

Using Trend Extraction and Spatial Trends to Improve Flood Modeling and Control

Jacob Hale, Suzanna Long, Vinayaka Gude and Steven Corns

Abstract

Effective management of flood events depends on a thorough understanding of regional geospatial characteristics, yet data visualization is rarely effectively integrated into the planning tools used by decision makers. This chapter considers publicly available data sets and data visualization techniques that can be adapted for use by all community planners and decision makers. A long short-term memory (LSTM) network is created to develop a univariate time series value for river stage prediction that improves the temporal resolution and accuracy of forecasts. This prediction is then tied to a corresponding spatial flood inundation profile in a geographic information system (GIS) setting. The intersection of flood profile and affected road segments can be easily visualized and extracted. Traffic decision makers can use these findings to proactively deploy re-routing measures and warnings to motorists to decrease travel-miles and risks such as loss of property or life.

Keywords: trend extraction, spatial and temporal trends, images

1. Introduction

Floods are the most frequently occurring natural disaster. A flood event occurs when stream flows exceed the natural or artificial confines at any point along a stream [1]. This is often due to heavy rainfall, ocean waves coming on shore, rapid snow melting, or failure of manmade structures such as dams or levees [2]. From 1998–2017, flood events affected more than two billion people globally [3]. Disasters of this frequency and magnitude are typified by extreme costs to governments. In 2019, historic flooding across Missouri, Arkansas, and the Mississippi River basin resulted in an estimated cost of 20 billion dollars [4]. These estimates typically do not reflect indirect costs such as added travel-miles and the subsequent loss of time. Further, floods are among the deadliest natural disasters. From 2010–2020, floods resulted in the fatalities of 1089 people in the United States [5]. A majority of these deaths were comprised of motorists. Therefore, urban planners such as traffic decision makers are tasked with proactively deploying resources that minimize motorist risk exposure. At present, traffic decision makers rely on static flash flood inundation profiles related to discrete rainfall events. These profiles are often created through multiagency cooperation efforts such as [6]. Some studies have begun to generate dynamic flood inundation data visualizations based on these profiles [7].

Additionally, integrated approaches that use machine learning and geographic information systems (GIS) to track changes in critical infrastructure over time are emerging as powerful decision support tools [8]. However, there is limited use of state-of-the-art time series prediction models to generate dynamic data visualizations in a GIS setting for improved flood management. This book chapter explores the integration of publicly available data and machine learning models to address this gap in the literature.

Precise determination of when and where to deploy re-routing measures is a complex task. One approach that improves planning effectiveness is to integrate time series characteristics of river behavior and corresponding spatial flood profile. In this chapter, a univariate time series prediction of river stage is conducted that improves the temporal resolution and accuracy of publicly available forecasts. This prediction is then tied to a corresponding spatial flood inundation profile in a GIS setting. The resulting geospatial deep learning model provides a data visualization tool that traffic decision makers can use to proactively manage road closures in the event that a flood is likely to occur. The first section provides an overview of relevant river behavior that causes flooding. State-of-the-art trend extraction and prediction techniques are then presented and tied to geospatial use cases. The methodology section presents the data used, time series prediction model selected, and geoprocessing procedures required for data visualization using GIS software. Next, an illustrative example is provided for a frequently flooded intersection in Missouri. A discussion section is provided that positions the findings in the context of improving traffic management in the event of a flood. Lastly, a conclusion is given that summarizes the key findings and outlines model limitations and future work.

2. A geospatial deep learning approach

Two key characteristics of streams that relate to flood events are stream stage and streamflow. Stream stage refers to height (ft) of the stream and streamflow corresponds to discharge (ft^3/s) or alternatively, volumetric flowrate. Typically, governmental organization such as the United States Geological Survey maintain a network of sensors that monitor these characteristics over time for various stream segments. The National Weather Service classifies flood categories into four groups based on stream stage: Action Stage, Flood Stage, Moderate flood Stage, and Major Flood Stage [9]. These values vary for a given segment of stream based on analysis of previous floods, local topography, and underlying geological properties.

Given that stage is monitored over time, the use of time series forecasting methods to predict stage values is appropriate. There are two modeling approaches that are useful in this context: statistical and computational intelligence. Statistical models use historical data to identify underlying patterns to predict future values [10]. Some commonly used techniques for flood forecasting include simple exponential smoothing [11], autoregressive moving average [12], and autoregressive integrated moving average [13]. However, one shortcoming of these approaches is lack of scalability as the quantity and complexity of data increases [14]. An alternative approach that addresses these issues is computational intelligence. A key feature of computational intelligence approaches is the capacity to manage complexity and non-linearity without needing to understand underlying processes [15]. In summary, statistical methods rely on precise underlying relationships and exhibit decreased performance as the number of variables increases whereas computational intelligence approaches identify patterns using large amounts of training data to

establish a model capable of accurate predictions [16]. Some commonly used flood forecasting computational intelligence models include support vector machines [17], artificial neural networks [18], and deep learning [19]. Further, they have demonstrated superior performance when compared to conventional statistical modeling approaches for flood prediction studies. LSTM models have explicitly shown promising results in time series contexts. Therefore, LSTM models provide a state-of-the-art trend extraction and prediction technique regarding stream stage values.

Stream stage values are categorized based on resulting flood severity. The physical reality of these categories is the spatial extent of the flooding event often referred to as a flood inundation map [20]. These maps provide decision makers with a useful visual reference to determine what specifically has been affected by a flood event. An area of research, data visualization, and practical application that has not been fully investigated is the integration of computational intelligence stream stage predictions with geospatial flood inundation maps. The methodology provided in the following section addresses this gap.

3. Methodology

This section consists of three parts: LSTM prediction of stream stage, data required, and geoprocessing procedures. First, a brief overview of LSTM will be given. This will include explanatory figures and relevant mathematical formulas. Second, data required to conduct the LSTM prediction of stream stage will be procured. Flood inundation imagery and road network data will also be obtained. Lastly, data will be uploaded to a GIS software and processed for end use by traffic decision makers. An illustrative example is presented in the next section.

3.1 LSTM prediction of stream stage

Stream stage prediction is a time series forecasting procedure that is dependent on previous data to predict future values. As the quantity and quality of data continues to increase, more powerful computational approaches can be applied to prediction problems. The results of the literature review demonstrated that deep learning approaches, namely LSTM networks, are increasingly being applied to these problems.

Deep learning is an extension of the conventional neural network by adding additional layers and layer types. **Figure 1** provides a visual comparison of the two approaches [21]. The simple neural network (left) consists of a single input layer, hidden layer, and output layer. Alternatively, the deep learning neural network (right) has one input layer followed by three successive hidden layers that ultimately feed into a final output layer. This configuration has generated superior performance in capturing complex relationships.

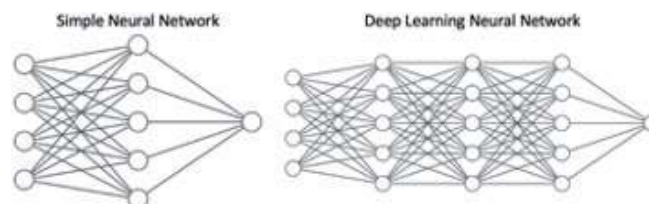


Figure 1.
Simple neural network vs. deep learning neural network.

However, neither approach retains previous time step information. Recurrent neural networks (RNNs) were introduced to address this limitation. LSTM networks are the deep learning variant of RNNs. All figures and mathematical formulation are borrowed from [15]. The primary benefit of LSTM networks is the capacity to retain longer term information. This is accomplished by removing and adding information determined by a series of ‘gates’ and vector operations. **Figure 2** provides a visual representation of an LSTM cell. The first gate, illustrated in yellow, generates a value between 0 and 1 using the current input (x_t) and output from the previous step (y_{t-1}) that determines how much information is passed on (forget gate). A zero corresponds to no information transfer whereas a one represents a complete transfer.

The result of this procedure (f_t) is presented mathematically in Eq. (1) as a sigmoid neural network layer where U (weights) and W (recurrent connections) are matrices.

$$f_t = \sigma(x_t U^f + y_{t-1} W^f) \quad (1)$$

Next, a decision must be made regarding what information needs to be stored. This is accomplished by applying an additional sigmoid layer (red, i_t). New values are then added to the cell state (\hat{C}_t) by using a tanh layer (green). Eqs. (2) and (3) present these procedures mathematically.

$$i_t = \sigma(x_t U^i + y_{t-1} W^i) \quad (2)$$

$$\hat{C}_t = \tanh(x_t U^g + y_{t-1} W^g) \quad (3)$$

The line at the top of the cell is known as the cell state (C_t) and has interactions with all components. Information has the opportunity of being forgotten when the old state (C_{t-1}) is multiplied by the result of the first forget gate (f_t). The product of the second (red) and third (green) gates are then added which results in new information being provided to the cell state and is represented by Eq. (4).

$$C_t = f_t C_{t-1} + i_t \hat{C}_t \quad (4)$$

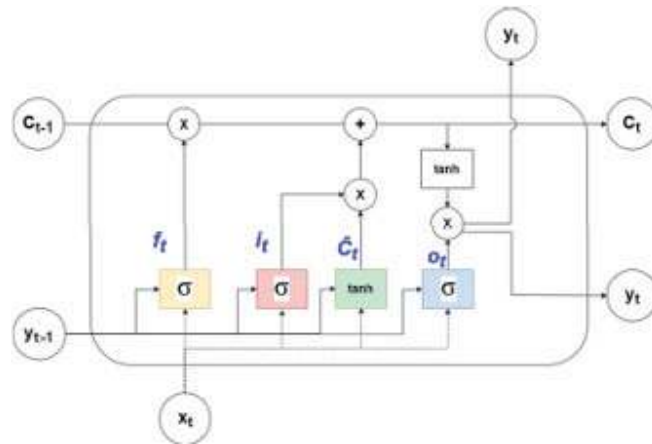


Figure 2.
LSTM network cell.

Lastly, the output layer of the LSTM cell determines the forecast for the current time step. A sigmoid layer (blue) and tanh layer are multiplied to generate an output (y_t). This final step is represented by Eqs. (5) and (6).

$$o_t = \sigma(x_t U^0 + y_{t-1} W^0) \quad (5)$$

$$y_t = \tanh(C_t) \times o_t \quad (6)$$

The result of this computational procedure is a time series forecast of future values. However, a large amount of data must be gathered to use as a model input. This data is presented in the next section.

3.2 Data required

Historic stream stage height for the location further explained in Section 4 must first be gathered. 113,994 data points were procured that correspond to 15-minute intervals from May 19, 2016 (5 PM) – September 1, 2019 (4 PM). Stage height is herein referred to as ‘gauge height’ to account for the source of the data. This data is represented graphically in **Figure 3** [22].

Using USGS’ flood inundation mapper (FIM), these gauge heights can be tied to a specific flood inundation profile [23]. The FIM is a publicly available tool that provides resulting flood inundation maps for one-foot gauge height increments in image format (.tif). A sliding bar that accomplishes this is available on the online user interface and is presented in **Figure 4**.

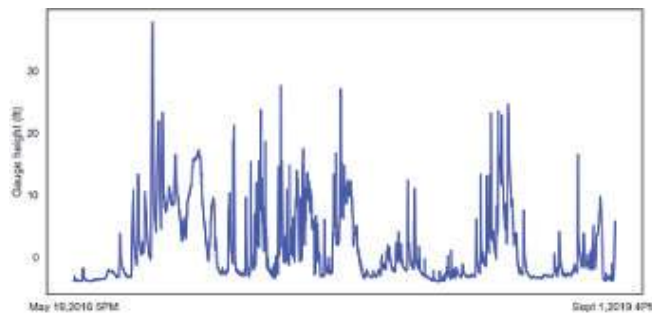


Figure 3.
Stream stage height for example locations.



Figure 4.
FIM sliding gauge height tool.

An example of a flash flood inundation profile being uploaded to a GIS software is provided in **Figure 5**. Purple lines correspond to road network data derived from the National Transportation Dataset [24]. Blue raster (grids of pixels) imagery denote the depth of water at discrete locations where darker blue reflects deeper water. Useful geoprocessing techniques that generate actionable decision support tools are presented in the next section.

3.3 Geoprocessing procedures

Traffic decisions makers are tasked with identifying flood affected road segments. In **Figure 5**, it can be observed that the flood inundation profile does overlap certain road segments. Relying on visual inspection alone is time consuming and prone to inaccuracies due to human error. A solution to this issue is the application of a set of straightforward geoprocessing tools that are built-in to most GIS softwares: conversion and intersection.

Some tools do not allow raster and vector data layer interoperability. Therefore, it is necessary to convert one of the data layers to establish a consistent data type. One approach is to convert the raster layer into a vector layer using the conversion tool within ArcGIS. **Figure 6** illustrates the result of this operation. The flood inundation profile has been converted into several points at 1-m increments. This spatial resolution can be modified by the user. The road network has been changed from its previous color to improve readability.

Once the raster layer has been converted into vector format, it is eligible for use as an input layer for the intersection tool. The intersection tool generates a point at



Figure 5.
Flood inundation profile example.

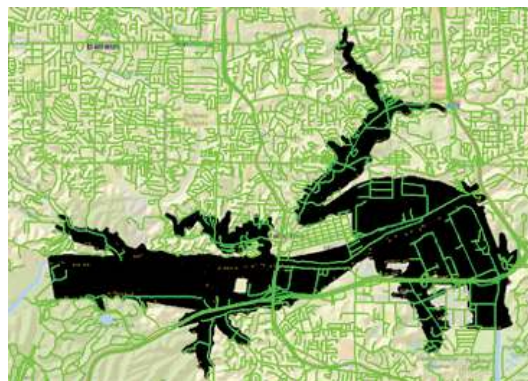


Figure 6.
Raster layer conversion example.

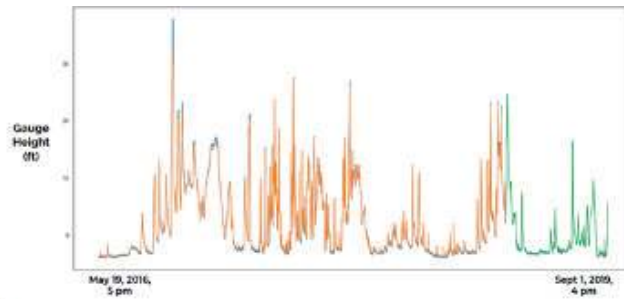


Figure 9.
LSTM training and testing results.

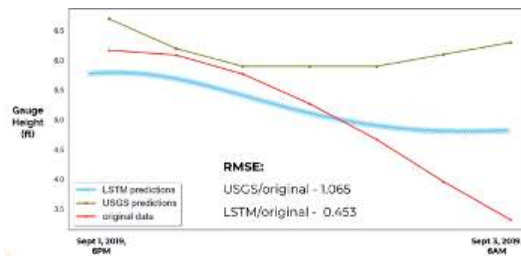


Figure 10.
USGS and LSTM prediction comparison.

the end of the dataset at less than 3.5 feet. The green line corresponds to the USGS prediction. This prediction initially overshoots the original data before briefly correcting and then diverging significantly from the observed trend. Lastly, the blue line represents the LSTM prediction. At first, this prediction captures the downward trend missed by the USGS prediction. Ultimately, the prediction flattens out and diverges from the original observations but to a lesser extent when compared to the USGS prediction. Root Mean Squared Error (RMSE) values for each of the predictions are provided to further demonstrate the difference in model performance. The RMSE value of 0.453 reported by the LSTM model represents superior accuracy compared to the 1.065 value reported by the USGS prediction. Therefore, the LSTM model presented here improves on the accuracy of publicly available forecasts and can be used as an input for the flood inundation tool.

Valley Park has 43 flood inundation profiles available in one-foot increments from 11–54 feet. The highest stage value recorded at this location is 44.11 feet on December 31, 2015. **Figure 11** provides the flood inundation profile for 45 feet to approximate this event. Note that 45 feet is used instead of 44. This is due to the



Figure 11.
Flood inundation profile for 45 ft. stage value for Valley Park, Missouri.



Figure 12.
Flood affected road segments for flood inundation profile corresponding to 45 ft. stage value for Valley Park, Missouri.

flood inundation profile incremental limitation and opting for a rounding approach that provides a more conservative risk assessment. The inundation profile is then converted to point format and intersected with the road network as illustrated by **Figure 12**.

5. Discussion

At present, urban planners such as traffic decision makers rely on static flood inundation maps and post hoc planning to reroute traffic in the event that a flood occurs. This approach puts motorists already in-transit at risk to rapidly changing road conditions. To address these risks, a field of research has emerged to provide decision makers with real-time decision-making tools. However, using time series prediction models that capture river characteristics and integrating them with flood inundation profiles has received limited attention. The methodology provided here addresses this gap.

Traffic decision makers can use the data visualization presented in **Figure 12** as a powerful decision support tool. The flood affected road segments can be easily identified (orange) and rerouting measures can be promptly dispatched. With the improved temporal resolution and accuracy of the LSTM prediction of stage height, traffic decision makers can deploy resources proactively to avoid unnecessary risk to motorists and improve traffic flow. Concluding remarks, limitations, and future work are presented in the next section.

6. Conclusion

Flash floods are a frequent and devastating natural disaster. The impetus to manage these events belongs to local decision makers that work in a resource constrained environment. To improve their decision-making effectiveness, a framework was presented that integrates machine learning and geospatial data to extract spatial and temporal trends using publicly available data. An illustrative example was provided to demonstrate the effectiveness of the framework provided. Valley Park, Missouri is located near the intersection I-44 and State Route 141. These roads represent major traffic throughputs and persistent flooding of the Meramec River has jeopardized the safety of motorists and the flow of commercial goods. Using 113, 994 river stage observations procured from a nearby sensor, an LSTM network was developed to improve the accuracy of publicly available forecasts. The result was an improvement in both the frequency and accuracy of forecasts provided.

Once the stage value is predicted it can be tied to a spatial flood inundation profile using the publicly available FIM. Using the flood inundation profile for 45 feet observed at Valley Park as a proxy for the historic crest at this location, data visualization of flood affected road segments was generated in a GIS setting. The key benefit of this output is the ease with which traffic decision makers can use the results presented to inform urban planning and decision making. Traffic decision makers can use the resulting data visualization presented here to guide real-time decision making in the event that a river stage value is predicted to reach a flood event stage for a specified river segment. Despite the usefulness of the findings, there remain a number of model limitations that represent areas of future work.

Model limitations can be divided into two categories: data gathering and model extension. Deep learning models are dependent on large amounts of data. Therefore, sensors that collect data need to be installed and active for an extended period. The cost to install and maintain an enlarged sensor network might be prohibitive for some locations. Due to this fact, model implementation is limited to river locations where sensors are already installed. Additionally, FIM coverage is confined to a small number of locations nationwide. Similarly, to sensor coverage, if there are not already-available flood inundation maps, then the model cannot be applied to those locations. Model extension includes options to improve the model in a material way. One recommendation would be to determine the best locations for road signage that will provide optimal re-routing to motorists given a finite amount of signage. Another approach would involve working with local decision makers to determine re-routing effectiveness based on how quickly resources are deployed given model predictions. Areas of future work not related to model extensions include alternative prediction approaches in river networks with no sensors and refinement of the model to account for flash floods. Each of these components represent considerable opportunity for model enrichment that further improve the decision-making effectiveness for traffic management professionals.

The results presented here demonstrate the utility of using machine learning models and geospatial data to generate data visualization tools that key stakeholders can use to improve planning effectiveness. As data becomes increasingly available, use of comparably sophisticated methods can be applied to a suite of natural disaster phenomena. The outcome of such an undertaking will be the widespread use of data visualization tools that will reduce the risk motorists are exposed to and mitigate the accompanying economic fallout.

Acknowledgements

This work was partially funded by the Missouri Department of Transportation, Award Number TR201912 and the Mid-America Transportation Center, Award Number 25-1121-0005-130.

Conflict of interest

The authors declare no conflict of interest.

Author details

Jacob Hale¹, Suzanna Long^{1*}, Vinayaka Gude² and Steven Corns¹

1 Department of Engineering Management and Systems Engineering,
Missouri University of Science and Technology, Rolla, MO, United States

2 Department of Arts and Media, Louisiana State University Shreveport,
Shreveport, LA, United States

*Address all correspondence to: longsuz@mst.edu

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] United States Geological Survey. Floods: Things to Know [internet]. 2019. Available from: https://www.usgs.gov/special-topic/water-science-school/science/floods-things-know?qt-science_center_objects=0#qt-science_center_objects [accessed 2019-03-15]
- [2] National Severe Storms Laboratory. Flood Basics [internet]. 2019. Available from: <https://www.nssl.noaa.gov/education/svrwx101/floods/> [accessed 2019-03-16]
- [3] World Health Organization. Floods [internet] Available from: https://www.who.int/health-topics/floods#tab=tab_1 [accessed 2019-05-20]
- [4] National Oceanic and Atmospheric Administration. 2010-2019: A landmark decade of U.S. billion-dollar weather and climate disasters [internet]. 2020. Available from: <https://www.climate.gov/news-features/blogs/beyond-data/2010-2019-landmark-decade-us-billion-dollar-weather-and-climate> [accessed 2020-10-15]
- [5] National Weather Service. NWS Preliminary US Flood Fatality Statistics [internet]. 2020. Available from: <https://www.weather.gov/arx/usflood> [accessed 2020-10-15]
- [6] Dietsch, B.J., and Strauch, K.R., 2019, Flood-inundation maps of the Meramec River from Eureka to Arnold, Missouri, 2018: U.S. Geological Survey Scientific Investigations Report 2019-5004, 12 p. Available from: <https://doi.org/10.3133/sir20195004>.
- [7] Jha MK, Afreen S. Flooding urban landscapes: Analysis using combined hydrodynamic and hydrologic modeling approaches. *Water (Switzerland)*. 2020;12(7).
- [8] Shirowzhan S, Sepasgozar SME, Li H, Trinder J, Tang P. Comparative analysis of machine learning and point-based algorithms for detecting 3D changes in buildings over time using bi-temporal lidar data. *Automation in Construction*. 2019Sep;105. Available from: <https://doi.org/10.1016/j.autcon.2019.102841>
- [9] Advanced Hydrological Prediction Service. Meramec River at Valley Park [internet]. 2019. Available from: <https://water.weather.gov/ahps2/hydrograph.php?wfo=lsx&gage=vllm7> [accessed 2019-06-02]
- [10] Lago J, De Ridder F, De Schutter B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Appl Energy* [Internet]. 2018;221(February):386-405. Available from: <https://doi.org/10.1016/j.apenergy.2018.02.069>
- [11] Papacharalampous G, Tyrallis H. Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. *J Hydrol* [Internet]. 2020;590(May):125205. Available from: <https://doi.org/10.1016/j.jhydrol.2020.125205>
- [12] Liu J, Wang J, Pan S, Tang K, Li C, Han D. A real-time flood forecasting system with dual updating of the NWP rainfall and the river flow. *Nat Hazards*. 2015;77(2):1161-82.
- [13] Adnan, Muhammad, X, Yuan, O, Kisi et al. Application of soft computing models in streamflow forecasting. *Proceedings of the Institution of Civil Engineers – Water Management*. 2019;172(3):123-124. Available from: <https://doi.org/10.1680/jwama.16.00075>
- [14] Mosavi A, Ozturk P, Chau KW. Flood prediction using machine learning models: Literature review. *Water (Switzerland)*. 2018;10(11):1-40.

- [15] Gude V, Corns S, Long S. Flood Prediction and Uncertainty Estimation Using Deep Learning. *Water* (Switzerland). 2020;12(3). from: https://waterdata.usgs.gov/nwis/uv?site_no=07019130[accessed 2019-10-16] usgs 2019
- [16] Bzdok D, Altman N, Krzywinski M. Points of Significance: Statistics versus machine learning. *Nat Methods* [Internet]. 2018;15(4):233-4. Available from: <http://dx.doi.org/10.1038/nmeth.4642>
- [17] Tehrany MS, Pradhan B, Jebur MN. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J Hydrol* [Internet]. 2014;512:332-43. Available from: <http://dx.doi.org/10.1016/j.jhydrol.2014.03.008>
- [18] Shafizadeh-Moghadam H, Valavi R, Shahabi H, Chapi K, Shirzadi A. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. *J Environ Manage.* 2018;217:1-11.
- [19] Tien Bui D, Hoang ND, Martínez-Álvarez F, Ngo PTT, Hoa PV, Pham TD, et al. A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Sci Total Environ* [Internet]. 2020;701:134413. Available from: <https://doi.org/10.1016/j.scitotenv.2019.134413>
- [20] Papaioannou G, Vasiliades L, Loukas A. Multi-Criteria Analysis Framework for Potential Flood Prone Areas Mapping. *Water Resour Manag.* 2015;29(2):399-418.
- [21] Publication-ready NN-architecture schematics [internet]. 2021. Available from: alexlenail.me/NN-SVG/index.html [accessed 2021-1-27]
- [22] United States Geological Survey. USGS 07019130 Meramec River at Valley Park, MO [internet]. 2019. Available from: https://waterdata.usgs.gov/nwis/uv?site_no=07019130[accessed 2019-10-16]
- [23] United States Geological Survey. Flood Inundation Mapper [internet]. 2019. Available from: <https://fim.wim.usgs.gov/fim/>[accessed 2019-10-18]
- [24] United States Geological Survey. USGS National Transportation Dataset (NTD) Downloadable Data Collection [internet]. 2019. Available from: <https://catalog.data.gov/dataset/usgs-national-transportation-dataset-ntd-downloadable-data-collectionde7d2>[accessed 2019-10-22]
- [25] KMOV4. Photos: Before & After Meramec River flooding [internet]. 2017. Available from: https://www.kmov.com/news/photos-before-after-meramec-river-flooding/article_fc16115e-12e2-54e6-bf84-ba1732a4dcdbd.html [accessed 2019-10-25]