, one between 256 to 2048 hours from January to March, and one occurring below 128 hours in June. In general, both LSTM and ARIMA are effective at capturing longer term variability, but LSTM is more effective at capturing high-frequency fluctuations and nonlinearities in the dataset.

In terms of computational cost, ARIMA requires very little resources: the `auto.arima` function in R requires approximately 40 seconds for one year of data on a personal computer with a 3.00 GHz CPU. Conversely, training and testing a single LSTM model takes approximately 20-30 minutes on dual NVIDIA P100 12GB PCI-e based GPUs.

5 Conclusion

In this study, we implemented an LSTM to account for spatio-temporal correlations in a 10-year spatially distributed time series dataset collected by a groundwater monitoring well network. We evaluated the performance of the LSTM-based gap filling method by creating synthetic data gaps of various lengths (24, 48 and 72 hours) so that the accuracy of the filled data could be quantified in terms of error statistics and how well the original nonlinear dynamics are captured. The performance of the LSTM-based gap-filling method is compared to that of a traditional, popular gap-filling method, ARIMA.

In general, both ARIMA and LSTM are reasonably effective at filling in gaps of 24, 48, and 72 hours. The relative errors are mostly within the range of instrument measurement error. The models both capture long-term trends in data, except during some time windows with highly dynamic fluctuations. ARIMA is found to be suitable for time series with less dynamic behavior. LSTMs excel in dealing with high-frequency dynamics or nonlinearities, although they do require more training data and computational power. Availability of sufficient training data is critical for the success of LSTM methods, as with any DNN-based learning methods. In the gap-filling use case studied here, 2 years of training data yielded similar results compared with 4 and 6 years of training data. A general guideline is to have training data that covers various scenarios of inter-annual, seasonal, daily and even sub-daily dynamics, which is best assessed by understanding the nature of physical processes and drivers underlying the time series data.

Wavelet analysis could provide useful insights to the dynamic signatures of the data and the change in composition of their important frequencies over time, which can serve as a prior basis for selecting an appropriate gap-filling method. For example, the ARIMA method would work well if the dynamics are dominated by seasonal cycles, while more sophisticated approaches like LSTMs could work better if there is evidence of daily and subdaily fluctuations. There may also be challenging situations for LSTMs, such as the highly dynamic time windows in our case study. The capability of learning over long sequences differentiates LSTM from other RNNs that do not have such memory. Depending on the mixture of long- and short-term variability inherent in the time series, different LSTM configurations can be further explored and evaluated to achieve better performance in capturing more complex dynamics. Capturing such dynamics is essential for generating the most valuable

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions

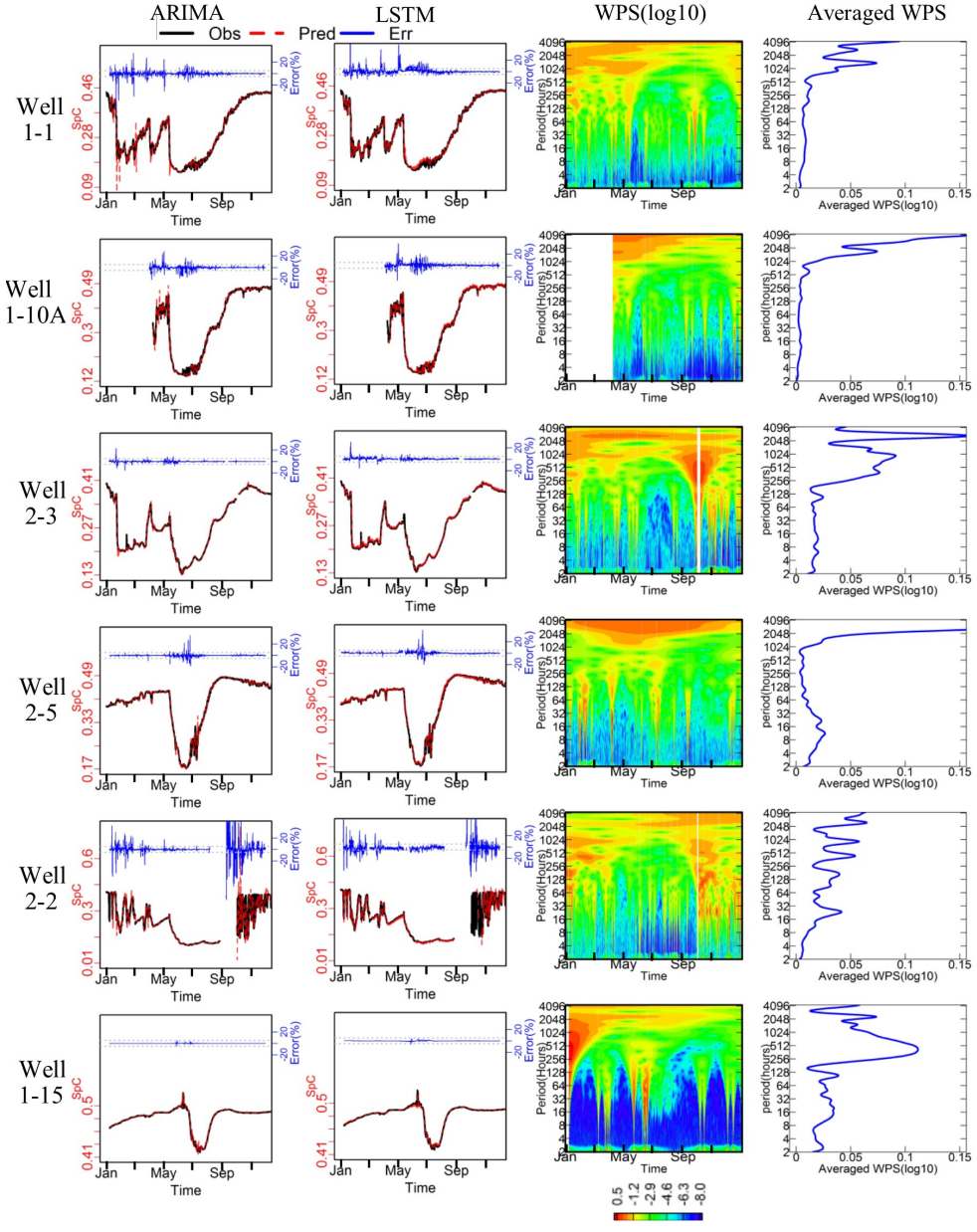

Figure 8. Columns 1 and 2 show time series of ARIMA and LSTM (red), respectively, in filling 24-hour gap lengths, compared with observations (black) and relative error (blue). The ARIMA model takes 72-hour inputs, whereas the input and output window sizes for the LSTM model are 72 and 6 hours, respectively. The LSTM model is trained on 4 years of data from well 2-3. Column 3 is the spectrogram of each well for the year 2011, column 4 is the averaged WPS for year 2011.

insights to advance our understanding of dynamic complex systems. Future research could involve time series from multiple locations to explicitly account for spatial correlations.

Code and data availability. The well observations have been made accessible at https://sbrsfa.velo.pnnl.gov/datasets/?UUID=14febd81-05b6-47fb-be52-439c4382decd

Author contributions. HR and EC developed scripts and performed the analyses. BK contributed on interpretation of the results. XC conceived and designed the study. All authors contributed to writing the manuscript.

Competing interests. The authors declare that they have no conflicts of interest.

Acknowledgements. This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Subsurface Biogeochemical Research Program (SBR). A portion of methodology development was supported by the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for DOE under contract DE-AC05-76RL01830. This research was performed using PNNL Institutional Computing at Pacific Northwest National Laboratory. This research was also supported in part by the Indiana University Environmental Resilience Institute and the *Prepared for Environmental Change* grand challenge initiative.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-196
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 16 May 2019

Hydrology and Earth System Sciences
Discussions

References

Arntzen, E. V., Geist, D. R., and Dresel, P. E.: Effects of fluctuating river flow on groundwater/surface water mixing in the hyporheic zone of a regulated, large cobble bed river, River Research and Applications, 22, 937–946, 2006.

Banerjee, S., Carlin, B. P., and Gelfand, A. E.: Hierarchical modeling and analysis for spatial data, CRC press, 2014.

Calculli, C., Fassò, A., Finazzi, F., Pollice, A., and Turnone, A.: Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in Apulia, Italy, Environmetrics, 26, 406–417, 2015.

Chen, X., Murakami, H., Hahn, M. S., Hammond, G. E., Rockhold, M. L., Zachara, J. M., and Rubin, Y.: Three-dimensional Bayesian geostatistical aquifer characterization at the Hanford 300 Area using tracer test data, Water Resources Research, 48, 2012.

Chen, X., Hammond, G. E., Murray, C. J., Rockhold, M. L., Vermeul, V. R., and Zachara, J. M.: Application of ensemble-based data assimilation techniques for aquifer characterization using tracer data at Hanford 300 area, Water Resources Research, 49, 7064–7076, 2013.

Cheng, T., Haworth, J., and Wang, J.: Spatio-temporal autocorrelation of road network data, Journal of Geographical Systems, 14, 389–413, https://doi.org/10.1007/s10109-011-0149-5, https://doi.org/10.1007/s10109-011-0149-5, 2012.

Cheng, T., Haworth, J., Anbaroglu, B., Tanaksaranond, G., and Wang, J.: Spatiotemporal data mining, in: Handbook of regional science, pp. 1173–1193, Springer, 2014.

Connor, J. T., Martin, R. D., and Atlas, L. E.: Recurrent neural networks and robust time series prediction, IEEE transactions on neural networks, 5, 240–254, 1994.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E.: Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets, Journal of the American Statistical Association, 111, 800–812, 2016.

Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M., and Niemi, J.: Estimation and prediction in spatial models with block composite likelihoods, Journal of Computational and Graphical Statistics, 23, 295–315, 2014.

Faruk, D. Ö.: A hybrid neural network and ARIMA model for water quality time series prediction, Engineering Applications of Artificial Intelligence, 23, 586–594, 2010.

Finley, A. O., Banerjee, S., and Gelfand, A. E.: spBayes for large univariate and multivariate point-referenced spatio-temporal data models, arXiv preprint arXiv:1310.8192, 2013.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G.: Could Machine Learning Break the Convection Parameterization Deadlock?, Geophysical Research Letters, 45, 5742–5751, https://doi.org/10.1029/2018GL078202, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078202, 2018.

Ghil, M., Allen, M., Dettinger, M., Ide, K., Kondrashov, D., Mann, M., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., et al.: Advanced spectral methods for climatic time series, Reviews of geophysics, 40, 3–1, 2002.

Grant, G. E. and Dietrich, W. E.: The frontier beneath our feet, Water Resources Research, 53, 2605–2609, 2017.

Graves, A.: Generating sequences with recurrent neural networks, arXiv preprint arXiv:1308.0850, 2013.

Graves, A., Mohamed, A.-r., and Hinton, G.: Speech recognition with deep recurrent neural networks, in: Acoustics, speech and signal processing (icassp), 2013 ieee international conference on, pp. 6645–6649, IEEE, 2013.

Griffith, D. A.: Modeling spatio-temporal relationships: retrospect and prospect, Journal of Geographical Systems, 12, 111–123, 2010.

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear processes in geophysics, 11, 561–566, 2004.

Grossmann, A. and Morlet, J.: Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM journal on mathematical analysis, 15, 723–736, 1984.

Güler, C. and Thyne, G. D.: Hydrologic and geologic factors controlling surface and groundwater chemistry in Indian Wells-Owens Valley area, southeastern California, USA, Journal of Hydrology, 285, 177–198, 2004.

Han, P., Wang, P. X., Zhang, S. Y., and Zhu, D. H.: Drought forecasting based on the remote sensing data using ARIMA models, Mathematical and Computer Modelling, 51, 1398–1403, 2010.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R.: Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580, 2012.

Ho, S., Xie, M., and Goh, T.: A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction, Computers & Industrial Engineering, 42, 371–375, 2002.

Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., et al.: Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.

Hocke, K. and Kämpfer, N.: Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram, Atmospheric Chemistry and Physics Discussions, 8, 4603–4623, 2008.

Hyndman, R. J., Khandakar, Y., et al.: Automatic time series for forecasting: the forecast package for R, 6/07, Monash University, Department of Econometrics and Business Statistics, 2007.

Kamarianakis, Y. and Prastacos, P.: Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches, Transportation Research Record: Journal of the Transportation Research Board, pp. 74–84, 2003.

Kamarianakis, Y. and Prastacos, P.: Space–time modeling of traffic flow, Computers & Geosciences, 31, 119–133, 2005.

Katzfuss, M. and Cressie, N.: Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets, Journal of Time Series Analysis, 32, 430–446, 2011.

Katzfuss, M. and Cressie, N.: Bayesian hierarchical spatio-temporal smoothing for very large datasets, Environmetrics, 23, 94–107, 2012.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, CoRR, abs/1412.6980, http://arxiv.org/abs/1412.6980, 2014.

Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets, Nonlinear Processes in Geophysics, 13, 151–159, 2006.

Kondrashov, D., Shprits, Y., and Ghil, M.: Gap filling of solar wind data by singular spectrum analysis, Geophysical research letters, 37, 2010.

Lin, C. Y., Abdullah, M. H., Praveena, S. M., Yahaya, A. H. B., and Musta, B.: Delineation of temporal variability and governing factors influencing the spatial variability of shallow groundwater chemistry in a tropical sedimentary island, Journal of hydrology, 432, 26–42, 2012.

Liu, J., Shahroudy, A., Xu, D., and Wang, G.: Spatio-temporal lstm with trust gates for 3d human action recognition, in: European Conference on Computer Vision, pp. 816–833, Springer, 2016.

Längkvist, M., Karlsson, L., and Loutfi, A.: A review of unsupervised feature learning and deep learning for time-series modeling, Pattern Recognition Letters, 42, 11 – 24, https://doi.org/https://doi.org/10.1016/j.patrec.2014.01.008, http://www.sciencedirect.com/science/article/pii/S0167865514000221, 2014.

Ma, X., Tao, Z., Wang, Y., Yu, H., and Wang, Y.: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data, Transportation Research Part C: Emerging Technologies, 54, 187–197, 2015.

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

Pfeifer, P. E. and Deutrch, S. J.: A three-stage iterative procedure for space-time modeling phillip, Technometrics, 22, 35–47, 1980.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, http://www.nature.com/articles/s41586-019-0912-1, 2019.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85 – 117, https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.003, http://www.sciencedirect.com/science/article/pii/S0893608014002135, 2015.

Shen, C.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resources Research, 54, 8558–8593, https://doi.org/10.1029/2018WR022643, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022643, 2018.

Song, S., Lan, C., Xing, J., Zeng, W., and Liu, J.: An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data., in: AAAI, vol. 1, pp. 4263–4270, 2017.

Song, X., Chen, X., Stegen, J., Hammond, G., Song, H.-S., Dai, H., Graham, E., and Zachara, J. M.: Drought Conditions Maximize the Impact of High-Frequency Flow Variations on Thermal Regimes and Biogeochemical Function in the Hyporheic Zone, Water Resources Research, 2018.

Stockwell, R. G., Mansinha, L., and Lowe, R.: Localization of the complex spectrum: the S transform, IEEE transactions on signal processing, 44, 998–1001, 1996.

Strobl, R. O. and Robillard, P. D.: Network design for water quality monitoring of surface freshwaters: A review, Journal of environmental management, 87, 639–648, 2008.

Stroud, J. R., Stein, M. L., and Lysen, S.: Bayesian and maximum likelihood estimation for Gaussian processes on an incomplete lattice, Journal of computational and Graphical Statistics, 26, 108–120, 2017.

Sun, A. Y.: Discovering State-Parameter Mappings in Subsurface Models Using Generative Adversarial Networks, Geophysical Research Letters, 45, 11,137–11,146, https://doi.org/10.1029/2018GL080404, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080404, 2018.

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., and Zhong, Z.: Combining Physically Based Modeling and Deep Learning for Fusing GRACE Satellite Data: Can We Learn From Mismatch?, Water Resources Research, 55, 1179–1195, https://doi.org/10.1029/2018WR023333, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023333, 2019.

Taylor, C. J. and Alley, W. M.: Ground-water-level monitoring and the importance of long-term water-level data, 1217-2002, US Geological Survey, 2002.

Vacha, L. and Barunik, J.: Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis, Energy Economics, 34, 241–247, 2012.

Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L. J., Guillén, A., Marquez, L., and Pasadas, M.: Hybridization of intelligent techniques and ARIMA models for time series prediction, Fuzzy sets and systems, 159, 821–845, 2008.

Wang, G., Garcia, D., Liu, Y., De Jeu, R., and Dolman, A. J.: A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations, Environmental Modelling & Software, 30, 139–142, 2012.

Wett, B., Jarosch, H., and Ingerle, K.: Flood induced infiltration affecting a bank filtrate well at the River Enns, Austria, Journal of Hydrology, 266, 222–234, 2002.

Wikle, C. K., Berliner, L. M., and Cressie, N.: Hierarchical Bayesian space-time models, Environmental and Ecological Statistics, 5, 117–154, 1998.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Łukasz Kaiser, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J.: Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, CoRR, abs/1609.08144, http://arxiv.org/abs/1609.08144, 2016.

You, Q., Jin, H., Wang, Z., Fang, C., and Luo, J.: Image captioning with semantic attention, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4651–4659, 2016.

Zachara, J. M., Long, P. E., Bargar, J., Davis, J. A., Fox, P., Fredrickson, J. K., Freshley, M. D., Konopka, A. E., Liu, C., McKinley, J. P., et al.: Persistence of uranium groundwater plumes: contrasting mechanisms at two DOE sites in the groundwater–river interaction zone, Journal of contaminant hydrology, 147, 45–72, 2013.

Zachara, J. M., Chen, X., Murray, C., and Hammond, G.: River stage influences on uranium transport in a hydrologically dynamic groundwater-surface water transition zone, Water Resources Research, 52, 1568–1590, 2016.

Zhang, G. P.: Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing, 50, 159–175, 2003.