AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023018

A Statistical Approach to Mapping Flood Susceptibility in the Lower Connecticut River Valley Region

Jason Giovannettone[1], Tom Copenhaver[2], Margot Burns[3], and Scott Choquette[4]

[1]Dewberry, Fairfax, VA, USA, [2]Dewberry, Denver, CO, USA, [3]RiverCOG, Essex, CT, USA, [4]Dewberry, New Haven, CT, USA

Key Points:
- Elevation, distance to water, and surficial materials had the highest contributions to flood susceptibility throughout the study area
- The contribution of elevation and land use to flood susceptibility increased substantially when comparing the urban to the rural subregion
- Very high and high susceptible areas add over 6% of nonwater and wetland area to the SFHA, including 8% more developed area

Correspondence to:
J. Giovannettone,
jgiovannettone@dewberry.com

Citation:
Giovannettone, J., Copenhaver, T., Burns, M., & Choquette, S. (2018). A statistical approach to mapping flood susceptibility in the Lower Connecticut River Valley Region. *Water Resources Research*, *54*, 7603–7618. https://doi.org/10.1029/2018WR023018

Received 26 MAR 2018
Accepted 17 AUG 2018
Accepted article online 24 SEP 2018
Published online 12 OCT 2018

Abstract Flood susceptibility in the Lower Connecticut River Valley Region attributable to nonclimatic flood risk factors is mapped using a quantitative method using logistic regression. Flood risk factors considered include elevation, slope, curvature (concave, convex, or flat), distance to water, land cover, vegetative density, surficial materials, soil drainage, and impervious surface. Values of factors at point locations were correlated to whether a location was located within or outside of the U.S. Federal Emergency Management Agency 100-year Special Flood Hazard Area (SFHA). The Lower Connecticut River Valley Region was divided into urban, rural, and coastal subregions to assess the differences in factor contributions to flood susceptibility between different region types; for each region flood risk factors were extracted from 4,000 points, of which an equal number were within or outside of the 100-year SFHA. Logistic regression coefficients were obtained. It was found that *elevation* and *distance to water* have the greatest contribution to flood susceptibility in the urban and coastal subregions, whereas distance to water and *surficial materials* dominate in the rural subregion. The contribution of *land use* to flood susceptibility increased by over 200% between the rural and urban regions. Probabilities of flooding were computed using each regional logistic regression equation. Several areas classified as *very high risk* (80–100%) and *high risk* (60–80%) were located outside of the SFHA and included several types of infrastructure critical for human health, safety, and education. This study demonstrates the utility of logistic regression as an efficient methodology to map regional flood susceptibility.

Plain Language Summary Flooding is one of the most severe and potentially devastating natural disasters that can occur. Floods can come in many forms, including river, coastal, and flash flooding. Whenever and wherever any of these types of flooding occur, long-term planning and adaptation, preparedness, and response time are all critical factors in reducing the overall impacts. Awareness of areas that are currently prone and will remain prone to flooding in the future is essential to consider in both short-term and long-term planning. Such awareness comes from an understanding of a combination not only of regional climatic factors but also of nonclimate factors that relate to natural, physical, and development characteristics. The current study estimates the risk of flooding throughout the Lower Connecticut River Valley Region (LCRVR) based on site and regional characteristics not related to climate. Several methods were considered to estimate flood risk; the method that was finally selected for this study involves a statistical approach in which a data set having one or more independent variables that produce a binary value of no or yes (0 or 1, respectively) for the dependent variable is analyzed. The independent variables in this case include several nonclimate factors related to flood risk that could potentially affect the region and for which sufficient data were available and are referred to as flood risk factors. Flood risk factors considered include elevation, land slope, land curvature (concave, convex, or flat), distance to water body, land cover, density of vegetation, surface geology, ability of the soil to drain water, and the percent of impervious surface (e.g., pavement). The objective is to link each of the flood risk factors to the dependent variables, which in this case is the occurrence of flooding for a flood event that is estimated to occur on average once in every 100 years. It was found that the overall quality of recent satellite images of the LCRVR during large flood events was not sufficient for the current analysis; therefore, it was decided to use the U.S. Federal Emergency Management Agency 100-year Special Flood Hazard Area (SFHA) to indicate areas where flood inundation would occur. The advantage of using the SFHA and the selected statistical modeling methodology is that they allow the contribution of each flood risk factor within the SFHA to be estimated and then applied to the entire study region to identify additional areas outside of the SFHA that have high flood risk. The LCRVR was divided into three subregions (urban, rural, and coastal) to accentuate the differences in the contributions of each flood risk factor to flood risk between an urban and a rural area and between inland and coastal

areas; for each subregion 4,000 point locations were randomly chosen from which to extract data for each flood risk factor. An equal number of these points were selected in locations that were within and outside of the SFHA for each subregion. Site data for each flood risk factor were extracted and associated with a 1 if the location was within the SFHA and a 0 otherwise. The resulting relations between each flood risk factor and flood occurrence were analyzed so that regression coefficients could be estimated for each factor, the magnitude of which indicates the relative strength of each flood risk factor's influence on flooding in a subregion. It was found that *elevation* and *distance to water* have the most influence on flood risk in the urban and coastal subregions, whereas distance to water and *surface geology* dominate in the rural subregion. The contribution of elevation and *land use* were also found to increase the most between the rural and urban subregions. The coefficients for each subregion are then used to assign probabilities of flooding to all locations over a grid covering that subregion. The results for each subregion were combined to create an overall flood probability map of the LCRVR. Probabilities were classified *very low risk* (0–20%), *low risk* (20–40%), *medium risk* (40–60%), *high risk* (60–80%), and *very high risk* (80–100%). It was observed that several areas classified as *very high risk* and *high risk* were located outside of the SFHA. Several types of infrastructure critical for human health, safety, and education were finally overlaid on the flood risk map to identify those assets that are most vulnerable to the 100-year flood and may therefore require additional flood risk mitigation.

1. Introduction

Flooding is one of the most severe and potentially devastating natural disasters. Flooding occurs in many forms, including river, coastal, and flash flooding, and arises from a variety of processes such as snow melt, severe precipitation events, storm surge, and on a more long-term scale, sea level rise. Whenever any of these types of flooding occur, long-term planning and adaptation, preparedness, and response time are all critical factors in reducing the overall impacts. The severity of flooding has increased over the last several decades in the northeast and throughout the Mississippi and Ohio River valleys (Peterson et al., 2013) because of a combination of factors related to the development of urban areas along rivers and coasts and potentially climate change, which have contributed to the total cost of flood damage escalating as well (Doocy et al., 2013). Awareness of areas that will be more prone to flooding because of these changes is essential to consider in long-term planning, whereas it can also inform short-term strategies, such as the development of early warning mechanisms (Li et al., 2018; Lopez et al., 2017; Rahman et al., 2013). Such awareness comes from an understanding of a combination not only of climatic factors impacting the region but also of nonclimate factors (e.g., urbanization) that relate to regional and site characteristics as well (Mahmoud & Gan, 2018; Miller & Hutchins, 2017; Zhu et al., 2007).

Various types of hydrological models can be used to model flood susceptibility (Devi et al., 2015) and can be categorized as physically based (Abbott et al., 1986; Gassman et al., 2007), conceptual (Crawford & Linsley, 1966), or data-driven (Gogoi & Chetia, 2011; Kia et al., 2012; Lee et al., 2012; Matori et al., 2014; Siddayao et al., 2014; Ullah & Choudhury, 2013) models. Physically based models rely on an understanding of complex physical processes and represent a mathematically idealized form of the real thing. These models use variables that are functions of both space and time and are measurable. Finite difference equations are used to model the hydrological processes associated with the movement of water. Even though physically based models do not require a large amount of hydrological and meteorological data for calibration, a large number of parameters are required to describe the physical characteristics of the catchment being modeled, including soil moisture, water depth, topography, and river network dimensions. Physically based models are versatile and have the advantage of using parameters that have a physical interpretation, but much time and resources are required to develop such models. There are a myriad of examples of physically based models, two of which include the Soil and Water Assessment Tool (Gassman et al., 2007) and the MIKE Systeme Hydrologique European model (Abbott et al., 1986).

Conceptual models are similar to physically based models in that they attempt to describe all of the component hydrological processes, albeit in a more simplified and less physical process manner. They are composed of interconnected reservoirs that are recharged by sources such as infiltration, percolation, and rainfall and emptied by runoff, evaporation, and drainage, and other types of sinks. The parameters that make up a conceptual model are assessed by analysis of field data and calibration. Unlike physically based model,

conceptual models require an extensive amount of meteorological and hydrological data for calibration, in addition to sophisticated analysis tools, which is not within the scope of the current project. One of the first conceptual models developed was the Stanford Watershed Model IV by Crawford and Linsley (1966).

In contrast to physically based and conceptual models, data-driven or empirical models rely completely on observations and an understanding of the hydrological and meteorological variables and regional characteristics that influence flood susceptibility with no consideration given to the physics of meteorological or hydrological processes. Many types of data-driven models use linguistic variables whose values include words or phrases, rather than the conventional numerical variables used in the models described above. Examples of linguistic data-driven models used for hydrological modeling purposes include (1) fuzzy logic (FL; Gogoi & Chetia, 2011; Hundecha et al., 2001; Sen & Altunkaynak, 2004), (2) artificial neural networks (ANN; Dawson & Wilby, 2001; Kia et al., 2012; Kovacevic et al., 2018), (3) Adaptive Neuro-Fuzzy Interface System (ANFIS; Ullah & Choudhury, 2013; Yaseen et al., 2018; Zounemat-Kermani & Teshnehlab, 2008), and (4) analytical hierarchy process (AHP; Matori et al., 2014; Richardson & Amankwatia, 2018; Siddayao et al., 2014).

The objective in most data-driven models is to produce a list of relative weights for whatever variables and local characteristics have been identified as affecting flood susceptibility; these weights can then be used to produce a flood susceptibility map. The method used to derive these weights represents the major difference between the various forms of data-driven models.

The first type of linguistic data-driven model is FL and is set up using membership functions and rules for factors related to flood susceptibility, hereafter referred to as flood risk factors. A membership function for each factor incorporates various classifications (e.g., high, medium, and low) of that factor. After the variables are partitioned into their different *fuzzy* classes, an IF … THEN type of rule is set up to establish the response of any combination of these fuzzy classes. For example, Gogoi and Chetia (2011) used a fuzzy rule-based model to forecast runoff in the Jiadhal Basin in Northeast India. The authors used three flood risk factors (total monthly rainfall, mean monthly temperature, and previous month's discharge) and three categories (e.g., high, medium, and low) to describe projected runoff, resulting in a total number of $3^3 = 27$ rules. Sets of values for each variable were then tested against these rules to identify rules that are fulfilled to a point that exceeds a certain threshold value. The identified rules are then used to project runoff based on values of the identified flood risk factors.

The second type of data-driven model is the ANN. ANNs consist of layers of nodes or neurons, which include an input layer (number of neurons equals the number of flood causative factors), an output layer (number of neurons equals the number of types of desired outputs), and one or more hidden layers where algorithms are used to model the complex relations that are expected to exist between each flood risk factor and the influence that they have on the output. In the context of flooding, outputs would be water levels and/or flow. Kia et al. (2012) used ANN to predict water levels and flood inundation using seven potential flood risk factors: rainfall, slope, elevation, flow accumulation, soil, land use/cover, and geology.

Alternatively, the third linguistic model type is the ANFIS, which uses a combination of the numeric power of neural networks and the verbal power of FL. Such a model contains features of both types of models such as learning and optimization abilities and IF … THEN rule thinking to map an input space to an output space. An example of this method was developed for the Barak River basin in Northeast India by Ullah and Choudhury (2013). Issues with using an ANN, ANFIS, or any other method that incorporates neural networks relate to their complexity and the substantial computing power that is required to run the networks. The quality of the resulting predictions in many cases has also been found to be inferior to other model types (Shortridge et al., 2016) and especially so when the data that are used to validate the model contain values outside of the range of data used to train the model.

The final type of linguistic data-driven model is the AHP. An AHP identifies potential flood risk factors, and their associated weights using expert opinions combined with geographical, statistical, and historical data. For example, Matori et al. (2014) and Siddayao et al. (2014) used an AHP in performing spatial assessments of flood susceptibility in northern Malaysia and the northern Philippines, respectively. Flood risk factors included rainfall, geology, soil type, land use, population density, distance from river bank, and site elevation and slope. The authors in both studies consulted with experts in their study areas and used the survey results to develop weights for each factor. They then combined the resulting weights with a Geographical

Information System to produce a color-coded map representing various levels of risk for each respective study region. The advantage of this method is that the final product is a flood susceptibility map based on the combined experience of several years of flooding events from various type of experts who are familiar with the region. The disadvantage is that the results can be based on subjective and conflicting opinions, especially when there are many flood risk factors being considered. This can be mitigated, however, when using the overall factor weighting mechanisms that are typically used in an AHP.

In contrast to the linguistic models, statistically based data-driven models use mathematical equations that are derived from concurrent input and output data (e.g., unit hydrograph). Regression and correlation models are two examples that attempt to find the functional relationship between the input and output time series. Other more quantitative types of data-driven models include multivariate statistical analysis (Allaire et al., 2015; Sharma et al., 2015; Singh et al., 2009; Wallis, 1965) and multivariate logistic regression (MLR; Park et al., 2017; Pradhan & Lee, 2010; Tehrany et al., 2014), or some combination of these. These methods rely on numerical expressions that characterize the relationships between the independent flood risk factors and flood inundation (Lee et al., 2012). The use of multivariate statistical analysis typically requires several strict assumptions to be made prior to the analysis and requires the relation between flooding and each flood risk factor to be considered independently from any potential relations between factors to develop weights for each factor. MLR can be used to solve this issue by examining the relations between a dependent variable (e.g., whether a location is flooded or not flooded) and any number of independent variables (e.g., flood risk factors; Pradhan & Lee, 2010). An advantage of MLR is that a separate analysis is not required to estimate the weight of each flood risk factor as this functionality is already built into such coding environments as R (R Development Core Team, 2018). Another advantage of MLR is that the variables can be continuous and/or categorical and is straightforward to implement.

Though somewhat ad hoc, after considering all of the advantages and disadvantages of the three major types of models described above (physically based, conceptual, and data driven) and due to the fact that one of the major objectives of the current study was to develop an accurate flood susceptibility mapping methodology that requires little resources in terms of time and money and can be applied not only to the study region used in the current study but also on a larger scale, it was decided to use a data-driven model of the Lower Connecticut River for the current project. In addition, it was decided to use MLR over the other types of data-driven models because of the fact that sufficient data were already available for a number of potential flood risk factors throughout the Lower Connecticut River; therefore, a quantitative relationship between these risk factors and flood inundation, which would provide more accuracy than the linguistic models, would be possible without expending significant additional resources in obtaining the required data. For these reasons, MLR was selected to model flood susceptibility for the current study.

2. Data and Methods

The Lower Connecticut River Valley Region (LCRVR) is located in the southeastern central part of the state of Connecticut and is focused around the confluence of the Connecticut River and Long Island Sound (Figure 1). Whereas the Connecticut River is tidally influenced throughout the study region, there are many smaller rivers and tributaries where the flood threat is primarily driven by local fluvial flooding. This region is also extremely heterogeneous in terms of the various land characteristics that can influence flood susceptibility. For these reasons, and the fact that the state of Connecticut hosts a large and relatively complete database of land and water characteristics throughout the state, the LCRVR was selected as the study region for the current study.

Even though the methodology used to develop the flood susceptibility map of the LCRVR is based on the method used in Tehrany et al. (2014), there are features of this work that differentiate it from previous studies. These studies, for example, all took place outside of the United States and involved land areas substantially smaller than the LCRVR. Because of the small size of each study region, these studies assumed that the study regions were homogeneous in terms of the influence of various regional characteristics on flood susceptibility. In contrast, the LCRVR is the first region within the United States for which the methodology described here has been used and is sufficiently large spatially that the assumption of homogeneity across the study area is less valid than it was in the international studies. The current study, therefore, includes different types of *subregions* (e.g., coastal, rural, and urban) for which separate flood

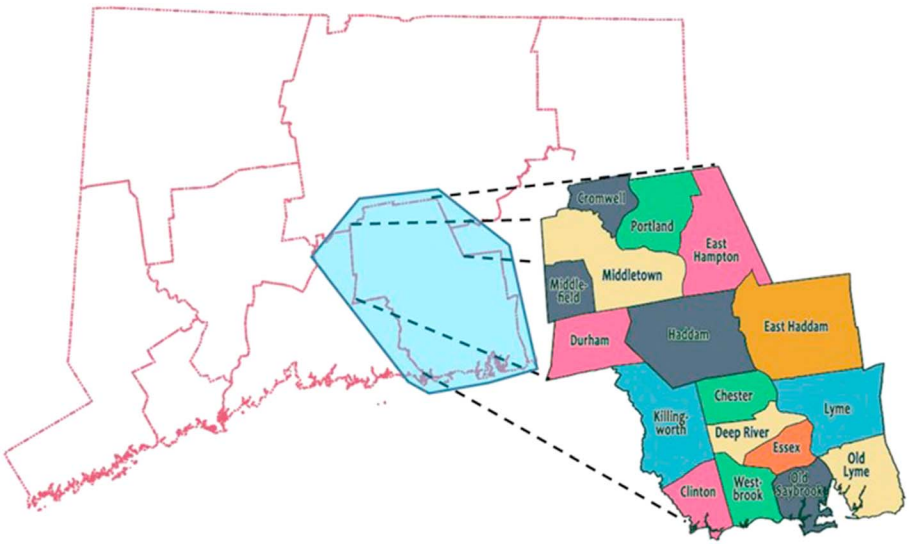

Figure 1. Map showing the location of the Lower Connecticut River Valley Region and the *area of influence* (shaded blue) within the state of Connecticut.

susceptibility analyses are performed and between which comparisons can be made on the influence of subregional characteristics (e.g., land use).

2.1. Flood Risk Factors

There are several types of nonclimatic data that are required as independent variables when using MLR to estimate flood susceptibility; these independent data represent parameters that may contribute to flooding in a region and are referred to as flood risk factors. Flood risk factors that are used for flood susceptibility mapping should be measurable and collected throughout the entire study region but should not represent information that is spatially uniform. Several risk factors may be prominent in one region but not in another; for example, the influence of flood factors will vary when comparing inland versus coastal regions or rural versus urban regions. In general, there is no agreement on which flood risk factors are the standard for any flood susceptibility analysis; however, there are factors that are more prominently used than others. Some of the most common factors are listed in Table 1 along with the citations for a few of the studies in which they were identified as influential. A subset of these flood risk factors was chosen for the present study after considering the availability, period of record, and completeness of each data set as applied to the study region: elevation, slope, land curvature, land cover, distance to water body, vegetation density, percent impervious surface, soil drainage class, and surficial materials. Several of these flood risk factors are related to each other so that some correlation is to be expected. Such correlation is common when a study is performed using MLR because the final objective is to develop a logistic regression that includes all factors that are expected to contribute to flooding and for which sufficient data are available. A potential issue occurs if detailed comparisons are made between the contributions of each flood risk factor; any correlation needs to be teased out if such comparisons are going to be made. Because the main objective of the current study is to provide a logistic regression equation that can be applied to the entire region, in addition to making some simple comparisons or observations related to each flood risk factor's contribution, no attempt was made to estimate these potential correlations.

Sources of flood risk factors for the LCRVR include the U.S. Geological Survey (USGS), the Connecticut Department of Energy and Environmental Protection (DEEP), the U.S. Department of Agriculture-National Resources Conservation Service, and the Federal Emergency Management Agency (FEMA). Abbreviations, sources, and the resolution/scale of each data set are given in Table 2.

All flood risk factor data were collected over the entire study region and compiled into spatial databases using the ArcGIS 10.2 software (Environmental Systems Research Institute, 2014). Flood risk factors *slope* and *curvature* were derived from the elevation data set, whereas the *distance to water* risk factor was computed as the minimum distance as the crow flies between each cell and the nearest water body as

Table 1
Flood Risk Factors and Examples of Studies in Which Each Has Been Considered

Flood risk factors	Literature
Temperature	Gogoi and Chetia (2011)
Previous month's discharge	Gogoi and Chetia (2011)
Population density	Siddayao et al. (2014), Sinha et al. (2008), and Zhang et al. (2005)
Distance from riverbank	Siddayao et al. (2014)
Landform: slope/elevation/curvature	Matori et al. (2014), Siddayao et al. (2014), Tehrany et al. (2014), Lawal et al. (2012), Saini and Kaushik (2012), Sinha et al. (2008), and Zhang et al. (2005)
Distance from access road	Qureshi and Harrison (2003)
Land-use zoning	Lawal et al. (2012) and Qureshi and Harrison (2003)
Drainage density	Lawal et al. (2012) and Saini and Kaushik (2012)
Proximity to drainage	Sinha et al. (2008)
Soil type/drainage	Matori et al. (2014), Tehrany et al. (2014), Lawal et al. (2012), Saini and Kaushik (2012), and Yahaya et al. (2010)
Distance from urban areas	Qureshi and Harrison (2003)
Precipitation/rainfall	Matori et al. (2014), Tehrany et al. (2014), Lawal et al. (2012), Gogoi and Chetia (2011), Yahaya et al. (2010), Zhang et al. (2005), and Qureshi and Harrison (2003)
Land cover/use and vegetation	Matori et al. (2014), Tehrany et al. (2014), Saini and Kaushik (2012), and Yahaya et al. (2010)
Geology	Matori et al. (2014) and Tehrany et al. (2014)
Timber type/size/density	Tehrany et al. (2014)

depicted on the USGS 7.5-min topographic quadrangle maps for the state of Connecticut (DEEP, 2005). All data sets were resampled using linear interpolation to a 30-m × 30-m grid comprised of 2,142 columns (north and south) and 1,957 rows (east and west) for a total of roughly 4.2 million points.

Prior to using each data set in the flood susceptibility analysis, each numerical flood risk factor was divided into classes. This is accomplished using the quantile method (Papadopoulou-Vrynioti et al., 2013; Tehrany et al., 2014; Umar et al., 2014), which partitions each numerical data set (e.g., elevation [0.0–277.5 m], slope [0.0–120.7°], vegetation density [0.0–93.0%], distance to water body [0.0–2,352.7 m], and percent impervious service [0.0–96.1%]) into classes containing the same number of features or pixels per class; partitioning the data in this manner ensures that data are included and that a regression coefficient can be determined for each flood risk factor class. For the purposes of this study, each of the numerical flood risk factor data sets

Table 2
Flood Risk Factors and Flood Event Data With Data Source and Resolution/Scale

Flood risk factors	Source (year)	Resolution/ scale	URL for data access
Land cover (LAND)	USGS (2011)	30 m	https://www.mrlc.gov/
Elevation (ELEV); slope (SLOPE); curvature (CURV)	USGS (2014)	30 m	https://earthexplorer.usgs.gov/
Distance from water (DIST)	DEEP (2005)	1:24,000	https://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707
Soil drainage (SOIL)	USDA-NRCS (2017)	varies	https://sdmdataaccess.nrcs.usda.gov/
Vegetation density (VEG)	USGS (2011)	30 m	https://www.mrlc.gov/
Impervious surface (IMP)	USGS (2011)	30 m	https://www.mrlc.gov/
Surficial materials (GEO)	DEEP (2005)	1:24,000	https://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707
FEMA 100-year NFHL	FEMA (2016)	1:12,000	https://fema.maps.arcgis.com/home/index.html

Note. USGS = U.S. Geological Survey; DEEP = Connecticut Department of Energy and Environmental Protection; USDA-NRCS = U.S. Department of Agriculture-National Resources Conservation Service; FEMA = Federal Emergency Management Agency; NFHL = National Flood Hazard Layer.

Table 3
Regression Coefficients for Each Risk Factor Class

Factor	Class	Logistic coefficient (C/R/U)	Factor	Class	Logistic coefficient (C/R/U)
a_0	—	5.18/5.06/20.24	DIST (m)	0.00–39.21	—/—/—
ELEV (m)	−2.65–2.84	—/—/—		39.22–117.64	−1.19/−2.16/−1.60
	2.85–20.42	−4.11/−2.17/−14.87		117.65–196.06	−2.01/−3.32/−2.64
	20.43–40.19	−20.48/−1.71/−15.70		196.07–274.48	−2.89/−3.63/−2.59
	40.20–56.67	−18.79/−1.59/−16.27		274.49–392.12	−3.00/−3.99/−3.20
	56.68–75.35	—/−1.40/−16.41		392.13–509.75	−4.63/−4.75/−3.57
	75.36–92.93	—/−1.54/−16.60		509.76–627.39	−4.45/−5.03/−3.87
	92.94–109.40	—/−2.22/−17.26		627.40–784.24	−5.61/−4.89/−4.07
	109.41–128.08	—/−2.53/−18.24		784.25–1,019.51	−19.61/−4.60/−3.91
	128.09–152.25	—/−2.84/−17.52		1,019.52–2,352.71	−17.33/−3.92/−2.68
	152.26–277.50	—/−3.72/−18.00	SOIL	not rated	—/—/—
CURV	Convex (−6.05 – −0.66)	—/—/—		excessively drained	−0.28/0.16/−2.24
	Flat (−0.65–0.65)	0.22/0.07/−0.46		somewhat excessively	−0.19/−0.53/−1.57
	Concave (0.66–6.05)	−0.89/1.79/0.99		well drained	−0.18/0.05/−1.43
SLOPE	0.00–0.47	—/—/—		moderately well	0.03/0.70/−1.33
	0.48–1.89	−0.29/−0.08/−0.10		somewhat poorly	—/2.52/0.30
	1.90–3.31	−0.11/−0.01/−0.41		poorly drained	1.02/1.48/−0.65
	3.32–4.73	−0.40/−0.62/−0.85		very poorly drained	0.60/1.02/0.68
	4.74–6.62	−0.97/−0.57/−1.06	IMP (%)	0.00–0.00	—/—/—
	6.63–8.52	−1.25/−0.92/−1.42		0.01–1.96	−0.89/−1.51/−0.27
	8.53–10.88	−0.79/−0.82/−1.37		1.97–4.70	0.02/−0.21/−0.20
	10.89–14.20	−0.88/−1.39/−2.65		4.71–10.98	−0.19/−0.27/−0.32
	14.21–19.40	−1.29/−1.14/−2.17		10.99–18.82	−0.28/−1.14/−0.34
	19.41–120.72	−0.70/−2.02/−2.40		18.83–28.62	−0.34/−0.44/−0.03
VEG (%)	0.00–0.00	—/—/—		28.63–38.82	−0.21/−0.23/−0.39
	0.01–32.00	−0.20/0.20/0.12		38.83–49.80	0.06/−0.07/−0.57
	32.01–55.00	−0.11/0.29/0.37		49.81–63.92	0.16/−1.32/−1.22
	55.01–70.00	−0.42/−0.34/0.41		63.93–99.61	−0.42/−0.31/−0.71
	70.01–80.00	0.00/0.35/0.32	GEO	thin till	—/—/—
	80.01–86.00	−0.57/0.15/0.77		sand/gravel/talus	0.90/0.89/0.82
	86.01–88.00	−1.07/0.67/0.86		fines	—/1.77/1.05
	88.01–89.00	−1.04/0.42/0.83		floodplain alluvium	16.31/3.11/2.91
	89.01–90.00	−1.26/−0.27/0.33		swamp deposits	0.08/1.37/1.41
	90.01–93.00	−1.93/−0.31/−0.18		thick till	−0.58/−2.03/−0.73
LAND	developed, open space	—/—/—		End Moraine deposits	0.08/−1.81/—
	dev., low intensity	−0.08/−0.04/−0.23		artificial fill	17.30/14.71/2.07
	dev., med.-high intensity	−0.34/0.04/−0.34		salt/tidal marsh deposits	1.18/13.38/—
	barren (rock/sand/clay)	0.94/−1.16/−16.55		beach deposits	2.39/—/—
	forest	0.00/−0.65/−0.95			
	shrub/scrub	−1.89/−1.77/−1.03			
	grassland/herbaceous	−0.20/−0.86/−0.69			
	pasture/hay	−0.10/−1.24/−0.38			
	cultivated crops	1.22/−0.47/−0.93			
	wetlands (woody/emer.)	0.05/0.35/−0.03			

Note. C = coastal; R = rural; U = urban.

was divided into 10 categories using the classifications given in Table 3; examples of the spatial distribution of two numerical flood risk factors are shown in Figures 2a and 2b for *elevation* and distance to water, respectively. Regarding the other flood risk factor data sets, land curvature was divided into three classes of concavity (not shown); *land cover* was divided into 10 classes (Figure 2c); *soil drainage* was divided into eight classes (not shown); and *surficial materials* was divided into 10 classes (Figure 2d).

2.2. Flood Inundation

The overall objective is to develop relations between flooding and all dependent flood risk factors. Therefore, a method is required to compare the values of each factor at a point with whether flooding would be expected or not expected to occur at that point for a specific flood (annual) return period. Because of the

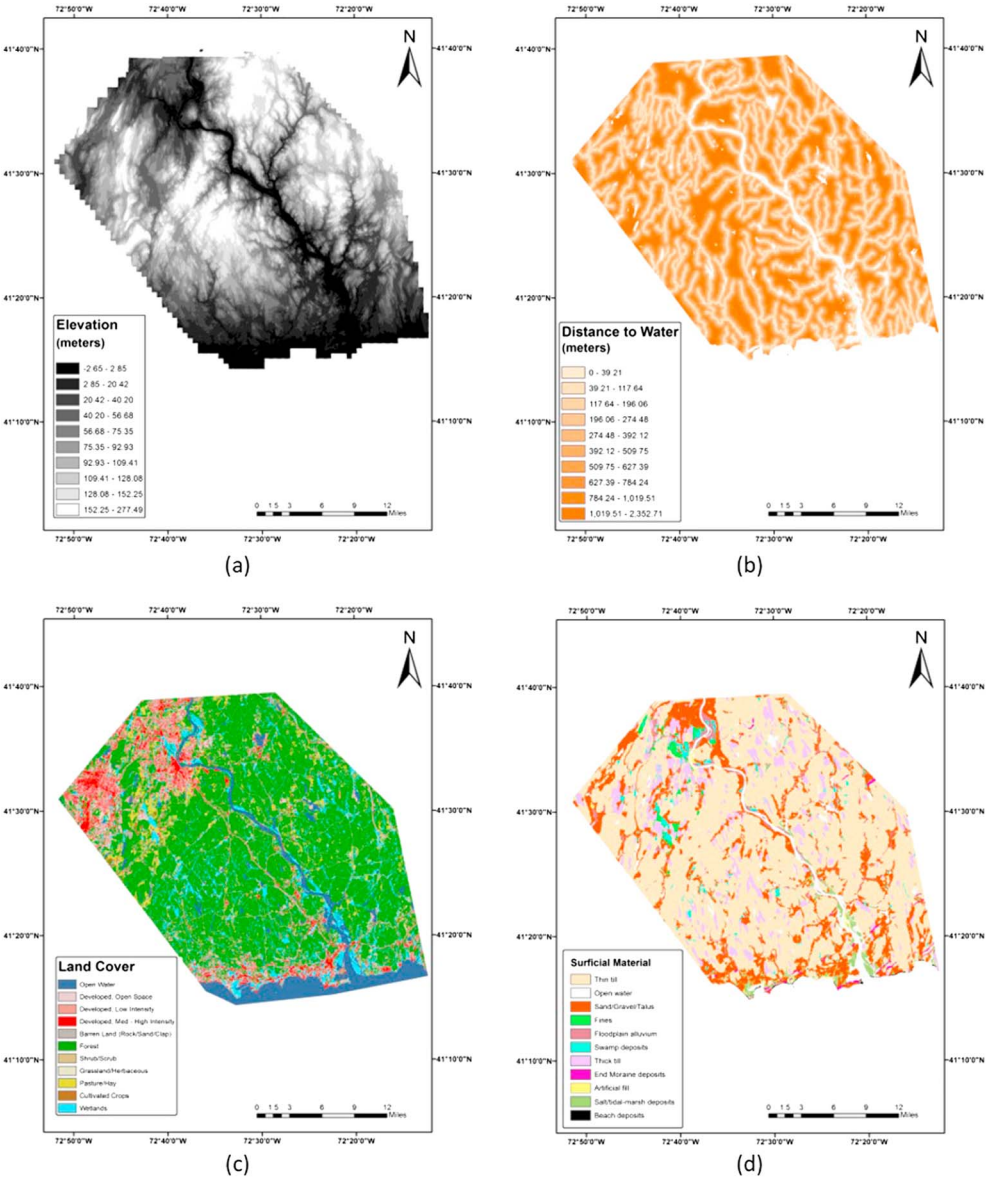

Figure 2. Spatial distribution of flood risk factors: (a) elevation (ELEV), (b) distance to water (DIST), (c) land cover (LAND), and (d) surficial materials (GEO).

fact that there has not been a flood event in the region greater in magnitude than a 1 in 25-year discharge for which USGS/National Aeronautics and Space Administration Landsat satellite images of sufficient quality are available, in addition to noting that the flood inundation delineation for all recent, but minor, flood events falls almost entirely within the boundary of the FEMA 100-year Special Flood Hazard Area (SFHA), it was decided to compare flood risk factors to flood inundation as defined by the FEMA 100-year SFHA (Federal Emergency Management Agency, 2016) for the region (Figure 3) to initially train the statistical model. Flood inundation data from the SFHA were compiled into a spatial database and resampled to a 30-m × 30-m grid identical to those used for the flood risk factors.

It should be noted that the SFHA has received much scrutiny because of its past dependence on one-dimensional hydraulic models and low-resolution elevation data. For example, Blessing et al. (2017) found that the SFHA missed near 75% of flood claims made by those affected within several municipalities of the southeastern suburbs of Houston, Texas, during five major flood events between the years 1999 and 2009, although the version of the SFHA used in Blessing et al. (2017) would have been updated

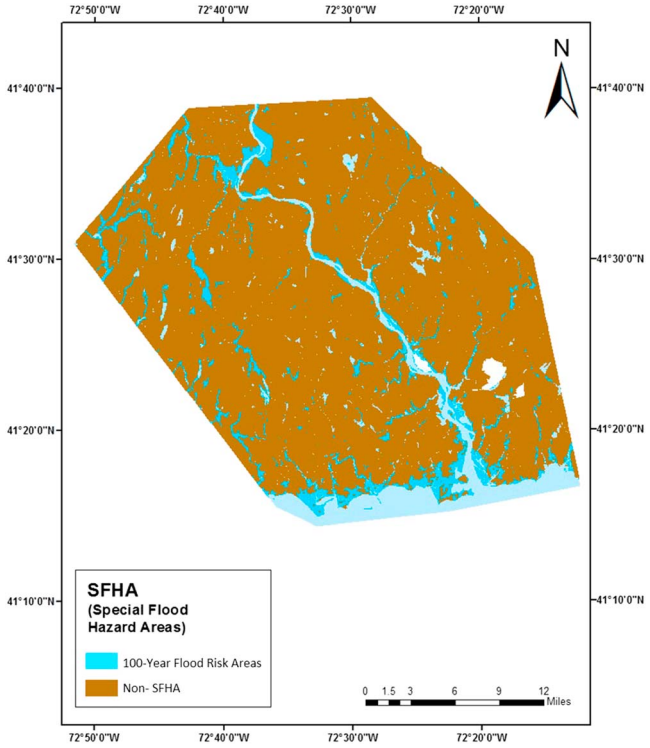

Figure 3. The 100-year Federal Emergency Management Agency SFHA within the Lower Connecticut River Valley Region. Light blue represents open water, whereas dark blue represents land areas within the SFHA.

prior to 1999 and would have employed lower-quality hydrologic and hydraulic models and lower-resolution elevation data than is currently used. In addition, the SFHA only takes into account riverine and coastal flooding, while many coastal events such as Hurricane Harvey are dominated by pluvial flooding. It should be noted that one limitation of the SFHA is that where there are combined effects of riverine and coastal flooding, the modeling that is used to develop the SFHA treats them as independent drivers, which may result in an inappropriate characterization of flood risk in some areas (Moftakhari et al., 2017). In another study where a high-resolution hydrodynamic model was developed for the entire conterminous United States using the well-accepted Height Above Nearest Drainage methodologies (Wing et al., 2017), it was found that the model matched up to 86% of the extent of the most current version of the SFHA, which employs higher-quality one-dimensional and two-dimensional hydraulic modeling tools and higher-resolution elevation data (down to 1 m) from the USGS National Elevation Dataset. Because of the improved performance of the SFHA in capturing areas that would be potentially impacted by a 100-year flood event and the fact that the SFHA is the only resource currently available within the LCRVR that provides an estimate of spatial flood inundation from an extreme flood event, the SFHA was assumed to provide a sufficiently accurate depiction of 100-year spatial flood inundation due to riverine and coastal events within the study region.

2.3. Logistic Regression

Logistic regression was implemented to develop a specific formula that measures the probability of flood inundation throughout the LCRVR during the 100-year flood event as defined in Figure 3. This is accomplished by designating several points throughout the LCRVR as testing points from which the logistic regression will be derived. Because of the large size of the LCRVR and in order to reduce the bias caused by one portion of the region on another part of the region, this was accomplished by first dividing the LCRVR into three separate subregions that represent urban, rural, and coastal environments (Figure 4). These subregions were selected based on land cover characteristics, particularly level of development, as depicted in Figure 2c; the relatively urban area of Middletown, CT, is observed in the northwest portion

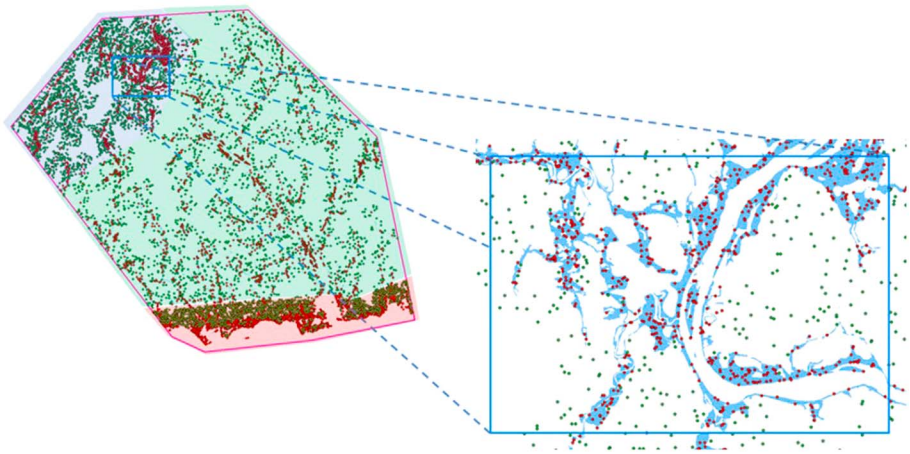

Figure 4. Map of the Lower Connecticut River Valley Region along with a zoomed-in area showing the distribution of sampling points used to train the logistic model. Green points represent locations where flooding did not occur, while red points represent locations where flooding did occur. Areas shaded in blue, green, and red, represent urban (U; blue), rural (R; green), and coastal (C; red) subregions, respectively.

of the region, while development can also be seen along the coast in the southern portion of the region; the remainder of the region is predominantly rural. A total of 4,000 points was randomly chosen from each subregion with the stipulation that an equal number of those points (2,000 per subregion) were within (green dots in Figure 4) or outside (red dots in Figure 4) of the FEMA 100-year SFHA. A total of 12,000 points, therefore, was chosen from which to extract flood inundation and flood risk factor data.

Flood data for all points consisted of either a 0 or a 1 to represent whether a location was not flooded or flooded, respectively; these values represented the dependent variable (L) in the logistic regression:

$$\ln\left(\frac{p}{1-p}\right) = L = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n, \tag{1}$$

where p is the probability of flooding. All flood risk factor data at each location were categorized into classes according to the class ranges designated in Table 3 and represented the independent variables (x_1 to x_n; n = 9) in equation (1). In some cases, the land cover, soil class, and/or surficial materials risk factors were classified as *open water* and/or the distance to water was equal to 0 even though the location was located outside of any particular body of water. This apparent artifact is attributable due to differences in the resolution of each data set, which can cause a slight shift in the boundaries of water bodies when the data sets are processed (*snapped* and *clipped*) within ArcGIS. The result is that extracted values from some layers will occur over open water, while extracted values from other layers will occur over the land that is adjacent to the same body of water. These points were justifiably eliminated from the analysis, which resulted in the total number of points being utilized in the urban, rural, and coastal subregional data sets, respectively, to be 3,815; 3,708; and 3,776. The independent and dependent variables were then analyzed using the function glm(…, family = *binomial*) in R to determine the regression intercept (a_0) and the coefficients (a_1 to a_n; n = 9) for each flood risk factor in equation (1).

The final step in the development of the logistic model for flood susceptibility is to estimate the model's goodness of fit. One common method that works well for binary data is the Hosmer-Lemeshow (H-L) goodness of fit test (Hosmer et al., 2013). The H-L test computes a test statistic that compares the predicted values of the model with observations and that approximately follows a chi-square distribution. The resulting p value is then computed as the right-hand tail probability of the distribution. A low p value ($<$0.05) suggests that the model fit is poor, while a high p value suggests that the null hypothesis that there is no relation between flooding and the flood risk factors can be rejected. Refer to Hosmer et al. (2013) for more details on the H-L test. The H-L test was implemented in R using the *hoslem.test* function.

After the coefficients of the logistic regressions are determined for each flood risk factor class, the probability of flooding at each grid cell is calculated from the first two members of equation (1) using the following equation:

$$p = {e^L}\Big/{(1+e^L)}, \tag{2}$$

which is used to create the final flood risk map. It should be noted that all flood risk factors are used but that for each flood risk factor only one coefficient is used that corresponds to the appropriate factor class (see Table 3) at each map grid cell.

2.4. Critical Infrastructure

The final step in the development of the flood susceptibility map involves identifying locations with vulnerable critical infrastructure, which included the following:

- dams;
- military compounds;
- airports;
- hospitals and other health-related facilities;
- fire and police stations;
- emergency operations centers;

Table 4
Critical Infrastructure Data Sets Used in the Current Study With Data Source and URL

Infrastructure	Source (year)	URL for data access
Airports	DEEP (2005)	https://www.ct.gov/deep
Bridges	National Bridge Inventory (Federal Highway Administration, 2016)	https://www.arcgis.com/home/item.html?id=775f08232eb1424189a4e8091edf893e
Dams	DEEP (1996)	https://www.ct.gov/deep
EOCs	RiverCOG (2017)	https://www.rivercog.org
Fire and police stations	RiverCOG (2017)	https://www.rivercog.org
Health	USDHHS (2012)	https://maps3.arcgisonline.com/ArcGIS/rest/services/A-6/HHS_IOM_Health_Resources/MapServer/
Land use and zoning	RiverCOG (2017)	https://www.rivercog.org
Military	MAGIC (2010)	https://magic.lib.uconn.edu/connecticut_data.html
Railroads	DEEP (2005)	https://www.ct.gov/deep
Routes	DEEP (2006)	https://www.ct.gov/deep
Schools	RiverCOG (2017)	https://www.rivercog.org
Town halls	RiverCOG (2017)	https://www.rivercog.org

Note. DEEP = Connecticut Department of Energy and Environmental Protection; NBI = National Bridge Inventory; FHWA = Federal Highway Administration; EOC = Emergency Operations Center; RiverCOG = The Lower Connecticut River Valley Council of Governments; USDHHS = U.S. Department of Health and Human Services; MAGIC = University of Connecticut Libraries' Map and Geographic Information Center.

- private and public K–12 schools;
- town halls;
- major routes;
- bridges; and
- railroads.

Data sets and sources related to critical infrastructure throughout the LCRVR and that were used in the current study are given in Table 4. All critical infrastructure data sets were clipped to the boundaries of the LCRVR and overlaid onto the final flood susceptibility map.

3. Results

3.1. Flood Risk

The coefficients from the logistic regression are listed in Table 3 for each class of each flood risk factor over the three subregions; the greater the magnitude of the coefficient, the stronger the impact of that risk factor class on flood inundation in the LCRVR. The *p* values computed for the logistic models in the coastal, rural, and urban subregions using the H-L test were approximately 0.76, < 0.01, and 0.60. Because of their high *p* values, there is no evidence of poor fit within the coastal and urban subregions, which are the two areas of highest concern in the LCRVR due to their relatively high population densities. The fit is much less reliable for the more sparsely populated rural subregion. The low *p* value indicates that the rural subregion is sufficiently large so that there is substantial variation in the relationship of each flood risk factor to flood inundation throughout its area.

In order to make a simple comparison of the results between subregions, especially due to the high variation in the relationships of the flood risk factors to flood inundation in the rural subregion, the regression coefficients for all flood risk factors were averaged for each subregion, the results of which are shown in Figure 5a. There are initially three flood risk factors that stand out as having a dominant correlation with flood susceptibility throughout the LCRVR: elevation (ELEV), distance to water (DIST), and surficial materials (GEO). Elevation has the most influence on flood susceptibility in the urban and coastal subregions because of the fact that both subregions are dominated by lower elevations, whereas elevation has less influence within the rural subregion where substantially higher elevations dominate. Distance to water has a large influence on flood susceptibility in all subregions because of the number of water bodies located throughout the LCRVR, which include a myriad of small lakes, ponds, and tributaries, in addition to the Connecticut River and Long Island Sound. Surficial materials has greater influence on flood susceptibility in the rural subregion and coastal subregions where much

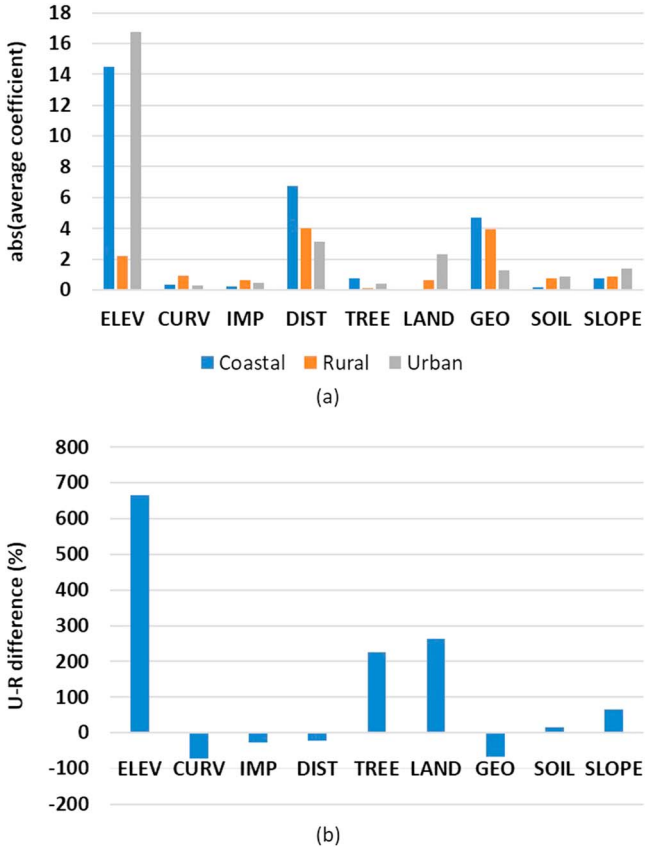

Figure 5. (a) Average absolute value of the logistic regression coefficients computed for each flood risk factor for the coastal (blue), rural (orange), and urban (gray) subregions, and (b) the percent difference between the urban (U) and rural (R) coefficients for each flood risk factor.

of the materials deposited from previous flood events are still present, whereas these same materials have likely been removed within the more urban Middletown area as development has occurred. To get an idea of additional impacts or sensitivity of urbanization on the contribution of each flood risk factor, Figure 5b shows a plot of the percent change in the contribution of each flood risk factor between the urban and rural subregions. Two flood risk factors stand out as having the largest impact: elevation (already discussed) and land cover. Assuming that elevation within the urban subregion has not changed substantially due to urbanization and that any differences in the contribution of elevation between the subregions can be attributed to natural differences in topographic features, Figure 5b shows that recent changes in land cover have had the most impact on changes in flooding behavior between the rural and urban subregions.

The results of the logistic regression for the initial set of data points were then applied to all map grid cells in the LCRVR to produce a flood susceptibility map for the entire region applicable to the 100-year flood event (Figure 6a). Flood susceptibility values are plotted as the percent chance that each 30-m × 30-m grid cell will be inundated and then classified into five categories according to the color scale shown in the figure: *very low risk* (0–20%), *low risk* (20–40%), *medium risk* (40–60%), *high risk* (60–80%), and *very high risk* (80–100%). The largest areas of *very high* and *high* risk are located along the Connecticut River and its major tributaries as well as along the coast. There are also several isolated areas of high susceptibility associated with smaller streams and creeks.

Finally, it is observed that when looking at the transitions between the different subregions, particularly between the coastal and rural subregions, the values are not continuous and there is a slight difference

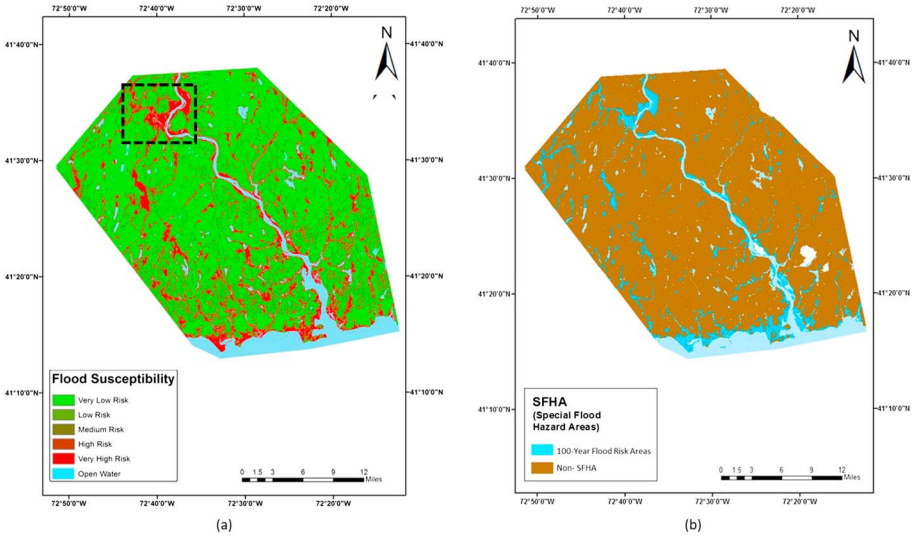

Figure 6. Flood susceptibility map for the Lower Connecticut River Valley Region for the Federal Emergency Management Agency 100-year flood event. Levels represent probabilities of flooding: *very low*: 0–20%; *low*: 20–40%; *medium*: 40–60%; *high*: 60–80%; *very high*: 80–100%. Dashed box (inset) shown in Figure 7. (b) The map showing the spatial extent of the SFHA is repeated for comparison purposes.

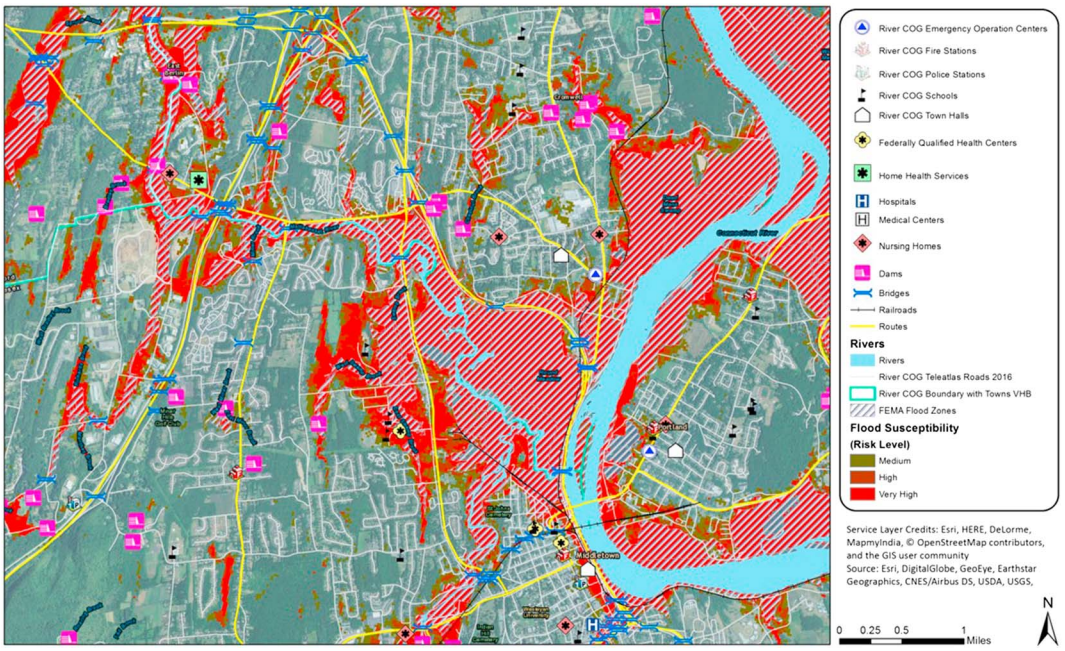

Figure 7. Locations of various vulnerable critical infrastructure relative to areas of *medium* (dark green), *high* (dark red), and *very high* (red) flood susceptibility; map is zoomed in on the city of Middletown, CT, and surrounding area (dashed box in Figure 6). The 100-year FEMA Special Flood Hazard Area (hatched) is also included for reference and comparison. USDA = U.S. Department of Agriculture; USGS = U.S. Geological Survey; COG = Council of Governments; FEMA = Federal Emergency Management Agency; CNES = Centre National d'Etudes Spatiales.

across the subregion boundary. This difference is a statistical artifact of splitting the region into three subregions and computing different values for the coefficients of each flood risk factor class; for example, the rural and urban sets of factor coefficients listed in Table 3 were used to separately compute the flood maps for the rural and coastal zones, respectively. The result is a small discontinuity between the subregions, albeit this discontinuity seems to manifest itself more in the lower susceptibility categories as opposed to the areas of very high susceptibility risk. If the entire LCRVR was analyzed as one subregion, these discontinuities would disappear, but the results would include a substantial bias from the urban subregion in determining flood susceptibility in the coastal subregion, which would likely produce inaccuracies that are much more substantial than the current discontinuities. The only other way to eliminate these discontinuities would be to use a sufficient number of subregions so that the discontinuities between each are minimal, which is unrealistic, and the choice of how subregions were chosen would be difficult to defend.

When comparing the susceptibility map to the map of the FEMA 100-year SFHA (repeated in Figure 6b for comparisons purposes), it is important to understand key distinctions between the two. The FEMA 100-year SFHA is limited to the subwatersheds of $>$2.59k m^2. Other limiting issues with the FEMA 100-year SFHA are (1) the age of the underlying studies (often more than two decades old) and (2) their focus on only areas where development either already existed or was imminently to be and so was then anticipated. By using the statistical modeling described herein it was possible to identify the contribution of flood risk factors within the existing FEMA 100-year SFHA and apply such factors to the entire study region to identify additional areas outside of the FEMA flood hazard area that are susceptible to inundation by a flood event having a 1% chance of occurring in any given year. It should be noted that there also were areas (not shown) within the SFHA that were not identified as very high or high susceptibility in the present analysis because of the fact that values of the dominant flood risk factors in these locations are different than those identified throughout the remainder of the SFHA.

Geographical Information System spatial analyses were made to compare the susceptibility mapping to FEMA's SFHA map using the University of Connecticut's Center for Land, Education, and Research 2010 Land Cover 30-m data set (Center for Land Use Education and Research Land Cover, College of Agriculture and Natural Resources, University of Connecticut, 2010). Twenty-five percent of the region's FEMA mapped

flood zones are developed, which represents approximately 8% of the overall developed area in the region. When subtracting waterbodies and wetlands at the areas designated as very high, high, or *medium*, an additional 115 km^2 are added to areas identified as susceptible. In the very high and high classified areas only, this previously unidentified susceptible acreage adds greater than 6% of the region's nonwater and wetland area to a flood susceptibility zone, including an additional 8% of the region's developed area.

One important disclaimer about the flood susceptibility map is that it was created for present-day conditions and is only to be used for increasing engineering and stakeholder awareness; it is not intended to replace the FEMA mapping for regulatory or flood insurance decisions. It should also be noted that the scale of the flood susceptibility map and data are most appropriately used at the regional scale. However, use of the data at the municipal scale should allow local stakeholders to examine areas of special concern for planning purposes.

3.2. Critical Infrastructure

Data sets for several types of critical infrastructure (listed in Table 4) were obtained and overlaid onto the final flood susceptibility map for the LCRVR. An area surrounding and including the City of Middletown, Connecticut, was chosen for further scrutiny because of the presence of a large very high susceptibility zone (Figure 7). Several dams, bridges, and a large portion of the major routes and railroad in the Middletown vicinity are included within the high and very high susceptibility areas of 100-year flood inundation. It is also concluded that there are some areas identified as having medium to very high flood susceptibility to the 100-year flood that were not included in the FEMA 100-year SFHA. These differences exist primarily in an area on the west and south sides of Middletown—as can be seen in Figure 7 by the red and dark green shaded areas that are located outside of the hatched areas. These differences could have a major impact on the perceived vulnerability of critical infrastructure located in these areas.

4. Conclusions

The current study estimated flood susceptibility in the LCRVR attributable to nonclimatic factors using a method that involved performing a logistic regression for three subregions (urban, rural, and coastal) to determine the relations between several flood risk factors and flood inundation at the 100-year return period, which was defined by the FEMA 100-year SFHA, in each subregion. It was found that elevation and distance to water have the most influence on flood susceptibility in the urban and coastal subregions, while distance to water and surficial materials have the greatest influence in the rural subregion. It was also determined that urbanization has had the most influence on the contribution of land cover to 100-year flood susceptibility when compared to the rural subregion; development within the urban subregion has increased the contribution of *land use* by over 200%. The difference in the contribution of elevation to flood susceptibility between the urban and rural subregions was greater than that for land use, but it is assumed that this is likely not because of urbanization but rather attributable to natural differences in topographic features between the two subregions. Because there is still sufficient room for continued growth and development within the urban subregion, future significant increases in the effects of changing land cover on flood susceptibility in the area are possible.

The logistic regression equation was then used to create an overall flood susceptibility map for each subregion of the LCRVR onto which various types of critical infrastructure and regional existing land use and zoning data were overlaid. Differences between the 100-year susceptibility map developed here and the FEMA 100-year SFHA were observed. Most importantly, developed residential and commercial areas within the region fall within the medium to very high flood susceptibility (hot spot) areas beyond what is designated as the FEMA 100-year SFHA. Although the regional data is not at a scale large enough for local determinations, these hot spot areas warrant further consideration for future localized flood susceptibility mapping if future suitable data sets become available and further consideration at the municipal planning level.

One important disclaimer about the flood susceptibility map is that it was created for present-day conditions and is only to be used for planning purposes. There are several prominent factors that could affect the future flood susceptibility map: changes in impervious area (through urbanization), a higher sea level (for coastal areas), and changes in climatic factors (e.g., heavier precipitation). A future flood susceptibility map can be created by studying how each of these types of factors are expected to change. However, it is expected that the present-day flood susceptibility map provides an excellent relative foundation from which to consider future changes.

Acknowledgments
This work was funded by the U.S. Department of Housing and Urban Development (HUD) Community Development Block Grant—Disaster Recovery (CDBG-DR) Program—Federal grant B-13-DS-09-0001 and administered by the Connecticut Department of Housing (DOH)—grant 6202. Additional funding was obtained through a subaward agreement with the University of Connecticut's (UCONN) Institute for Resilience and Climate Adaption (CIRCA)—grant PO#43280 and the Connecticut Department of Energy and Environmental Protection (DEEP)—grant PS#2014-14249. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of University of Connecticut or Connecticut Department of Energy and Environmental Protection. All data used in the current study can be found at the URLs provided in the tables and list of references.

References

Abbott, M. B., Bathrust, J. C., Cunge, J. A., O'Connell, P. E., & Rasmussen, J. (1986). An introduction to the European Hydrological System—Systeme Hydrologique Europeen, "SHE", 2: Structure of a physically-based, distributed modeling system. *Journal of Hydrology*, *87*(1-2), 61–77. https://doi.org/10.1016/0022-1694(86)90115-0

Allaire, M. C., Vogel, R. M., & Kroll, C. N. (2015). The hydromorphology of an urbanizing watershed using multivariate elasticity. *Advances in Water Resources*, *86*, 147–154. https://doi.org/10.1016/j.advwatres.2015.09.022

Blessing, R., Sebastian, A., & Brody, S. D. (2017). Flood risk delineation in the United States: How much loss are we capturing? *Natural Hazards Review*, *18*(3), 04017002. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000242

Center for Land Use Education and Research Land Cover, College of Agriculture and Natural Resources, University of Connecticut. (2010). Land cover 30-m data set. Location. Retrieved from https://clear.uconn.edu/projects/landscapeLIS/landcover.htm

Crawford, N. H., & Linsley, R. K. (1966). Digital simulation in hydrology: Stanford Watershed Model IV. (Tech. Rep. 39, p. 210). Stanford, CA: Department of Civil Engineering, Stanford University.

Connecticut Department of Energy and Environmental Protection. (1996). Connecticut GIS datasets. Location. Retrieved from https://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707

Connecticut Department of Energy and Environmental Protection. (2005). Connecticut GIS datasets. Location. Retrieved from https://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707

Connecticut Department of Energy and Environmental Protection. (2006). Connecticut GIS datasets. Location. Retrieved from https://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707

Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography: Earth and Environment*, *25*(1), 80–108. https://doi.org/10.1177/030913330102500104

Devi, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic Procedia*, *4*, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The human impact of floods: A historical review of events 1980–2009 and systematic literature review. *PLOS Currents Disasters*, *1*, 29. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Environmental Systems Research Institute. (2014). ArcGIS desktop: Release 10.2.2. Redlands, CA: Environmental Systems Research Institute.

Federal Emergency Management Agency. (2016). FEMA GeoPlatform, National Flood Hazard Layer (NFHL). Location. Retrieved from https://fema.maps.arcgis.com/home/index.html

FHWA (Federal Highway Administration). (2016). National Bridge Inventory (NBI) bridges dataset. Location. Retrieved from https://www.arcgis.com/home/item.html?id=94c41e96db0d4b85b9eb622923e0a0e8

Gassman, P. W., Reyes, M. R., Green, C. H., & Arnold, J. G. (2007). The soil and water assessment tool: Historical development, applications, and future research directions. *Transactions of the American Society of Agricultural and Biological Engineers*, *50*, 1211–1250.

Gogoi, S., & Chetia, B. C. (2011). Fuzzy rule-based flood forecasting model of Jiadhal River basin, Dhemaji, Assam, India. *International Journal of Fuzzy Mathematical Systems*, *1*, 59–71.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. New York: John Wiley. https://doi.org/10.1002/9781118548387

Hundecha, Y., Bardossy, A., & Theisen, H.-W. (2001). Development of a fuzzy logic-based rainfall-runoff model. *Journal des Sciences Hydrologiques*, *46*(3), 363–376. https://doi.org/10.1080/02626660109492832

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, *67*(1), 251–264. https://doi.org/10.1007/s12665-011-1504-z

Kovacevic, M., Ivanisevic, N., Dasic, T., & Markovic, L. (2018). Application of artificial neural networks for hydrological modelling in karst. *Gradjevinar*, *70*(1), 1–10. https://doi.org/10.14256/JCE.1594.2016

Lawal, D., Matori, A., Hashim, A., Yusof, K., & Chandio, I. (2012). Detecting flood susceptible areas using GIS-based analytic hierarchy process. In *Proceedings of the International Conference on Future Environment and Energy* (Vol. 28, pp. 1–5). Singapore: IACSIT Press.

Lee, M. J., Kang, J. E., & Jeon, S. (2012). Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 895–898). Munich, Germany: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/IGARSS.2012.6351414

Li, H., Lei, X., Shang, Y., & Qin, T. (2018). Flash flood early warning research in China. *International Journal of Water Resources Development*, *34*(3), 369–385. https://doi.org/10.1080/07900627.2018.1435409

Lopez, M. G., Baldassarre, G. D., & Seibert, J. (2017). Impact of social preparedness on flood early warning systems. *Water Resources Research*, *53*, 522–534. https://doi.org/10.1002/2016WR019387

Mahmoud, S. H., & Gan, T. Y. (2018). Urbanization and climate change implications in flood risk management: Developing an efficient decision support system for flood susceptibility mapping. *Science of the Total Environment*, *636*, 152–167. https://doi.org/10.1016/j.scitotenv.2018.04.282

Matori, A. N., Lawal, D. U., Yusof, K. W., Hashim, M. A., & Balogun, A.-L. (2014). Spatial analytic hierarchy process model for flood forecasting: An integrated approach. *IOP Conference Series: Earth and Environmental Science*, *20*, 012029. https://doi.org/10.1088/1755-1315/20/1/012029

Miller, J. D., & Hutchins, M. (2017). The impacts of urbanisation and climate change on urban flooding and urban water quality: A review of the evidence concerning the United Kingdom. *Journal of Hydrology: Regional Studies*, *12*, 345–362. https://doi.org/10.1016/j.ejrh.2017.06.006

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Compounding effects of sea level rise and fluvial flooding. *Proceedings of the National Academy of Sciences of the United States*, *114*, 9785–9790.

Papadopoulou-Vrynioti, K., Bathrellos, G. D., Skilodimou, H. D., Kaviris, G., & Makropoulos, K. (2013). Karst collapse susceptibility mapping considering peak ground acceleration in a rapidly growing urban area. *Engineering Geology*, *158*, 77–88. https://doi.org/10.1016/j.enggeo.2013.02.009

Park, S., Hamm, S.-Y., Jeon, H.-T., & Kim, J. (2017). Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS. *Sustainability*, *9*(7), 1157–1176. https://doi.org/10.3390/su9071157

Peterson, T. C., Heim, R. R. Jr., Hirsch, R. M., Kaiser, D. P., Brooks, H., Diffenbaugh, N. S., et al. (2013). Monitoring and understanding changes in heat waves, cold waves, floods and droughts in the United States: State of knowledge. *Bulletin of the American Meteorological Society*, *94*(6), 821–834. https://doi.org/10.1175/BAMS-D-12-00066.2

Pradhan, B., & Lee, S. (2010). Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environmental Earth Sciences*, *60*(5), 1037–1054. https://doi.org/10.1007/s12665-009-0245-8

Qureshi, M., & Harrison, S. (2003). Application of the analytic hierarchy process to riparian revegetation policy options. *Small-Scale Forest Economics, Management and Policy*, *2*(3), 441–458. https://doi.org/10.1007/s11842-003-0030-6

R Development Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rahman, M. M., Goel, N. K., & Arya, D. S. (2013). Study of early flood warning dissemination system in Bangladesh. *Journal of Flood Risk Management*, *6*(4), 290–301. https://doi.org/10.1111/jfr3.12012

Richardson, C. P., & Amankwatia, K. (2018). GIS-based analytic hierarchy process approach to watershed vulnerability in Bernalillo County, New Mexico. *Journal of Hydrologic Engineering*, *23*(5), 04018010. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001638

RiverCOG (The Lower Connecticut River Valley Council of Governments). (2017). Lower Connecticut River Valley Region critical infrastructure dataset. Location. Retrieved from https://www.rivercog.org

Saini, S. S., & Kaushik, S. P. (2012). Risk and vulnerability assessment of flood hazard in part of Ghaggar basin: A case study of Guhla block, Kaithal, Haryana, India. *International Journal of Geomatics and Geosciences*, *3*, 10.

Sen, Z., & Altunkaynak, A. (2004). Fuzzy awakening in rainfall-runoff modeling. *Hydrology Research*, *35*(1), 31–43. https://doi.org/10.2166/nh.2004.0003

Sharma, S. K., Gajbhiye, S., & Tignath, S. (2015). Application of principal component analysis in grouping geomorphic parameters of a watershed for hydrological modeling. *Applied Water Science*, *5*(1), 89–96. https://doi.org/10.1007/s13201-014-0170-1

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, *20*(7), 2611–2628. https://doi.org/10.5194/hess-20-2611-2016

Siddayao, G. P., Valdez, S. E., & Fernandez, P. L. (2014). Analytic hierarchy process (AHP) in spatial modeling for floodplain risk assessment. *International Journal of Machine Learning and Computing*, *4*(5), 450–457. https://doi.org/10.7763/IJMLC.2014.V4.453

Singh, P. K., Kumar, V., Purohit, R. C., Kothari, M., & Dashora, P. K. (2009). Application of principal component analysis in grouping geomorphic parameters for hydrologic modeling. *Water Resources Management*, *23*(2), 325–339. https://doi.org/10.1007/s11269-008-9277-1

Sinha, R., Bapalu, G. V., Singh, L. K., & Rath, B. (2008). Flood risk analysis in the Kosi River Basin, North Bihar using multi-parametric approach of analytical hierarchy process (AHP). *Journal of the Indian Society of Remote Sensing*, *36*(4), 335–349. https://doi.org/10.1007/s12524-008-0034-y

Tehrany, M. S., Lee, M.-J., Pradhan, B., Jebur, M. N., & Lee, S. (2014). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences*, *72*(10), 4001–4015. https://doi.org/10.1007/s12665-014-3289-3

Ullah, N., & Choudhury, P. (2013). Flood flow modeling in a river system using adaptive neuro-fuzzy inference system. *Environmental Management and Sustainable Development*, *2*(2), 54–68. https://doi.org/10.5296/emsd.v2i2.3738

Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N., & Tehrany, M. S. (2014). Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *Catena*, *118*, 124–135. https://doi.org/10.1016/j.catena.2014.02.005

United States Department of Agriculture-Natural Resources Conservation Service. (2017). Soil data access. Location. Retrieved from https://sdmdataaccess.nrcs.usda.gov/

United States Department of Health and Human Services. (2012). Healthcare facilities dataset. Location. Retrieved from https://maps3.arcgisonline.com/ArcGIS/rest/services/A-16/HHS_IOM_Health_Resources/MapServer/

University of Connecticut Libraries' Map And Geographic Information Center. (2010). Connecticut GIS data, Connecticut military installations. Location. Retrieved from https://magic.lib.uconn.edu/connecticut_data.html

U.S. Geological Survey. (2011). National Land Cover Database (NLCD). Location. Retrieved from https://www.mrlc.gov/

U.S. Geological Survey. (2014). EarthExplorer elevation dataset. Location. Retrieved from https://earthexplorer.usgs.gov/

Wallis, J. R. (1965). Multivariate statistical methods in hydrology—A comparison using data of known functional relationship. *Water Resources Research*, *1*(4), 447–461. https://doi.org/10.1029/WR001i004p00447

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). Validation of a 30 m resolution flood hazard model of the conterminous United States. *Water Resources Research*, *53*, 7968–7986. https://doi.org/10.1002/2017WR020917

Yahaya, S., Ahmad, N., & Abdalla, R. F. (2010). Multicriteria analysis for flood vulnerable areas in Hadejia-Jama'are River Basin, Nigeria. *European Journal of Scientific Research*, *42*, 71–83.

Yaseen, Z. M., Ramal, M. M., Diop, L., Jaafar, O., Demir, V., & Kisi, O. (2018). Hybrid adaptive neuro-fuzzy models for water quality index estimation. *Water Resources Management*, *32*(7), 2227–2245. https://doi.org/10.1007/s11269-018-1915-7

Zhang, H., Zhang, J., & Han, J. (2005). The assessment and regionalization of flood/waterlogging disaster risk in middle and lower reaches of Liao River of Northeast China. In *Proceedings of the Fifth Annual IIASA-DPRI Forum on Integrated Disaster Risk Management* (pp. 103–118). Beijing, China: IIASA-DPRI.

Zhu, T., Lund, J. R., Jenkins, M. W., Marques, G. F., & Ritzema, R. S. (2007). Climate change, urbanization, and optimal long-term floodplain protection. *Water Resources Research*, *43*, W06421. https://doi.org/10.1029/2004WR003516

Zounemat-Kermani, M., & Teshnehlab, M. (2008). Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Applied Soft Computing*, *8*(2), 928–936. https://doi.org/10.1016/j.asoc.2007.07.011