Author/s:
Yousefpour, N;Downie, S;Walker, S;Perkins, N;Dikanski, H

Title:
Machine Learning Solutions for Bridge Scour Forecast Based on Monitoring Data

Date:
2021

Citation:
Yousefpour, N., Downie, S., Walker, S., Perkins, N. & Dikanski, H. (2021). Machine Learning Solutions for Bridge Scour Forecast Based on Monitoring Data. Transportation Research Record: Journal of the Transportation Research Board, 2675 (10), pp.036119812110126-036119812110126. https://doi.org/10.1177/03611981211012693.

Machine Learning Solutions for Bridge Scour Forecast Based on Monitoring Data

Negin Yousefpour, PhD, PE
Doreen Thomas Fellow, The University of Melbourne
Senior Engineer, Arup
Melbourne, Australia
Email: negin.yousefpour@unimelb.edu.au

Steve Downie, PhD
Associate
Arup
London, UK
Email: steven.downie@arup.com

Steve Walker
Associate Principle
Arup
London, UK
Email: steve.walker@arup.com

Nathan Perkins
Analyst
Arup
Melbourne, Australia
Email: nathan.perkins@arup.com

Hristo Dikanski, PhD
Engineer
Arup
London, UK
Email: hristo.dikanski@arup.com

Revised Version Word Count: 7,500 words + 7 tables

Submitted [08/01/2019]
Revised [10/15/2020]
Revised [01/20/2021]

ABSTRACT

Bridge scour has been a challenge throughout the US and other countries. Despite the scale of the issue, there is still a substantial lack of robust methods for scour prediction to support reliable, risk-based management and decision-making. Throughout the past decade, the use of real-time scour monitoring systems has gained an increasing interest by DOTs (department of transportation) across the US. This paper introduces three distinct methodologies for scour prediction using advanced Artificial Intelligence (AI)/Machine Learning (ML) technics based on real-time scour monitoring data. Scour monitoring data included the riverbed and river stage elevation time-series at bridge piers gathered from various sources. Deep learning algorithms showed promising in predicting bed elevation and water level variations as early as a week in advance. Ensemble neural networks proved successful in predicting the maximum upcoming scour depth, using the observed sensor data at the onset of a scour episode, and based on bridge pier, flow, and riverbed characteristics. In addition, two of the common scour empirical models were calibrated based on the observed sensor data using the Bayesian inference method, showing significant improvement in prediction accuracy. Overall, this paper introduces a novel approach for scour risk management of bridges by integrating emerging AI/ML algorithms with real-time monitoring systems for early scour forecast.

Keywords: Bridge Scour, Artificial Intelligence, Scour Prediction, Scour Monitoring, Deep Learning, Neural Networks, Bayesian Inference, LSTM, Early Warning Systems

INTRODUCTION

Scour is the number one cause of bridge failure in the US, with approximately 260,000 scour critical bridges across the country. Scour accounts for a significant number of bridge collapses not only in the US (60%), but also in other countries such as New Zealand (62%) and South Africa (21%) [1]. It has been also a major concern for transportation authorities in the UK and Australia.

Despite various research efforts, especially in the past two decades, there is still a significant need for developing innovative solutions for more reliable scour predictions and scour risk management. Many studies have applied ML methods for predicting the maximum scour depth, such as self-adaptive evolutionary extreme learning machine [2], artificial neural networks (ANNs) [3,4], the adaptive neuro-fuzzy inference system [5,6], genetic programming (GP) [7], support vector machines [8,9], and gene expression programming (GEP) [10]. These studies have aimed at predicting the maximum scour depth by learning from patterns in field data (and occasionally laboratory data) obtained from direct measurements of scour at bridge piers in various years and locations. However, the common drawback is that the models are trained using data from often geographically limited and scattered locations, therefore their reliability cannot be validated for bridges outside the domain of training data. Given the uncertainty of scour as a physical phenomenon involving complex flow-structure-soil interaction and the lack of reliability in existing predictive models, there is an evident need for regular monitoring of scour at critical bridges with higher risk of scour failure.

In the US, Federal Highway Administration (FHWA) requires state DOTs to monitor scour-vulnerable bridges on a regular basis and implement countermeasures to protect them from scour failure. Throughout the past decade, many states have implemented real-time bridge scour monitoring programs. Similarly scour monitoring programs have been devised and implemented by transport authorities in other countries. There are various scour monitoring devices that provide direct measurement of the bed level or depth of scour. Examples of such devices are the followings [11-14]:

- *Single-use devices:* Installed vertically in the riverbed, and when scour depth reaches the installation depth, the device simply floats out to the surface, triggering an electrical switch.
- *Pulse or radar devices:* use a radar signal or an electromagnetic pulse to determine changes in the material properties, as a signal is propagated through a changing physical medium. Thus, a pulse is sent along a buried measuring probe, determining the depth of the interface between water and soil, enabling the estimation of scour depth. Such devices can be used to monitor the variation of scour over time.
- *Buried or driven rod systems:* are based on a gravity-based physical probe that rests on the riverbed and moves downward as scour develops. As the device is based on a gravity sensor, it remains at the lowest depth of scour after each flood, so it is not suitable for monitoring of scour development over time.
- *Fiber-Bragg Grating devices:* piezo-electric devices, measuring strain along an embedded cantilever rod to generate electrical signals, which indicate the progression of scour along the rod.
- *Electrical conductivity devices:* use the difference in electrical conductivity of different materials to determine the depth of the water-sediment interface.
- *Sonar (Soundwave) devices:* are one of the most commonly used scour monitoring sensors and work on the same principle as electromagnetic devices. Waves are reflected through materials of different densities, thus establishing the depth of scour. Sonar sensors operate by releasing a sonic pulse from an emitter, which travels until it is reflected to a receiver due to the change in medium, such as at the sediment-water interface. The time taken for the signal to propagate from the emitter to the receiver can be used to establish the distance to the water-sediment interface. Portable sonar can be a useful bridge inspection tool but does not provide a continuous record of scour development. Fixed sonar has several limitations; the most important one is that sonar measures scour depth within a narrow area around a bridge pier and if the instrument is not mounted above the deepest scour location, it may give a false sense of security.

  Sonar instrumentation is often used in combination with a Water Stage Sensor (Stage), an ultrasonic distance finder, aimed from a fixed reference point to the water surface. This can help establish the water surface elevation.

There are also monitoring methods that identify scour with measurements of structural dynamic properties of the bridge. Limited research has been conducted, trying to determine response of the structure to varying depth of scour [13,15-17]. A disadvantage of this type of sensors is that they do not provide information on the scour depth. They are often used in parallel with depth-measuring instrumentation. Instruments used to measure the structural response to scour include tiltmeters and accelerometers. Tiltmeters measure the relative rotation of a structural element and can be used to detect differential settlement. Accelerometers can help determine changes in the stiffness of the bridge pier-foundation system, caused by scour around the bridge foundation.

This study introduces a novel approach using AI/ML to forecast the trend of scour, based on sonar and stage monitoring data around bridge piers. Three methodologies were developed providing algorithms that can 1) make predictions on variation of bed level (therefore scour depth) on a real-time basis and serve as an early warning system for detecting a major scour event in advance, 2) estimate the maximum upcoming scour depth from the onset of a major scour event, 3) calibrate the scour empirical models based on real-time data and then use them to predict the upcoming scour depth with systematic uncertainty assessment. Integrating the proposed technics with the scour monitoring programs can significantly reduce the scour failure risk, especially for large-scale bridges in flood-prone areas.

DATA

Data Collection and Characteristics

Despite difficulties in finding legitimate scour monitoring data, such as data quality, confidentiality, and access, various reliable databases from different sources such as Alaska DOT, Minnesota DOT, and USGS were gathered and used for training and validation of AI algorithms. Majority of the training database contained sensor readings from Alaska DOT's scour monitoring program managed by USGS. This dataset includes sonar and stage water readings for more than 20 bridges in Alaska from 2003 to 2017. In addition, training data includes sonar and stage readings for two piers of Winona bridge in Minnesota between 2012 to 2017.

The received sensor data contains real time readings of bed elevation (sonar) and water level (stage). The list of bridge monitoring data gathered in this study is provided in **Table 1**. **Figure 1** provides plots of sonar and stage data received for bridge 742 in Alaska. It should be noted that sensor readings were not available between November to April for Alaskan bridges due to frozen waters. **Figure 2** shows the installed sensors for bridge 230 and the corresponding historic bed elevation obtained from survey and sounding across the years. Also, **Figure 3** provides a schematic of the installed sensors.

In addition to sensor data, relevant information on bridge and river characteristics were collected to inform the analyses using several sources, including National Bridge Inventory (NBI), National Bridge Scour Database, and USGS online reports [18,19]. The key data extracted for the selected bridges is summarised in **Table 2.**

Data Quality and Pre-processing

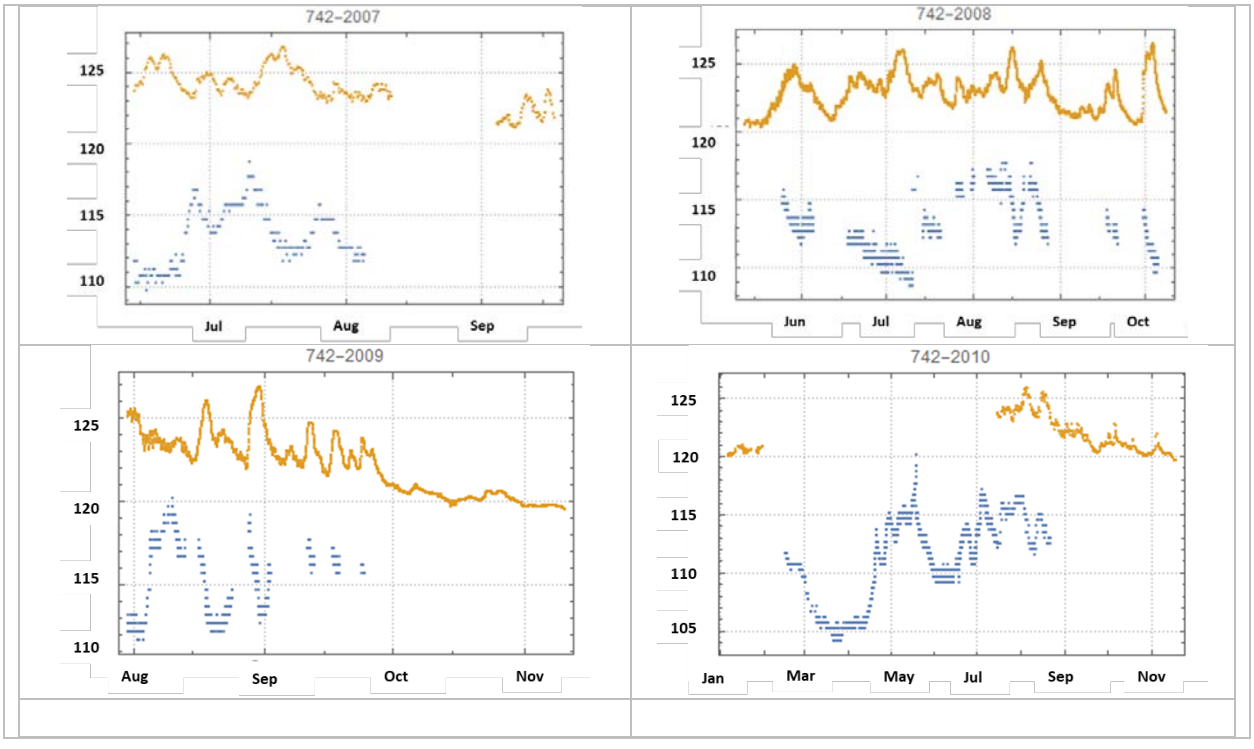
Prior to using the sensor data for model development, extensive pre-processing was performed to generate smooth continuous time series with synchronised steps between sonar and stage readings. Data series were up or down sampled to provide a uniform 1hr time step. In some cases, a portion of data was missing due to issues with the sensors (damages due to debris, etc.) or the readings were noisy and erroneous. Also, sensor reading intervals were not uniform across all the bridges and years of monitoring. **Table 3** provides a summary of data quality issues and pre-processing technics applied to address them, prior to using them for AI/ML. The following main steps were taken for pre-processing of data. The workflow is provided in **Figure 4**.

- Apparent bias shifts in the sensors were corrected based on manual inspection.
- Outliers were removed via median filtering
- Timestamps were made uniform (1hr) throughout the dataset and synchronized between sonar and stage

- Denoising/filtering technics including Moving Average (MA) and Low Pass Filter (LPF) were applied to remove high frequency noises and to capture the actual trend
- Missing data were imputed using polynomial interpolation and Gaussian Process.

TABLE 1 List of Bridges with Stage and Sonar Data

Bridge Number	Bridge Name	Years Included
202	Tanana River at Nenana	2003 – 2017
212	Kashwitna River	2004 – 2017
215	Montana Creek	2004 – 2017
230	Sheridan Glacier No. 3	2005 – 2017
524	Tanana River Big Delta	2007 – 2017
527	Salcha River	2003 – 2017
539	Knik River	2003 – 2017
573	Tazlina River	2011 – 2017
634	Twenty Mile River	2011 – 2017
670	Kasilof River	2003 - 2006
742	Chilkat River	2007 – 2017
983	Red Cloud River	2005 – 2017
999	Glacier Creek	2010 – 2017
1020	Quartz Creek	2015 – 2017
1187	Copper River (Flag Pt E)	2014 – 2017
1243	Nenana River at Windy	2003 – 2017
1383	Lowe River at Keystone	2006 – 2017
5900-19	Mississippi River - Winona Bridge/Pier 19	2012 – 2017
5900-20	Mississippi River - Winona Bridge/Pier 20	2012 - 2017



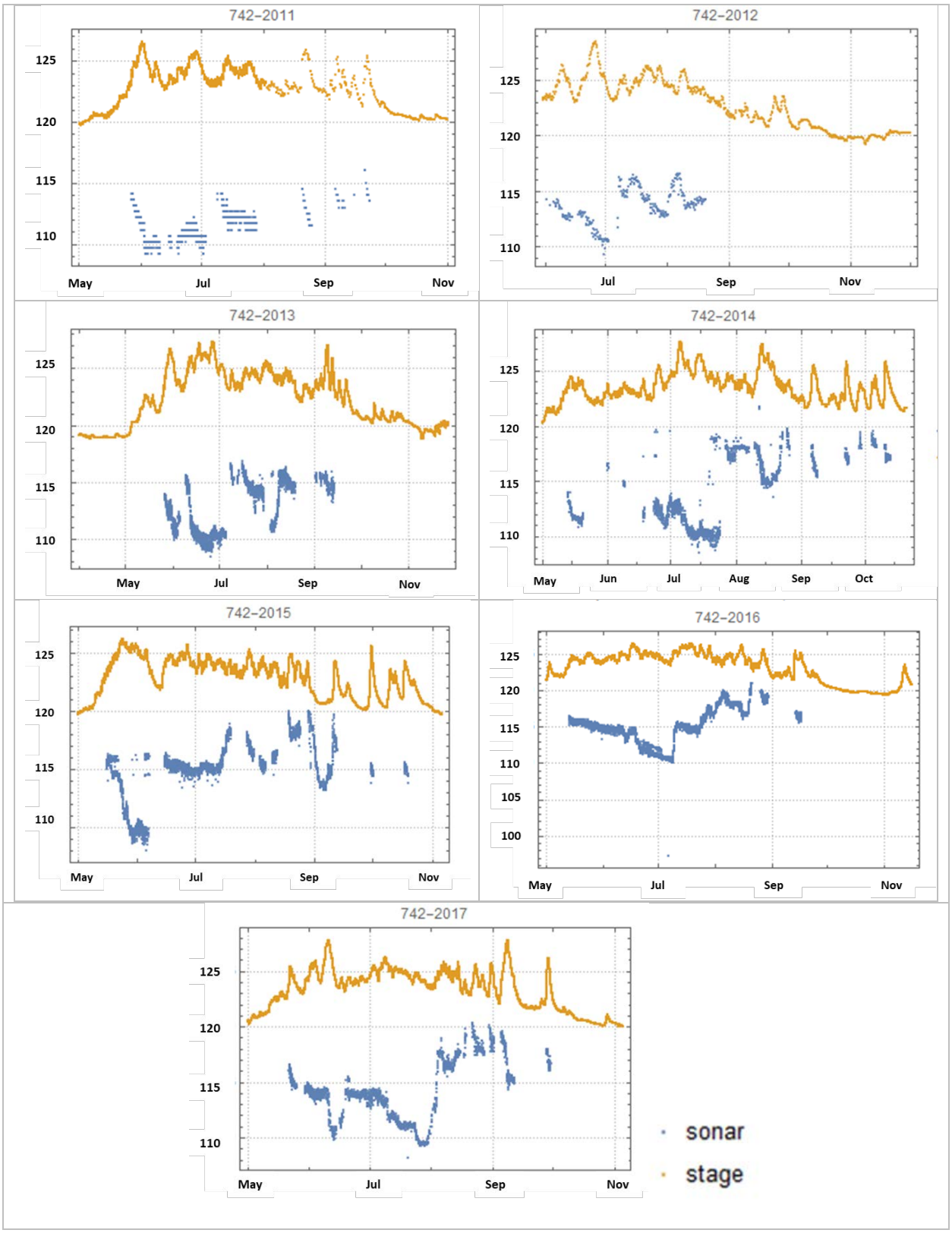

742–2011
742–2012
742–2013
742–2014
742–2015
742–2016
742–2017
125
120
115
110
105
100
May
Jun
Jul
Aug
Sep
Oct
Nov
sonar
stage

Figure 1 Synchronized stage water and sonar data from 2007 to 2017 – Bridge 742 (x axis is time, y axis is elevation in ft)

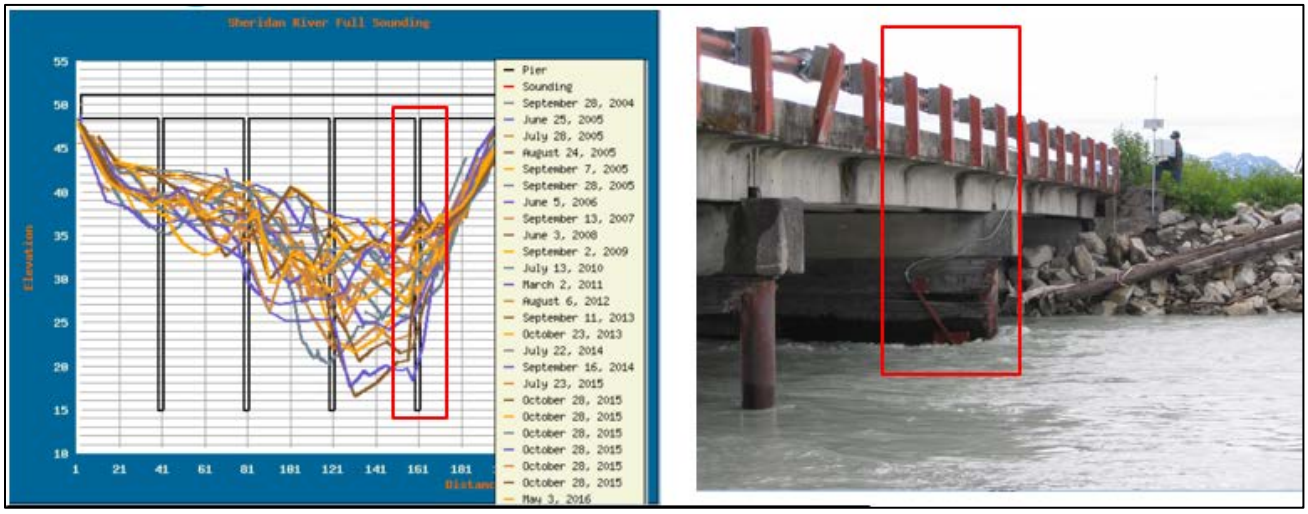


Figure 2 Riverbed profile variations (from sounding) over years (left) and installed scour sensors (right) – bridge 230 [20]

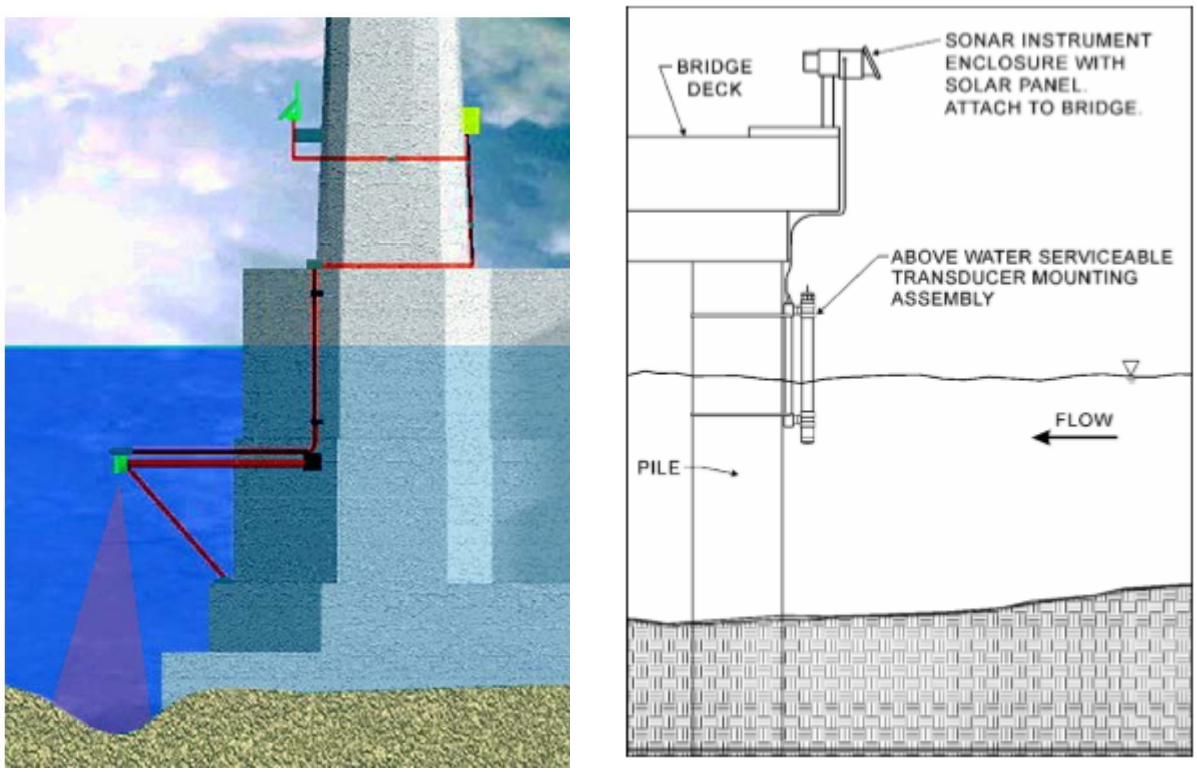


Figure 3 Schematic examples of installed sonar sensors at bridge piers [21,12]

TABLE 2 Summary of Key Scour-Related Characteristics for Selected Bridges

Bridge NBI ID	River/Bridge Name	Pier Width ft (m)	Pier Spacing ft (m)	Pier Length ft (m)	Manning's Coef.	Valley Slope	D50 inch (mm)	Angle of Attack deg	Discharge 100 RP ft3/s (m3/s)	Discharge 500 RP ft3/s (m3/s)
202	Tanana River at Nenana	10 (3)	510 (155.5)	48 (14.6)	0.03	0.008	0.6 (15)	15	162,000 (4,590)	203,000 (5,750)
212	Kashwitna River	5 (1.5)	70 (21.3)	28 (98.5)	0.035	0.002	0.4 (10)	0	3,920 (110)	4,650 (130)
215	Montana Creek	5 (1.5)	62 (19.5)	28 (8.5)	0.035	0.004	0.4 (10)	0	8,270 (230)	11,700 (330)
230	Sheridan Glacier No. 3	5 (1.5)	40 (12.2)	30 (8.5)	0.03	0.0015	0.4 (10)	20	16,300 (460)	20,200 (570)
524	Tanana River Big Delta	5 (1.5)	95 (29.0)	35 (9.4)	0.03	0.0002	055 (14)	35	86,700 (2,450)	95,600 (2,700)
527	Salcha River	4 (1.22)	140 (42.7)	35 (10.7)	0.03	0.002	0.4 (10)	10	50,600 (1,430)	64,900 (1,830)
539	Knik River	4.2 (1.31)	180 (54.9)	26 (7.9)	0.046	0.0006	0.2 (5)	0	79,400 (2,250)	104,000 (2,940)
557	Lowe River Lower Cross	2.5 (0.76)	100 (30.5)	38 (11.6)	0.04	0.008	0.4 (10)	0	22,100 (620)	285,000 (8,070)
573	Tazlina River	15 (4.6)	175 (53.4)	35 (10.8)	0.036	0.002	3.54 (90)	35	79,400 (2,250)	10,9000 (3,080)
634	Twenty Mile River	5 (1.5)	80 (24.4)	30 (9.1)	0.035	0.0005	0.4 (10)	0	32,500 (920)	35,600 (1,000)
670	Kasilof River	5 (1.5)	110 (33.5)	30 (9.1)	0.035	0.001	0.4 (10)	0	14,500 (410)	16,900 (470)
742	Chilkat River	5 (1.5)	50 (15.2)	24 (7.3)	0.035	0.006	0.4 (10)	0	32,600 (920)	42,900 (1,210)
983	Red Cloud River	5 (1.5)	30 (9.4)	38 (8.5)	0.045	0.02	0.4 (10)	0	1,580 (410)	1,960 (50)
999	Glacier Creek	5 (1.5)	55 (16.8)	30 (9.1)	0.04	0.015	0.4 (10)	20	7,020 (190)	10,200 (280)
1020	Quartz Creek	5 (1.5)	28 (8.5)	13.7 (4.2)	0.03	0.002	0.4 (10)	0	NaN	NaN
1187	Copper River	5 (1.5)	210 (64)	30 (9.1)	0.03	0.00015	0.4 (10)	35	54,400 (1,540)	69,000 (1,950)
1243	Nenana River at Windy	5 (1.5)	190 (58.2)	32 (9.8)	0.05	0.0042	0.4 (10)	0	13,000 (360)	15,400 (430)
1383	Lowe River at Keystone	5 (1.5)	100 (30.5)	36 (11)	0.04	0.005	0.4 (10)	50	21,300 (600)	27,600 (780)
5900-19	Mississippi River-Winona	8 (2.5)	345 (105.5)	30 (9.4)	0.03	0.00001	0.4 (10)	0	NaN	NaN
5900-20	Mississippi River-Winona	11.5 (3.5)	185 (57)	30 (9.4)	0.03	0.00001	0.4 (10)	0	NaN	NaN

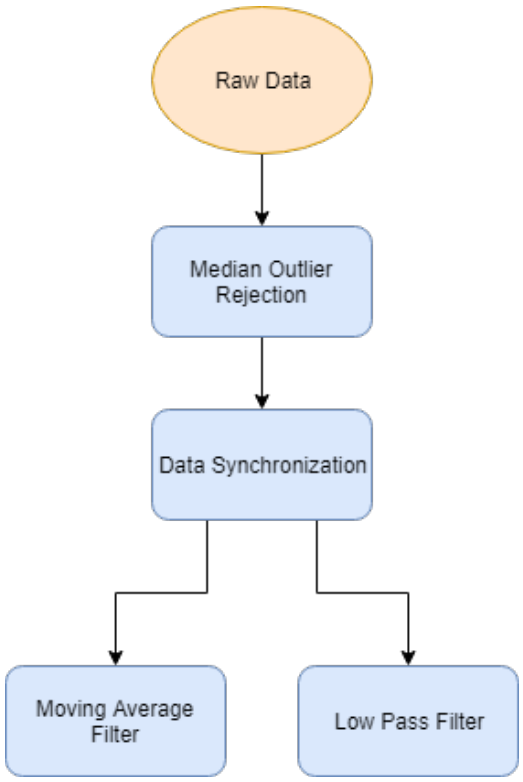


Figure 4 Basic data processing workflow

TABLE 3 Data quality issues and applied pre-processing technics

Data Quality Issue	Applied Processing Technic
Sample rate variability and misalignment between stage and sonar	Each point in the original data is a dual entry of [timestamp, value]. Sequential timestamped entries are not uniformly distributed, i.e., there isn't a standard length between the time steps (typically stage is at a higher sample rate than sonar).
Non-uniform sampling	There are some timestamps that differ from the majority sampling rate, these were detected and culled.
Outlier and spikes	Median filtering (non-linear filter that is effective in removing spikes in data), also applied some low pass filtering.
High frequency noise	Low pass filtering using a range of frequencies.
Bias shifts	A change in bias or a change in datum and units of the data. Manual inspection was performed for correction.
Interruptions to data recording/ Missing data	Missing data occurs throughout the data, particularly in the sonar readings. All data records are composed of shorter sub-sequences generally interspersed with small runs of missing data. Data imputation and interpolation technics were applied.

METHODOLOGY AND ALGORITHMS

Three distinct methods/algorithms have been developed using advanced AI/ML technics to predict scour depth at bridge piers based on stage and sonar data. The theoretical details of the methods are not provided in this paper. Readers are encouraged to refer to provided references for more background. The following sections discuss each method.

Method 1: Bed Elevation and Stage Water Level Prediction using Long-Short Term Memory (LSTM) Networks

LSTM networks were developed in this study with the aim of forecasting trends of bed elevation variation during critical scour episodes. The core idea was to develop an early warning system using algorithms trained on the sensor data that can predict future based on the past.

LSTM networks are a type of Recurrent Neural Networks (RNNs) with special internal memory units that can be updated, erased or read out [22,23]. These networks have proven successful for recognizing temporal patterns in time-series data. In addition to feed-forward connections, computational units (neurons) in RNNs have recurrent connections where the output of a unit is fed back to itself with a weight and a time delay, which provides it with a memory of past activations. Stacking memory units in such networks enables learning higher levels of temporal patterns in sequential data. LSTMs have been applied for forecast in time-series data such as stock market trends, text, language and voice recognition [24]. **Figure 5** provides a diagram of the LSTM architecture.

A grid search was performed to find the best configuration for LSTMs and to find the optimum values for the hyperparameters. The hyperparameters include the number of LSTM units, number of hidden neurons, optimization algorithm, initial gradient decay rate, maximum number of epochs, training and test data size, and number of sequences (sequence length). **Table 4** provides the range of values tested for each hyperparameter and the selected value for each.

LSTM models were trained with a bigger portion of the sequential data (training dataset) and their performance was validated using the remaining of the data (test dataset). The target forecast window was defined to cover both short (1hr in advance) and longer term (7-10 days in advance) predictions. For training the LSTMs, different methods of data slicing and batching were experimented, including: 1) single sequence using continuous data across all years, 2) single sequence using one-year data, 3) multiple sequences from synchronized data across various years, and 4) dividing the data into smaller sequences with a specified length (48 hrs). These sequences were then grouped into a number of mini-batches. The LSTM model was then trained over all these mini-batches in each epoch. An example of training and test sequences is shown in **Figure 6**.

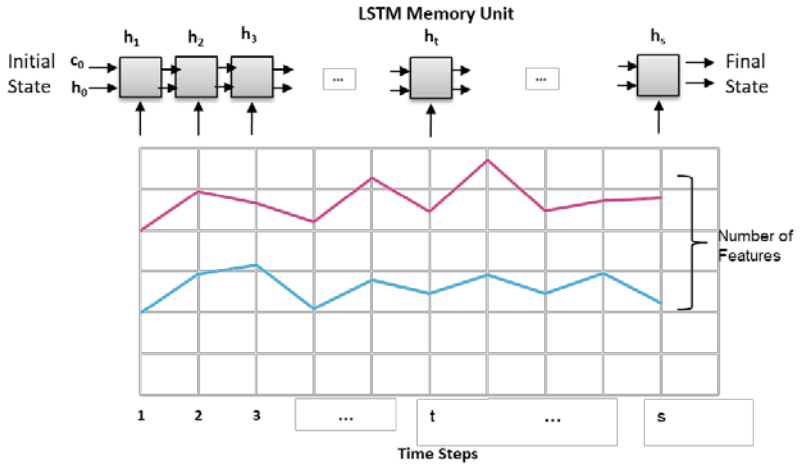


Figure 5 LSTM network architecture, a) Unfolded in time, b) LSTM Unit Memory

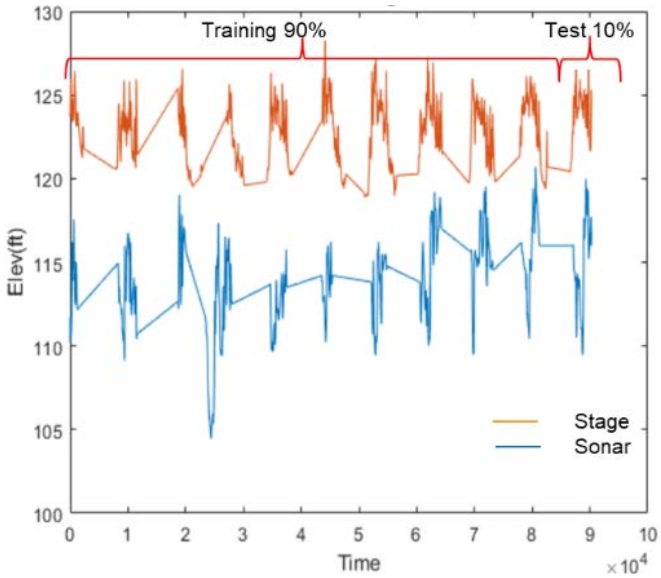


Figure 6 Example of training and test data sequences showing single continuous sequence across eleven years, from 2007 to 2017 (bridge 742), time step= 1hr

TABLE 4 LSTM model hyperparameters

Parameter	Tested Ranges	Selected Value
Number of LSTM Units	1, 2, 5, 10	1
Number of Hidden Neurons	20, 50, 100, 200	100
Optimization algorithm	SGDM, Adam and RMSProp	Adam
Initial Gradient Decay Rate	0.01, 0.005, 0.001	0.005
Max Number of Epochs	50 - 250	Variable depending on the dataset and batch size
Training/Test Data Size	98/2,95/5,90/10	95/5
Number of Sequences (Sequence Length)	One (all years of data in one sequence), No of Years (11), sequence length=48 hrs	-

Method 2: Maximum Scour Prediction using Physically Driven Neural Networks

This method aims at providing a holistic predictive model for the maximum depth of scour, not only based on observed scour and stage data but also considering the actual physics of the scour process. The advantage of this method over the first method is that, unlike LSTMs that are bridge-specific, the models can be trained by data from various bridges and subsequently can make predictions for any bridge at any location.

Multilayer Perceptron networks (MLPs) are feed-forward neural networks proven successful for prediction and approximation in high-dimensional problems [25-27]. Three-layer MLP networks were developed to predict the maximum scour depth based on selected physical and engineering features (input parameters) governing the scour process. The input parameters included flow, riverbed, and bridge characteristics which are the key components of scour empirical models, such as HEC-18 [28,29] and CIRIA [30] (see **Equations 2** to **9**).

Stepwise parameter selection method was implemented for feature selection. This method includes a forward selection and a backwards elimination of input parameters for the MLP networks [25]. Also, the number of hidden neurons (NH) was a driving factor, as it depends on the number of input parameters. Two sets of analyses were performed with both variable number of number of hidden neurons and a range of fixed values.

The model performance was evaluated using adjusted R^2. The adjusted R^2 penalize the model performance based on the number of input parameters/feature as provided in following equation:

$$R^2_{adj} = \frac{(n-1)R^2-p}{n-1-p} \qquad (1)$$

where:

R^2_{adj} = Adjusted R^2

n = No. of data points presented to the network in the training dataset

R^2 = R^2 for the reduced model

p = No. of network parameters (weights and biases) for the reduced model

The following parameters were gathered as provided in **Table 2**.

- Pier width (m)

- Pier spacing (m)
- Manning's coefficient
- Valley slope
- Pier spacing (m)
- Pier length (m)
- Median riverbed particle size - D50 (mm)
- Flow angle of attack (deg)

In addition, the flow depth and scour depth variations during a bridge major scour episode (usually one major scour episode within a year corresponding to a major flood event) were extracted from the collected sonar and stage data. Around 70 scour episodes were identified across all bridges in the database from 2007 to 2017. **Figure 7** shows an example of a major scour episode. The criteria for identifying scour was if the lowest bed elevation is at least 3ft (1m) below the as-built bed elevation of a bridge (the elevation based on which the bridge foundation was originally designed). Also, in this diagram, t_0 is the starting point of the scour episode (peak point) and t_m is the time of maximum scour (or the target bed elevation). Based on a certain forecast window (dt), t_n is defined as the time of last available sonar reading, equal to t_m-dt.

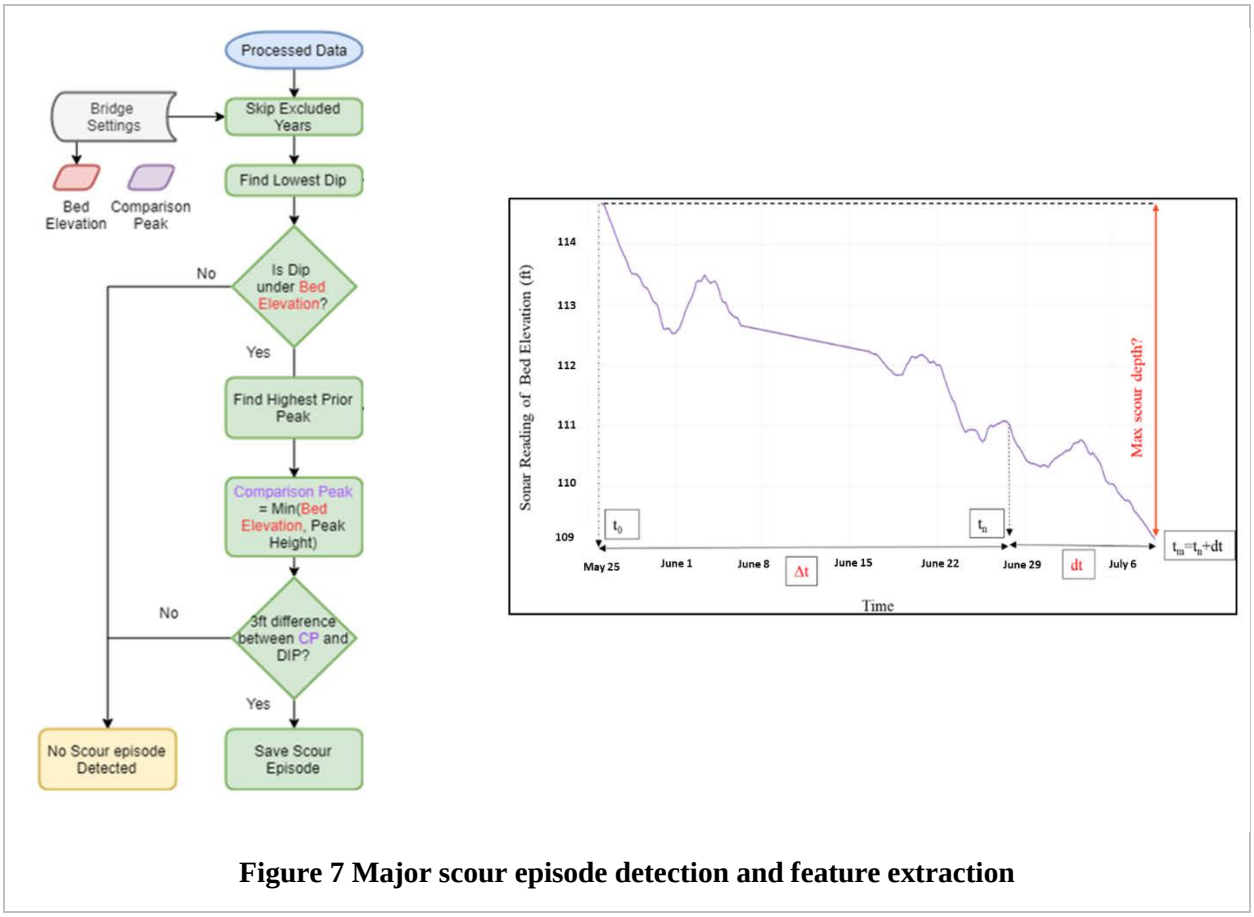


Figure 7 Major scour episode detection and feature extraction

The following additional input parameters were extracted from each scour episode:

- Flow depth at t_0 = Stage_t_0 – Sonar_t_0
- Flow depth at t_n = Stage_t_n – Sonar_t_n
- Scour depth at t_0 = Sonar_t_0 – As built elev
- Scour depth at t_n = Sonar_t_n – As built elev
- Scour episode length, $\Delta t = t_n - t_0$ (hrs)

- Forecast window, dt = 7days or 40% of Δt (hrs), whichever is smaller (when a scour event happens rapidly, 7 days can be too long relative to the total length of scour episode, therefore 40% of total time is suggested as an alternative)

Figure 8 provides a diagram of the developed MLP networks, showing the architecture, as well as input and target/output parameters. Previous studies on the optimum architecture for MLP networks have shown that one hidden layer is sufficient [25]. The optimum number of hidden units depends on the number of input and model parameters (neural network weights and biases). This was determined through series of experiments on the network performance.
MLPs were trained using random subsampling/Monte Carlo sampling method [25]. Thousands of networks were trained from scratch by randomly sampling data records for training (60%) and test (40%) datasets. Networks were trained using training dataset and their performance were evaluated over the test dataset.
Also, Bayesian Regularization was implemented in the optimization algorithm to enhance the accuracy of predictions [31]. This method automatically detects the most important model parameters (weights) and avoids overfitting to training data. Out of the large pool of generated networks, the ones with the best performance were selected to form an ensemble of predictive models. The accuracy was evaluated based on the average over the network predictions within the ensemble and the uncertainty of the predictions due to network initialization and data subdivision was captured.

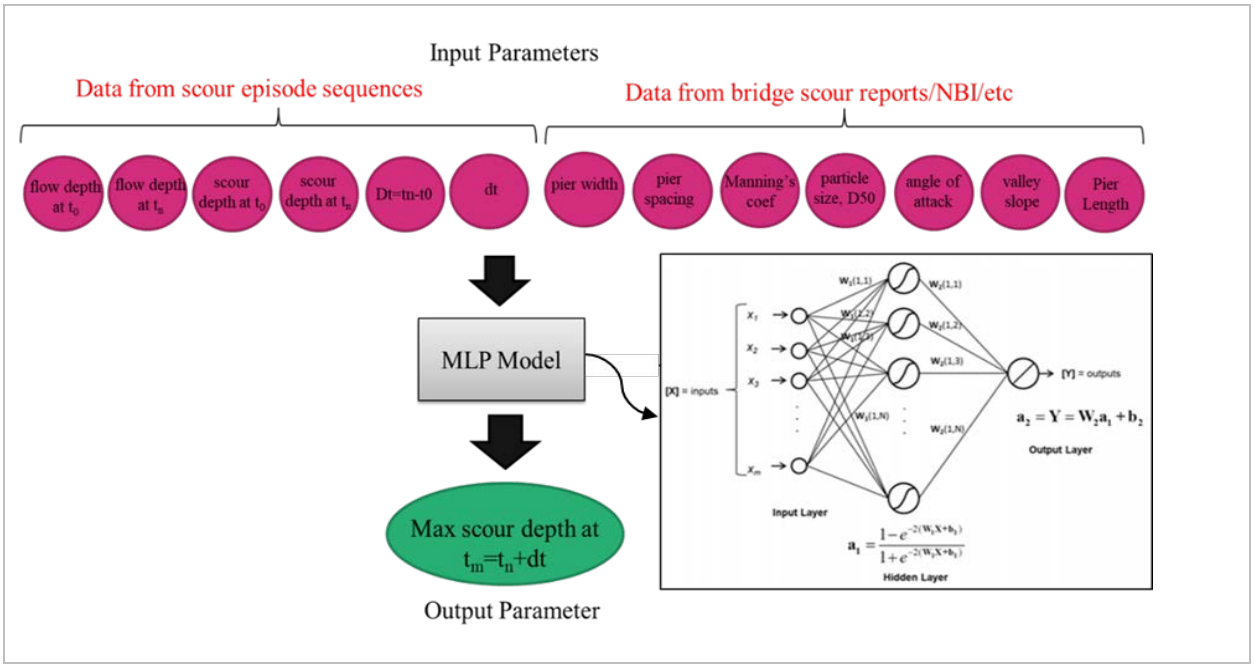


Figure 8 Physically driven neural network approach, showing MLP model architecture and the input and output parameters

Method 3: Scour Prediction and Uncertainty Quantification using Bayesian-Calibrated Empirical Models

Existing studies on empirical scour models show that the majority can result in significant error in scour depth prediction. Most of these models have been calibrated based on local, limited datasets and are not necessarily accurate when applied for an arbitrary bridge [32-34]. Despite the limitations of traditional scour empirical models, their role in management of scour risk is deeply-rooted.
In this study, two scour empirical models were calibrated using the Bayesian inference approach and based on the collected scour monitoring data. The uncertainty over the scour model parameters was modelled by

representing them via an initial, assumed probability distributions (prior distribution). Parameters treated in this manner are referred to as "latent" variables. Using the Bayesian framework, the prior distributions over the latent variables were updated based on the observed (training) data. The result was a joint probability distribution, conditional on the observed data, over all the latent variables. Markov-chain Monte Carlo (MCMC) method was implemented to generate this joint posterior distribution. For more details on the theory and implementation of Bayesian regression and the MCMC method, readers are referred to [26,35,36].

Probabilistic predictions of scour depth for a given river stage can be made by repeatedly sampling (Monte Carlo) from the joint posterior distribution over the latent parameters to generate a set of predicted scour depths. **Figure 9** provides a graphical representation of the Bayesian model structure and prediction of the bed elevation (sonar) using calibrated empirical models.

For implementing this method, the stage water data needs to be predicted in advance. This is possible using a combination of recorded rainfall in the river catchment and a hydrological model.

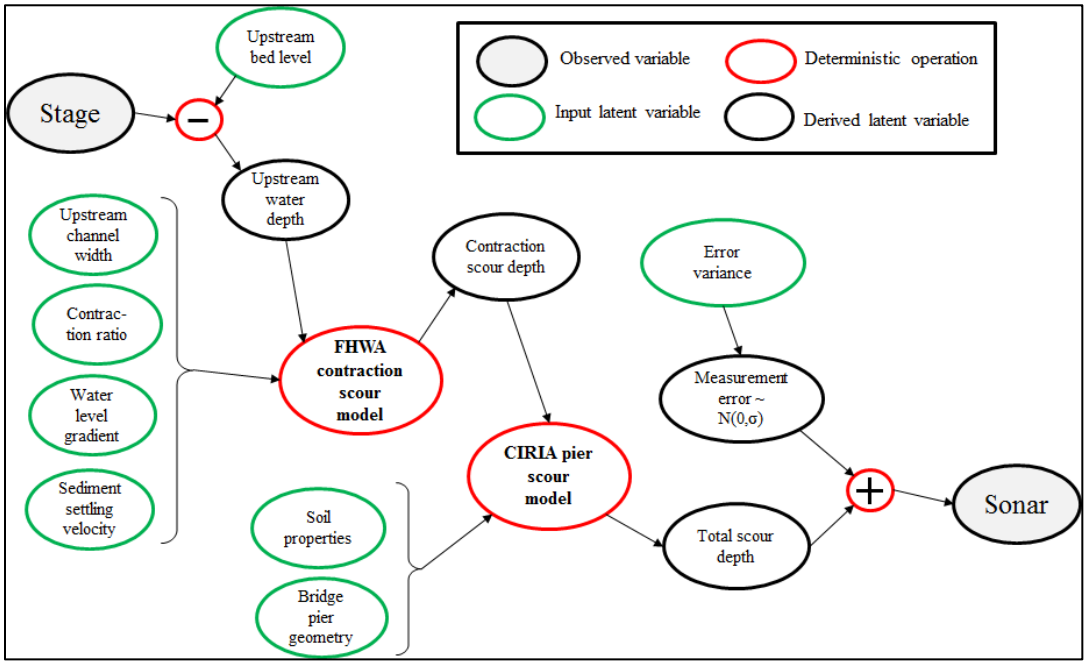


Figure 9 Graphical representation showing Bayesian model structure.

The empirical equations applied in this study for estimating local and contraction scour are described in the following sections.

CIRIA Local Scour Model

Local scour around a pier is assessed as the product of the pier width (D) and a number of factors depending on the flow depth (F_{depth}), angle of attack (F_{angle}), flow velocity ($F_{velocity}$), pier shape (F_{shape},) and debris condition (F_{debris},) as shown in **Equations 2** to **6**. F_{shape} varies between 1.15 and 2.0 and F_{debris} varies between 1.0 and 1.48. S_F is a safety factor related to the assumed probability of exceedance [30].

$$d_s = D \times \left(F_{angle} \times F_{shape} \times F_{depth} \times F_{velocity} \times F_{debris}\right) \times S_F$$	(2)
$$F_{angle} = \left(\cos\alpha + \left(\frac{L}{D}\right)\sin\alpha\right)^{0.62}$$	(3)

$F_{depth} = 0.55\left(\frac{Y_B}{D}\right)^{0.60}$, for $\frac{Y_B}{D} \leq 2.7$ $F_{depth} = 1.0$, for $\frac{Y_B}{D} > 2.7$	(4)
$F_{velocity} = 0$, for $\frac{V}{V_c} \leq 0.375$ $F_{velocity} = 1.6\left(\frac{V}{V_c}\right) - 0.6$, for $0.375 < \frac{V}{V_c} < 1$ $F_{velocity} = 1$, for $\frac{V}{V_c} \geq 1$	(5)
$V_c = -\sqrt{32}\log_{10}\left(\frac{k_s}{12Y_B} + \frac{0.222\upsilon}{Y_B V_*}\right)V^*$	(6)

In the equations above L is pier length, α is the angle of attack, V is the mean channel velocity. V_c is the critical velocity for sediment motion, Y_B is the flow depth, υ is the kinematic viscosity of water (equal to $1.14e^{-6}$ at 15°C), k_s is the effective roughness, taken as D_{50} (median size of bed material) for uniform sediment and $2*D_{50}$ for well-graded sediment, and V^* is the shear velocity at threshold of sediment movement.

FHWA (HEC-18) Contraction Model

Depending on the flow velocity either **Equation 7** (live-bed contraction scour) or **Equation 8** (clear-water conditions) is applied. The depth of contraction scour (d_g) is estimated using **Equation 9** [29].

$\frac{Y_B}{Y_u} = \left(\frac{Q_2}{Q_1}\right)^{\frac{6}{7}}\left(\frac{W_u}{W_B}\right)^{k_1}$	(7)
$Y_B = \left(\frac{K_u Q^2}{D_m^{\frac{2}{3}} W_B^2}\right)^{\frac{3}{7}}$	(8)
$d_g = Y_B - Y_e$	(9)

In the equations above Y_u is the average flow depth in the upstream main channel, Y_B is the average flow depth in the contracted section, Y_e is the existing flow depth in the contracted section before scour, which can be approximated by Y_u for sand and gravel beds. Q is the discharge through the bridge, Q_1 is the discharge in the upstream channel, Q_2 is the discharge in the contracted channel, W_u is the width of the upstream main channel, W_B is the width of the main channel in the contracted section, k_1 is an exponential constant, D_m is the diameter of the smallest non-transportable particle taken as $1.25*(D_{50})$, K_u is a constant equal to 0.025 (SI unit) or 0.0077 (English units).

Table 5 provides a summary of latent and observed model parameters and adopted prior distributions (U represents the uniform probability distribution).

TABLE 5 Latent and observed model parameters (variables) for Bayesian regression

Variable name	Variable Type	Units	Prior Distribution	Observed value(s)
Upstream bed elevation	Latent	ft	U (31,36.5)	-
Upstream channel width	Latent	ft	U (150,500)	-
Contraction ratio	Latent	-	U (1.0,1.3)	-
Sediment settling velocity	Latent	ft/s	U (1.64,32.8)	-
Water surface friction slope	Latent	-	U (0,0.02)	-
Pier length	Latent	ft	U (5,10)	-
Pier shape factor (Fshape)	Latent	-	U (1.0,1.2)	-
Manning coefficient	Latent	-	U (0.02,0.06)	-
Friction velocity	Latent	ft/s	U (0.14,0.3)	-
Pier angle of attack	Latent	-	U (0,π/3)	-
(Nikuradse) Sand grain roughness	Latent	-	U (0.005,0.02)	-
Error standard deviation	Latent	ft	U (0.5,3.0)	-
Pier width	Observed	ft	-	4.3
Scoured bed elevation	Observed	ft	-	sonar sensor
Water level	Observed	ft	-	stage sensor

RESULTS AND DISCUSSIONS

Method 1: LSTM Predictions

Figure 10 shows examples of the LSTMs' performance over test datasets for 1hr ahead predictions, using single sequence training method. At each time step, the LSTM model predicts the stage water and bed elevation, provided the actual sensor readings at the previous time steps. **Figure 11** shows the LSTM predictions of stage water and bed elevation for 7 days in advance, based on multiple sequence training methods. Seven to ten days was considered as a reasonable forecast window for early warnings of a bridge scour event. Across the forecast window, instead of the actual readings, the LSTM predicted values for a time-step are input to the next time step.

The 1hr ahead forecasts were highly accurate, showing Root Mean Squared Error (RMSE) close to zero, and as expected, the accuracy decayed as the forecast window was extended to 7-10 days. The LSTMs' performance showed to vary, depending on the bridge and quality of the data. The average RMSE on bed elevation predictions over 7-10 days forecast window was estimated between 1-5 ft.

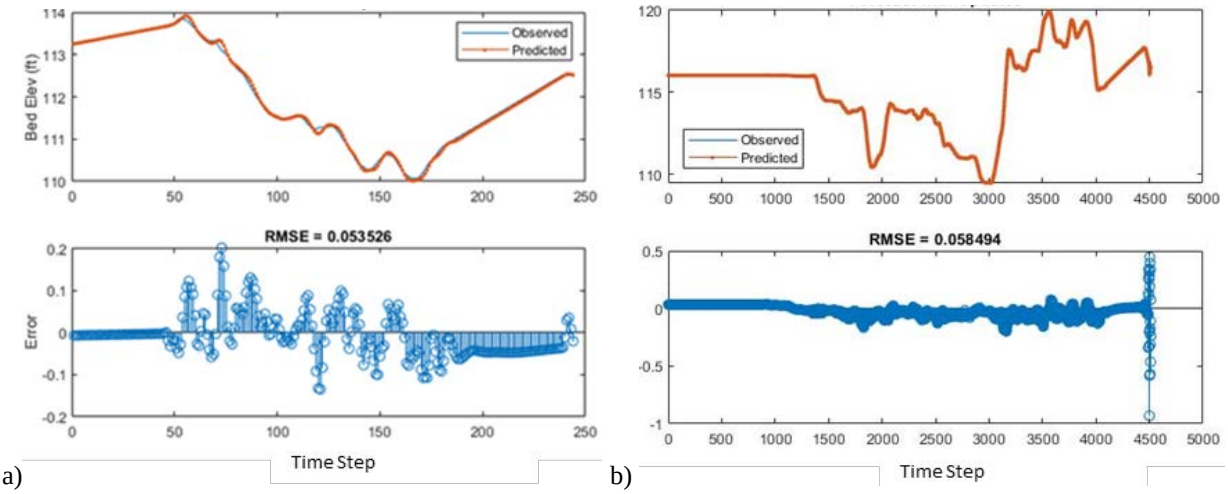
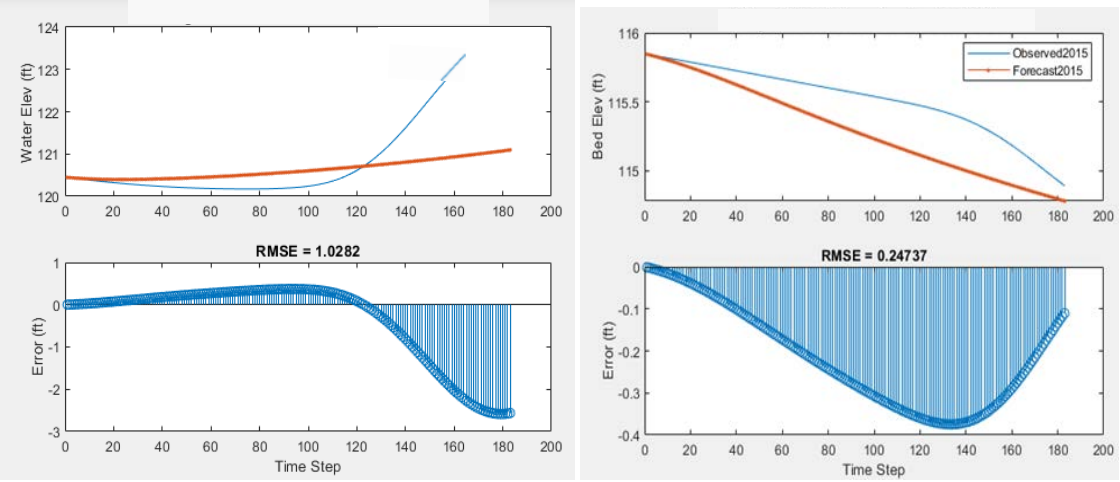


Figure 10 LSTM performance over test datasets, trained using single sequences with, a) 1 year of sensor data, b) 11 years of sensor data – 1hr forecast window, bridge 742 [time-step=1 hr]

a)



b)

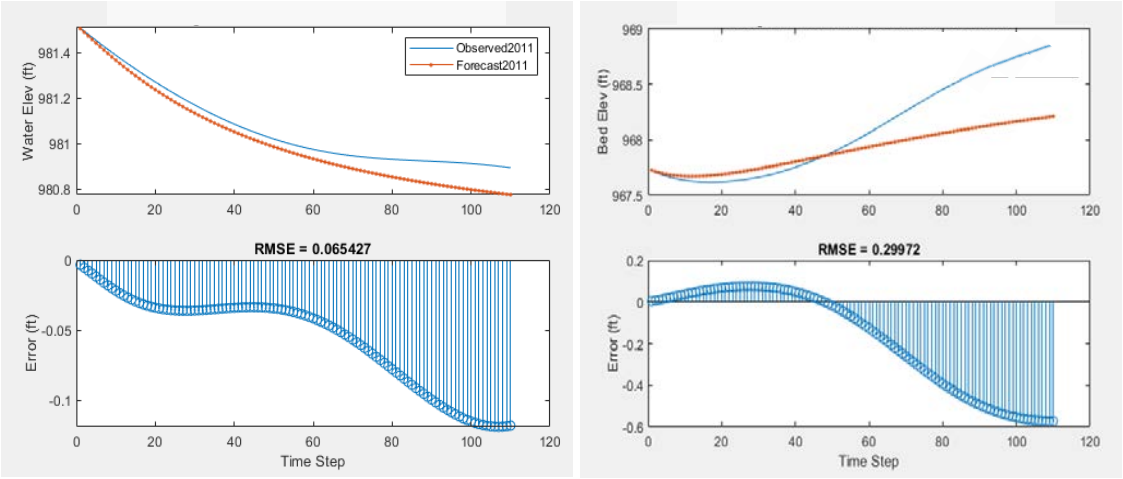


Figure 11 LSTM performance over test datasets, trained using synchronized sequences, a) bridge 742-2015, b) bridge 524-2011 – 7 to 10 days forecast window [time-step=1 hr]

These results show a great potential in real-time assessment of scour around bridge piers using LSTMs. It should be noted that in general, LSTMs are extremely "data-hungry." The size of the database available for training in this study was relatively small in deep learning terms. The performance of the developed LSTMs is expected to improve further, as more data is received and included in the training process in the future phases of this study.

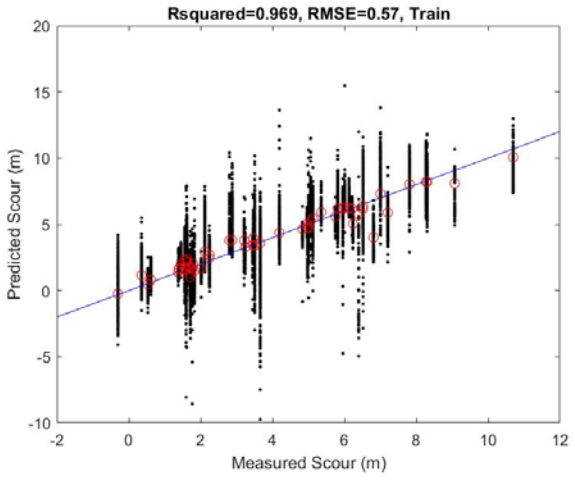
Method 2: MLP Ensemble Predictions

Stepwise parameter selection (forward selection and backward elimination) analyses were carried out to optimize the MLP performance by removing the redundant parameters and identifying the most relevant input parameters for maximum scour depth prediction. The following input parameters were selected for the final configuration of the MLP networks:

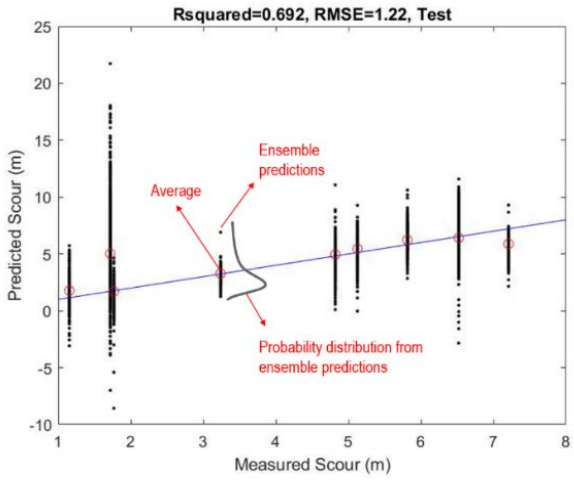
- Flow depth at t_0
- Flow depth at t_n
- Scour depth at t_0
- Scour depth at t_n
- Manning's coefficient
- Valley slope

- Aggregate size, D_{50}

Figure 12 shows the performance of the MLP ensembles for maximum scour depth prediction over training and test datasets. The black dots show individual predictions from each network in the ensemble, and the red circles show the average predictions. Summary statistics for the MLP ensemble performance, including the R^2 of predicted versus measured values, as well as *RMSE* are provided in **Table 6**. This includes the performance over a control data set which was set aside during model development. R^2 of 0.7 for the test dataset and 0.91 for the control dataset (equivalent to 1.2m and 0.9m prediction error, respectively) show that this approach can be used to make reasonably accurate predictions on maximum scour depth.
The MLP ensembles method shows to outperform the commonly-used empirical models reported in [34]. SSE (sum of squared errors divided by sum of squared measured values × 100%), the performance metric used in [34] is estimated 6.67% over the test dataset and 3.79% over the control dataset. Comparing this with the performance of the best reported empirical model [37] with 6.8% SSE proves this claim.
Moreover, as shown in **Figure 12**, the range of predictions by ensemble networks for each scour episode allows for uncertainty quantification and probabilistic inferences, such as estimating confidence intervals and percentile values.



a)



b)

Figure 12 MLP ensemble predictions over a) training dataset and b) test dataset, showing R^2 and RMSE of predictions. Blue line is the equity line.

TABLE 6 Performance of the MLP ensembles over data subsets

Dataset	***R^2***	***RMSE, ft (m)***
Training Subsets (average)	0.97	2 (0.6)
Test Subsets (average)	0.70	3.9 (1.2)
Control Subset (average)	0.91	2.9 (0.9)

Method 3: Predictions by the Bayesian Empirical Models

Unlike neural networks, which are capable of learning complex, non-linear patterns in data, linear models are less likely to benefit from large and diverse sets of training data. Thus, the empirical models should be calibrated on a case by case basis, with limited potential for learnings to transfer between bridges. The sonar readings from bridge 539 showed good quality, hence the initial models were developed and evaluated using this dataset.

After generating the joint posterior probability distribution of all latent parameters of the empirical models, the models were used to make posterior predictions on bed elevation. This was performed by drawing a large number of samples, using Monte Carlo method, from the joint posterior probability distribution of latent parameters, for each value of stage water and then using empirical scour model to estimate the scour depth. Using the generated predictions, the 5th, 50th and 95th percentile values from the posterior predictive distribution of bed level (i.e. sonar readings) were obtained. The range between the P5 and P95 values can be considered as the 90% credible interval for the predicted scour depth. This range was compared against the actual sonar readings (see **Figure 13** and **14**).

Global Fit

A "global" model was trained using the whole dataset between 2008 to 2012. The model predictions on the training datasets and the test dataset taken from 2014 are shown in **Figure 13**.

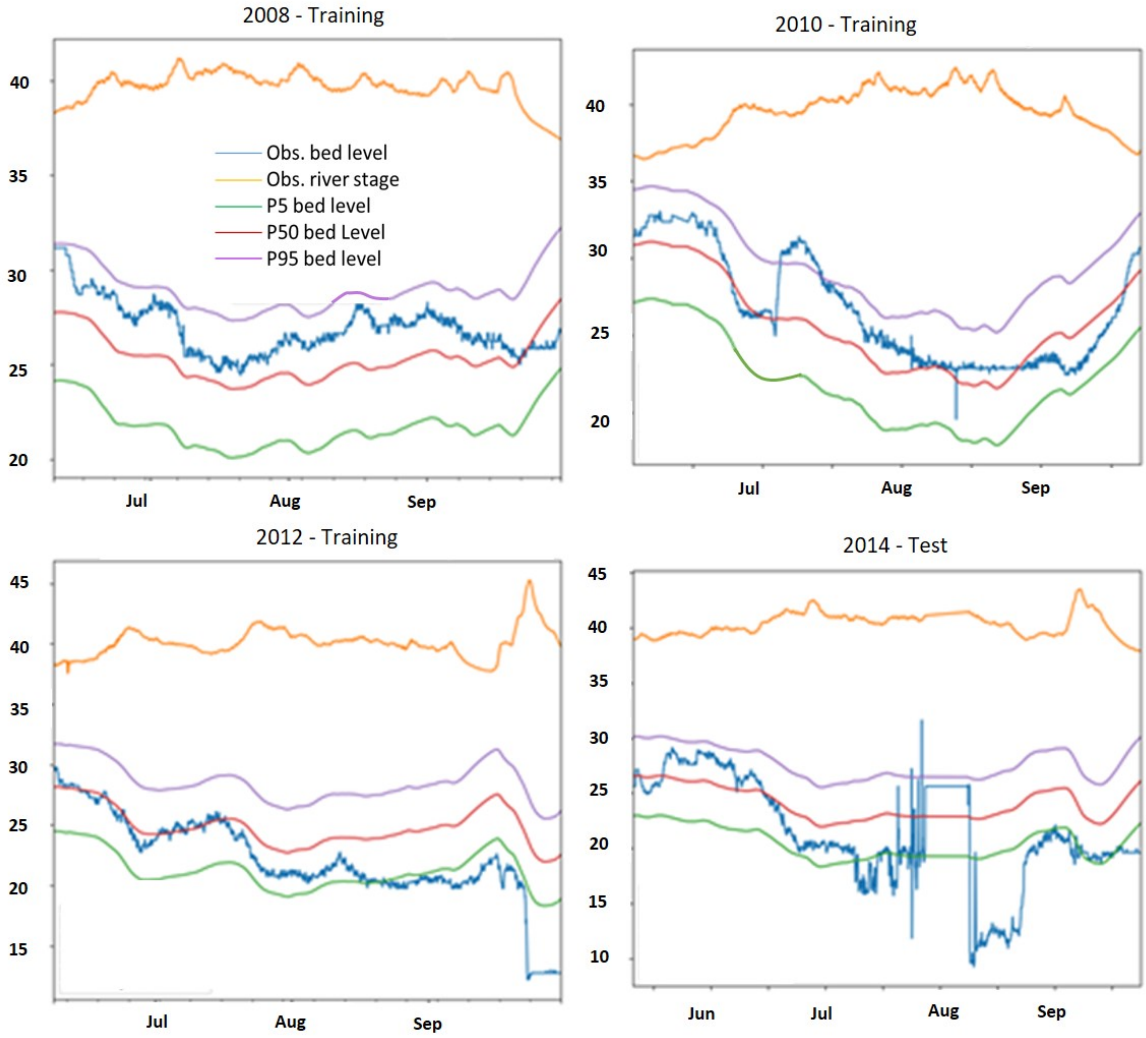


Figure 13 Model predictions on 2008-2012 training dataset and test dataset from 2014 (x axis is time, y axis is elevation in ft).

Seasonal/Local Fit

In this analysis, it was of interest to know what quantity of observed data would be required to make useful predictions on unseen data. For example, is it possible to deploy temporary sensors during one season only to calibrate a model which can then be used reliably in future years? If this was the case, then the cost of monitoring could be significantly reduced.

Figure 14 shows an example of prediction results using this method. Training data was extracted from the date range from 15th June to 31th August 2012. Model predictions on the training dataset are reasonably accurate, with the observed data mostly falling inside the 90% credible interval. Predictions on test data shows a general tendency to underestimate bed elevations in 2008 but shows better agreement in other years. The most noticeable discrepancies with the observed data coincides with sections where the actual readings are seemingly erroneous, for example the sudden bump in the sonar readings in early June 2010, or fluctuations in August 2014.

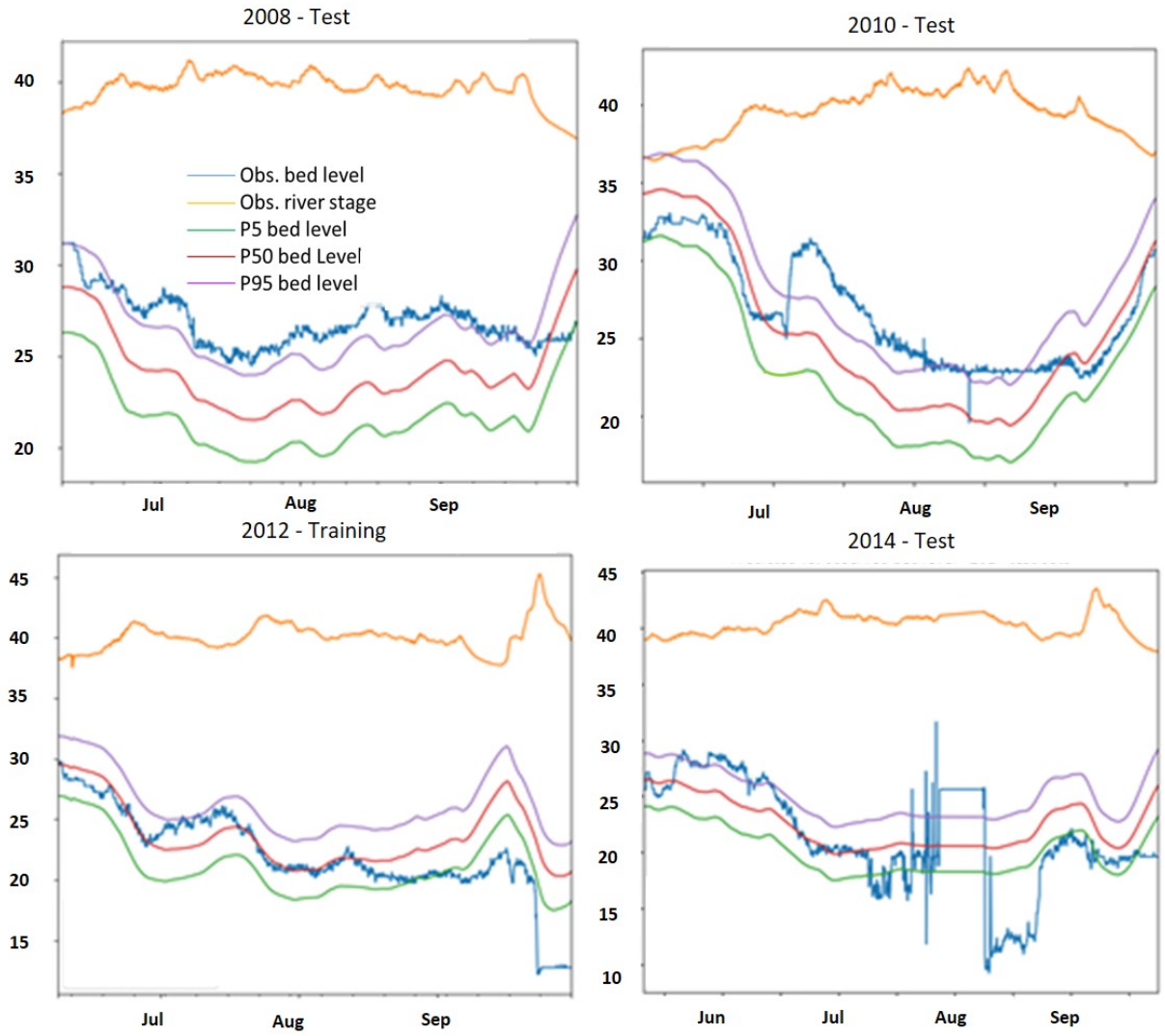


Figure 14 Model predictions on 2012 training dataset and test dataset from years 2008, 2010 and 2014 (x axis is time, y axis is elevation in ft).

The *RMSEs* for the "global" and "seasonal" models, based on the P50 (median) values from the posterior predictive distributions are shown in **Table 7**. Qualitatively, there is no strong evidence that training over multiple years could result in more accurate predictions. The predictions are generally accurate to within 2-3 ft, with the largest errors generally explained by measurement errors rather than modelling errors.

TABLE 7 RMSE in predicted bed elevation, ft (m) for models trained against long term data (2008 - 2012) and single season of data (2008 and 2012).

Test Period	Training Period		
	Global: 2008-2012	**Local: 2008**	**Local: 2012**
2008	2.0 (0.6)	1.2 (0.4)	3.3 (1.0)
2010	2.0 (0.6)	2.3 (0.7)	2.9 (0.90)
2012	3.9 (1.2)	5.3 (1.5)	3.1 (0.90)
2014	4.9 (1.5)	6.0 (1.8)	3.8 (1.2)

Implementation Remarks

Each of the three introduced methods is designed for a specific implementation/application purpose:

Method 1: *Real-time Scour Early Warning System*. The application of this method is straightforward, as the AI models will be trained using historical data for a bridge and then retrained as new monitoring data is fetched on a real-time basis. The training process can be reduced to several hours on a CPU or GPU cluster or other high-performance computing (HPC) systems, depending on the amount of data and the training process (e.g. number of epochs, number of mini-batches, and sequence length). This method provides real-time scour depth forecast (as well as stage water forecast).

Method 2: *Holistic Scour Prediction*. This method can be applied for any bridge at any location once the models are trained with a well-populated database from bridges in the area. The training on a HPC system is estimated to take up to several hours depending on the size of the database. The predicted maximum scour depth using this method can inform decision making and scour risk assessments.

Method 3: *Bayesian Empirical Scour Prediction*. This method can be implemented to predict upcoming scour depth at a bridge and provide systematic uncertainty assessments. The Bayesian calibration process can be time-intensive depending on the HPC system. Hydro-meteorological models will be required to forecast the stage water height (flow depth) as an input to the scour empirical equations.

SUMMARY AND CONCLUSIONS

New methodologies were developed using advanced AI/ML technics for early forecast of scour and maximum scour depth prediction based on real-time monitoring data. Sonar and stage sensor data were collected for a number of bridges in the US, along with other physical characteristics relevant to scour process, such as bridge geometry, river flow, and riverbed condition.

As the first approach, recurrent neural network models, LSTMs, were developed and trained to predict the trend of bed elevation and stage water variations over a predefined forecast window of 1hr to 10 days in advance, to provide a real-time early warning system. The main advantage of this approach is the ability to predict the stage water level in parallel to scour (flood forecast). Validation results revealed a good potential for this approach to be implemented for real-time scour forecast, with an average prediction error of 1-5 ft. It should be noted that accuracy of predictions using this approach may vary in a case by case basis, depending on the quality of the training data, scour process variations, and how well the network is trained over the collected dataset. Authors expect to further improve the algorithms introduced in this study in the upcoming phases.

As the second approach, ensembles of feedforward neural networks were developed and trained using both observed sensor data as well as bridge, flow and riverbed characteristics. This approach provides a holistic, physically-driven predictive model that can make reasonably accurate predictions on the maximum scour depth during a high-flood season. Unlike the first approach, the developed algorithms are robust in making predictions for bridges from different locations and do not require individual calibration and training for each bridge; once trained with spatially diverse database of bridges, they theoretically can make reasonable forecast for any selected bridge. In addition, using the random subsampling (ensemble) method, a probabilistic framework was introduced to capture the uncertainty and facilitate risk-based solutions for scour management.

Third approach implemented semi-empirical scour models and calibrated them using Bayesian regression based on the observed sensor data. This approach proved successful in predicting the scour depth with acceptable accuracy, while allowing for systematic uncertainty quantification. This method is specifically advantageous when attempting to extrapolate to extreme conditions (for example, floods with return periods larger than 100 years), which are highly unlikely to be represented in the available training data. Another advantage of this approach is the familiarity of bridge management authorities with the applied empirical models. Current state of practice requires bridge operators to perform regular surveys of

bed profiles and apply empirical models to estimate scour failure risk for a range of pre-defined flood events. One disadvantage of this approach is that, unlike the first and second approach, it requires the river stage data to be available throughout the forecast window, hence requires a flood forecast model in parallel.

The commonly used empirical scour models have been developed based on pre-defined regression models and calibrated with local, limited data. These models have proven unreliable, too conservative in some cases, yet underpredicting scour depth in other cases. The developed AI/ML models in this study showed to outperform these traditional methods. These advanced algorithms in combination with real-time monitoring systems can provide bridge owners with robust, cost-effective tools for scour risk management. Considering the significant costs and causalities associated with bridge scour failure, implementing such state-of-the-art technologies can be a revolution in the bridge management practice.

ACKNOWLEDGMENTS

This project was funded by Arup Global Research group/Arup University. Many academics and practitioners helped throughout this research study. Authors would like to specifically acknowledge the great contribution of Dr Robin Beebee and Jeff Conaway from Alaska Science Center of USGS, and Dr Solomon Woldeamlak from MnDOT for providing the monitoring data for this study. In addition, authors appreciate the support, technical advice, and reviews received from Professor Jean-Louis Briaud, Dr Sharid Amiri, and Dr Ken Fishman.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design:

N. Yousefpour, S. Downie, S. Walker; data collection: N. Yousefpour, H. Dikanski; analysis and interpretation of results: N. Yousefpour, S. Walker, S. Downie, N. Perkins, H. Dikanski; draft manuscript preparation: N. Yousefpour. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Annandale, G. W. 2006. *Scour Technology*, Mechanics and Engineering Practice, McGraw-Hill, New York.

2. Ebtehaj, I., Sattar, A. M., Bonakdari, H., & Zaji, A. H. 2017. Prediction of scour depth around bridge piers using self-adaptive extreme learning machine. *Journal of Hydroinformatics*, *19*(2), 207-224.

3. Lee T. L., Jeng, D. S., Zhang, G. H. and Hong, J. H. (2007). "Neural Network Modeling for Estimation of Scour Depth around Bridge Piers." *Journal of Hydrodynamics, Ser. B*, 19(3), 378-386.

4. Kaya A. 2010. "Artificial Neural Network Study of Observed Pattern of Scour Depth around Bridge Piers." *Computers and Geotechnics*, 37(3), 413-418.

5. Bateni S. M., Borghei, S. M. and Jeng, D. S. 2007. "Neural Network and Neuro-Fuzzy Assessments for Scour Depth around Bridge Piers." *Engineering Applications of Artificial Intelligence*, 20(3), 401-414.

6. Zounemat-Kermani M., Beheshti, A. A., Ataie-Ashtiani, B. and Sabbagh-Yazdi, S. R. 2009. "Estimation of Current-Induced Scour Depth around Pile Groups Using Neural Networks and Adaptive Neuro-Fuzzy Inference System." *Applied Soft Computing*, 9(2), 746-755.

7. Azamathulla, H. M., Ghani, A. A., Zakaria, N. A., & Guven, A. 2010. Genetic programming to predict bridge pier scour. *Journal of Hydraulic Engineering*, *136*(3), 165-169.

8. Pal, M., Singh, N. K., & Tiwari, N. K. 2011. Support vector regression based modeling of pier scour using field data. *Engineering Applications of Artificial Intelligence*, *24*(5), 911-916.

9. Kim, I., Fard, M. Y., & Chattopadhyay, A. 2015. Investigation of a bridge pier scour prediction model for safe design and inspection. *Journal of bridge engineering*, *20*(6), 04014088.

10. Khan, M., Azamathulla, H. M., & Tufail, M. 2012. Gene-expression programming to predict pier scour depth using laboratory data. *Journal of Hydroinformatics*, *14*(3), 628-645.

11. NCHRP Report 396. 1997. Instrumentation for measuring scour at bridge piers and abutments, Transportation Research Board, Washington DC.

12. NCHRP Synthesis of Highway Practice 396. 2009. Monitoring scour critical bridges, Transportation Research Board, Washington DC.

13. Briaud, J-L, Hurlebaus S, Chang, K., Yao, C., Sharma, H., Yu, O-Y., Darby, C., Hunt, B.E., and Price, G.R., 2011. Real-time Monitoring of Bridge Scour Using Remote Monitoring Technology, Texas A&M University/TTI - Report 0-6060-1.

14. Prendergast, L. and Gavin, K. 2014. A review of bridge scour monitoring techniques. Journal of Rock Mechanics and Geotechnical Engineering, 6(2). 138-149.

15. Elsaid A and Seracino R. 2014. Rapid assessment of foundation scour using the dynamic features of bridge superstructure. *Construction and Building Materials*; 50:42e9.

16. Foti S and Sabia D. 2011. Influence of foundation scour on the dynamic response of an existing

bridge. *Journal of Bridge Engineering*;16(2):295e304.

17. Kariyawasam, K., Middleton, C., Madabhushi, G., Haigh, S., Talbot, J. 2020. Assessment of bridge natural frequency as an indicator of scour using centrifuge modelling. *Journal of Civil Structural Health Monitoring,* 10, 861-881.

18. USGS, 2001. Methodology and Estimates of Scour at Selected Bridge Sites in Alaska.

19. USGS. 2004. Summary and Comparison of Multiphase Streambed Scour Analysis at Selected Bridge Sites in Alaska.

20. USGS. 2019. Alaska Science Center. https://www.usgs.gov/centers/asc/science/streambed-scour-bridges-alaska?qt-science_center_objects=0#qt-science_center_objects.

21. Lagasse, P.F., L.W. Zevenbergen, J.D. Schall, and P.E. Clopper. 2001. "Hydraulic Engineering Circular 23; Bridge Scour and Stream Instability Countermeasures", 2nd ed. Federal Highway Administration, Washington, D.C.

22. Hochreiter, S. and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural computation*, 9(8):1735–1780.

23. Ordonez F.J. and Roggen D. 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors Journal*, 16 (1), 115.

Prendergast LJ, Hester D, Gavin K, O'Sullivan JJ. 2013. An investigation of the changes in the natural frequency of a pile affected by scour. Journal of Sound and Vibration;332(25):6685e702.

24. Du S., Pandy, M., and Xing C. 2017. Modelling approaches for time series forecasting and anomaly detection. Stanford University.

25. Bishop, C. M. 1995. Neural Networks for pattern recognition. Oxford: Oxford University Press

26. Yousefpour, N. 2013. *Comparative Deterministic and Probabilistic Modelling in Geotechnics*, PhD Dissertation, Texas A&M University, College Station, Texas.

27. Yousefpour, N., Medina-Cetina, Z., and Briaud, J-L. 2014. Evaluation of Unknown Foundations of Bridges Subjected to Scour: Physically Driven Artificial Neural Network Approach, *Journal of Transportation Research Record*, 2433 (1), 27-38.

28. Richardson, E.V. and Davis, S.R. 2001. Evaluating Scour at Bridges- Fourth Edition, FHWA NHI 01-001, HEC-18, Washington, D.C.

29. Arneson, L., Zevenbergen, L., Lagasse, P. & Clopper, P. 2012. Evaluating Scour at Bridges. 5th ed. US Department of Transportation, FHWA.

30. Kirby, A., Roca, M., Kitchen, A., Escarameia, M. & Chesterton, O. 2015. Manual on Scour at Bridges and Other Hydraulic Structures. 2 ed. CIRIA, London, UK.

31. Okut H. 2016. Bayesian Regularized Neural Networks for Small *n* Big *p* Data. In *Artificial Neural Networks - Models and Applications*, INTECH.

32. Deng, L. and Cai, C. 2010. Bridge Scour: Prediction, Modelling, Monitoring and Countermeasures-Review. *Practice Periodical on Structural Design and Construction*, 15, 125-134.

33. Busari A. O., Gbadebo A.O., Jimoh I.O., Takeet O. E., and Enifu, J.E. 2013. Evaluation of Performance of Empirical Models for the Prediction of Local Scour at Bridge Piers (Case Study on Chanchaga Bridge, Minna, Nigeria), *International Journal of Engineering Research & Technology*, Vol. 2, Issue 3.

34. Sheppard, D., Melville, B. & Demir, H. 2014. Evaluation of Existing Equations for Local Scour at Bridge Piers. *Journal of Hydraulic Engineering*, 140, 14-23.

35. Robert C. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York, NY.

36. Hoff P. D. 2009. *A First Course in Bayesian Statistical Methods*. Springer, New York, NY.

37. Sheppard, D. M., and Miller, W. 2006. Live-bed local pier scour experiments. *Journal of Hydraulic Engineering*, 10.1061/(ASCE)0733-9429(2006)132:7 (635), 635–642