IMPUTATION OF CONTIGUOUS GAPS AND EXTREMES OF SUBHOURLY GROUNDWATER TIME SERIES USING RANDOM FORESTS

Dipankar Dwivedi,[1,*] Utkarsh Mital,[1] Boris Faybishenko,[1] Baptiste Dafflon,[1] Charuleka Varadharajan,[1] Deborah Agarwal,[2] Kenneth H. Williams,[1] Carl I. Steefel,[1] & Susan S. Hubbard[1]

[1]*Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

[2]*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

*Address all correspondence to: Dipankar Dwivedi, Earth Environmental Sciences Area Geosciences Division Building 074A, Room 0102 M/S 74R316C Lawrence Berkeley National Laboratory 1 Cyclotron Road, M.S. 74R316C Berkeley, CA 94720 USA, E-mail: ddwivedi@lbl.gov

Original Manuscript Submitted: 4/16/2021; Final Draft Received: 8/19/2021

Machine learning can provide sustainable solutions to gap-fill groundwater (GW) data needed to adequately constrain watershed models. However, imputing missing extremes is more challenging than other parts of a hydrograph. To impute missing subhourly data, including extremes, within GW time-series data collected at multiple wells in the East River watershed, located in southwestern Colorado, we consider a single-well imputation (SWI) and a multiple-well imputation (MWI) approach. SWI gap-fills missing GW entries in a well using the same well's time-series data; MWI gap-fills a specific well's missing GW entry using the time series of neighboring wells. SWI takes advantage of linear interpolation and random forest (RF) approaches, whereas MWI exploits only the RF approach. We also use an information entropy framework to develop insights into how missing data patterns impact imputation. We discovered that if gaps were at random intervals, SWI could accurately impute up to 90% of missing data over an approximately two-year period. Contiguous gaps constituted more complex scenarios for imputation and required the use of MWI. Information entropy suggested that if gaps were contiguous, up to 50% of missing GW data could be estimated accurately over an approximately two-year period. The RF-feature importance suggested that a time feature (months) and a space feature (neighboring wells) were the most important predictors in the SWI and MWI. We also noted that neither SWI nor MWI methods could capture the missing extremes of a hydrograph. To counter this, we developed a new sequential approach and demonstrated the imputation of missing extremes in a GW time series with high accuracy.

KEY WORDS: *groundwater, gap filling, modeling, extremes, information entropy, sequential imputation*

1. INTRODUCTION

The application of watershed numerical models for predicting water quantity and water quality in both space and time requires high-resolution spatial and temporal data for enhanced predictive capabilities, given the extreme heterogeneity of hydrological processes (e.g., Arora et al., 2016b, 2019a; Dwivedi and Mohanty, 2016; Li et al., 2017; Likens, 2001; Zachara et al., 2016). Recently, sensor networks and environmental observatories have revolutionized watershed science by providing a range of high-resolution spatial and temporal data (Hubbard et al., 2020; Reichstein et al., 2019; Varadharajan et al., 2019). Despite this progress, datasets are often incomplete, containing unspecified or missing entries owing to various reasons such as equipment failures, improper system maintenance, power outages, and extreme weather conditions. Groundwater (GW) time series is a prominent example of time-series data with missing entries. Such missing entries are usually estimated to appropriately constrain numerical models in order to capture the spatiotemporal variability of watershed processes. This process of estimating, or gap-filling, missing data is commonly referred to as "imputation."

Despite its importance, there are no standardized methods for estimating missing GW values. The literature includes several approaches for imputing missing entries in a GW time series based on the availability of annual, monthly, and daily GW data. In general, conventional statistical approaches, such as autoregression or spatial interpolation, work well for imputing GW data at seasonal time scales (Dax and Zilberbrand, 2018; Moritz and Bartz-Beielstein, 2017) because strong inherent seasonality and comparable interannual hydrographs of GW time series make it relatively unchallenging for estimating missing GW entries. However, these multiple autoregressive approaches are not very efficient for long, contiguous gaps. Other methods include spectral analyses, such as singular spectrum analysis and multichannel spectrum analysis, for estimating or imputing missing GW entries through exploring time and frequency domains interchangeably (e.g., Aissia et al., 2017; Kondrashov et al., 2005). Then again, these spectral techniques tend to perform less satisfactorily if the GW time series deviates from a normal/Gaussian distribution and includes extreme values.

Missing extreme values and long contiguous gaps are considerable challenges when imputing missing GW data. These challenges require leveraging missing GW data information from other related (but diverse) datasets (e.g., precipitation, evapotranspiration, temperature). More recently, machine-learning-based approaches have gained momentum for imputing the GW time series because of their flexibility and versatility in dealing with diverse data. Khedri et al. (2020) compared several machine-learning (ML) approaches, including artificial neural networks, fuzzy logic, adaptive neuro-fuzzy inference system, neural net group method of data handling, and support vector machines, for short-term (one to three months) groundwater level predictions. They considered precipitation, temperature, and evapotranspiration as input features for estimating monthly missing GW data. Müller et al. (2019) used a suite of machine-learning-based models (e.g., multilayer perceptron, convolutional, recurrent, and long short-term memory neural networks) to predict daily GW levels at several monitoring wells in California.

Although the approaches mentioned above are efficient and reasonable for imputing missing GW levels, these approaches have been primarily used to impute annual, monthly, and—on a very few occasions—daily GW data. To date, there is no robust imputation technique to gap-fill subdaily to subhourly GW time-series data, even though they are often used for constraining high-resolution watershed ecohydrological models. Imputation techniques to fill subdaily to subhourly data have not been investigated partly because modern sensor networks have only recently

started catering to high-resolution GW temporal data (Varadharajan et al., 2019). In addition to long contiguous gaps and missing extremes, another key point is the missing data pattern that is equally important to consider for accurately imputing missing GW entries. However, it has not been thoroughly investigated.

The missing data pattern poses a significant challenge not only toward imputing GW data but also for standardizing a GW imputation strategy (Aguilera et al., 2020; Aissia et al., 2017; Yozgatligil et al., 2013). For example, a single missing record can be gap-filled by either estimating averaged values before and after the timestamp (i.e., Unix digital time) or spatially correlating nearby GW records (Aggarwal and Zhai, 2012). However, a GW time series typically does not have just a few gaps that are easy to impute but has several missing entries or contiguous gaps. Indeed, imputation accuracy depends upon the missing data fraction, missing data patterns, and variability of the time series. A highly variable time series is more difficult to impute than a relatively uniform time series. As an example, if a ten-year time series has 50% missing data versus a one-year time series with only 20% missing data, it may be easier to fill gaps in the ten-year time series if it has a less challenging missing data pattern. A long time series with 50% missing entries may still include adequate information about the temporal variability of GW data. However, if missing data is centered around extremes of a hydrograph, and the GW time series does not contain complete information about temporal variability, then imputing extremes can be more challenging than other parts of the hydrograph. In this work, we use information theory (e.g., Arora et al., 2019b; Cui and Singh, 2015; Dwivedi et al., 2018a; Singh, 2010a) to assess how much information about temporal variability is contained in a GW time series and infer imputation performance (or reliability) beforehand. Specifically, information theory helps us infer the effects of missing data fraction and missing data pattern on imputation. With this in mind, we will elaborate the use of information entropy for assessing imputation reliability in the Discussion section (Section 4).

Here, we explore (a) single-well imputation (SWI), which gap-fills missing GW data in a well using the same well's time-series data, and (b) multiple-well imputation (MWI), which gap-fills a specific well using GW data from other wells located in that well's proximity. The SWI method takes advantage of linear interpolation (LI) and random forest (RF) approaches, whereas the MWI method uses only the RF approach. We choose the RF regressor because it is computationally efficient for large datasets and can handle many input variables. In addition, RF informs the relative feature importance of predictor variables (Du et al., 2020; Pedregosa et al., 2011). Specifically, our goal is to develop a robust framework to gap-fill subhourly GW data based on the following missing data patterns: (a) random gaps, (b) contiguous gaps away from extremes, and (c) contiguous gaps around extremes. We use information theory to investigate what percentage of missing data can effectively be imputed. Subsequently, we develop a strategy to sequentially impute missing data (including extremes) from the GW time series across multiple wells. For developing the sequential imputation strategy, we hypothesize that (a) MWI is more effective than SWI, and (b) imputing along a sequence (ascending order of missing data fraction) across wells, and then leveraging each gap-filled well data progressively to impute the next well, enables high imputation accuracy. The rationale behind the first part of the hypothesis is that neighboring wells can provide additional information for imputing missing GW data in a time series. The second part of the hypothesis benefits from gleaning information from wells with small gaps, likely to be imputed more effectively than wells with large gaps. Finally, this is the first study that (a) assesses imputation reliability using information entropy, and (b) employs a sequential approach in order to impute contiguous gaps around extremes.

2. METHODOLOGY

2.1 Field Site and Datasets

2.1.1 Site Location and Hydrologic Characteristics

We conducted this study using the GW level data collected in monitoring wells at the East River watershed (southwestern Colorado), which is a high elevation drainage basin located between 38.8° to 38.9° N and 106.8° to 106.9° W (Fig. 1). The East River watershed is a study site of the Watershed Function Scientific Focus Area project at Berkeley Lab, funded by the U.S. Department of Energy (Arora et al., 2020; Hubbard et al., 2018). There are several intensive sampling sites in the East River watershed that serve to give us a better understanding of how mountainous watersheds respond to climatic perturbations. The East River floodplain is one of several such intensive sampling sites, extending over 11 km of East River and encompassing multiple meanders (Fig. 1). The floodplain is hydrologically connected to the East River system through an underlying unconfined alluvial aquifer (Dwivedi et al., 2017, 2018b). The East River predominantly receives flow from snowmelt in late spring (mid-April) to early summer (June), followed by monsoonal rainfall from mid to late summer (mid-June to mid-October).

2.1.2 Water Level Data

To capture the inherent spatial and temporal variability of a floodplain environment, we measured water levels at approximately 15 min intervals from November 1, 2015 to August 24, 2019 in seven monitoring wells using pressure transducers (model Onset HOBO U2). The data were downloaded from the sensors periodically, corrected for changes in atmospheric pressure, and

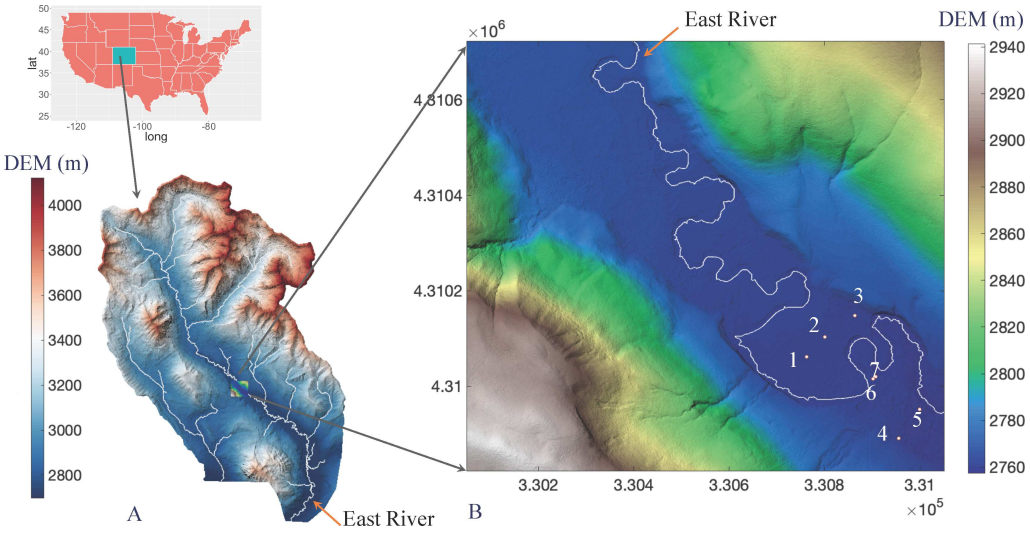

FIG. 1: (a) The study site is located in the East River watershed, a high elevation catchment in southwestern Colorado. (b) Seven monitoring wells (WLE1 to WLE7, marked as 1 to 7) are located in the East River watershed floodplain.

converted to GW level depth. Further, real-time kinematic GPS measurements of the well locations and elevations were used to infer GW level elevation (Dafflon and Dwivedi, 2020). We resampled the GW time-series data uniformly at 30 min intervals for consistency. The layout of GW wells is shown in Fig. 1(b). All the wells are located within a 1.4 m-wide range in elevation, from 7.7 to 270 m apart from each other, and from 3 to 89 m away from the river [Fig. 1(b)]. The in-well depth to the sensor varies between 0.65 and 1.9 m depth (for more details, see Appendix A; Table A1).

GW levels at seven wells (WLE1 to WLE7) within the meandering floodplain are shown in Fig. 2. All the wells have gaps ranging from several days to several months. Moreover, during certain periods, the wells have synchronous missing records (i.e., missing at the same time). The period in which wells have synchronous missing records resulted from improper system maintenance, while the other gaps came about for various reasons, including sensor failure, the GW table being deeper than the sensor, and at times improper system maintenance.

2.2 Imputation Strategy

To impute missing data, we use SWI and MWI. SWI exploits LI and RF approaches (hereafter, SWI-LI and SWI-RF) to gap-fill missing GW data in a well using the same well's time-series data. MWI takes advantage of the RF approach (hereafter, MWI-RF) to gap-fill a specific well using GW data from other wells located in that well's proximity. In SWI-LI, a linear interpolation assumes that the values are equally spaced and estimates missing data points by fitting a linear function within the range of a discrete set of available data points (Fig. 3). Below, we provide

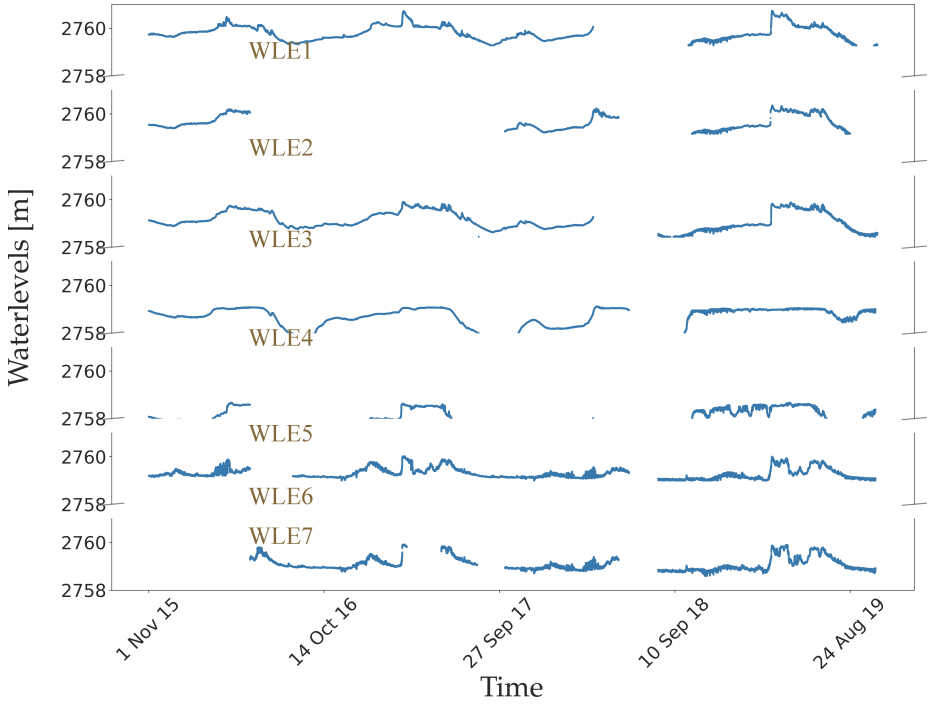

FIG. 2: GW data in wells WLE1 to WLE7 showing gaps at different times

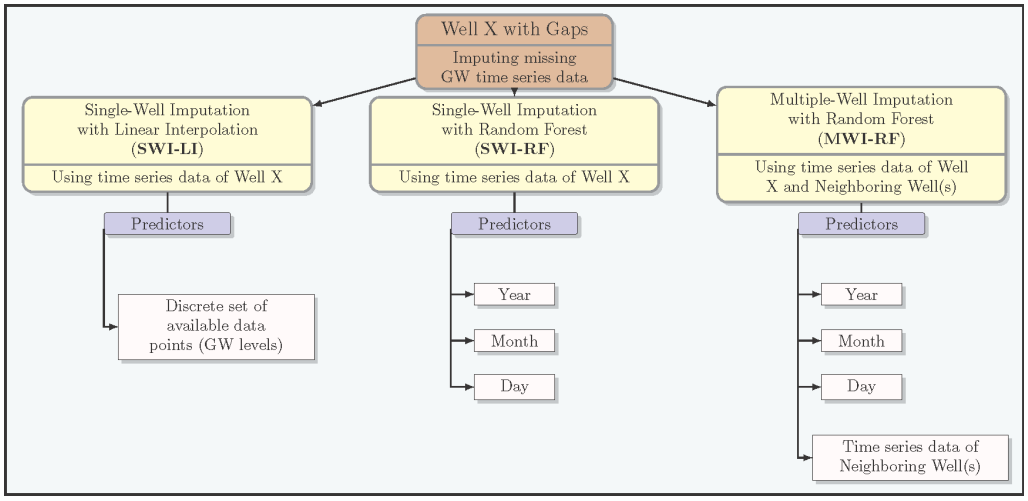

FIG. 3: Schematic of the imputation strategy to gap-fill missing GW time-series data

a brief overview of RF and a performance metric (i.e., Nash–Sutcliffe efficiency) for evaluating the imputation performance.

2.2.1 Random Forest

RF is based on an aggregation of decision trees (Breiman, 1996), which are flowchartlike structures that recursively partition the input feature space into smaller subspaces. Data in each subspace are modeled using simple linear fits. For regression, recursive partitioning of subspaces is carried out such that the mean-squared error between the tree output and the observed output is minimized. The RF model trains each decision tree on a different subset of data points obtained by sampling the training data with a replacement.

Furthermore, each tree is trained by considering a different subset of randomly selected input features. The final output of the RF is obtained by aggregating the results of all decision trees. For regression problems, aggregation is done by considering the mean. For completeness, a short description of the RF regressor is given below.

A binary tree is grown, and at each node, the data are split into two daughter nodes based on a splitting criterion. The predicted value at a node is the average response variable for all observations in the node.

Given data $D = \{(X_i, Y_i), i = 1, \ldots, n\}$, a random forest model $\hat{f}$ is defined as follows:

$$\hat{Y} = \hat{f}(X) \tag{1}$$

where $X = (X_1, X_2, \ldots, X_p)$ are p predictor variables (e.g., year, month, day, GW levels at neighboring wells); Y is a continuous response variable (e.g., GW predictions).

In this study, we use the RandomForestRegressor (RF Regressor) toolbox from Python's sklearn package (Pedregosa et al., 2011). Some examples using RF approach include Giannakou et al. (2020), Mital et al. (2020), Stockman et al. (2019), and Oppel and Schumann (2020).

2.2.2 Evaluating Imputation Accuracy: Nash-Sutcliffe Efficiency (NSE)

The overall imputation performance was evaluated by computing the Nash–Sutcliffe efficiency (*NSE*) on test data given by

$$NSE = 1 - \frac{\sum_{i=1}^{N}(y_i^o - y_i^m)^2}{\sum_{i=1}^{N}(y_i^o - \bar{y}^o)^2} \tag{2}$$

where N is the test set size, y_i^o is the ith observed value, y_i^m represents the corresponding modeled value, and $\bar{y}^o$ is the mean of all observed values in the test set.

The *NSE* is widely used for evaluating model performances in hydrology. It calculates the relative magnitude of a model's residual variance compared to the measured data variance. It indicates how well the observed versus modeled data fit the 1:1 line (Moriasi et al., 2007). The *NSE* is dimensionless and ranges from $-\infty$ to 1. An *NSE* value equal to 1 implies that the modeled values match the observations perfectly, an *NSE* value equal to 0 implies that the modeled values are only as good as the mean of observations, and a negative *NSE* value implies that the mean of observations is a better predictor than the modeled values. Positive *NSE* values are desirable, and higher values imply greater accuracy of the model.

2.2.3 Information Entropy

Next, we describe an information entropy framework that we use to develop insights into various aspects of the imputation process, such as missing data patterns and gaps around extremes that affect imputation performance. Imputation performance depends on the missing data patterns and variability in the GW time series. Variability in time-series data is characterized by several measures such as randomness, periodicity, discontinuity, and systematic variation. Several descriptive statistical measures can quantify variability in a GW time series, such as range, mean, standard deviation, and coefficient of variation. In contrast, Shannon entropy (also known as information entropy) exploits a probabilistic approach to measure a random variable's spatial and temporal variability. Shannon entropy holds an advantage over descriptive statistics because entropy measures the uncertainty associated with a random variable and is a nonparametric quantity (Hockett, 1953; Shannon, 1948). The entropy approach has been extensively used in hydrology for describing spatial and temporal variability in data (e.g., Arora et al., 2016a; Dwivedi and Mohanty, 2016; Mays et al., 2002; Singh, 1997).

For a discrete random variable X, the equation for Shannon entropy H is given by

$$H = -\sum_{n=1}^{N} p(X_n)\log[p(X_n)] \tag{3}$$

where n is the index of a discrete data interval (or bin), varying from 1 to N; X_n is the outcome within interval n; and $p(X_n)$ is the probability of X_n. The probabilities are computed by counting the number of observations that fall into various data intervals or bins. The logarithmic function can be natural (base e), binary (base 2), or common (base 10), resulting in information entropy with units of napiers (or nats), bits, or decibels, respectively (Shannon, 1948). However, interpretations do not depend upon the choice of a particular base. We use the natural log function to compute entropy (in nats) in this study.

2.3 Numerical Experiments for GW Imputation

As previously stated, our goal is to develop a robust framework to gap-fill subhourly GW data. Figure 4 shows the missing data pattern of GW levels at the wells located in the East River floodplain. For the entire duration of the study (i.e., from November 2015 to August 2019), wells WLE2, WLE5, and WLE7 are missing a significant portion of data: approximately 47%, 35%, and 29%, respectively. All other wells have less than 15% missing data. When analyzing missing patterns and gaps, the entire duration can be divided into three periods [Fig. 4(b)]. First, from November 2015 to March 2018 (Period 1), where well WLE2 has the highest portion of missing data (57%), followed by wells WLE7 (37%), and WLE5 (34%); during this same period, wells WLE1, WLE3, and WLE4 have less than 0.6% missing data, and contiguous gaps are primarily a few hours [Fig. 4(c)]. Second, from April 2018 to November 2018 (Period 2), when all wells have synchronous gaps for a significant block: wells WLE1, WLE3, and WLE 5 are missing almost all data, followed by WLE2 (60%), WLE7 (60%), WLE4 (45%), and WLE6 (45%) [Fig. 4(d)]. Finally, from November 2018 to August 2019 (Period 3), WLE1, WLE2, and WLE5 have less than 20% missing data, and all other wells have only a small fraction of gaps ($< 6\%$), as shown in Fig. 4(e).

To develop a robust framework to gap-fill subhourly GW data, we start by synthetically creating gaps in the time-series data of wells WLE1 and WLE3. Subsequently, we impute these gaps using SWI and MWI methods and compare imputed data with the original data. We use wells WLE1 and WLE3 because these two wells have an almost complete GW time series in Period 1. We design three numerical experiments considering (a) random gaps, (b) contiguous gaps away from extremes, and (c) contiguous gaps around extremes. These missing data patterns represent

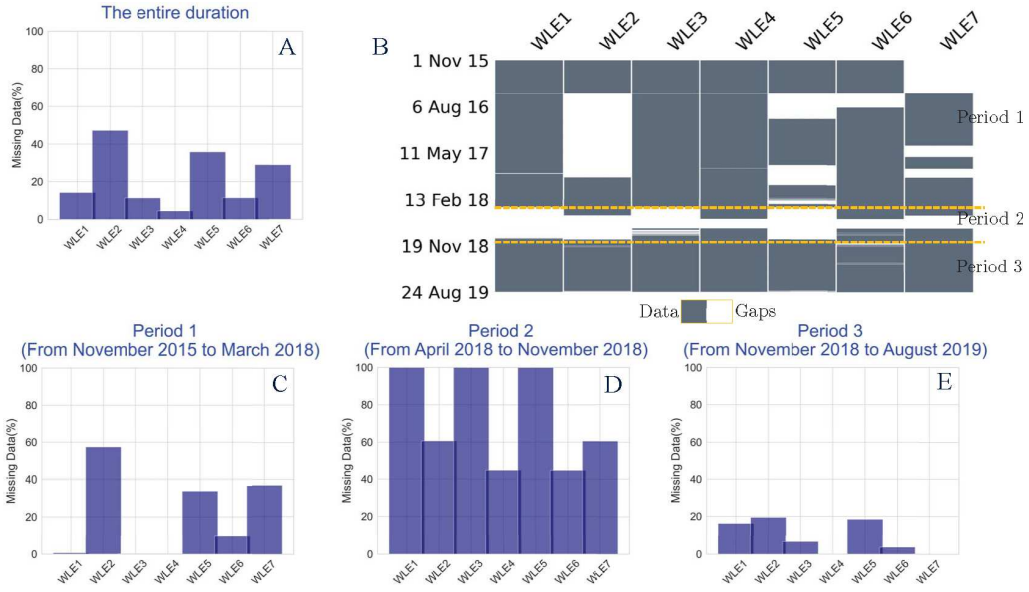

FIG. 4: (a) Total percentage of missing data in each monitoring well for the entire duration. (b) The entire duration can be divided into three periods based on large contiguous, synchronous, and small gaps. For these periods, (c), (d), and (e) show the total percentage of missing data in each monitoring well.

the simplest to the most complex scenarios in GW time series data. The first numerical experiment (E1) involves 10%, 20%, 50%, and 90% missing-at-random gaps (Fig. 5). The second numerical experiment (E2) involves 10%, 20%, and 50% contiguous gaps not centered around peaks or troughs, i.e., away from extremes (Fig. 6). The third numerical experiment (E3) involves 10%, 20%, and 50% gaps that are contiguous and located around the extremes of the hydrograph (Fig. 7). The three numerical experiments are summarized in Table 1.

2.4 Sequential Imputation of Missing Data across Multiple Wells

To test the first part of our hypothesis, that MWI is more effective than SWI, we evaluate both methods using numerical experiments as described above. The second path of the hypothesis sets the stage for the sequential imputation strategy. It states that following a sequence along the percentage (lowest to highest) of missing data across wells can enable high imputation accuracy. To test this, below we provide the following key steps.

1. Divide GW time-series data into different periods such that missing data in different wells have the least possible synchronous gaps.

2. Find fractions of missing data in different wells for each period.

3. Sequentially impute missing data in different wells. The sequence will follow the percentage of missing data across wells. In other words, first, we will impute the well with the smallest fraction of missing data using MWI. Next, we will impute the well with the second smallest fraction of missing data.

4. If certain periods include contiguous synchronous gaps across all the wells, then use SWI for the well with the least missing data. Next, apply step (3).

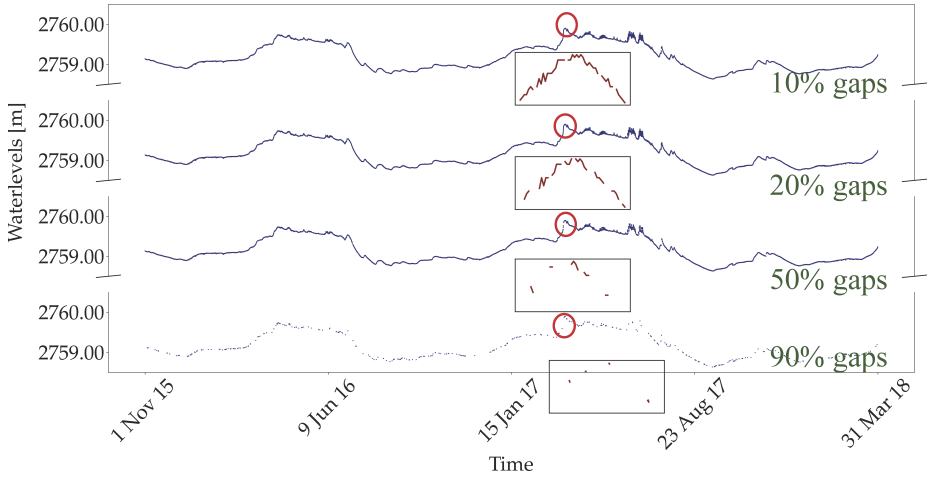

FIG. 5: Missing-at-random synthetic gaps (10%, 20%, 50%, and 90%) are shown for WLE3 (Period 1). Similarly, random gaps for WLE1 were also created. Data are dense and uniformly spaced at 30 min, so insets are shown to visualize gaps for different scenarios.

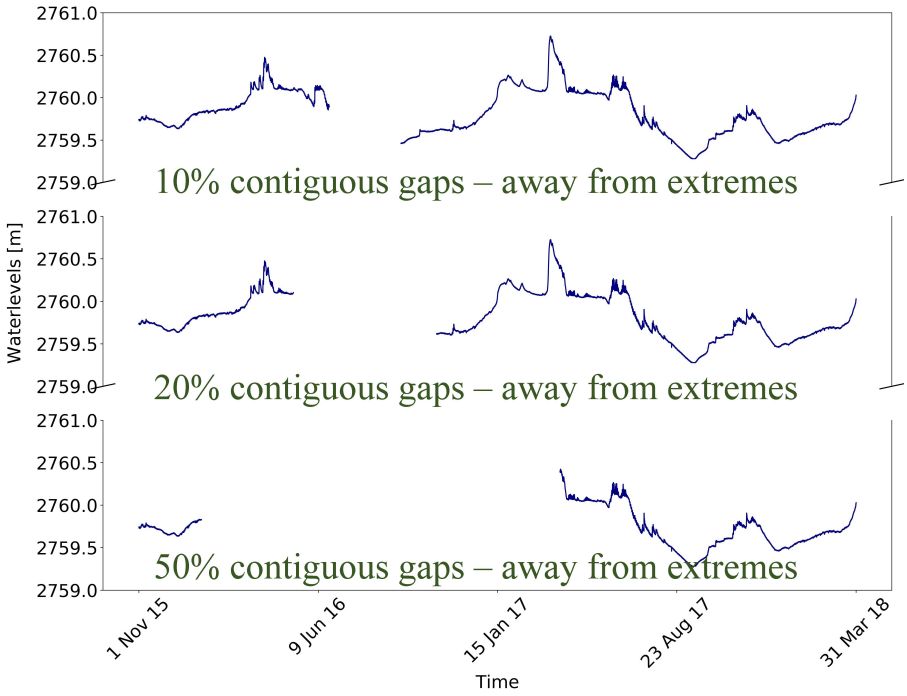

FIG. 6: Contiguous, but away from the hydrograph extremes, synthetic gaps (10%, 20%, and 50%) are shown for WLE3. Similarly, contiguous, but away from the hydrograph extremes, synthetic gaps for WLE1 were also created.

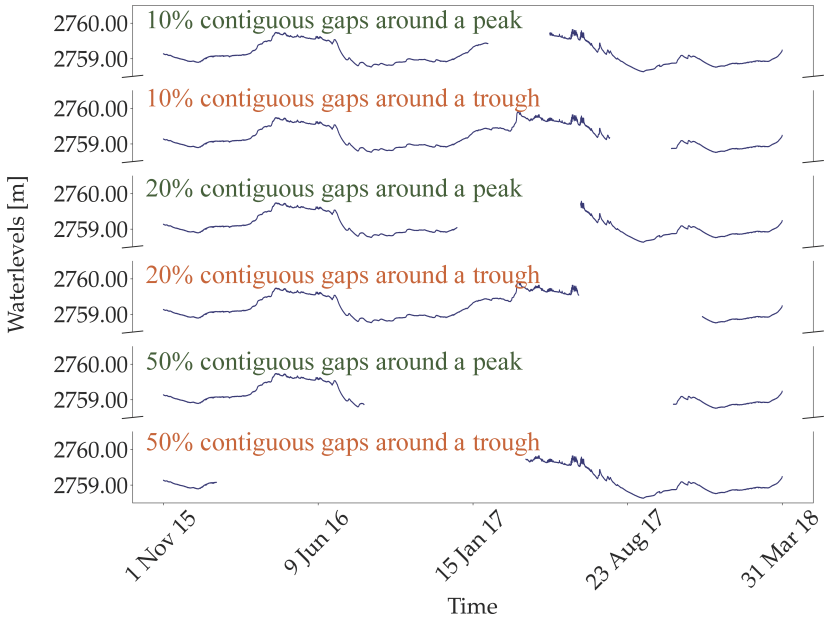

FIG. 7: Contiguous synthetic gaps (10%, 20%, and 50%) missing around peaks and troughs of the hydrograph are shown for WLE3. Similarly, synthetic contiguous gaps missing around peaks and troughs of the hydrograph were created for WLE1.

TABLE 1: Numerical experiments for GW imputation

Numerical experiments	Gaps (% missing data)	Missing data patterns
E1	10%, 20%, 50%, 90%	Random gaps
E2	10%, 20%, 50%	Contiguous gaps away from extremes
E3	10%, 20%, 50%	Contiguous gaps around extremes

TABLE 2: Sequentially imputing GW wells for each period

Period	Well sequence	Neighboring wells used as predictors	Fraction of missing data
Period 1	WLE4*		0.18
	WLE1	WLE4	0.58
	WLE6	WLE4, WLE1	9.70
	WLE5	WLE4, WLE1, WLE6	33.91
	WLE7	WLE4, WLE1, WLE6, WLE5	36.82
	WLE2	WLE4, WLE1, WLE6, WLE5, WLE7	57.38
Period 3	WLE7*		0.02
	WLE4	WLE7	0.02
	WLE6	WLE7, WLE4	3.70
	WLE1	WLE7, WLE4, WLE6	16.16
	WLE5	WLE7, WLE4, WLE6, WLE1	18.36
	WLE2	WLE4, WLE1, WLE6, WLE5, WLE7	57.38
Period 2†	WLE6*		4.15
	WLE4	WLE6	4.15
	WLE7	WLE6, WLE4	5.61
	WLE2	WLE6, WLE4, WLE7	5.61
	WLE1	WLE6, WLE4, WLE7, WLE2	9.27
	WLE5	WLE6, WLE4, WLE7, WLE2, WLE1	9.27

*SWI (LI and RF, both applicable);
†Period 1 and Period 3 were combined before imputing Period 2 to leverage maximum possible information.

To demonstrate the efficacy of the sequential imputation, we choose E3, because this numerical experiment represents the most complex scenario for imputation. We impute well WLE3 with contiguous synthetic gaps missing around the hydrograph extremes (i.e., experiment E3). Following the sequential imputation steps, we divide the entire GW monitoring duration for the seven wells into three periods, as shown in Fig. 4(b). We then sequentially impute the missing data in different wells for Period 1 and Period 3 (Table 2). To obtain better imputation quality, we sequentially impute Period 3 before Period 2 because the gaps are small ($< 10\%$) in Period 3. For Period 2, first we combine Period 1 and Period 3 because that will allow more data for training RF, then we sequentially impute different wells, starting with well WLE6, and follow the sequence summarized in Table 2. It is worth mentioning that a different set of periods (and sequence) can also be chosen to achieve high accuracy. However, periods and well sequence are chosen so the MWI can be used with neighboring wells with gaps as small as possible.

3. RESULTS

3.1 E1: GW Imputation with Random Gaps

We used the SWI method, using LI and RF approaches to impute missing GW time-series data for E1 with 10%, 20%, 50%, and 90% random gaps in WLE1 and WLE3 for Period 1. The SWI-LI assumed the values as equally spaced and estimated missing data points by fitting a linear function within the range of a discrete set of nearby available data points. In the SWI-RF, the missing data were modeled using a set of predictor variables. The predictors constituted the feature space, which in this case included Unix timestamps, days, months, and years of the same well (Fig. 3). The results showed that both SWI-LI and SWI-RF techniques could impute missing values for random gaps ranging from 10% to 90% with high accuracy (*NSE* $\sim$ 0.99). WLE1 and WLE3 showed similar performances for both methods and gaps ranging from 10% to 90%. Hence, for brevity, we show a 90% imputed missing GW time series for WLE3 and compare imputed data with observations using SWI-RF [Figs. 8(a) and 8(b)]. As is evident from Figs. 8(a) and 8(b), the water level in WLE3 started rising in March and peaked around mid-April to June; the water level subsequently decreased until October. The results indicate that both techniques could capture temporal variability appropriately when gaps were random, even though they were as high as 90%. The RF Regressor also revealed feature importance while imputing the GW time series with random gaps. Months (60%) were the most important predictor, followed by timestamps (25%), years (10%), and days (5%).

3.2 E2: GW Imputation with Contiguous Gaps Away from Extremes of the Hydrograph

We used SWI-LI and SWI-RF methods to impute missing GW time-series data with 10%, 20%, and 50% contiguous gaps away from extremes in WLE1 and WLE3 for Period 1. SWI-LI failed to impute missing values for both wells (WLE1 and WLE3) effectively, even with 10% contiguous gaps; *NSE* values were negative. SWI-RF also failed to impute missing values with the desired accuracy and had *NSE* values less than zero for all three missing data scenarios. However, SWI-RF slightly outperformed the linear interpolation approach.

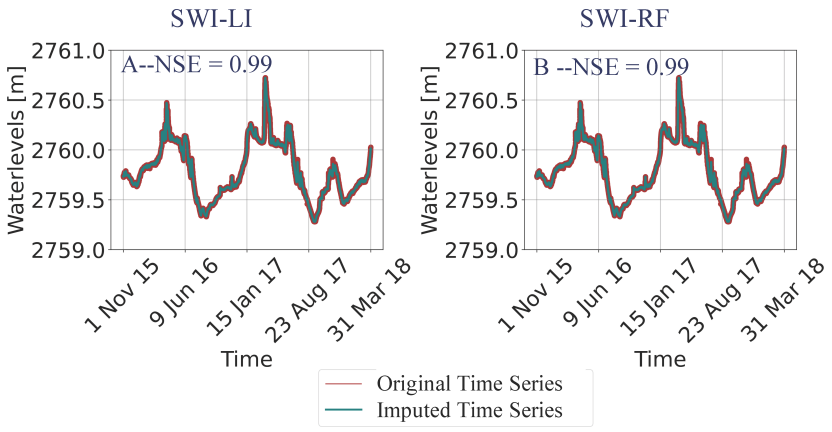

FIG. 8: SWI-LI and SWI-RF methods imputed randomly missing GW data with high accuracy (*NSE* 0.99), even when gaps were as high as 90%

Subsequently, we imputed 10%, 20%, and 50% contiguous gaps in WLE3 using MWI-RF, in which the feature space included timestamps, days, months, years, and water levels in WLE1. Figure 9 shows observed and imputed time series of WLE3 with 10%, 20%, and 50% contiguous gaps away from extremes. Figure 9 demonstrates that MWI-RF was able to capture temporal variability adequately for all three scenarios (10%, 20%, and 50% gaps). MWI-RF achieved *NSE* values 0.94, 0.93, and 0.70 for 10%, 20%, and 50% contiguous gaps, respectively. MWI-RF performance slightly improved when we removed data around peaks with 50% gaps while computing *NSE* values because E2 focuses on gaps away from extremes. We observe here that MWI-RF outperforms SWI-RF, corroborating the first part of the overarching hypothesis. It was also noted that neighboring wells were the most important feature, accounting for 57%, 57%, and 44% of variability for 10%, 20%, and 50% contiguous gaps, respectively (Fig. 10). For 10% and 20% contiguous gaps, timestamps were the second-most important feature. However, months were the second-most important feature for 50% contiguous gaps. Years were the third-most important feature, accounting for 6% to 10% variability for 10% to 50% contiguous gaps (Fig. 9). Note that imputed data in WLE1 showed similar results as WLE3 and for brevity are not shown. Incorporating days and hours in the feature space slightly improved *NSE* values.

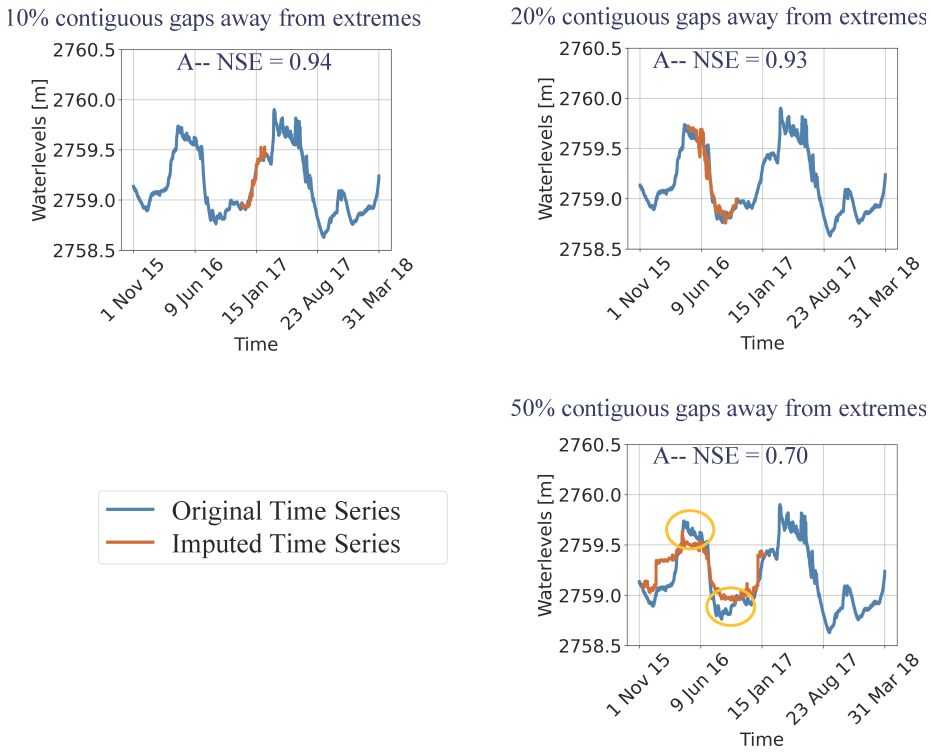

FIG. 9: MWI-RF was able to impute 10%, 20%, and 50% contiguous but missing-at-random GW data in WLE3 with high accuracy, with *NSE* values ranging from 0.70 to 0.94. Because this scenario focuses on contiguous gaps away from extremes, we computed *NSE* values after removing the encircled portion as shown.

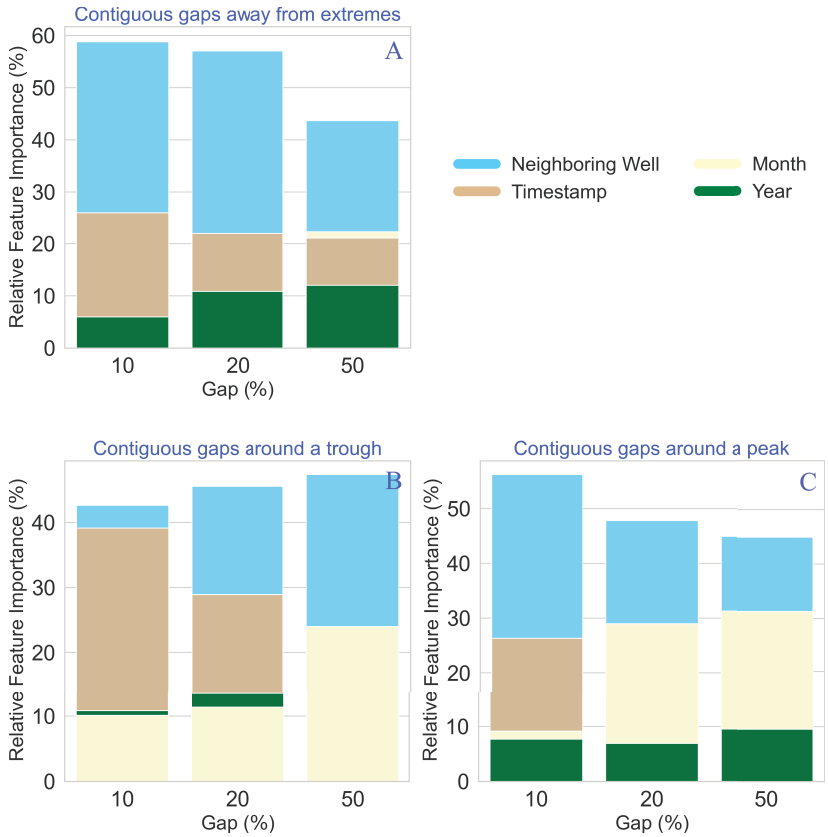

FIG. 10: RF relative feature importance for (a) contiguous gaps away from extremes, (b) around a trough, and (c) around a peak suggests that neighboring wells are the most important feature, followed by months and timestamps together and years for imputing missing GW data. The relative importance of neighboring wells grows as gaps in a time series increase.

3.3 E3: GW Imputation with Contiguous Gaps around Extremes of the Hydrograph

The numerical experiment E3 reflected the most challenging scenario for GW imputation. Here, contiguous gaps were located around the extremes (i.e., peaks and troughs) of the hydrograph. SWI-LI and SWI-RF approaches were not able to impute missing GW data satisfactorily.

MWI-RF was used to impute 10%, 20%, and 50% contiguous gaps missing around peaks and troughs of the hydrograph in WLE3, in which RF feature space included timestamps, days, months, years, and water levels in WLE3. Figure 11 shows observed and imputed data for 10%, 20%, and 50% contiguous gaps around the extremes for well WLE3. As Fig. 11 indicates, MWI-RF was able to capture hydrologic variability reasonably well except around the hydrograph extremes. *NSE* values were positive and ranged from 0.32 to 0.88 for 10% and 50% contiguous gaps.

As shown in Fig. 10, MWI-RF feature importance suggested that generally neighboring wells were the most important feature, followed by months, timestamps, and years. The relative feature

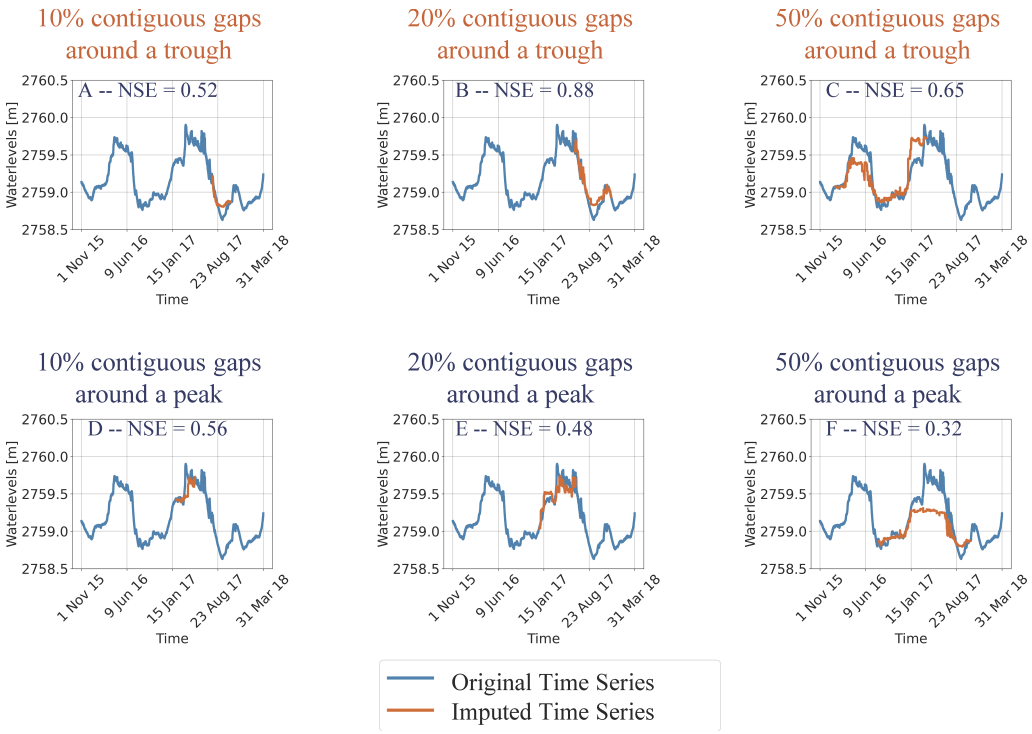

FIG. 11: MWI-RF was able to impute 10%, 20%, and 50% contiguous GW data around troughs (top panel) and peaks (bottom panel) of the hydrograph with reasonable accuracy, with *NSE* values ranging from negative to 0.87. However, MWI-RF failed to capture peaks and troughs.

importance of neighboring wells typically increased when gaps were more prominent, accounting for 40% to 60% variability for 10% to 50% missing data.

3.4 Sequential Imputation of Missing GW Data

As can be seen from the preceding sections, MWI outperforms SWI, which corroborates the first part of our overarching hypothesis. Although MWI-RF could not capture the hydrograph extremes, it captured the hydrologic variability reasonably well for other parts of the hydrograph. To capture these extremes and test the second part of our overarching hypothesis, we sequentially imputed WLE3 for 10%, 20%, and 50% of contiguous synthetic gaps (i.e., E3).

We used MWI-RF to impute these contiguous gaps missing around the hydrograph extremes in well WLE3, where feature space included timestamps, days, months, years, and water levels in the other six neighboring wells. To leverage data from neighboring wells, we first sequentially imputed the missing data for these six wells following the imputation strategy laid out in Table 2. Figure 12 shows the observed and imputed 10%, 20%, and 50% of contiguous, but missing around extremes, gaps at well WLE3 for Period 1. We observe that sequential imputation was able to adequately capture these contiguous gaps around peaks and troughs of WLE3. *NSE* values improved significantly as compared to the nonsequential case and ranged from 0.66 to 0.94.

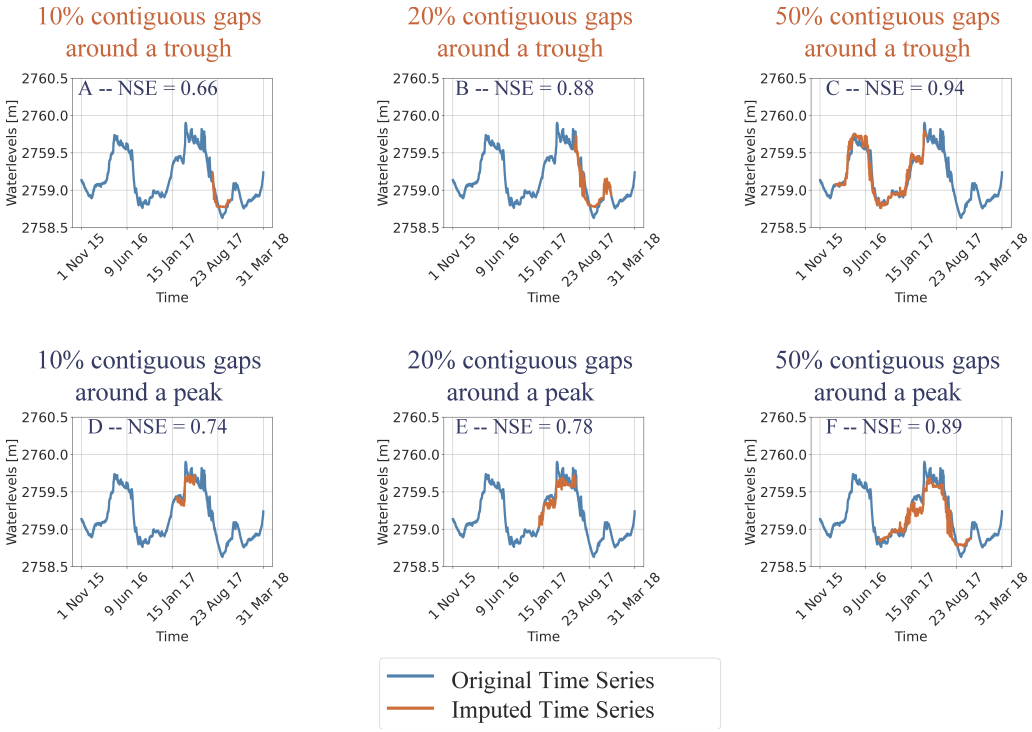

FIG. 12: The sequential imputation strategy demonstrates that MWI-RF was able to impute 10%, 20%, and 50% contiguous GW data around troughs (top panel) and peaks (bottom panel) of the hydrograph with high accuracy, with *NSE* values ranging from 0.66 to 0.94

Figure 13 shows a comparison of *NSE* values for 10%, 20% and 50% contiguous gaps around extremes. *NSE* values are higher for each case, except for 20% gaps around a trough. However, the *NSE* value was already high for this case, so it did not matter if the sequential imputation could not improve this performance.

4. DISCUSSION

4.1 GW Imputation with Random Gaps

SWI could impute missing data only for numerical experiment E1, which, with random gaps ranging from 10% to 90%, can be considered as the simplest missing data scenario. SWI-LI and SWI-RF methods both resulted in almost a perfect match, even with 90% synthetic gaps for the time-series data of WLE1 and WLE3. To understand the mechanism behind such an effective imputation of randomly missing data, we computed the information entropy of WLE1 and WLE3. Figure 14(a) shows how entropy changed with the number of observations of WLE3 time-series data. In particular, we computed the mean and standard deviation of entropy by randomly selecting the number of observations (e.g., 5%, 10%).

As described in Section 2.2.3, entropy is a measure of variability, so the entropy value will increase when the number of observations increases (Singh, 2010a,b). However, the entropy will

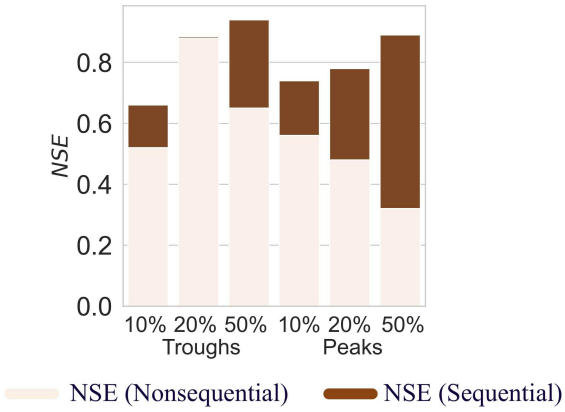

FIG. 13: *NSE* values are consistently higher for sequential imputation than nonsequential imputation for contiguous gaps around extremes

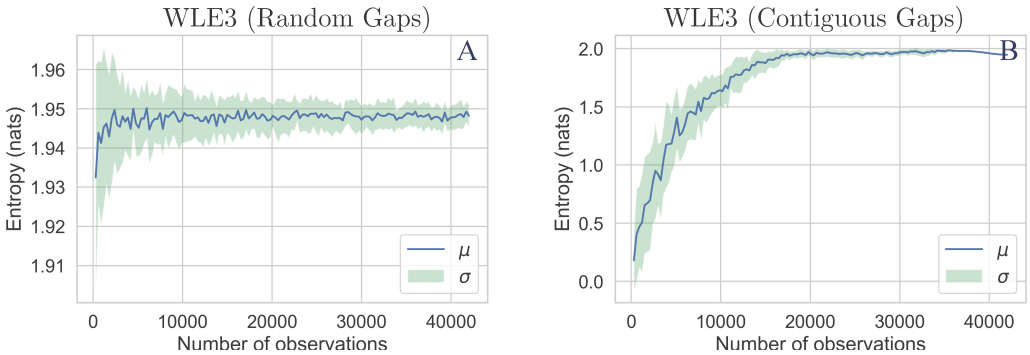

FIG. 14: Information entropy as a function of the number of observations can reveal how many data points are needed to capture complete temporal variability of a GW time series

reach a maximum value when the number of observations has already captured the data's inherent temporal variability (Arora et al., 2019b; Cui and Singh, 2015). Accordingly, it is clear from Fig. 14(a) that the entropy increases initially, with a wider spread (i.e., band) around it (i.e., higher standard deviation). The entropy reaches a maximum value when the number of observations is approximately 10%. With a higher number of observations, the spread in entropy value decreases. Essentially, this means that 10% of randomly selected GW data inherits the entire temporal variability contained in the complete GW time-series data of WLE1 and WLE3 for Period 1. Therefore, RF could learn the variability of GW data from a mere fraction of 10% and impute 90% missing data effectively. In addition, RF revealed that months were the most important predictors for random gaps (E1). Although days and hours slightly improved *NSE* values, daily and subhourly fluctuations were relatively minor. Therefore, a simple interpolation technique could also impute 90% of randomly missing GW data effectively.

4.2 GW Imputation with Contiguous Gaps

We imputed missing GW data for numerical experiments E2 and E3 that represented contiguous missing data patterns (10%, 20%, and 50%), both located away from hydrograph extremes and centered around extremes. The MWI-RF method was able to impute missing GW data reasonably well for both E2 and E3. However, SWI-LI and SWI-RF could not impute missing data for E2 and E3. As previously noted in Section 3.1, months were important predictors for imputing random gaps (i.e., in E1). Therefore, SWI-LI and SWI-RF would not be effective if a GW time series included missing data spanning multiple days and months. More information would be needed to impute missing data in contiguous gap scenarios. Accordingly, the MWI-RF derived additional information from neighboring well(s) and achieved high accuracy in imputing missing data for both wells WLE1 and WLE3 for 10%, 20%, and 50% gaps. Furthermore, neighboring wells provided critical information for imputation as their relative feature importance was relatively high for all scenarios (10%, 20%, and 50% gaps) in both E2 and E3 (Fig. 10).

To further explore MWI-RF performance, we computed the information entropy of WLE1 and WLE3 with contiguous gaps of variable window sizes. Figure 14(b) shows how entropy changes with the number of observations of WLE3 time-series data. As Fig. 14(b) indicates, entropy values change if the number of observations are less than 50%. To capture the temporal variability of the GW time series in WLE1 and WLE3, we would need at least 50% GW time-series data for WLE1 and WLE3 for effectively imputing missing entries in Period 1.

Although the MWI-RF was able to impute contiguous gaps with high accuracy for both E2 (contiguous gaps away from extremes) and E3 (contiguous gaps around extremes), it was not able to capture extremes of the hydrograph for E3. To understand the poor performance of the MWI-RF for E3, we need to recognize the underlying philosophy of ML approaches. Two aspects are important for imputing any data using ML approaches: (a) temporal variability in time-series data, and (b) availability of training data for an ML model to capture data's critical characteristics (Alpaydin, 2014; Dwivedi et al., 2013; Haykin, 2010; Shen, 2018). Although information entropy suggested that contiguous data with up to 50% gaps contained the entirety of temporal variability for Period 1, it does not distinguish between E2 and E3. E3 has another level of complexity above E2. Specifically, E3 does not provide sufficient training data for the MWI-RF to learn peaks and troughs of the hydrograph. If we had a more extended GW time series, unlike our numerical experiments, then MWI-RF could use sufficient training data to "learn" critical characteristics.

4.3 Sequential Imputation of GW with Contiguous Gaps around Extremes

The sequential imputation strategy leverages information from neighboring wells even if they have missing data. We have seen in preceding sections that the SWI method can impute up to 90% missing data—if gaps are random—with high accuracy, and MWI-RF can effectively impute up to 50% of missing data for approximately a two-year period. Table 2 provides the missing data fractions of different GW wells for three distinct periods. For Period 1, we note that all the wells except WEL2 had $< 37\%$ either random gaps or contiguous gaps missing away from extremes. The inferences drawn from information entropy (Sections 4.1 and 4.2) indicate that imputation quality would be high. Although WLE2 had gaps slightly more than 50%, *NSE* values were significantly higher in the sequential versus the nonsequential case, suggesting that GW data from five other wells compensated for relatively larger gaps in WLE2 using MWI-RF. However, the sequential imputation would perform better if these gaps were less than 50%.

We make similar observations for Period 2 and Period 3 (i.e., gaps $<$ 20%). Following this, we see that the MWI-RF captured peaks and troughs reasonably well when we sequentially imputed WLE3. *NSE* values were significantly higher for the sequential imputation than for the nonsequential case, when we imputed WLE3 for period 1, using MWI-RF with only well WLE3 (Fig. 13). Part of the reason for low *NSE* values was that nonsequential imputation could not inform peaks and troughs to adequately train MWI-RF. Overall, the high imputation quality in sequential imputation suggests that an extended GW time series that allows MWI-RF to "learn" critical characteristics is likely to perform effectively.

5. SUMMARY AND CONCLUSIONS

Monitoring and predictions of GW levels are critical for managing water resources and constraining watershed numerical models. However, the GW time-series data usually have missing records, owing to various unavoidable reasons such as malfunctioning monitoring transducers and physical disturbances. Several techniques in the literature impute missing GW data, ranging from simple interpolation to statistical modeling to advanced ML methods. Although these techniques can impute GW reasonably well, their applications can differ and be contingent upon "missing data" patterns. In this study, we explored techniques to gap-fill subhourly GW data, including missing extremes, based on the following missing data patterns: (a) random gaps, (b) contiguous gaps away from extremes, and (c) contiguous gaps around extremes. We used the SWI-LI, SWI-RF, and MWI-RF to impute missing GW data. We also used an information entropy framework to understand imputation performance based on missing data patterns and percent-missing data. The results showed that SWI-LI and SWI-RF methods imputed up to 90% missing data—if gaps were random—with high accuracy (*NSE* $\sim$ 0.99). Information entropy revealed that efficient imputation resulted because as small as 10% of data could also capture the complete temporal variability of a GW time series dataset. Moreover, months were found to be the most important features as, indicated by the SWI-RF, followed by timestamps.

Contiguous gaps required MWI-RF for effective imputation, while SWI-LI and SWI-RF performed poorly even when gaps were as small as 10%. The MWI-RF revealed that neighboring wells, followed by months, timestamps, and years, were important features for imputing GW time-series data with contiguous gaps. The MWI-RF and information entropy indicated that up to 50% of missing data could be estimated reasonably well over an approximately two-year period with high accuracy. Although the MWI-RF imputed up to 50% contiguous gaps with high accuracy using a single neighboring well, it failed to capture gaps located around extremes. To capture these extremes, we developed a sequential imputation strategy to estimate the missing data in GW time series across multiple wells. We successfully demonstrated the strategy using seven wells in the East River floodplain.

We note that our approach is highly transferable across sites and may apply to a suite of other environmental variables such as temperature, soil moisture, and precipitation. Further, whereas we used RF to demonstrate our GW imputation strategy, other ML-based methods (such as support vector machines and artificial neural networks) should also work. An information entropy approach to identifying how much missing percentage can be imputed effectively is variable-agnostic and can be applied to any other environmental variables. Overall, we showed that machine-learning methods could provide sustainable solutions to gap-fill environmental datasets because our approach is transferable and applicable to a suite of other environmental variables.

ACKNOWLEDGMENT

This work was funded by the Watershed Function Scientific Focus Area and ExaSheds projects, supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Earth and Environmental Systems Sciences Division, under Award No. DE-AC02-05CH11231.

REFERENCES

Aggarwal, C.C. and Zhai, C., *Mining Text Data*, Springer Science & Business Media, 2012.

Aguilera, H., Guardiola-Albert, C., and Serrano-Hidalgo, C., Estimating Extremely Large Amounts of Missing Precipitation Data, *J. Hydroinf.*, vol. **22**, no. 3, pp. 578–592, 2020.

Aissia, M.A.B., Chebana, F., and Ouarda, T.B., Multivariate Missing Data in Hydrology – Review and Applications, *Adv. Water Resour.*, vol. **110**, pp. 299–309, 2017.

Alpaydin, E., *Introduction to Machine Learning Ethem Alpaydin*, Cambridge, MA: MIT Press, 2014.

Arora, B., Dwivedi, D., Hubbard, S.S., Steefel, C.I., and Williams, K.H., Identifying Geochemical Hot Moments and Their Controls on a Contaminated River Floodplain System Using Wavelet and Entropy Approaches, *Environ. Model. Software*, vol. **85**, pp. 27–41, 2016a.

Arora, B., Spycher, N.F., Steefel, C.I., Molins, S., Bill, M., Conrad, M.E., Dong, W., Faybishenko, B., Tokunaga, T.K., Wan, J., et al., Influence of Hydrological, Biogeochemical and Temperature Transients on Subsurface Carbon Fluxes in a Flood Plain Environment, *Biogeochemistry*, vol. **127**, nos. 2-3, pp. 367–396, 2016b.

Arora, B., Dwivedi, D., Faybishenko, B., Jana, R.B., and Wainwright, H.M., Understanding and Predicting Vadose Zone Processes, *Rev. Mineral. Geochem.*, vol. **85**, no. 1, pp. 303–328, 2019a.

Arora, B., Wainwright, H.M., Dwivedi, D., Vaughn, L.J., Curtis, J.B., Torn, M.S., Dafflon, B., and Hubbard, S.S., Evaluating Temporal Controls on Greenhouse Gas (GHG) Fluxes in an Arctic Tundra Environment: An Entropy-Based Approach, *Sci. Total Environ.*, vol. **649**, pp. 284–299, 2019b.

Arora, B., Burrus, M., Newcomer, M., Steefel, C.I., Carroll, R.W., Dwivedi, D., Dong, W., Williams, K.H., and Hubbard, S.S., Differential CQ Analysis: A New Approach to Inferring Lateral Transport and Hydrologic Transients within Multiple Reaches of a Mountainous Headwater Catchment, *Front. Water*, 2020. DOI: 10.3389/frwa.2020.00024

Breiman, L., Bagging Predictors, *Mach. Learn.*, vol. **24**, no. 2, pp. 123–140, 1996.

Cui, H. and Singh, V.P., Configurational Entropy Theory for Streamflow Forecasting, *J. Hydrol.*, vol. **521**, pp. 1–17, 2015.

Dafflon, B. and Dwivedi, D., Groundwater Level Elevation and Temperature at the Lower Montane in the East River Watershed, Colorado, Watershed Function SFA, 2020. DOI: 10.15485/1647040

Dax, A. and Zilberbrand, M., Imputing Missing Groundwater Observations, *Hydrol. Res.*, vol. **49**, no. 3, pp. 831–845, 2018.

Du, P., Bai, X., Tan, K., Xue, Z., Samat, A., Xia, J., Li, E., Su, H., and Liu, W., Advances of Four Machine Learning Methods for Spatial Data Handling: A Review, *J. Geovisualiz. Spatial Anal.*, vol. **4**, pp. 1–25, 2020.

Dwivedi, D. and Mohanty, B.P., Hot Spots and Persistence of Nitrate in Aquifers across Scales, *Entropy*, vol. **18**, no. 1, p. 25, 2016.

Dwivedi, D., Mohanty, B.P., and Lesikar, B.J., Estimating *Escherichia coli* Loads in Streams Based on Various Physical, Chemical, and Biological Factors, *Water Resour. Res.*, vol. **49**, no. 5, pp. 2896–2906, 2013.

Dwivedi, D., Steefel, I.C., Arora, B., and Bisht, G., Impact of Intra-Meander Hyporheic Flow on Nitrogen Cycling, *Proc. Earth Planet Sci.*, vol. **17**, pp. 404–407, 2017.

Dwivedi, D., Arora, B., Steefel, C.I., Dafflon, B., and Versteeg, R., Hot Spots and Hot Moments of Nitrogen in a Riparian Corridor, *Water Resour. Res.*, vol. **54**, no. 1, pp. 205–222, 2018a.

Dwivedi, D., Steefel, C.I., Arora, B., Newcomer, M., Moulton, J.D., Dafflon, B., Faybishenko, B., Fox, P., Nico, P., Spycher, N., et al., Geochemical Exports to River from the Intrameander Hyporheic Zone under Transient Hydrologic Conditions: East River Mountainous Watershed, Colorado, *Water Resour. Res.*, vol. **54**, no. 10, pp. 8456–8477, 2018b.

Giannakou, A., Dwivedi, D., and Peisert, S., A Machine Learning Approach for Packet Loss Prediction in Science Flows, *Future Gener. Comput. Syst.*, vol. **102**, pp. 190–197, 2020.

Haykin, S., *Neural Networks and Learning Machines*, London: Pearson Education India, 2010.

Hockett, C.F. Review of the Mathematical Theory of Communication by Claude L. Shannon and Warren Weaver, *Language*, vol. **29**, no. 1, pp. 69–93, 1953.

Hubbard, S.S., Williams, K.H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., Brodie, E., Carroll, R., Dafflon, B., Dwivedi, D., et al., The East River, Colorado, Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multiscale Hydrological–Biogeochemical Dynamics, *Vadose Zone J.*, vol. **17**, no. 1, pp. 1–25, 2018.

Hubbard, S.S., Varadharajan, C., Wu, Y., Wainwright, H., and Dwivedi, D., Emerging Technologies and Radical Collaboration to Advance Predictive Understanding of Watershed Hydrobiogeochemistry, *Hydrol. Processes*, vol. **34**, no. 15, pp. 3175–3182, 2020.

Khedri, A., Kalantari, N., and Vadiati, M., Comparison Study of Artificial Intelligence Method for Short Term Groundwater Level Prediction in the Northeast Gachsaran Unconfined Aquifer, *Water Supply*, vol. **20**, no. 3, pp. 909–921, 2020.

Kondrashov, D., Feliks, Y., and Ghil, M., Oscillatory Modes of Extended Nile River Records (AD 622–1922), *Geophys. Res. Lett.*, vol. **32**, no. 10, 2005.

Li, L., Maher, K., Navarre-Sitchler, A., Druhan, J., Meile, C., Lawrence, C., Moore, J., Perdrial, J., Sullivan, P., Thompson, A., et al., Expanding the Role of Reactive Transport Models in Critical Zone Processes, *Earth Sci. Rev.*, vol. **165**, pp. 280–301, 2017.

Likens, G.E., Biogeochemistry, the Watershed Approach: Some Uses and Limitations, *Marine Freshwater Res.*, vol. **52**, no. 1, pp. 5–12, 2001.

Mays, D.C., Faybishenko, B.A., and Finsterle, S., Information Entropy to Measure Temporal and Spatial Complexity of Unsaturated Flow in Heterogeneous Media, *Water Resour. Res.*, vol. **38**, no. 12, pp. 49-1–49-11, 2002.

Mital, U., Dwivedi, D., Brown, J.B., Faybishenko, B., Painter, S.L., and Steefel, C.I., Sequential Imputation of Missing Spatio-Temporal Precipitation Data Using Random Forests, *Front. Water*, 2020. DOI: 10.3389/frwa.2020.00020

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L., Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, *Transact. ASABE*, vol. **50**, no. 3, pp. 885–900, 2007.

Moritz, S. and Bartz-Beielstein, T., Imputets: Time Series Missing Value Imputation in R, *R J.*, vol. **9**, no. 1, p. 207, 2017.

Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., and Agarwal, D., Surrogate Optimization of Deep Neural Networks for Groundwater Predictions, *J. Global Opt.*, pp. 1–29, 2019.

Oppel, H. and Schumann, A.H., Machine Learning Based Identification of Dominant Controls on Runoff Dynamics, *Hydrol. Processes*, vol. **34**, no. 11, pp. 2450–2465, 2020.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., Scikit-Learn: Machine Learning in Python, *J. Mach. Learn. Res.*, vol. **12**, pp. 2825–2830, 2011a.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al., Deep Learning and Process Understanding for Data-Driven Earth System Science, *Nature*, vol. **566**, no. 7743, pp. 195–204, 2019.

Shannon, C., A Mathematical Theory of Communication, *Bell Syst. Technol. J.*, vol. **27**, no. 3, pp. 379–423, 1948.

Shen, C., A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, *Water Resour. Res.*, vol. **54**, no. 11, pp. 8558–8593, 2018.

Singh, V.P., Entropy Theory for Movement of Moisture in Soils, *Water Resour. Res.*, vol. **46**, no. 3, 2010a.

Singh, V.P., Entropy Theory for Derivation of Infiltration Equations, *Water Resour. Res.*, vol. **46**, no. 3, 2010b.

Singh, V., The Use of Entropy in Hydrology and Water Resources, *Hydrol. Processes*, vol. **11**, no. 6, pp. 587–626, 1997.

Stockman, M., Dwivedi, D., Gentz, R., and Peisert, S., Detecting Control System Misbehavior by Fingerprinting Programmable Logic Controller Functionality, *Int. J. Crit. Infrastruc. Prot.*, vol. **26**, Article 100306, 2019.

Varadharajan, C., Agarwal, D.A., Brown, W., Burrus, M., Carroll, R.W., Christianson, D.S., Dafflon, B., Dwivedi, D., Enquist, B.J., Faybishenko, B., et al., Challenges in Building an End-to-End System for Acquisition, Management, and Integration of Diverse Data from Sensor Networks in Watersheds: Lessons from a Mountainous Community Observatory in East River, Colorado, *IEEE Access*, vol. **7**, pp. 182796–182813, 2019.

Yozgatligil, C., Aslan, S., Iyigun, C., and Batmaz, I., Comparison of Missing Value Imputation Methods in Time Series: The Case of Turkish Meteorological Data, *Theor. Appl. Climatol.*, vol. **112**, no. 1, pp. 143–167, 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 Resour. Res.*, vol. **52**, no. 3, pp. 1568–1590, 2016.

APPENDIX A.

TABLE A1: Description of wells used for imputing missing data

Well names	Well names (in database)	Other well names (in database)	Absolute well elevations (m)	Average GW depths (m)
WLE1	UPW	UP1W	2760.64	0.64
WLE2	UPM	UP2	2760.74	1
WLE3	UPE	UP3E	2760.1	1.5
WLE4	DOW	DO1W	2759.43	1.89
WLE5	DOE	DO2E	2759.23	1.2
WLE6	MBA1	M1B1	2760.18	1
WLE7	MBA2	M1B2	2759.96	1.59