Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR029576

Application of Nonlinear Time Series and Machine Learning Algorithms for Forecasting Groundwater Flooding in a Lowland Karst Area

Bidroha Basu[1,2], Patrick Morrissey[2], and Laurence W. Gill[2]

[1]Departme of Civil, Structural and Environmental Engineering, Munster Technological University, Cork, Ireland, [2]Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Dublin, Ireland

Key Points:

- Early warning system for groundwater flooding in lowland karst developed using nonlinear modeling approaches
- Nonlinear autoregressive model with exogenous variables showed the best flood forecasting performance up to 60 days into the future
- Real-time telemetric monitoring of water level in the catchment can be fed into the model to provide an early warning flood warning tool

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. W. Gill,
Laurence.Gill@tcd.ie

Citation:

Basu, B., Morrissey, P., & Gill, L. W. (2022). Application of nonlinear time series and machine learning algorithms for forecasting groundwater flooding in a lowland karst area. *Water Resources Research*, *58*, e2021WR029576. https://doi.org/10.1029/2021WR029576

Received 9 JAN 2021
Accepted 9 JAN 2022

Author Contributions:

Conceptualization: Bidroha Basu, Laurence W. Gill
Data curation: Patrick Morrissey, Laurence W. Gill
Formal analysis: Bidroha Basu, Patrick Morrissey
Investigation: Bidroha Basu
Methodology: Bidroha Basu, Laurence W. Gill
Project Administration: Laurence W. Gill
Supervision: Laurence W. Gill
Writing – original draft: Bidroha Basu, Laurence W. Gill

Abstract In karst limestone areas interactions between ground and surface waters can be frequent, particularly in low lying areas, linked to the unique hydrogeological dynamics of that bedrock aquifer. In extreme hydrological conditions, however, this can lead to wide-spread, long-duration flooding, resulting in significant cost and disruption. This study develops and compares a nonlinear time-series analysis based nonlinear autoregressive model with exogenous variables (NARX), machine learning based near support vector regression as well as a linear time-series ARX model in terms of their performance to predict groundwater flooding in a lowland karst area of Ireland. The models have been developed upon the results of several years of field data collected in the area, as well as the outputs of a highly calibrated semi-distributed hydraulic/hydrological model of the karst network. The prediction of total flooding volume indicates that the performances of all the models are similarly accurate up to 10 days into the future. A NARX model taking inputs of the past 5 days' flood volume; rainfall data and tidal amplitude data across the past 4 days, showed the best flood forecasting performance up to 30 days into the future. Existing real-time telemetric monitoring of water level data at two points in the catchment can be fed into the model to provide an early warning flood warning tool. The model also predicts freshwater discharge from the inter-tidal spring into the Atlantic Ocean which hitherto had not been possible to monitor.

1. Introduction

The occurrence of groundwater flooding, and the associated damage and disruption that results, has become more frequently studied in recent decades. Several notable groundwater flooding events have taken place in recent history, many linked to karst catchments, whereby this phenomenon has become recognised as a hazard which must be accounted for when planning and managing new and existing developments (e.g., flood risk assessments; Cobby et al., 2009). For example, extensive karst-related groundwater flooding occurred in winter 2000/2001 in the chalk catchments of southern England and in the Somme Basin in France (Hughes et al., 2011; Pinault et al., 2005); as well as being reported in the Dinaric karst region of south-eastern Europe in poljes (Ristic, 1976) and the Unica river basin in Slovenia (Kovacic & Ravbar, 2010). In Ireland, there was considerable groundwater flooding in the karst of south-west Galway in 2009 and 2015/2016 (Morrissey et al., 2020; Naughton et al., 2017) which caused prolonged and wide-spread flooding. Whilst groundwater flooding is less hazardous when compared to fluvial flooding in terms of acute risk to life, it can persist for much longer durations resulting in significant cost and disruption. Such flood events are also likely to get more frequent and more serious as a result of changes to the climate over the next few decades, in particular the predictions of more intense rainfall during winter months (Blöschl et al., 2019; Noone et al., 2017).

In karst limestone areas interactions between ground and surface waters can be frequent, linked to the unique hydrogeological dynamics of that bedrock aquifer, with sinking and rising rivers/streams common and surface water features absent completely in many areas (Drew, 2008). If the fractures or conduits within the limestone, which regulate the main flow paths through such secondary porosity dominated rocks, are unable to drain the recharge fast enough during intense or prolonged rainfall events, this can result in groundwater surcharging from the network above ground level. This flood water can be contained temporarily within low-lying topographic depressions known as *turloughs* in Ireland, which represent the principal form of extensive, recurrent groundwater flooding in Ireland (Naughton et al., 2012) or, more generally in karst areas, in poljes (Bonacci, 2014). Such flood events in karst areas are a function of long periods of cumulative rainfall, as well as antecedent storage

Writing – review & editing: Patrick Morrissey, Laurence W. Gill

conditions and so prove to be difficult to characterize by normal hydrological flood risk approaches taken for fluvial flooding events (Morrissey et al., 2020).

Different approaches have been developed to modeling karst aquifers with their inherent complex spatial and temporal heterogeneities (White & White, 2005), from data driven (lumped parameter) models through to more physically based distributed hydrodynamic models, with various forms of semi-distributed models in between (Ghasemizadeh et al., 2012; Kovács & Sauter, 2007).

The data driven models relate an input rainfall signal to an output spring discharge using a transfer function (e.g., Dreiss, 1982; Jukic & Denic-Jukic, 2008). Due to the time-invariant limitations of some transfer function techniques, non-parametric transfer functions have been used to capture non-linear and non-stationary dynamics of karst aquifers more accurately based on, for example, the spectral (frequency) domain (Bailly-Comte et al., 2008; Labat et al., 2000; Larocque et al., 1998) and time/frequency domains using wavelets (Labat et al., 2001; Schuler et al., 2020).

More physically distributed, hydrodynamic modeling approaches require much higher levels of input information; aquifers can be subdivided into small-scale two- or three-dimensional grids with hydraulic parameters and system states defined throughout (Abusaada & Sauter, 2013; Gill et al., 2020; Hartmann et al., 2014). However, given that detailed knowledge of the inner structure of most karst networks is generally unknown (Borghi et al., 2016), the value of such highly parameterized models is a recurrent debate in hydro(geo)logy (Beven, 2006). In between this, more hybrid, semi-distributed models generally use conceptual reservoirs to control areal recharge processes, but then attempt to model the hydraulic dynamics of the high flow conduit network more realistically using specific pipe network morphometry (Chen & Goldscheider, 2014; Gill, Naughton, & Johnston, 2013). It is these types of models that seem to have been most successful in simulating groundwater-surface water interactions in karst catchments, for example, in poljes (Mayaud et al., 2019) and in turloughs (Gill, Naughton, & Johnston, 2013; McCormack et al., 2016).

Due to the limitation of linear models in the prediction/forecasting of hydrological systems, nonlinear models are gaining interest. Different techniques, including various machine learning approaches, have been used in conjunction with groundwater research, including studies on karst systems (Al-Fugara et al., 2020; Hu et al., 2008; Kurtulus & Razack, 2007; Naghibi et al., 2017). In situations where nonlinear models have been developed to understand hydrological systems, neural network-based machine learning algorithms have mainly been used. A popular nonlinear hydrological model is the neural network based nonlinear Auto Regressive model with exogenous variables (NN-NARX). This has been used to forecast future river flood levels based on real-time level gauges along a catchment with meteorological inputs (Lee & Resdi, 2016) as well as to forecast flooding in Taipei city (Taiwan) up to an hour in advance (Chang et al., 2014). Equally, three variants of NN-NARX models have been used on a recurrent neural network to predict multi-step ahead flood inundation depth using rainfall as the exogenous variable by Shen and Chang (2013). A few studies have focused on prediction/forecasting of groundwater dynamics using network based NN-NARX models. Guzman et al. (2017) considered 8 years of daily historical precipitation and groundwater level to train network based NN-NARX model using Levenberg-Marquardt and Bayesian Regularization algorithms at Mississippi River Valley Alluvial aquifer, which was then used to predict groundwater levels up to 90 days ahead. Wunsch et al. (2018) developed a network based NN-NARX model by considering precipitation and temperature as the exogenous variables to forecast groundwater levels in several wells in southwest Germany up to 26 days in the future. The main advantage of such network-based nonlinear models is that they do not require a detailed understanding of the complex hydrological system and are found to be robust in situations where uncertain information is present in the data (ASCE, 2000a; 2000b); their disadvantage is that the models do not provide any explicit mathematical expression to approximate the hydrological system (Lee & Resdi, 2016).

This study develops and compares nonlinear time-series analysis based NARX models, machine learning based nonlinear Support Vector Regression (SVR) model, and a linear time-series ARX model in terms of their performance to predict groundwater flooding in a lowland karst area of Ireland. The models have been developed upon the results of several years of field data collected in the area, as well as the outputs of a highly calibrated semi-distributed hydraulic/hydrological model of the karst network and its allogenic recharge catchments. A new formulation has been developed to extend the optimum NARX model with single exogenous variables to multiple exogenous variables. This study also demonstrates the concept of first identifying the optimal lags for each

variable as part of the model optimisation process (Wunsch et al., 2021) which affects the forecasted output, from which only the critical ones are then used as input variables.

2. Methods—Modeling Approaches and Performance Evaluation

In situations where the system to be modeled is essentially linear in nature, any input-output (IO) relation-based linear model, such as an auto regressive time series model with exogenous variable(s) (ARX), an impulse response function based model, or a state-space model can be used, provided the model order is correct and the noise in the measurements are accommodated (Billings, 2013). However, most natural hydrologic systems are found to be nonlinear in nature and hence a linear modeling approach may not be valid (Amorocho, 1967; Amorocho & Brandstetter, 1971; Jacoby, 1966; Jayawardena & Lai, 1994; Kavvas, 2003; Sivakumar & Singh, 2012). It is extremely difficult to relate any two nonlinear models and hence there is no unique nonlinear model present that can represent all types of nonlinear systems. However, a nonlinear time series model called Nonlinear Auto Regressive model with exogenous variable(s) (NARX) has been found to cover the broadest set of nonlinear systems (Billings, 2013). The NARX model or the NARMAX model, where the NARX model is combined with the Moving Average (MA) error model, was first proposed in 1981 (Billings and Leontaritis, 1981) and has been subsequently developed by several researchers (Billings & Fadzil, 1985; Billings et al., 1988; Billings et al., 1989; Chen et al., 1989; Chen et al., 1991; Neshat et al., 2018; Swain & Billings, 2001; Wei et al., 2004). The main advantages of this model are that it is easy to implement, flexible to identify the nonlinearity present in the data, and can provide information on the degree of nonlinearity that is present in the system. This study describes two NARX models (NARX1 and NARX2) adapted from Billing (2013), where the first model has only one exogenous variable and the second model has two exogenous variables. Furthermore, a machine learning based nonlinear model, SVR model was used. For comparison purposes, a linear ARX model has also been considered in the study.

2.1. Nonlinear Auto Regressive Model With K Exogenous Variable (NARX)

The NARX model with K exogenous variables can be expressed as,

$$\begin{aligned} y(t) = F[&y(t-1), y(t-2) \dots, y(t-n_y), u_1(t-d_1), u_1(t-d_1-1), \dots, u_1(t-d_1-n_{u_1} \\ &+1), u_2(t-d_2), u_2(t-d_2-1), \dots, u_2(t-d_2-n_{u_2}+1), \dots, u_K(t \\ &-d_K), u_K(t-d_K-1), \dots, u_K(t-d_K-n_{u_K}+1)] + e(t) \end{aligned} \tag{1}$$

where $y(t)$ is the system output variable at time t, $\{u_1, u_2, \cdots, u_K\}$ are the K exogenous/system input variables, $\{d_1, d_2, \cdots, d_K\}$ are the time delay corresponding to each exogenous variable respectively, n_y and $\{n_{u1}, n_{u2}, \cdots, n_{uK}\}$ are the maximum number of lags for the system output and the K system input respectively, $F[\cdot]$ is some nonlinear function to be modeled/estimated, and $e(t)$ is the noise at time t. In this study, the time delays were considered to be equal to zero $d_1 = d_2 = \cdots = d_K = 0$. It needs to be noted that not all the terms present in the model might be significant.

The challenge lies in selecting the nonlinear function $F[-]$ as well as the maximum lags n_y and $\{n_{u1}, n_{u2}, \cdots, n_{uK}\}$. Several model forms have been used for the function $F[\cdot]$ by researchers: polynomial models, radial basis functions, rational models, neural networks, wavelets being a few examples (Billings, 2013). This study considered a polynomial model for the development of the NARX model because of the following advantages (Billings, 2013): (a) polynomials are smooth functions, (b) any continuous function within a closed interval can be approximated by a polynomial function (Weierstrass approximation theorem), hence a wide range of nonlinear systems can be modeled using polynomial NARX, (c) parameters of the model can be estimated efficiently and with less computational cost, and (d) since the model is a polynomial function, it can be easily written in a simple expression, which enables the modellers to interpret the nonlinearity of the underlying system. However, the polynomial form of the model does have limitations in situations where the natural system is severely nonlinear, in which case rational functions or wavelet functions can be used.

Considering $F[\cdot]$ takes the form of a nonlinear polynomial function with ℓ being the highest nonlinear degree of the polynomial, Equation 1 can be represented as (Billings & Voon, 1984),

$$y(t) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(t) + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(t) x_{i_2}(t) + \cdots + \sum_{i_1=1}^{n} \cdots \sum_{i_\ell=i_{\ell-1}}^{n} \theta_{i_1 i_2 \cdots i_\ell} x_{i_1}(t) x_{i_2}(t) \cdots x_{i_\ell}(t) + e(t) \quad (2)$$

where

$$x_m(t) = \begin{cases} y(t-m) & 1 \le m \le n_y \\ u_1(t-(m-n_y+d_1-1)) & n_y+1 \le m \le n_y+n_{u_1} \\ u_2(t-(m-n_y-n_{u_1}+d_2-1)) & n_y+n_{u_1}+1 \le m \le n_y+n_{u_1}+n_{u_2} \\ \vdots & \\ u_K(t-(m-n_y-n_{u_1}-\cdots-n_{u_{K-1}})) & n_y+n_{u_1}+\cdots+n_{u_{K-1}}+1 \le m \le n \end{cases} \quad (3)$$

$$n = n_y + n_{u_1} + \cdots + n_{u_K} \quad (4)$$

and $\theta_{i_1 i_2 \cdots i_m}$ are model parameters. One point to be noted here is that the total number of potential terms M in the NARX model is equal to $M = (n_y + n_{u_1} + \cdots + n_{u_K} + \ell)!/[(n_y + n_{u_1} + \cdots + n_{u_K})! \times (\ell)!]$, which implies that the number of potential terms increases substantially with an increase in the degree of nonlinearity ℓ. In situations where the degree of nonlinearity ℓ is chosen to be unity, the developed model will be a linear model, while a value of ℓ greater than unity indicates the presence of nonlinear terms in the model. Previous studies (Billings, 2013; Billings et al., 1988; Billings et al., 1989; Chen et al., 1989) indicate that the degree of nonlinearity ℓ equal to 2 has the potential to develop NARX models exhibiting considerable output prediction accuracy. However, it should be noted that a higher degree of nonlinearity might be needed in situations where the complexity of the natural system is higher.

The parameters of the NARX model can be estimated using Orthogonal Least Squares (OLS) estimator (Billings, 2013; Billings et al., 1988). The first step is to represent Equation 2 into a function which is linear with the parameters, given by Equation 5.

$$y(t) = \sum_{i=1}^{M} \theta_i p_i(t) + e(t);\ t = 1, \cdots, N \quad (5)$$

where $\{y\ (t),\ t = 1,\cdots, N\}$ is the output sequence; $\{p_i\ (t) = p_i(\mathbf{x}\ (t));\ i = 1,\cdots, M\}$ are regressors from Equation 2 that are formed by combinations of predetermined model variables chosen from vector $\mathbf{x}(t) = [x_1\ (t), x_2\ (t),\cdots,x_n\ (t)]^T$; $\{\theta_i,\ i = 1,\cdots, M\}$ are the model parameters; and $\{e\ (t),\ t = 1,\cdots, N\}$ is the noise/error sequence. In order to solve Equation 5 using the OLS approach, it is necessary to assume that the model parameters are independent of the regressors, that is, $\partial p_i\ (t)/\partial\theta_j = 0\ \forall i = 1,\cdots, M;\ \forall j = 1,\cdots, M$.

Equation 5 can be rewritten in a matrix form as,

$$Y = P\Theta + \mathrm{E} \quad (6)$$

where

$$[Y]_{N\times 1} = [y(1), y(2), \cdots, y(N)]^T$$

$$[\Theta]_{M\times 1} = [\theta_1, \theta_2, \cdots, \theta_M]^T$$

$$[\mathrm{E}]_{N\times 1} = [e(1), e(2), \cdots, e(N)]^T$$

$$[P]_{N\times M} = [\mathrm{p}_1, \mathrm{p}_2, \cdots, \mathrm{p}_M] = \begin{bmatrix} p_1(1) & \cdots & p_M(1) \\ \vdots & \ddots & \vdots \\ p_1(N) & \cdots & p_M(N) \end{bmatrix} \quad (7)$$

The objective of the OLS algorithm is to transform the non-orthogonal regression equation, given in Equation 5, to another regression equation whose regressors are orthogonal to each other. The advantage of this conversion is

that the model parameters $\{\theta_i, i = 1,\cdots, M\}$ can be estimated independently in the transformed orthogonal regression form. The orthogonal regression equation can be expressed as,

$$y(t) = \sum_{i=1}^{M} g_i w_i(t) + e(t);\ t = 1, \cdots, N \tag{8}$$

where $\{w_i\,(t); i = 1,\cdots, M\}$ are orthogonal over the N data points, that is,

$$\sum_{t=1}^{N} w_i(t)w_j(t) = \begin{cases} \sum_{t=1}^{N} w_i^2(t) \neq 0 & i = j \\ 0 & i \neq j \end{cases} \tag{9}$$

Based on OLS algorithm, the orthogonal terms can be estimated as,

$$\begin{cases} w_1(t) = p_1(t) \\ w_2(t) = p_2(t) - a_{1,2}w_1(t) \\ w_3(t) = p_3(t) - a_{1,3}w_1(t) - a_{2,3}w_2(t) \\ \vdots \\ w_m(t) = p_m(t) - \sum_{r=1}^{m-1} a_{r,m}w_r(t); m = 2, 3, \cdots, M \end{cases} \tag{10}$$

where

$$a_{r,m} = \frac{\sum_{t=1}^{N} p_m(t)w_r(t)}{\sum_{t=1}^{N} w_r^2(t)};\ 1 \leq r \leq m-1 \tag{11}$$

Based on Equations 8 and 10,

$$g_i = \frac{\sum_{t=1}^{N} y(t)w_i(t)}{\sum_{t=1}^{N} w_i^2(t)}; i = 1, \cdots, M \tag{12}$$

The model parameters can be estimated from the g_i values as,

$$\begin{cases} \theta_M = g_M \\ \theta_{M-1} = g_{M-1} - a_{M-1,M}\theta_M \\ \theta_{M-2} = g_{M-2} - a_{M-2,M-1}\theta_{M-1} - a_{M-2,M}\theta_M \\ \vdots \\ \theta_m = g_m - \sum_{j=m+1}^{M} a_{m,j}\theta_j; m = M-1, M-2, \cdots, 1 \end{cases} \tag{13}$$

It should be noted that all the M terms in the NARX model (Equation 5) are not significant, and only a few terms are sufficient to obtain very accurate results. The reduction in error corresponding to each term of the NARX model, termed as Error Reduction Ratio (ERR), can be estimated as,

$$ERR_i = \frac{\left[\sum_{t=1}^{N} y(t)w_i(t)\right]^2}{\sum_{t=1}^{N} y^2(t) \times \sum_{t=1}^{N} w_i^2(t)}; i = 1, \cdots, M \tag{14}$$

The error-to-signal ratio (ESR) can be expressed as (Billings, 2013; Billings et al., 1988),

$$ESR = 1 - \sum_{i=1}^{M} ERR_i \tag{15}$$

Based on Equation 15 it can be noted that terms in the NARX model with higher values of ERR_i yields lower ESR. Hence, the terms with the highest ERR_i need to be retained for the selected model. In general, the final NARX model constitutes of M_o terms such that ESR goes below a certain threshold ρ. The threshold in this study is chosen to be 0.01 (Billings, 2013; Billings et al., 1988). It can be noted that if all the terms were considered for model development, the developed model would be susceptible to overfitting. By selecting terms with a high Error Reduction Ratio ensures that the significant terms are retained, while discarding the terms with low ERRs based on an ESR threshold ensures that the model is parsimonious. A higher threshold value leads to fewer terms in the model, while in situations where all the possible terms are considered, the ESR would become zero. The model performance should be evaluated on the validation set alone to ensure that the model is not overfitted. In situations where the model performance is poor in the calibration set, it might indicate that the developed model is under fitted and a lower ESR threshold is needed, while if the model performance is accurate in the calibration set but poor in the validation set, the developed model might be overfitted and higher ESR threshold may improve the overall model performance.

2.2. Support Vector Regression (SVR)

The SVR develops a relationship (Vapnik, 2013) between an input vector (predictor) $\mathbf{x}(t) = [x_1(t),\ldots, x_K(t)] \in \mathfrak{R}^K$, where K denotes the number of predictors, and output (predictand) $y(t) \in \mathfrak{R}$ corresponding to time $t = \{1,2, \ldots, N\}$.

The relationship can be expressed as,

$$y(t) = f(\mathbf{x}(t)) + e(t) \tag{16}$$

where $\{f(\cdot); \mathfrak{R}^K \rightarrow \mathfrak{R}\}$ is a nonlinear transformation function (Vapnik, 2013), and $e(t)$ is white noise whose expected value $E[e(t)]$ is zero.

Consider the function $\phi(\cdot)$ map $\mathbf{x}(t)$ at a higher p-dimensional space, where it is assumed that a linear relationship exists between $\phi(\mathbf{x}(t))$ and $y(t)$. The linear relationship can be expressed as,

$$[y(t)]_{1\times1} = [\phi(\mathbf{x}(t))]_{1\times p}[\boldsymbol{w}]_{p\times1} + [b]_{1\times1} \tag{17}$$

the parameters $[\boldsymbol{w}]_{p\times1}$ and b can be estimated by optimizing the following objective function,

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{e}) = \min_{\boldsymbol{w},b,\boldsymbol{e}} \left[\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\frac{1}{2}\sum_{t=1}^{N}(e(t))^2\right] \tag{18}$$

where

$$e(t) = y(t) - \phi(\mathbf{x}(t))\boldsymbol{w} - b \tag{19}$$

Minimization of the term $\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$ ensures that the model is not over-fitted, while minimization of the term $C\frac{1}{2}\sum_{t=1}^{N}(e(t))^2$ ensures that the model prediction error is not significantly high.

The SVR model can be solved by using the Lagrange multipliers (Vapnik, 2013), where the future projections of the predictand $y(t_f)$ can be obtained as,

$$y(t_f) = \sum_{t=1}^{N} \alpha_t K(\mathbf{x}(t), \mathbf{x}(t_f)) + b \tag{20}$$

where $\mathbf{x}(t_f)$ is the future projection of predictor vector corresponding to time t_f and $K(\mathbf{x}(i),\mathbf{x}(j)) = e^{-\gamma\mathbf{x}(i)-\mathbf{x}(j)2}, \gamma > 0$. Details of the SVR algorithm can be found in Vapnik (2013).

2.3. Auto Regressive Model With Exogenous Variables (ARX)

The linear auto regressive model with one exogenous variable can be defined as,

$$y(t) + a_1 y(t-1) + \cdots + a_{n_y} y(t-n_y) = b_1 u(t-1) + \cdots + b_{n_u} u(t-n_u) + e(t) \tag{21}$$

Details on the solution of the ARX model can be found in Brockwell et al. (1991).

2.4. Performance Evaluation

The performance of each model has been evaluated in terms of five performance measures:

1. Nash-Sutcliffe Efficiency (NSE):

$$NSE = 1 - \frac{\sum_{t=1}^{T} [\hat{y}(t) - y(t)]^2}{\sum_{t=1}^{T} [y(t) - \bar{y}]^2} \tag{22}$$

where y (t) is the observed output variable at time t, $\hat{y}$ (t) is the model predicted value of the output variable at time t, T is the number of daily data points, and $\bar{y} = \sum_{t=1}^{T} y(t)/T$ is the mean value of the observed output variable estimated over the total available records T.

2. Pearson correlation coefficient (r):

$$CORR = \frac{T \sum_{t=1}^{T} \hat{y}(t) \times y(t) - \left[\sum_{t=1}^{T} \hat{y}(t)\right] \times \left[\sum_{t=1}^{T} y(t)\right]}{\sqrt{\left[T \sum_{t=1}^{T} (\hat{y}(t))^2 - \left(\sum_{t=1}^{T} \hat{y}(t)\right)^2\right] \times \left[T \sum_{t=1}^{T} (y(t))^2 - \left(\sum_{t=1}^{T} y(t)\right)^2\right]}} \tag{23}$$

3. Kling-Gupta Efficiency (KGE):

$$KGE = 1 - \sqrt{(CORR - 1)^2 + (a - 1)^2 + (b - 1)^2} \tag{24}$$

where $CORR$ is the Pearson correlation coefficient,

$$a = \sqrt{\frac{T \sum_{t=1}^{T} (\hat{y}(t))^2 - \left(\sum_{t=1}^{T} \hat{y}(t)\right)^2}{T \sum_{t=1}^{T} (y(t))^2 - \left(\sum_{t=1}^{T} y(t)\right)^2}}, \; and \; b = \sqrt{\frac{\sum_{t=1}^{T} \hat{y}(t)}{\sum_{t=1}^{T} y(t)}} \tag{25}$$

4. Relative bias (Rbias):

$$Rbias = \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{y}(t) - y(t)}{y(t)} \times 100 \tag{26}$$

5. Relative Root Mean Square Error (RRMSE):

$$RRMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\hat{y}(t) - y(t)}{y(t)}\right)^2} \times 100 \tag{27}$$

the value of NSE and KGE ranges from [−∞,1], r ranges from [−1,1], while Rbias ranges from [−∞,∞] and RRMSE ranges from [0,−∞]. The values of NSE, KGE and r tend to unity while values of Rbias and RRMSE tends to zero as model predictions become closer to the observed values.

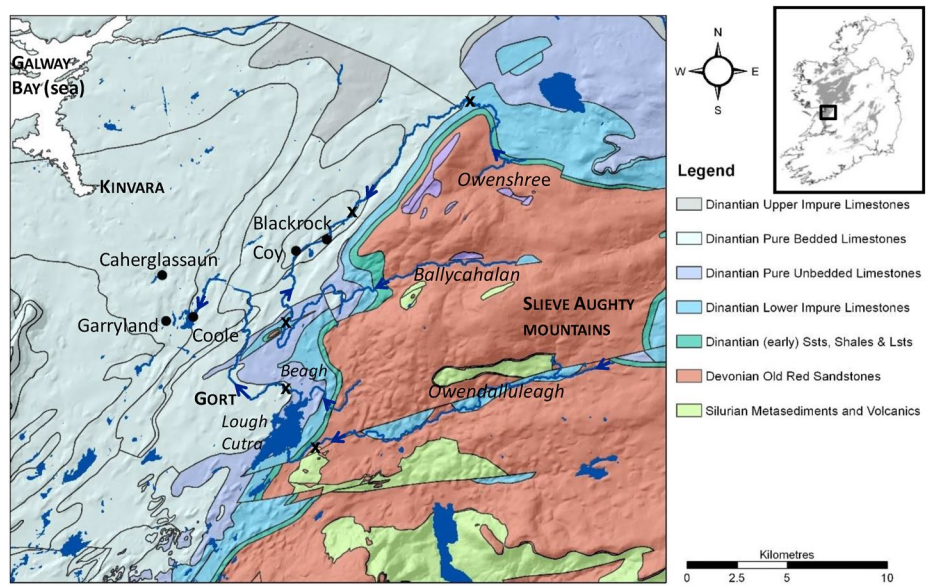

Figure 1. Geology and relief map of study area showing five turloughs (•), the main rivers draining the Slieve Aughty mountains (gauging stations marked **X**) and the main spring at Kinvara (inset shows limestone geology of Ireland and location of study area) (Gill et al., 2013).

3. Methods—Flood Data

3.1. Catchment Description

The karst aquifer is in Carboniferous limestone bedrock in a lowland setting in the west of Ireland, situated less than 100 m above the current sea level (Drew, 2018). The area, a catchment of approximately 500 km^2 receives allogenic runoff from about one-third of the catchment area from three main rivers draining the Old Red Sandstone Slieve Aughty mountains (see Figure 1). The lowland karst aquifer is fed by these sinking streams which disappear into underground fissures and conduits and then frequently reappear again in surface reaches or as turloughs in glacially formed depressions. The drainage takes the flows underground to the north-west to the Atlantic Ocean at Kinvara through a complex multi-level conduit system in this lowland network that has formed as a result of past glaciation cycles and their impact on karstification processes (Naughton et al., 2018).

The turloughs flood in winter providing temporary storage for huge volumes of water, forming a key component in the hydrogeological regime. They behave as surge tanks in the system, attenuating flow in the subterranean conduit network. In summer, however, these turloughs normally dry out. Turloughs are designated a Priority Habitat in Annex 1 of the European Habitats Directive (92/43/EEC) as they provide a habitat for many protected flora and fauna species. Equally, under the Water Framework Directive (2000/60/EC) turloughs are designated as groundwater dependent terrestrial ecosystems (GWDTEs).

3.2. Karst Network and Groundwater Flooding

3.2.1. Field Data

Continuous water level data at an hourly time scale were collected using pressure transducers with in-built data-loggers at the base of five key turloughs (Blackrock, Coy, Coole, Garryland, and Caherglassaun) in the lowland karst catchment between 2007 and 2018. This was initiated by an integrated multi-disciplinary research interest in the ecohydrology of the ephemeral wetlands (Naughton et al., 2012; Porst et al., 2012; Waldren et al., 2015) and then a groundwater flood study (Morrisey et al., 2020).

Depth-area-volume relationships for the turloughs and linked floodplains were derived from LiDAR mapping data for the catchment in the form of a Digital Elevation Model (DEM) with a grid spacing of 2 m and a vertical accuracy of ±0.15 m. The depth-area-volume relationship was developed based on a third order polynomial to obtain optimal fit of the data. Where such data was not available, further topographical survey data was obtained

from manual surveys carried out using a Trimble 4700 GPS system with a minimum accuracy of 0.01 m (horizontal and vertical direction). These topographical data were combined with the available LiDAR data in ArcGIS and a new integrated DEM was constructed using the Kriging method with a 2 m grid spacing.

Rainfall data were collected from two tipping bucket rain gauges positioned at 70 m above ordnance datum (mAOD) and 150 mAOD in the catchment to assess the spatial distribution of rainfall. These data were then related to the Gort Derrybrien gauge operated by Met Eireann, which provided a longer data set, in order to fill missing gaps, albeit only at a daily frequency.

Tidal levels in Kinvara Bay (within Galway Bay) into which the system drains (i.e., the downstream boundary condition for the model), were obtained at 15-min intervals from the Marine Institute of Ireland's Galway Port tide gauge which is located c.15 km to the north.

The allogenic flows into the lowland karst network from the three main rivers draining off the Slieve Aughty hills were constantly gauged at locations on the edge of the sandstone before disappearing into the karst limestone via pressure transducers set into the bed of the rivers operated by the Office of Public Works (OPW). Rating curves (i.e., flow against stage relationship) were developed over time for each of the three gauging stations on the rivers flowing off the mountains.

3.2.2. Hydraulic Model Development

A semi-distributed 1D model of the catchment was developed in the Infoworks ICM (Innovyse) urban drainage software to simulate groundwater-surface water interaction for turloughs (see Morrisey et al., 2020 for more details) and more specifically for this study, to generate a complete time series of the flooded volume data as described in Section 3.2.3. This was based on a previous 1D semi-distributed model of the five turloughs due to its ability to model the hydraulics of the karst conduit network in both open channel and pressurized pipe flow (Gill, Naughton, & Johnston, 2013). The groundwater-surface water turlough dynamics were modeled as storage ponds in the software which were configured with the same depth-volume characteristics as the surface topography, as derived from the DEM. Diffuse recharge from rainfall is modeled per sub-catchment via a series of reservoirs: rainfall-runoff, soil, and groundwater stores in series, all yielding different delayed discharges in parallel into the pipe network. All flows discharge into permeable pipes, one connected for each sub-catchment to represent the primary and secondary permeability. The allogenic recharge into the karst network from the rivers was input as point flow time series into the head of the pipe network. These data were derived by three separate rainfall-runoff models one for each river, using the MIKE11-NAM software (DHI Software). Certain key floodplains were also included as 2-D areas into the model to allow for overland flow between turloughs/floodplains which was not previously accounted for and therefore provided an accurate representation of flooding mechanisms across the whole catchment. The final 1-D/2-D model network was developed on the back of previous field investigations (tracer studies, caving records etc.), with further insights into the system was gained from the accumulating hydro-meteorological data across the catchment as well as water hydrochemistry (Gill et al., 2018). Time and frequency series analyses on the continuous water level measurements of the five turloughs in the linked network were used in order to elucidate the nature of the hydraulic pipe configurations at key points in order to improve the conceptual model (Gill, Naughton, Johnston, Basu, & Gosh, 2013).

3.2.3. Data Used for Non-Linear Flood Forecasting Model Development

The Flooded Volume (FV) data from the five turloughs (Coole, Caherglassaun, Garryland, Blackrock, and Coy), collected between 1 June 2007 to 31 March 2018 were converted to a daily scale by cumulating the respective data. The Total Flooded Volume (TFV) was then obtained by cumulating the FV for each of the five turloughs. Three of the five turloughs (Coole, Caherglassaun, and Garryland) are located at a similar low elevation within 7 km of the spring at the coast (with their bases between 0.9 and 1.8 mAOD); Blackrock and Coy turloughs are closely located at a more distant upper part of the catchment (within 12 km of the coast) with their bases between 8.2 and 11.0 mAOD and receive more direct runoff from the rivers draining the Slieve Aughty mountains. The locations of the five turloughs are shown in Figure 1. Rainfall data, at an hourly timescale, were also converted to a daily scale by cumulating the respective data. Subsequently, the rainfall anomaly was estimated by substracting the long-term mean rainfall from daily rainfall values. The mean subtracted rainfall has been used in the paper for further analysis. The maximum difference in tidal amplitude per day was determined from the tidal data collected at 15 min intervals.

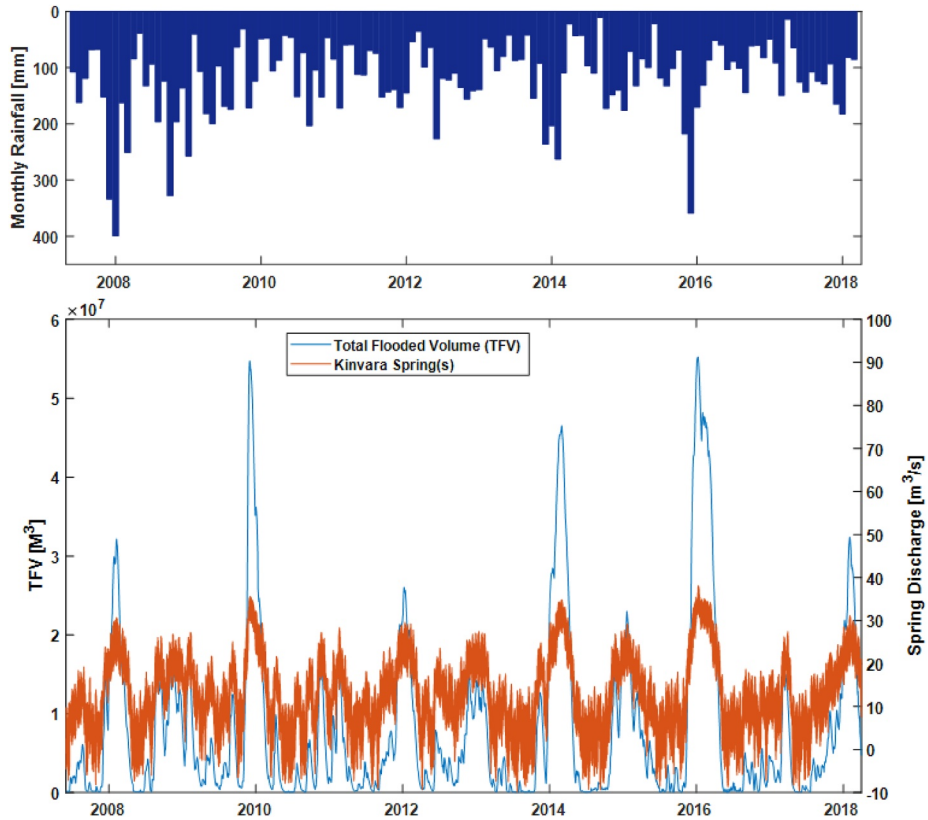

Figure 2. Plot of total flooded volume at the study turloughs (Blackrock, Coy, Coole, Garryland and Caherglassaun) and simulated spring discharge at Kinvara over the karst model calibration period (2007–2018).

4. Results

4.1. Karst System Groundwater Flooding Dynamics

The overall flooding dynamics of the lowland karst system between 2007 and 2018, as simulated using the Infoworks ICM hydraulic model, is shown in Figure 2. This shows the response of the total flooded volume in the catchments to rainfall, as well as the simulated final discharge from the spring out to Kinvara bay. In the northern temperate climate of Ireland, the seasons are defined as follows: Spring (March, April May), Summer (June, July, August). Autumn (September, October, November), Winter (December, January, February). The periodic fluctuations of the final spring discharge every day are caused by the impact of the tides on the outlet flows. The performance of the Infoworks model to predict the water levels in the individual five turloughs is shown in Figure S1 in Supporting Information S1, with NSEs and KGEs of >0.9 for all turloughs: a more extensive evaluation of the same hydraulic model over a longer time period can be found in Morrissey et al. (2020). Hence, whilst this semi-distributed model now works well to simulate the flooded volumes (and areas) of the catchment, it has taken several years to develop (requiring a lot of investigative studies to establish the network morphometry), requires detailed input data, and takes several days to run a simulation in 1D/2D and so would be impractical/unwieldy for use as a flood forecasting tool, particularly if developed from scratch every time.

This data was then used initially to define the nature of the relationship between the key variables: rainfall, TFV, and tidal amplitude, which has been used to develop the nonlinear time series and machine learning algorithms for forecasting groundwater flooding. These models are much quicker to run and easier to implement onto, for example, an online web portal using real-time water level, rainfall and tidal data which can run automatically without day-to-day additional desk-based processing requiring human input.

A system with a response function $y\,(t)$ can be considered nonlinear (Billings and Voon, 1984) when the cross-correlation function between $y'\,(t)$ and $y'\,(t)^2$ for any lag τ are nonzero (Equation 28), where $y'(t) = y(t) - \overline{y(t)}$.

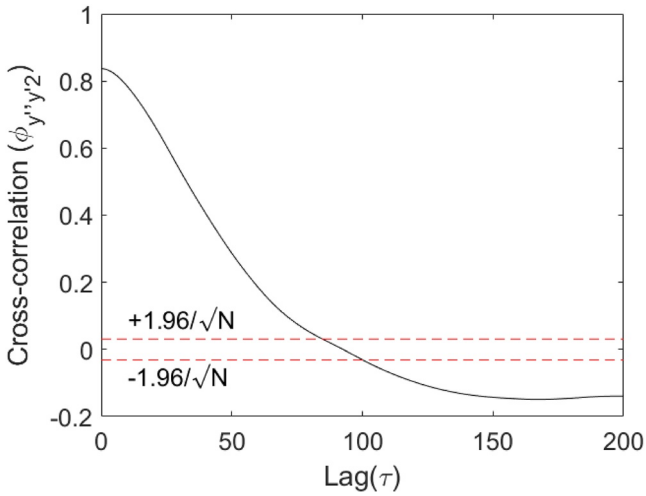

Figure 3. Nonlinearity detection function of total flooded volume.

$$\phi_{y',y'^2}(\tau) = \frac{\sum_{t=1}^{N-\tau}\left[y'(t)-\overline{y'^{(t)}}\right]\times\left[(y'(t+\tau))^2-\overline{(y'(t))^2}\right]}{\sqrt{\sum_{t=1}^{N}\left[y'(t)-\overline{y'^{(t)}}\right]}\times\sqrt{\sum_{t=1}^{N}\left[(y'(t))^2-\overline{(y'(t))^2}\right]}}$$

$$= \begin{cases} 0,\ \forall\tau = 0,1,2,\cdots & \Rightarrow linear\ system \\ \neq 0,\ \forall\tau = 0,1,2,\cdots & \Rightarrow nonlinear\ system \end{cases} \quad (28)$$

For large data (considerable high value of N), the 95% confidence interval can be chosen as $\pm 1.96/N$. The correlation function $\phi_{y',y'^2}(\tau)$ for the TFV up to 200 lags (Figure 3) indicates that the function does not converge to zero, indicating that the system generating the TFV is nonlinear.

The auto-correlation function of the TFV and the cross-correlation function between the TFV and the rainfall and tidal magnitude corresponding to different lags were estimated (Figure S2 in Supporting Information S1). The auto-correlation function of the TFV (Figure S2a in Supporting Information S1) indicates that the function decays at a very slow rate, indicating that the TFV is a long-memory process. For practical purposes, the present day's TFV is assumed to be dependent on the past 5 day's TFV. A higher lag can lead to a complex model and might not be necessary. The cross-correlation function between the TFV and rainfall (Figure S2b in Supporting Information S1) indicates that the correlation becomes insignificant after 17 days of lag. However, a threshold correlation value of 0.08 was considered as the minimum correlation between rainfall and TFV for the analysis, which leads to the selection of the rainfall lag up to the past 4 days. The cross-correlation function between TFV and tidal amplitude (Figure S2c in Supporting Information S1) was found to be insignificant. However, the tidal amplitude up to the past 4 days was also considered, similar to the rainfall. It needs to be noted that the auto- and cross-correlation functions measure the linear dependence between two random variables. In situations where the system is nonlinear, a physical understanding of the system needs to be considered along with the correlation functions.

4.2. NARX Models With Either Rainfall or Both Rainfall and Tidal Amplitude as the Exogenous Variable(s) and Total Flooded Volume (TFV) as Endogenous/Output Variable

The daily TFV data were subdivided into two sets, the calibration set, and the validation set for modeling. Data ranging from 1 June 2007 to 1 January 2015 (2,772 days consisting of initial 70% data) was considered as the calibration set data, data ranging from 2 January 2015 to 31 January 2016 (395 days consisting of 10% data) was considered as the training set and were used for model development and model parameter estimation, while the remaining 20% data (790 days from 1 February 2016 to 31 March 2018) was used to evaluate the model performance in terms of performance measures. It needs to be noted that the NARX model parameters are estimated based on the Error to Signal Ratio (ESR) value, and hence does not require the training set data. However, the training set data are necessary to identify the optimal model parameters of the SVR model.

Based on an understanding of the physical karst network and the autocorrelation function of TFV and the lagged cross-correlation function between TFV and rainfall, and TFV and tidal amplitude (see Section 4.1), the TFV for a given day was assumed to be dependent on the past TFV, the past and present day's rainfall and tidal amplitude. The TFV for a given day was considered to be dependent on the TFV of the previous 5 days from that day, and the rainfall and tidal amplitude of that day and the previous 4 days. The parameters in Equation 3 for the developed NARX model are: $n_y = 5; K = 2;\ d_1 = d_2 = 0;\ n_{u1} = n_{u2} = 5$. The input/exogenous variable rainfall is denoted by u_1 while the tidal amplitude is denoted by u_2. Henceforth, the NARX model was indicated as NARX2 to reflect that two exogenous variables were considered to develop the nonlinear model. The total number of possible terms increases from 136 to 816 when the degree of nonlinearity ℓ was increased from 2 to 3. Considering that 2,772 data points are available in this study to develop the model, the degree of nonlinearity was chosen to be 2. Furthermore, literature reveals that NARX models with a degree of nonlinearity equal to 2 usually have the potential to be sufficiently accurate (Billings, 2013). In situations where the performance of the developed model is poor, it would be necessary to develop a more complex model with a higher degree of nonlinearity or consider a rational or wavelet model instead of a polynomial model.

Table 1
Three Performance Evaluation Measures Nash-Sutcliffe Efficiency (NSE), Pearson Correlation Coefficient (r), and Kling-Gupta Efficiency (KGE) in D-Days Ahead Prediction in the Validation Period, Where Values Denoted in Bold Denotes the Best Result Out of the Three Developed Models (NARX2, SVR and ARX)

Error	Days	1	2	3	4	5	7	10	15	20	25	30	45	60	90
NSE	NARX2	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**	**0.98**	**0.97**	**0.96**	**0.95**	**0.92**	**0.91**	**0.90**
	SVR	0.98	0.97	0.96	0.94	0.92	0.90	0.86	0.83	0.80	0.77	0.75	0.71	0.67	0.65
	ARX	1.00	0.99	0.99	0.99	0.99	0.98	0.96	0.93	0.90	0.87	0.83	0.74	0.67	0.50
r	NARX2	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**	**0.99**	**0.98**	**0.98**	**0.97**	**0.96**	**0.96**	**0.96**
	SVR	0.99	0.98	0.98	0.97	0.97	0.96	0.94	0.93	0.92	0.91	0.91	0.91	0.92	0.93
	ARX	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.97	0.96	0.94	0.93	0.88	0.84	0.73
KGE	NARX2	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**0.97**	**0.95**	**0.93**	**0.89**	**0.85**	**0.81**
	SVR	0.96	0.95	0.93	0.91	0.88	0.84	0.79	0.75	0.71	0.68	0.65	0.59	0.54	0.50
	ARX	0.99	0.98	0.98	0.98	0.97	0.97	0.96	0.93	0.91	0.90	0.88	0.86	0.83	0.72

Note. SVR, support vector regression; NARX, nonlinear auto regressive model with K exogenous variables; KGE, Kling-Gupta efficiency; ARX, auto regressive model with exogenous variables; NSE, Nash-Sutcliffe efficiency.

The developed NARX2 model can be expressed as:

$$\begin{aligned} y(t) = {} & 1.7127 \times y(t-1) - 0.8462 \times y(t-2) + 0.3328 \times u_1(t-1) + 0.0063 \\ & \times u_1(t-2)^2 + 0.3135 \times u_1(t-2) + 0.0024 \times y(t-1) \times u_1(t-1) \\ & - 0.0021 \times y(t-4) \times u_1(t-1) + 0.0243 \times y(t-5) + 0.2841 \\ & \times y(t-3) - 0.1800 \times y(t-4) + 0.0081 \times u_1(t-1) \times u_1(t-2) \\ & + 0.4860 + e(t) \end{aligned} \tag{29}$$

In situations where the tidal amplitude is not available or the turlough is located far from the ocean where the effect of the tide is not significant, the NARX model can be developed by using rainfall as the sole exogenous variable, termed as the NARX1 model.

To evaluate the performance of the NARX2 model using the performance measures, the developed model was used to predict TFV at different time periods in the future corresponding to the validation set and the model predictions were compared with the observed TFV. The projections were obtained for 1, 2, 3, 4, 5, 7, 10, 15, 20, 25, 30, 45, 60, and 90 days ahead. Three performance measures NSE, KGE and *r* were provided in Table 1 and the error in projected time series plots along with the observed time series are shown in Figure 4. The error is estimated as the difference between model predicted TFV from the observed TFV for the validation period. A positive error indicates that the model is overpredicting the TFV, whereas a negative value denotes the model underpredicts the TFV for the hosen day. Figure 5 provides the difference between the model predicted TFV and the observed TFV corresponding to the entire validation period in terms of boxplots. A positive (negative) error value indicates that the model is overpredicting (underpredicting) the TFV. It needs to be noted that the observed TFV during the validation period ranges from 1.007×10^6 to 1155.35×10^6 m^3 per day, with a mean TFV of 208.8×10^6 m^3 per day. Low observed values are susceptible to relative performance measures such as Rbias and RRMSE, hence those two measures were discarded while estimating the performance measure for the entire validation period. However, the primary objective of developing the model is to estimate its effectiveness in quantification and prediction of peak flows. The observed TFV in the validation period was found to be less than 10×10^6 m^3 per day for 151 days (19.1%), less than 50×10^6 m^3 per day for 298 days (37.7%), with only 243 days (30.76%) exhibit TFV greater than the average TFV in the validation period. All of the five performance measures were estimated by considering those days in the validation period where the TFV was greater than 50×10^6 m^3 per day. The performance measures are provided in Table 2.

The robustness of the developed model (Equation 29) is investigated by performing an uncertainty analysis by perturbing the estimated model parameters. For this purpose, the lower and upper range of each of the 12 model parameters corresponding to 90% confidence interval (5% significance level on either side) were estimated. The values of the range of those model parameters are provided in Table S1 in Supporting Information S1. Subsequently, the lower and upper bound of each of the parameters has been used to develop the NARX model. The developed model can then be used to predict the TFV d-days ahead. It needs to be noted that in situations where

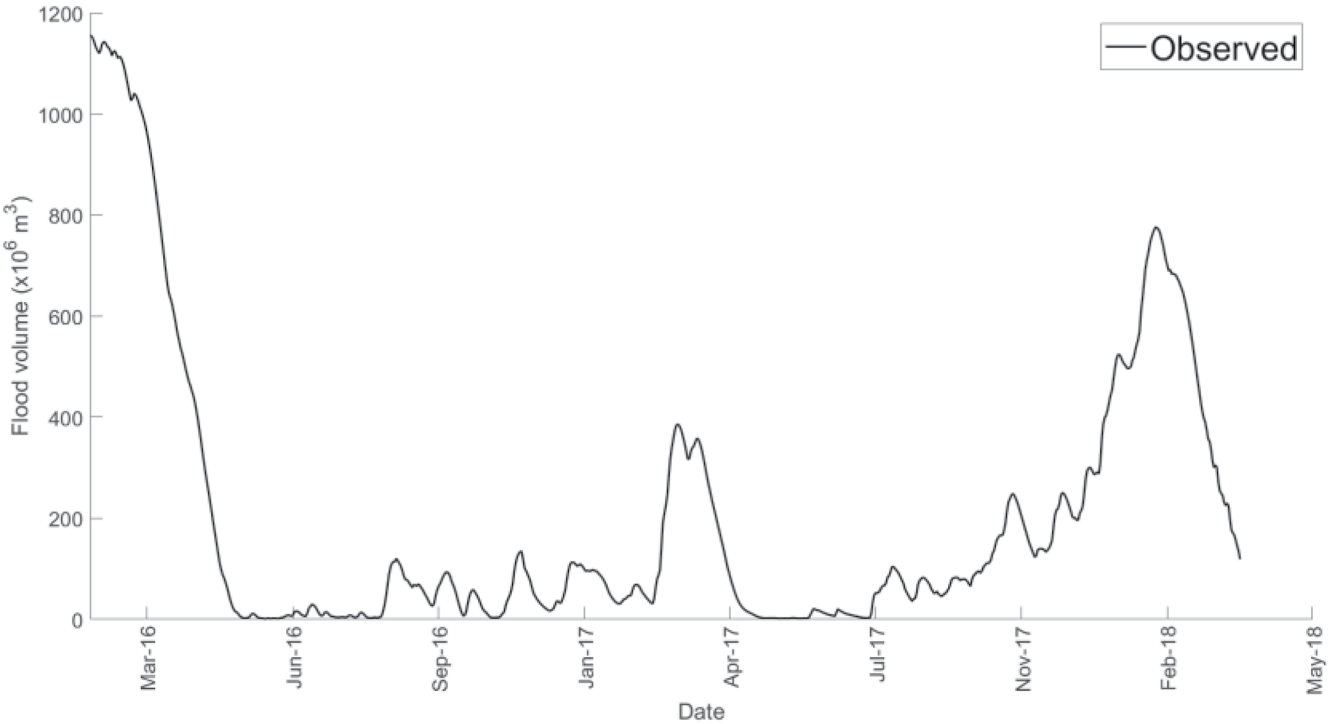

Figure 4. Observed time series of total flooded volume and error in model prediction, where a positive error indicate model is over-predicting and a negative error indicate under-predicting

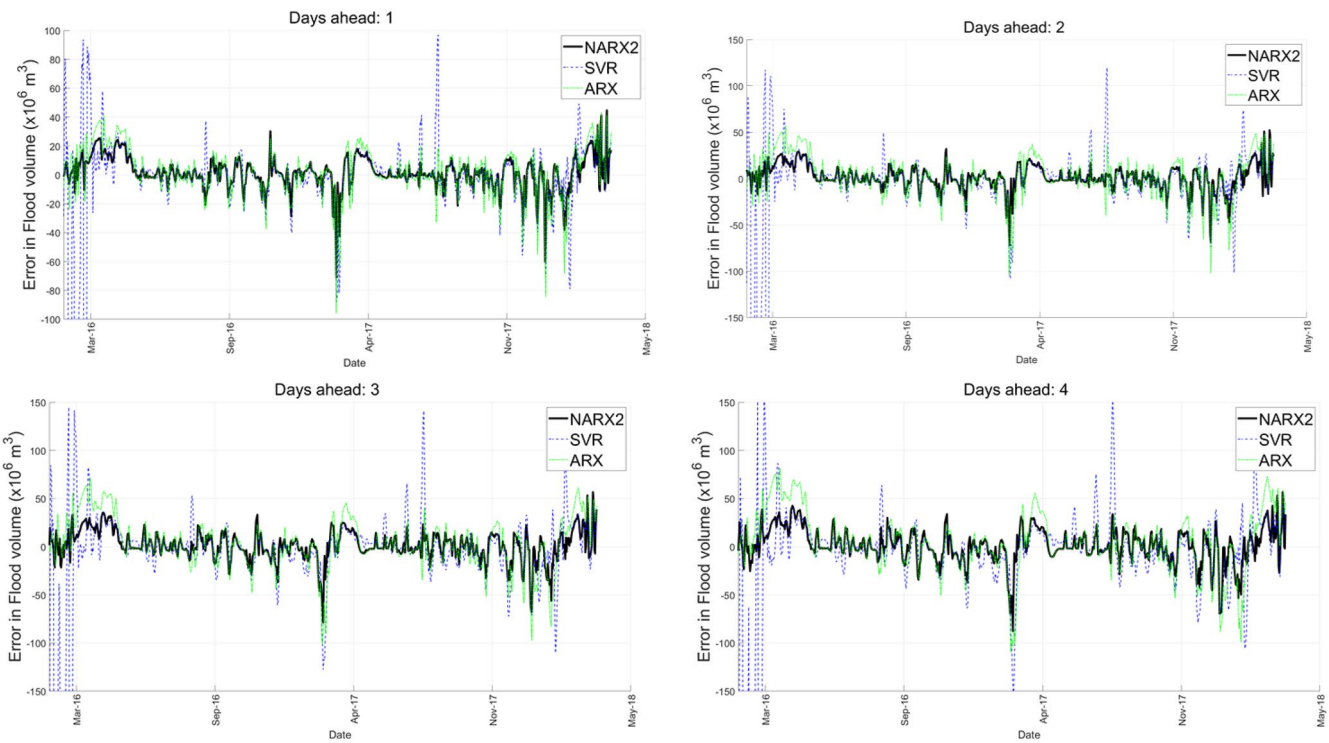

Figure 4. (Continued)

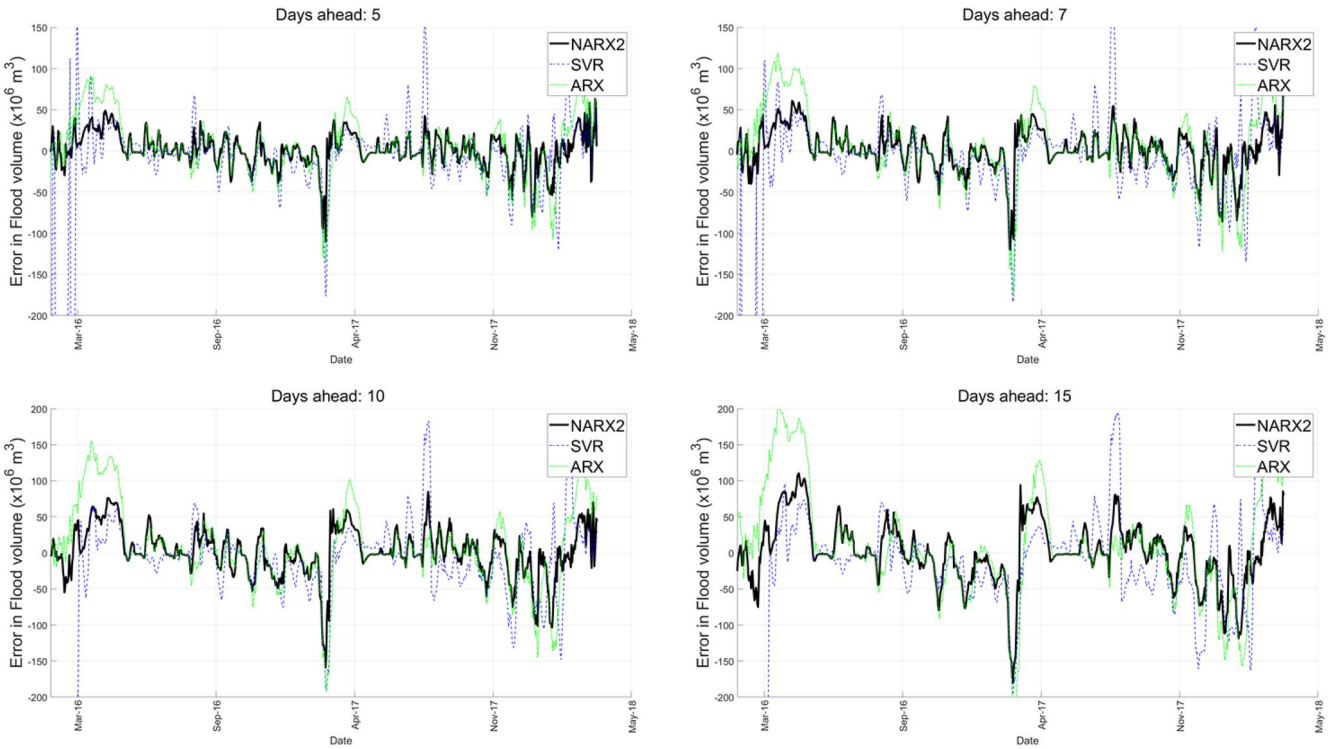

Figure 4. (Continued)

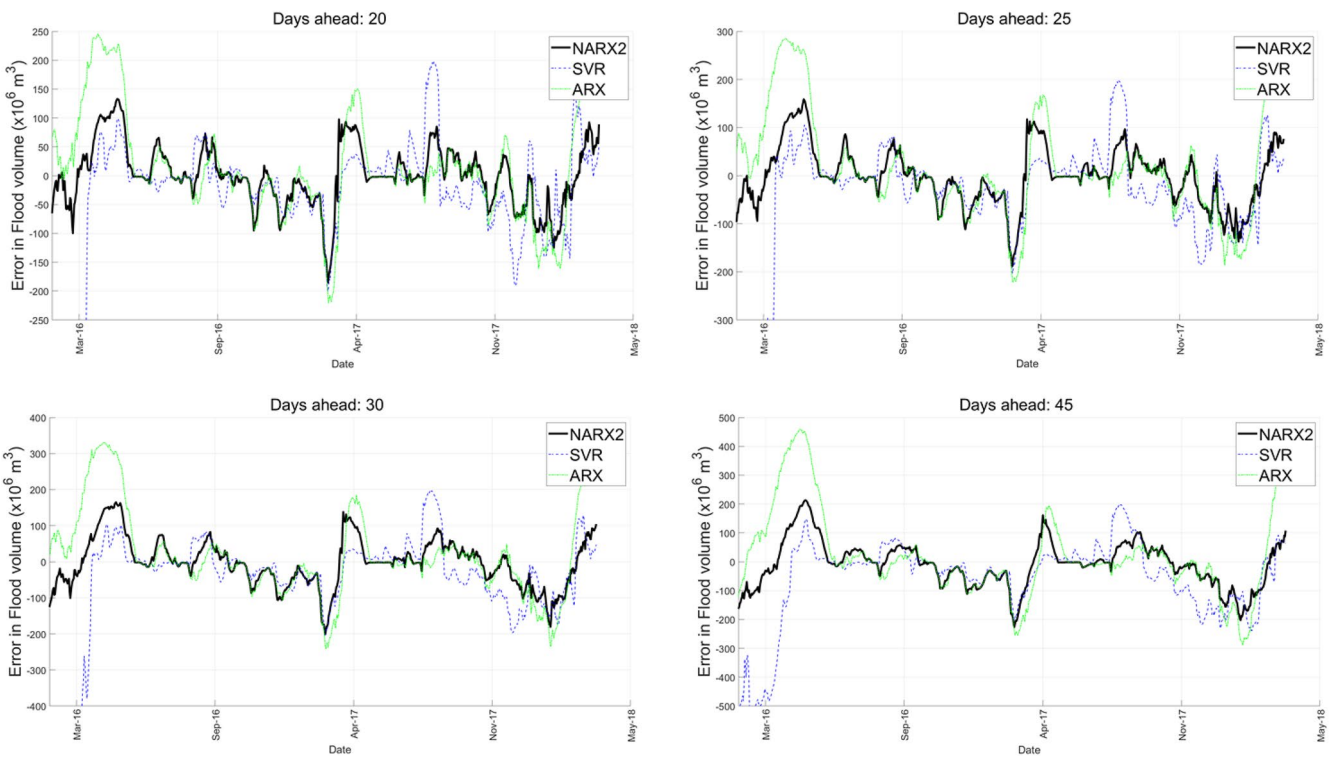

Figure 4. (Continued)

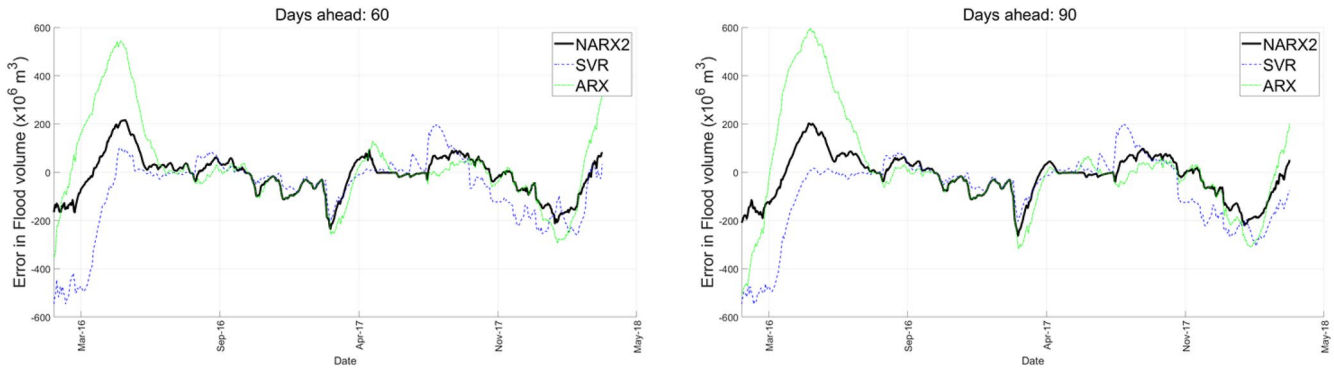

Figure 4. (Continued)

lower and upper bounds of all of the 12 model parameters were considered, a total of $2^{12} = 4,096$ combinations of model parameters can be generated, leading to 4,096 models. However, it should be noted that the Error Reduction Ratio (ERR) is high for the initial set of parameters and reduces considerably for the other parameters. It has been noted that the ERR was found to be high for the initial 5 terms, hence lower and upper bounds of the first 5 parameters has been considered while the model predicted values of the remaining 7 parameters were chosen for the simulation study. Since lower and upper bounds of only 5 parameters were chosen, a total of 32 (2^5) parameter combinations and models were generated to perform the model parameter uncertainty analysis. Once the 32 models were generated, each of those models was used to simulate the d-day ahead TFV for the validation period. The developed time series was plotted along with the observed and model predicted TFV in Figure 6. The figure provides the range of expected error when the top 5 important parameters are allowed to have an estimation error corresponding to 10% significance level. The figure indicates that the range of error is very low up until 10 days from which point the error starts to increase.

To understand the effectiveness of the nonlinear terms in model performance, a another NARX model has been developed where the degree of nonlinearity was chosen to be 1. This ensures that there are no nonlinear terms

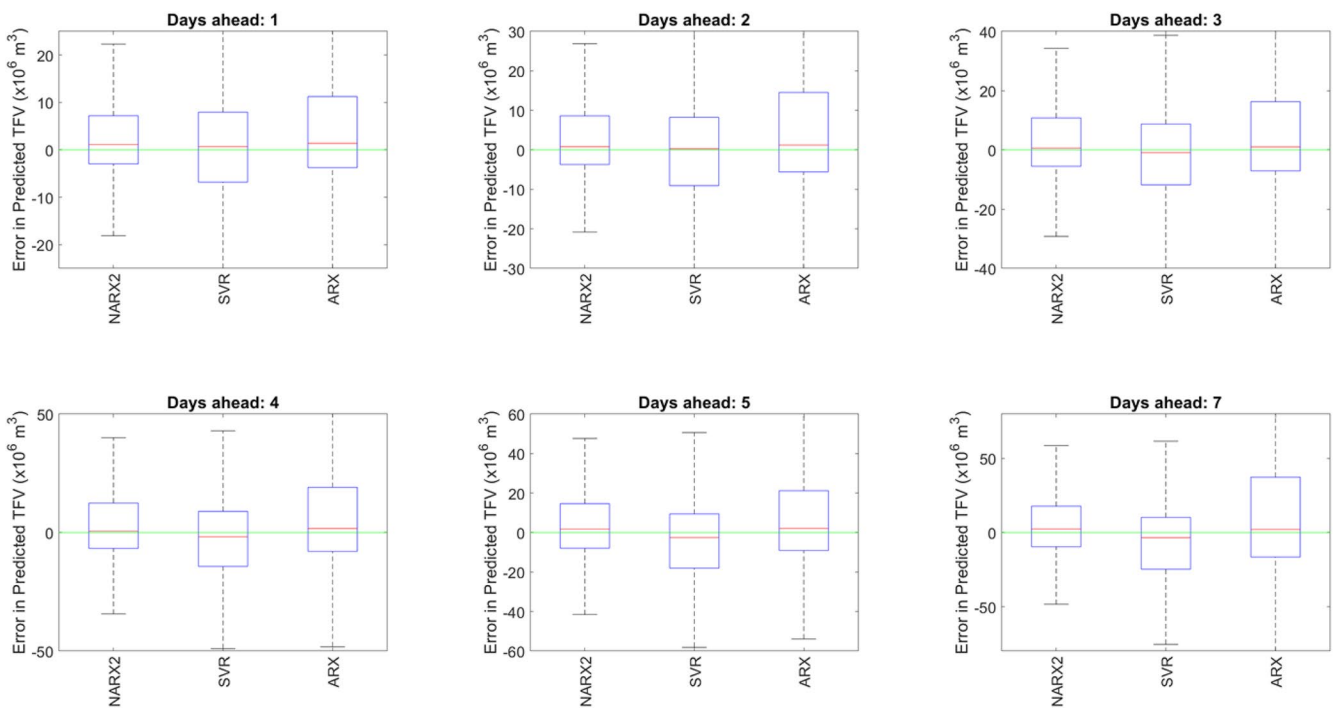

Figure 5. Boxplots of prediction error obtained using different models.

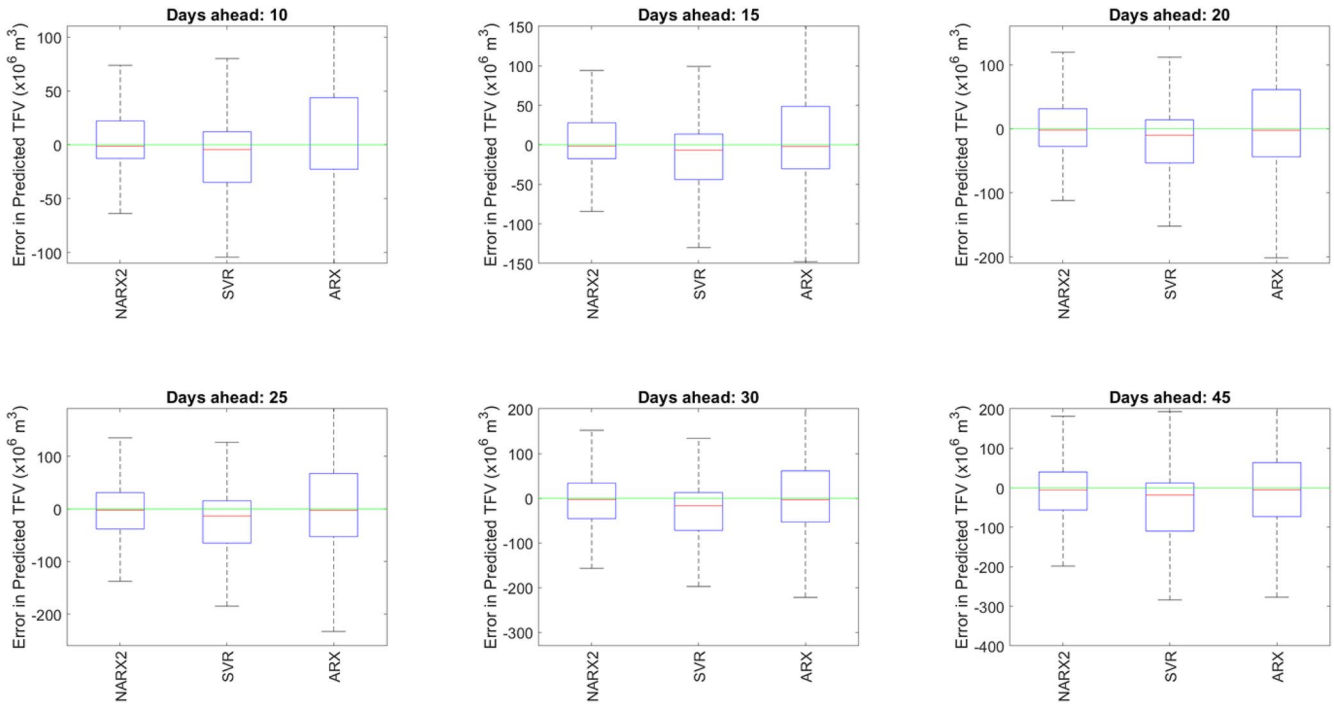

Figure 5. (Continued)

in the developed model. The linearized NARX model with two exogenous variables is termed as LNARX2. The linear model can be expressed as,

$$\begin{aligned} y(t) = {} & 1.7974 \times y(t-1) - 0.9496 \times y(t-2) + 0.4922 \times u_1(t-2) + 0.1181 \times u_1(t) \\ & + 0.1189 \times u_2(t) + 0.0354 \times y(t-5) + 0.3153 \times y(t-3) - 0.2023 \\ & \times y(t-4) - 0.0275 \times u_1(t-3) - 0.0213 \times u_1(t-4) + 0.0760 \\ & \times u_2(t-1) + e(t) \end{aligned} \tag{30}$$

Performance of the model has been provided in Tables S2 and S3 in Supporting Information S1 for comparison with the NARX2 model, while the errors are plotted in Figures S3 and S4 in Supporting Information S1.

4.3. SVR and ARX Models With Rainfall and Tidal Amplitude as Exogenous Variables

The machine learning based (SVR) nonlinear model was developed using the methodology described in Section 2.2. The modeling was performed by using the LS-SVMlab1.8 software package obtained from the

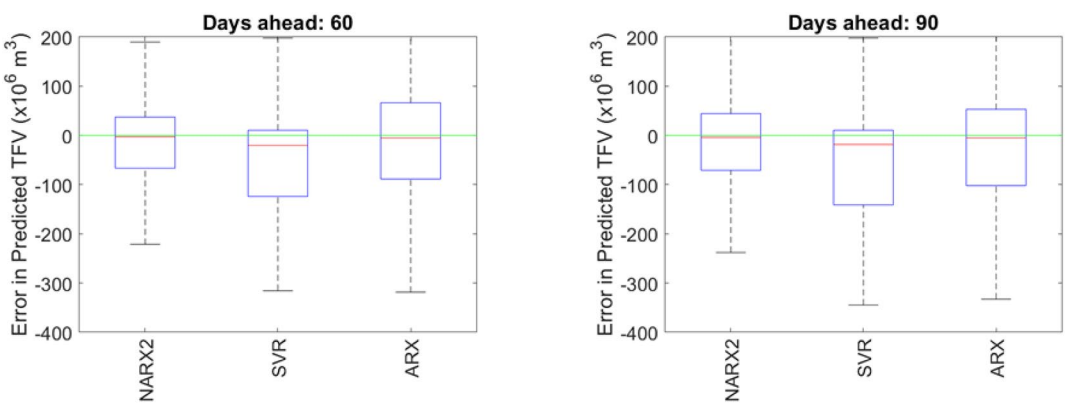

Figure 5. (Continued)

Table 2
Performance Measures NSE, r, KGE, Rbias and RRMSE in the Validation Period Excluding Days Where Observed Flows Were Less Than 50 × 10^6 m^3/Day

Error	Days	1	2	3	4	5	7	10	15	20	25	30	45	60	90
NSE	NARX2	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.97	0.96	0.95	0.94	0.91	0.90	0.88
	SVR	0.97	0.96	0.94	0.92	0.90	0.86	0.82	0.76	0.72	0.69	0.66	0.61	0.56	0.52
	ARX	1.00	0.99	0.99	0.99	0.98	0.97	0.95	0.92	0.88	0.84	0.80	0.72	0.67	0.57
r	NARX2	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.97	0.97	0.96	0.96	0.96
	SVR	0.98	0.98	0.97	0.96	0.95	0.94	0.92	0.90	0.89	0.88	0.88	0.89	0.90	0.91
	ARX	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.96	0.95	0.93	0.92	0.88	0.85	0.78
KGE	NARX2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.97	0.95	0.93	0.88	0.83	0.79
	SVR	0.95	0.94	0.92	0.89	0.86	0.82	0.77	0.72	0.67	0.63	0.59	0.52	0.45	0.41
	ARX	0.99	0.98	0.98	0.97	0.97	0.96	0.95	0.93	0.91	0.89	0.88	0.88	0.88	0.78
Rbias	NARX2	0.62	0.85	1.16	1.44	1.81	2.60	3.38	4.54	4.96	4.42	3.91	3.90	1.85	0.34
	SVR	−0.98	−1.72	−2.58	−3.52	−4.50	−6.30	−8.63	−9.81	−9.77	−10.56	−12.02	−10.49	−12.52	−13.22
	ARX	1.08	1.54	1.91	2.35	2.87	3.72	5.01	6.89	7.80	8.25	9.18	8.10	4.58	−1.23
RRMSE	NARX2	8.69	10.30	12.44	14.99	17.59	22.68	28.21	36.63	44.67	50.34	54.52	65.36	63.70	62.92
	SVR	16.93	20.80	25.28	30.29	35.02	41.90	49.84	58.07	65.30	70.82	73.42	79.61	80.20	80.63
	ARX	11.91	15.68	18.35	20.90	23.50	29.06	36.47	49.46	60.69	69.79	78.34	99.57	109.30	113.25

Note. SVR, support vector regression; NARX, nonlinear auto regressive model with exogenous variables; KGE, Kling-Gupta efficiency; ARX, auto regressive model with exogenous variables; NSE, Nash-Sutcliffe efficiency; RRMSE, relative root mean square error.

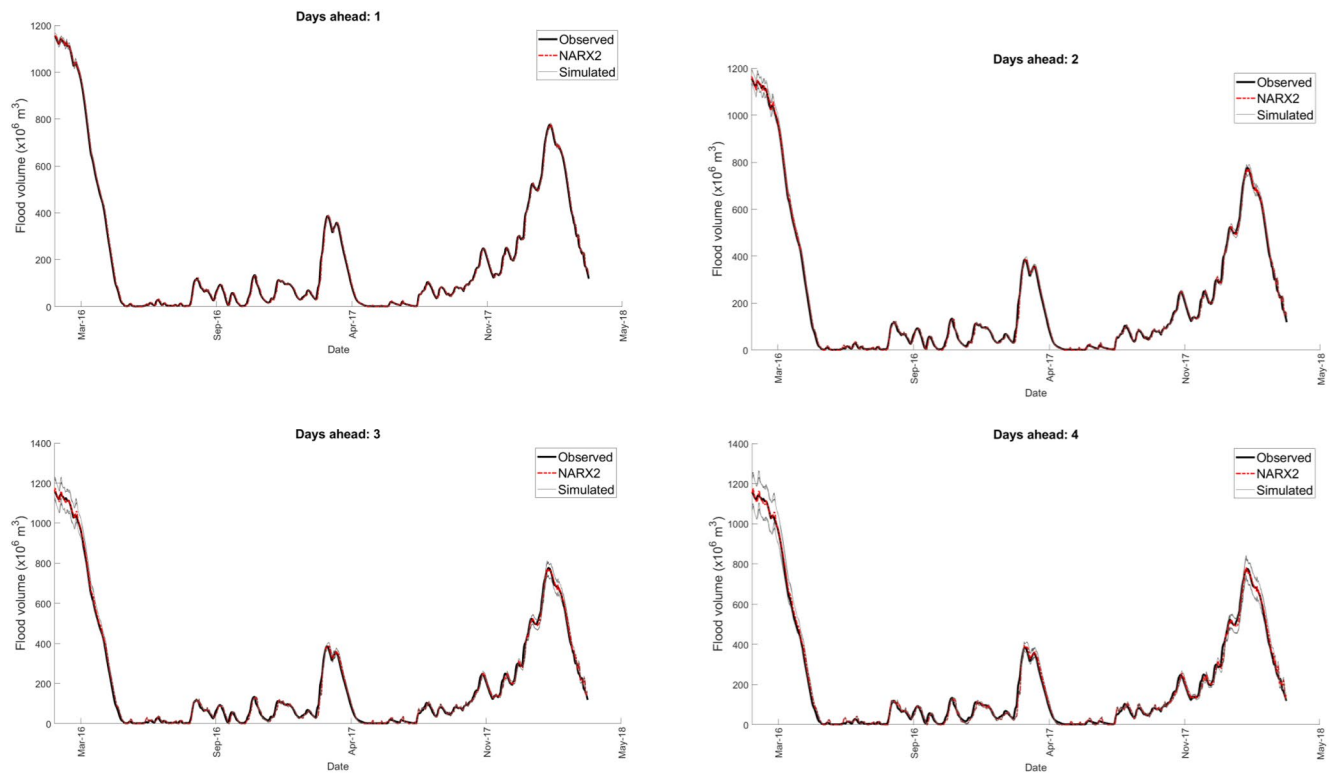

Figure 6. Range of d-days ahead total flooded volume projections based on the simulation study performed by considering the lower and upper bound of the most important five model parameters of (nonlinear auto regressive model with K exogenous variable 2 model.

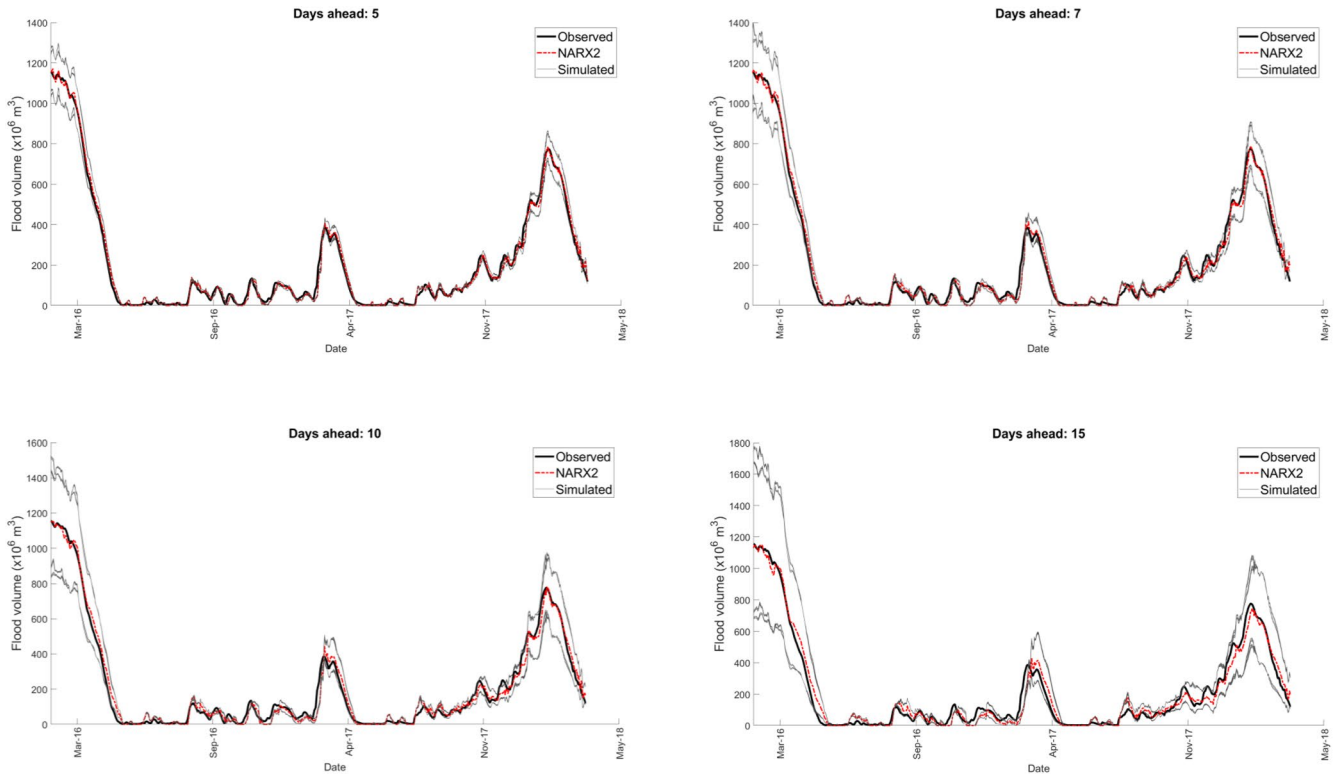

Figure 6. (Continued)

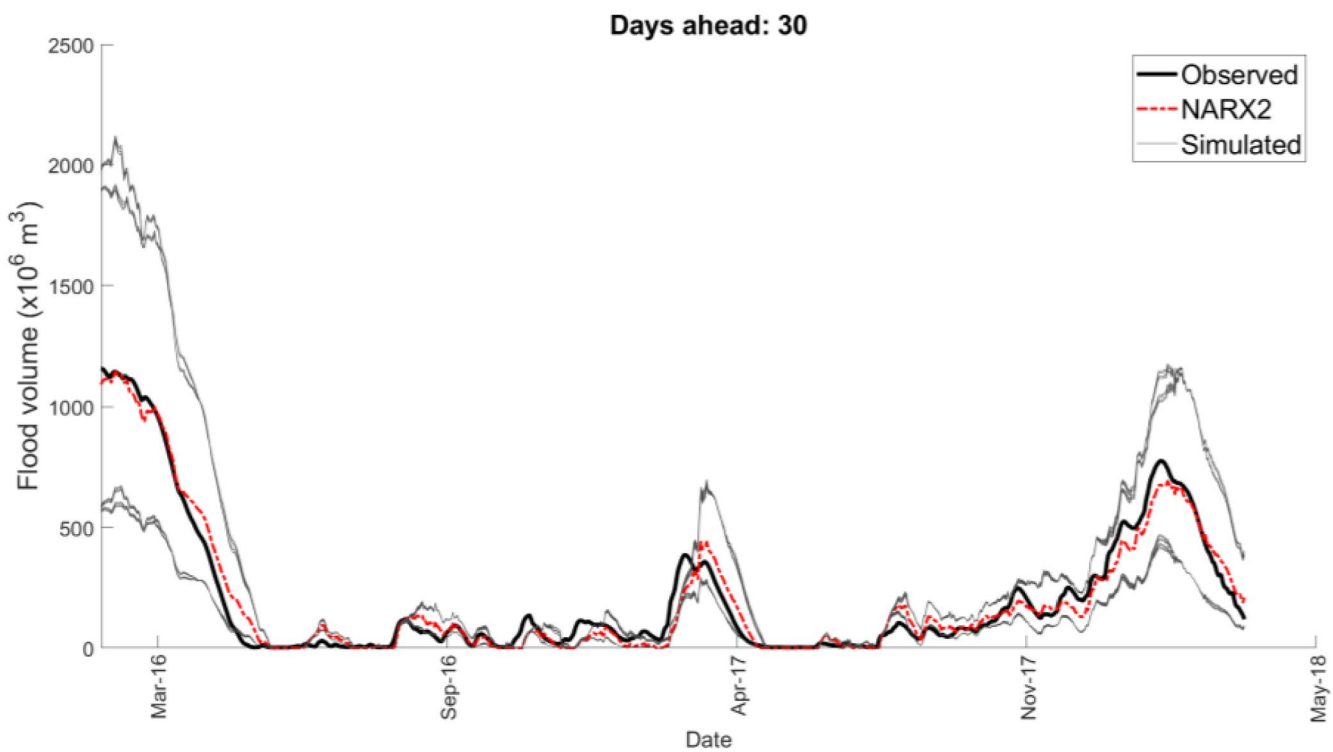

Figure 6. (Continued)

Table 3
Mean of Absolute Delay in Occurrence of Peak TFV Over a 7 Days Windows and the Number of Model Parameters in the Developed Model (NA: Not Applicable)

	NARX2	SVR	ARX
Days	12 terms	NA	13 terms
1	0.82	2.37	2.93
2	1.37	2.39	2.88
3	1.41	2.61	3.33
4	1.63	2.49	3.60
5	1.61	2.67	4.09
7	1.56	2.67	4.05
10	1.39	2.65	3.47
15	1.47	2.55	4.02
20	1.58	2.78	3.81
25	1.63	2.75	4.00
30	1.63	2.53	3.81
45	1.54	2.69	3.67
60	1.61	2.35	4.14
90	1.63	2.33	4.02

Note. NARX, nonlinear auto regressive model with K exogenous variables; SVR, support vector regression; ARX, auto regressive model With exogenous.

following link: https://www.esat.kuleuven.be/sista/lssvmlab/. Since the parameters γ and C are not known a priori, the optimal value of those parameters is estimated based on a grid search procedure. In the process, several combinations of those parameters were considered and the SVR models were developed, and the performance of the SVR model in predicting the TFV for the training period was noted using NSE. The optimal parameters yield the highest NSE value. In the grid search process, γ was chosen as 0.001, 0.01, 0.1, 1, 10, 100, 1,000, 10,000 and 100,000, while C was ranged from 1 to 100 at an increment of 1. The optimal value of γ and C was found to be 10 and 7 respectively. Subsequently, the developed SVR model was used to predict the TFV corresponding to the aforementioned 14 future time projections. The performance measures, predicted time series plots, and the errors in terms of boxplots are shown in Tables 1 and 2, Figures 4 and Figure 5 respectively.

Auto Regressive Model with Exogenous variables (ARX) model was developed using the procedures described in Section 2.4, while the modeling was performed by using the *Time Series Analysis* package available in MATLAB. The performance measures are shown in Tables 1 and 2 and Figures 3 and 4 for conducting a comparative analysis between the linear, nonlinear and machine learning models in predicting TFV at turloughs.

Similarly, the LSTM has been developed based on the Deep Learning package available in MATLAB and as described in the Supplemental Information. Extensive model parameter search was performed by considering a range of values for hidden layers, gradient threshold, initial learn rate and learn rate drop factor using the training set data and the developed LSTM model was used to predict d-days ahead TFV for the validation period data. Furthermore, to compare the results of all those models, the basic linear auto-regressive model with 1 parameter (AR1) has been developed. Performance measures of each of those models were compared in Tables S2 and S3 in Supporting Information S1, and by using Figures S3 and S4 in Supporting Information S1. It should be noted that the SVR and LSTM models consider the entire input database with all the chosen lags for their development. The ARX has the flexibility to develop the optimal model based on criteria such as AIC/BIC: AIC has been used while developing the ARX model in this study. The AR1 model has a sole model coefficient. Based on the results shown in Figures S3 and Tables S2 and S3 in Supporting Information S1, it can be noted that the performance of LSTM was considerably inferior when compared to the NARX, SVR and ARX models. Hence the LSTM model was not considered for further analysis.

4.4. Prediction of Occurrence of Peak Discharge Using NARX, SVR, and ARX Models

Along with the peak discharge, the time of occurrence of the peak is another important factor in the modeling. The error in predicting the peak time of occurrence for the predicted time series obtained using NARX2, SVR and ARX model corresponding to 1, 2, 3, 4, 5, 7, 10, 15, 20, 25, 30, 45, 60, and 90 days ahead prediction was estimated. It has been noted that during the validation period, a total of 61 peaks had been identified from the observed TFV data. Considering a 7 days window, the difference of occurrence of peak TFV were estimated and the mean of absolute difference was calculated. The values are provided in Table 3. It can be noted from the table that the error in mean absolute difference increases with an increase in d-day ahead prediction. The nonlinear NARX2 model had an error ranging from 0.82 to 1.63 days. The SVR model had an error from 2.33 to 2.78 days, while that for ARX is 2.88–4.14 days. Based on the results it can be concluded that the NARX2 model is closest in predicting the occurrence of peak TFV.

4.5. Refined NARX Model at a Fine Spatial Scale and Different Lags

Based on the results obtained using the different models, it can be seen that NARX model provides the highest performances. This approach was therefore adopted to produce a refined model based upon two separate NARX models, one to predict the total flooded volume from the two turloughs located in the upper part of the karst network (Blackrock and Coy) and the other to predict total flooded volume from the three turloughs located at

a lower altitude (Coole, Garryland and Caherglassaun) which are influenced by both the rainfall and the tidal dynamics. Blackrock and Coy turloughs are located >12 km from the intertidal spring outfall in Kinvara bay and are at a higher elevation and it has been shown in previous investigations using Fast Fourier Transforms (Gill, Naughton, Johnston, Basu, & Gosh, 2013) that the tidal amplitude does not appear to have a significant effect on their FVs. Hence, a NARX1 model has been chosen for the upper turloughs with rainfall being the only exogenous variable, whereas a NARX2 model with rainfall and tidal amplitude as the two exogenous variables was chosen to predict the TFV for the other three lower turloughs located closer to the sea. Another phenomenon that needs to be investigated is the effect of maximum lags in days for both TFV and the exogenous variables. For this purpose, the values of n_y as well as n_{u1} and n_{u2} (or n_{u1} alone) were varied from 1 to 5 days and the corresponding NARX models were developed. Since the NARX model does not require training set data for identification of optimal model parameters, and can be estimated solely based on the calibration set data using the Error to Signal Ratio (ESR), the entire data for both upper and lower karst regions were subdivided into two sets, the calibration set and validation set. The calibration set data ranged from 1 June 2007 to 17 July 2015 (2,969 days consisting of initial 75% data), while the remaining 25% data (988 days from 18 July 2015 to 31 March 2018) was used to evaluate the model performance as the validation set. The optimal maximum lags were identified based on the model performance by considering the entire validation set. The optimal maximum lags and the best model for the high-altitude turloughs were found to be $n_y = 2$, $n_u = 3$ and

$$\begin{aligned} y(t) = {} & 1.6357 \times y(t-1) - 0.6494 \times y(t-2) + 0.0027 \times u_1(t-2)^2 + 0.1058 \\ & \times u_1(t-1) + 0.0134 \times y(t-1) \times u_1(t-2) + 0.0031 \times u_1(t-1) \\ & \times u_1(t-2) + 0.0040 \times y(t-1) \times u_1(t-1) + 0.0843 \times u_1(t-2) \\ & - 0.0127 \times y(t-2) \times u_1(t-2) + 0.0011 \times u_1(t-1)^2 - 0.0019 \\ & \times y(t-2)^2 - 0.0030 \times y(t-1)^2 + 0.0049 \times y(t-1) \times y(t-2) \\ & + 0.0008 \times u_1(t)^2 - 0.0020 \times u_1(t) \times u_1(t-2) - 0.0031 \times y(t-2) \\ & \times u_1(t-1) + 0.1172 + 0.0010 \times y(t-2) \times u_1(t) - 0.0119 \times u_1(t) \\ & + 0.0009 \times u_1(t) \times u_1(t-1) - 0.0008 \times y(t-1) \times u_1(t) + e(t) \end{aligned} \tag{31}$$

While for the lower altitude turloughs the optimal maximum lags were $n_y = 4$, $n_{u_1} = n_{u_2} = 5$ and,

$$\begin{aligned} y(t) = {} & 1.5998 \times y(t-1) - 0.6861 \times y(t-2) + 0.1806 \times u_1(t-1) + 0.2641 \\ & \times u_1(t-2) - 0.1789 \times y(t-4) + 0.0003 \times y(t-1) \times u_1(t-3) \\ & + 0.4976 + 0.0062 \times u_1(t-1) \times u_1(t-4) + 0.2597 \times y(t-3) + 0.0002 \\ & \times y(t-1) \times u_1(t-4) + 0.0035 \times u_1(t-1)^2 + 0.0049 \times u_1(t) \\ & \times u_1(t-4) + 0.0543 \times u_1(t-3) + 0.0050 \times u_1(t) \times u_2(t) + e(t) \end{aligned} \tag{32}$$

the TFV prediction error for each of those karst systems based on the identified NARX models were shown in Table 4, and the observed and NARX model predicted time series plots are shown in Figure 7.

4.6. Monte Carlo Simulation Study to Investigate Effectiveness of NARX Model

In order to investigate the effectiveness of the NARX model, a hypothetical simulation study has been considered. In the simulation study, a hypothetical rainfall-runoff process was generated. The rainfall-runoff relationship considered in the study was based on Soil Conservation Service (SCS) curve number, where the data was generated at a daily scale. The daily rainfall was assumed to follow the Pearson Type 3 (PE3) distribution (Ye et al., 2018). The reason for this simulation study is to investigate the effectiveness of a NARX model in the identification of the presence of nonlinear terms in a generic input-output system where the degree of nonlinearity present in the system is already known hypothetically. The generated rainfall data and the estimated runoff data were subdivided into two sets - the calibration set and the validation set. The rainfall and runoff from the calibration set were then used to develop the nonlinear time series model NARX, machine learning model SVR and linear time series model ARX. Subsequently, the generated rainfall in the validation set was used as an input to each of those three developed models and the runoff was generated for the entire duration of the validation period. The model predicted runoff were then compared to the actual runoff obtained based on the SCS curve number based approach and the errors in prediction were quantified in terms of the three performance measures: NSE, r, KGE.

Table 4
Prediction Error of Total Flooded Volume of the Turloughs Located in the Upper and Lower Part of the Karst Catchment

	NARX1 (high-altitude turloughs)			NARX2 (low-altitude turloughs)		
Predict	NSE	r	KGE	NSE	r	KGE
1day	0.997	0.999	0.998	0.999	0.999	0.999
2days	0.996	0.998	0.997	0.998	0.999	0.998
3days	0.994	0.997	0.996	0.998	0.999	0.998
4days	0.993	0.996	0.994	0.997	0.999	0.997
5days	0.991	0.995	0.993	0.996	0.998	0.996
7days	0.986	0.993	0.990	0.995	0.997	0.993
10days	0.980	0.990	0.987	0.992	0.996	0.988
15days	0.971	0.986	0.983	0.986	0.993	0.979
20days	0.965	0.983	0.980	0.981	0.991	0.970
25days	0.960	0.980	0.979	0.976	0.988	0.959
30days	0.955	0.978	0.976	0.972	0.986	0.949
45days	0.938	0.969	0.962	0.962	0.983	0.917
60days	0.932	0.966	0.947	0.957	0.982	0.888
90days	0.922	0.961	0.921	0.944	0.979	0.850

Note. NARX, nonlinear auto regressive model with K exogenous variables; NSE, Nash-Sutcliffe efficiency; KGE, Kling-Gupta efficiency.

In order to generate the rainfall data, parameters of the PE3 distributions needed to be estimated. The PE3 distribution consists of three parameters: location parameter (μ), scale parameter (σ) and shape parameter (γ). Assuming $\gamma > 0$, the probability density function $f(x)$ can be expressed as,

$$f(x) = \frac{(x-\xi)^{\alpha-1} e^{-(x-\xi)/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \tag{33}$$

where $\alpha = 1/\gamma^2$, $\beta = \frac{1}{2}\sigma|\gamma|$ and $\xi = \mu - 2\sigma/\gamma$.

The L-moments based parameter estimation (Basu & Srinivas, 2013; Hosking & Wallis, 1997) has been used in the simulation study to estimate the PE3 parameters. Based on historical observations (Ye et al., 2018), both the L-CV and L-skewness of the rainfall was assumed to be 0.5, while the mean was considered to be 1 inch. The three parameters can be estimated as (Hosking & Wallis, 1997): $\mu = 1$, $\sigma = 1.14997$, $\gamma = 3.07934$.

Based on the estimated parameters, a rainfall time series of length 1,000 days have been generated (see Figure 8a).

Using the generated rainfall data, the runoff was estimated based on the SCS curve number approach. The SCS curve number assumes that the runoff Q (*inches*) can be estimated from the precipitation P (*inches*) as (Te Chow, 2010),

$$Q = \frac{(P - 0.2S)^2}{P + 0.8S} \tag{34}$$

where S(*inches*) is the potential maximum retention after the beginning of runoff and is given as,

$$S = \frac{1000}{CN} - 10 \tag{35}$$

CN is called curve number ranging from $0 < CN \leq 100$, with a completely impervious surface taking a value of 100 and natural surfaces lower numbers. Based on empirical studies (Te Chow, 2010) the curve number depends on land use and soil type. The present simulation study assumes a cultivated land without conservation treatment and soil type C: clay loams, for which the $CN = 88$ (Te Chow, 2010; Table 5.5.2, page 150).

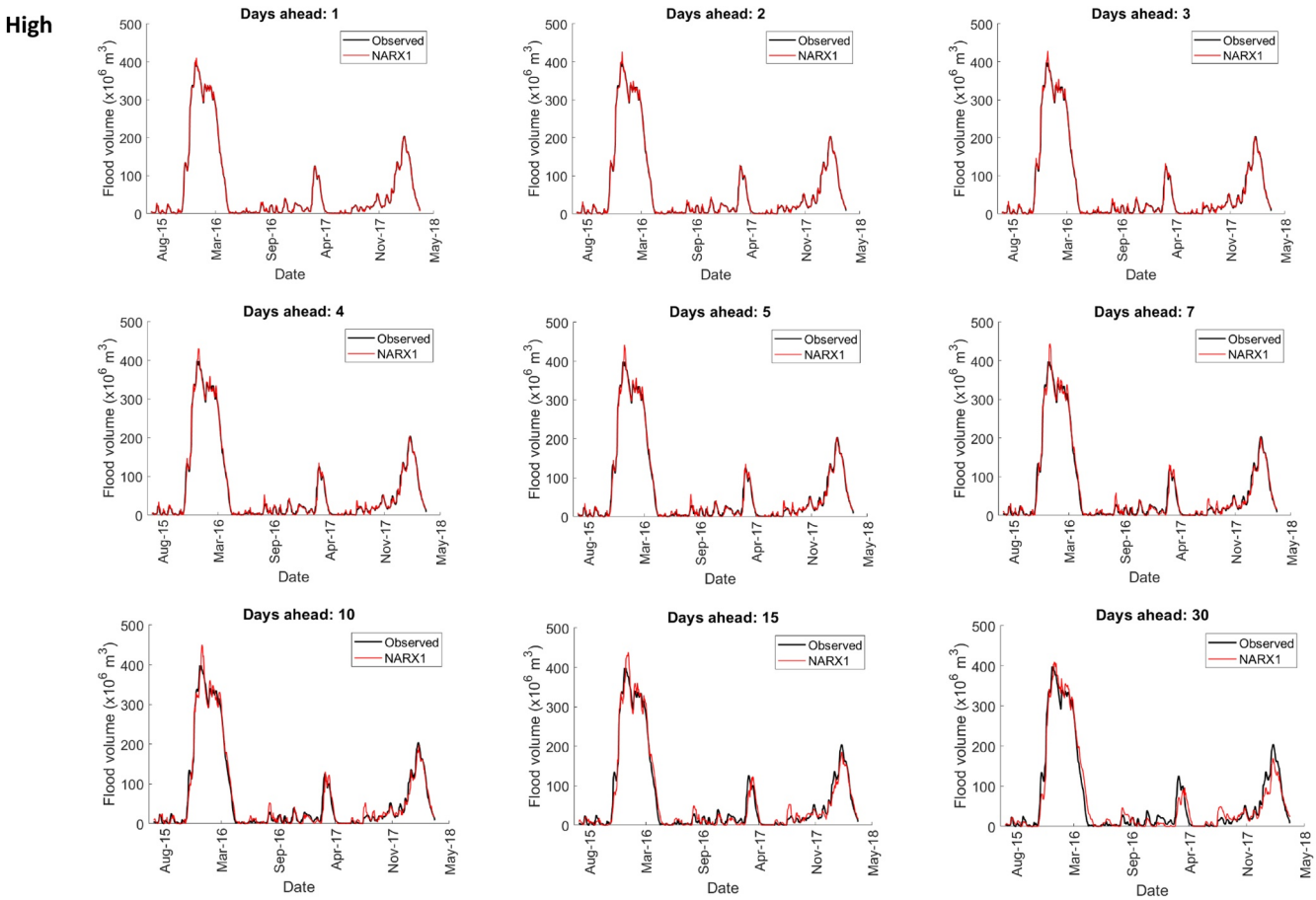

Figure 7. Nonlinear auto regressive model with K exogenous variable model predictions of the high-altitude and low-altitude turloughs corresponding to different time period ahead predictions.

The estimated runoff (in inches) using the SCS curve number and the simulated precipitation is shown in Figure 8b. It should be noted that the relationship between the runoff and precipitation is nonlinear in nature as per Equation 34.

The NARX model was developed by considering $n_y = 5$; $K = 1$; $l = 2$; $d_1 = d_2 = 0$; $n_{u1} = 5$. The selected NARX model for the simulation study was as follows:

$$y(t) = 0.6149 \times u_1(t) + 0.3648 + 0.0400 \times u_1(t)^2 - 0.0054 \times y(t-1) \times u_1(t) - 0.0004 \times y(t-1) + e(t) \quad (36)$$

It can be noted from the developed model that terms containing the current day's precipitation and the previous day's runoff got selected in the model. Furthermore, out of a possible 66 terms, only four most significant terms were selected in developing the NARX model which are: the present day's precipitation, the square of the present day's precipitation, the past day's runoff and the product of present day's precipitation and past day's runoff. Investigation of the SCS curve number Equation 34 indicates that the runoff depends on the square of the present day's precipitation. Expansion of the SCS curve number equation in an algebraic form clearly indicate similarities in the model terms from both the equations.

Prediction of runoff using the NARX, machine learning based SVR and linear ARX models are provided in Figure 8c with the errors in runoff prediction using each of those models shown in Table 5. Since the runoff in the simulation study is low, the Bias and RMSE had been estimated instead of the Rbias and RRMSE. The error measures clearly indicate that the NARX model is superior when compared to the SVR and ARX models in predicting runoff for the validation period. Furthermore, it can be noted that the linear model predicts negative

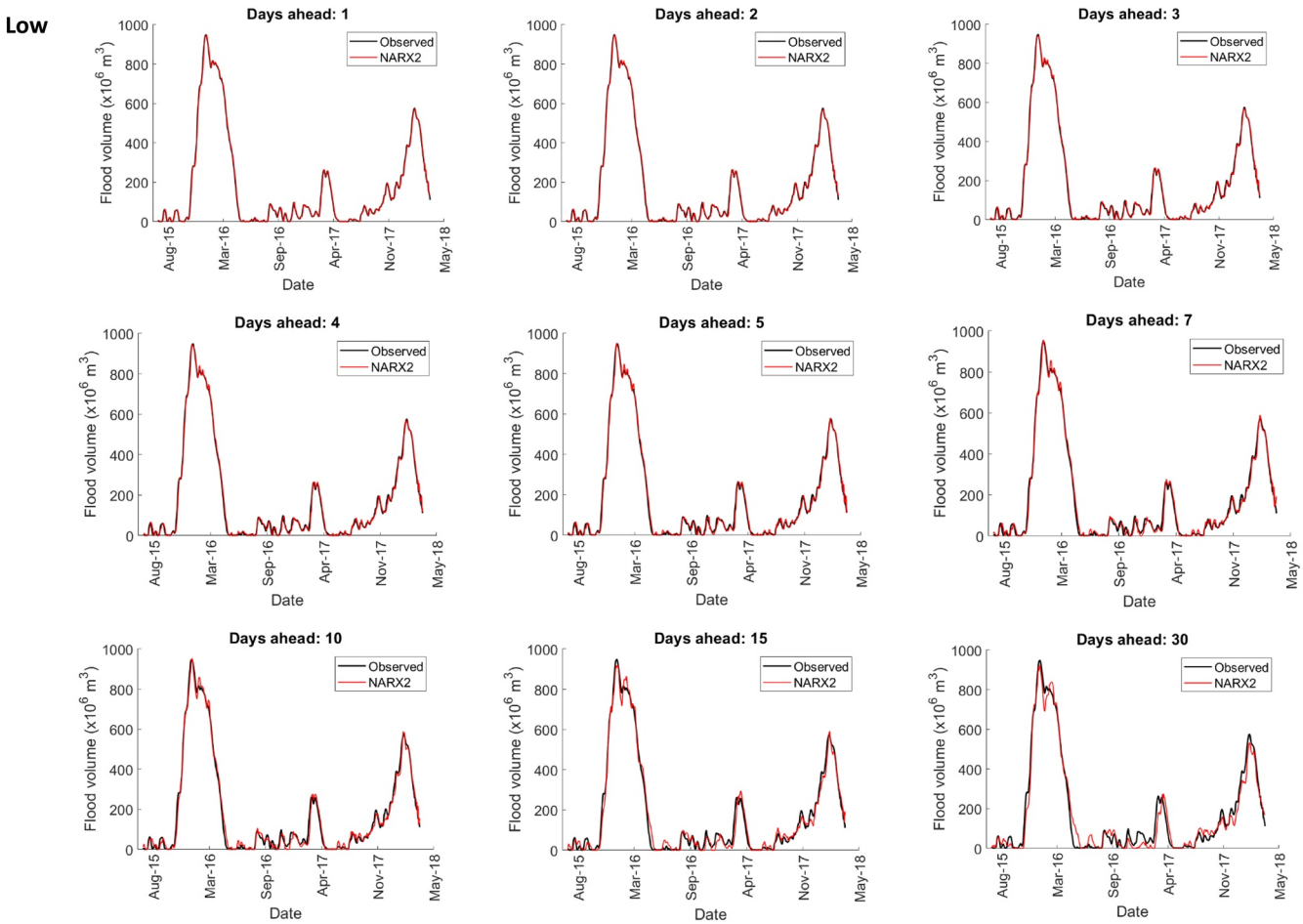

Figure 7. (Continued)

runoff for a considerable period of time and performs poorly. This is due to the fact that the true model used to generate the runoff data (SCS curve number) is nonlinear in nature. The SVR model provides considerable accuracy, as it can account for nonlinearity, however, the SVR model underpredicts the extreme events and over-predicts the low runoffs.

5. Discussion

The comparison of the performance measures (Table 1) between the nonlinear, machine learning and linear time-series models based on the prediction of TFV from the five turloughs indicates that the performances of all of the models are all similarly highly accurate up to 1–10 days into the future. The three error measures NSE, *r*, and KGE indicate that the performance of NARX is better than the other two models. The NARX2 model consists of 12 important terms and require all past 5 days TFV, past 3 days rainfall and the present day's tidal amplitude data.

In general, an increase in rainfall increases the TFV in the turloughs, as expected. The performance of the NARX2 model was found to be very reliable in predicting TF volume and time of occurrences of peak flow with predictions up to 60 days ahead (Figures 3 and 4). The performance of each model, however, does deteriorate when predicting more than 30 days ahead. However, even though the models were unsuccessful in predicting the values of TFV beyond 30 days, the models still were able to identify successfully the time of occurrences of peak flows at later future dates within a range of 1.63 days (Table 3). Figure 4 further indicates that the error in model prediction is higher during high values of TFV when compared to low TFVs; however, the error obtained based on NARX2 model are the lowest.

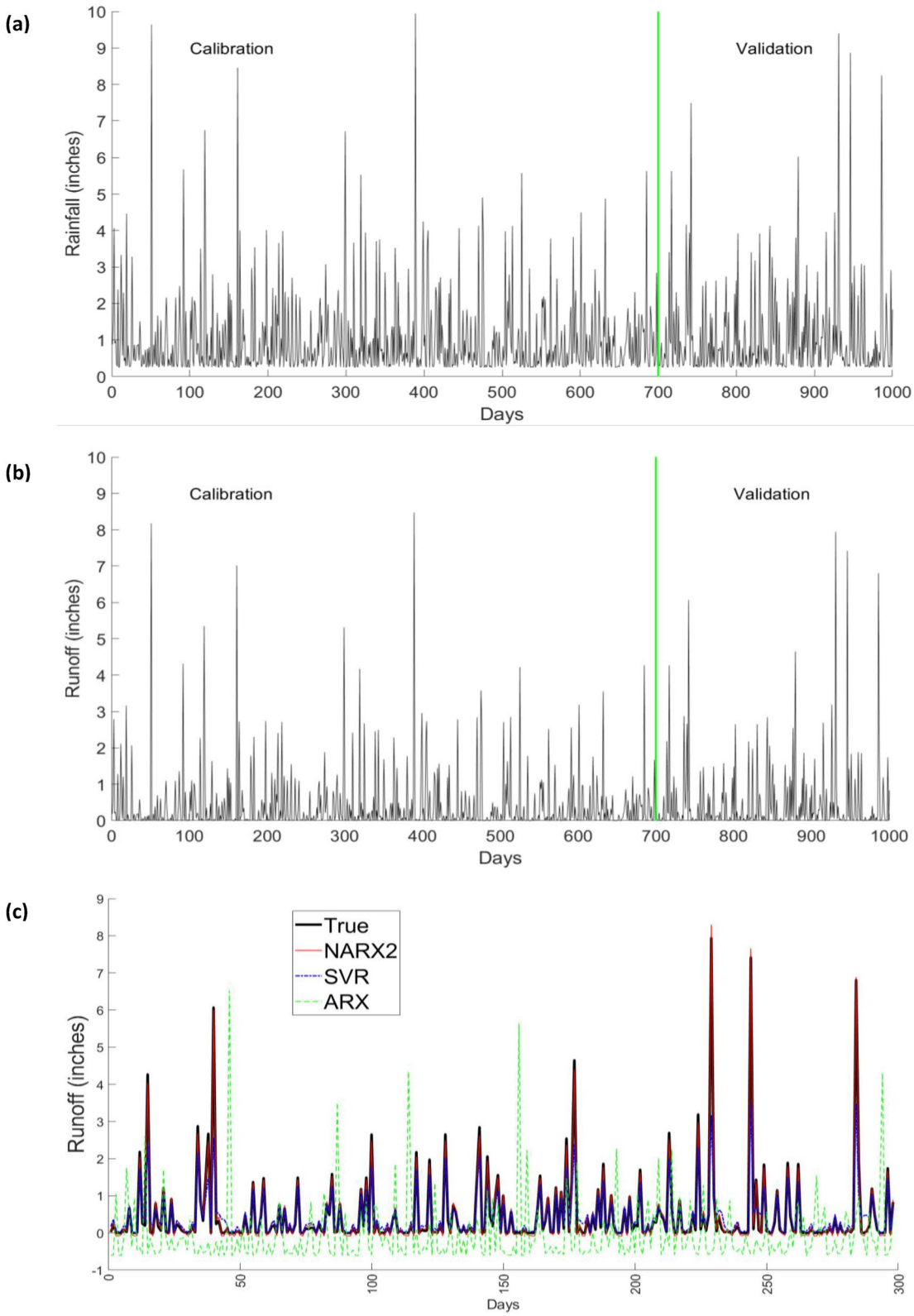

Figure 8. The (a) simulated daily rainfall time series using Pearson Type 3 distribution, (b) estimated runoff using SCS curve number based empirical rainfall-runoff model, and (c) comparison of estimated/true and predicted runoff using NARX, SVR and ARX models during the validation period in the Monte Carlo simulation based case study.

Table 5
Performance Measures of Runoff Prediction Using NARX, SVR and ARX Model in the Monte Carlo Simulation Study

Performance measure	NARX	SVR	ARX
NSE	0.994	0.742	−1.050
r	0.997	0.952	−0.038
KGE	0.992	0.532	−0.509
Bias	−0.38	−5.79	−52.11
RMSE	8.05	53.43	150.57

Note. NARX, nonlinear auto regressive model with K exogenous variables; SVR, support vector regression; ARX, auto regressive model with exogenous.

One major advantage of the NARX models is that they can identify the important variables necessary to develop the relationship between the input variables (past TFV, current, and past rainfall and tidal amplitude) and the output variable (current TFV). The developed NARX2 model selected the 12 most important terms out of a possible 136 terms. The flexibility in selecting such a limited number of terms allows the modeler to develop a nonlinear model that takes the form of an algebraic equation, as well as identifying the presence of redundant variables in the input data set. Hence, such a model is parsimonious and easy to use for obtaining future predictions. Furthermore, a well-developed NARX model can indirectly provide information on the complexity of the underlying physical phenomenon of the hydrological process. For example, in situations where a second-degree NARX model fails to provide sufficiently accurate predictions, but a third-degree NARX model can reach sufficient accuracy, that information can be used to understand and develop a physical or dynamical system model to represent the physical phenomenon of the system.

In the case of the machine-learning based (SVR) nonlinear model investigated, all the input variables with every past day's lagged data are used to develop the models. These types of models do not have the flexibility to identify redundant terms that might be present in the database, nor can they be used to obtain information on the physical properties of the underlying system. The performance of the linear ARX model was found to be accurate up to 15 days, but then starts to deteriorate thereafter. Even though the turlough system is nonlinear in nature (Figure 3), the high performance of a linear model along with a sufficiently accurate performance of a second-degree NARX model indicates that the complexity of the overall karst system in terms of the relationship between rainfall, surface flooded volume and tidal fluctuations can be assumed to be low.

Since the primary objective in development of the TFV model is to predict the peak flows with sufficient accuracy, five performance measures (NSE, *r*, KGE, Rbias and RRMSE) were estimated during the validation period by discarding the data corresponding to days where the observed TFV was less than 50×10^6 m^3 per day. It had been noted that the TFV was greater than the chosen threshold for 492 days out of 790 days in the validation period, and for 1,916 days out of 3,167 days in the combined calibration and training period. The NSE, *r* and KGE obtained using NARX2 model were found to be higher while estimating the peak flows when compared to that obtained based on SVR and ARX model, while the Rbias and RRMSE was found to be closer to zero for NARX2 model. The Rbias for SVR was found to be negative, indicating that the SVR model has, in general, a tendency to underpredict the TFV. Overall, it can be concluded that both linear and nonlinear times series models perform considerably better in predicting TFV up to 10 days into the future, whilst the proposed nonlinear time series NARX model performs better in predicting TFV from 15 days onwards. Based on the performance of the nonlinear and linear models from the real-world data and the Monte Carlo simulation studies, it can be noted that in situations where the degree of nonlinearity of the system is low, the performance of the linear and nonlinear models can be similar when only a few days ahead prediction is obtained, however, for highly nonlinear systems the linear model's performance will be considerably poor.

Identification of optimal lags for endogenous (TFV) as well as exogenous (rainfall, tidal amplitude) variables indicate that turloughs at different locations in the overall karst network can exhibit different optimal lags. The dynamics of the two turloughs located at the higher altitude require the past 2 days TFV and current and past 2 days exogenous variable (rainfall) for an accurate model, whereas the three turloughs grouped further down the system need the past 4 days TFV, current and the past 4 days rainfall and the current day's tidal amplitude value. Hence, proper identification of the degree of nonlinearity, a set of exogenous variables, and their associated maximum lags are clearly important when developing accurate models for prediction.

These predictive models of TFV can be used to create a flood warning system in a real-world scenario. It should be noted that the model requires ongoing inputs of information on the rainfall and tidal amplitude data for forecasting TFV. In order to develop a useful early flood warning system for the lowland karst area, the refined NARX model (developed in Section 4.5) needs to be fed with real-time data, continuously updating the flooded volume of the catchment. In this catchment, depths are being monitored in real time at Blackrock and Coole turlough via pressure transducers recently installed by Geological Survey Ireland, which update data to a public site via

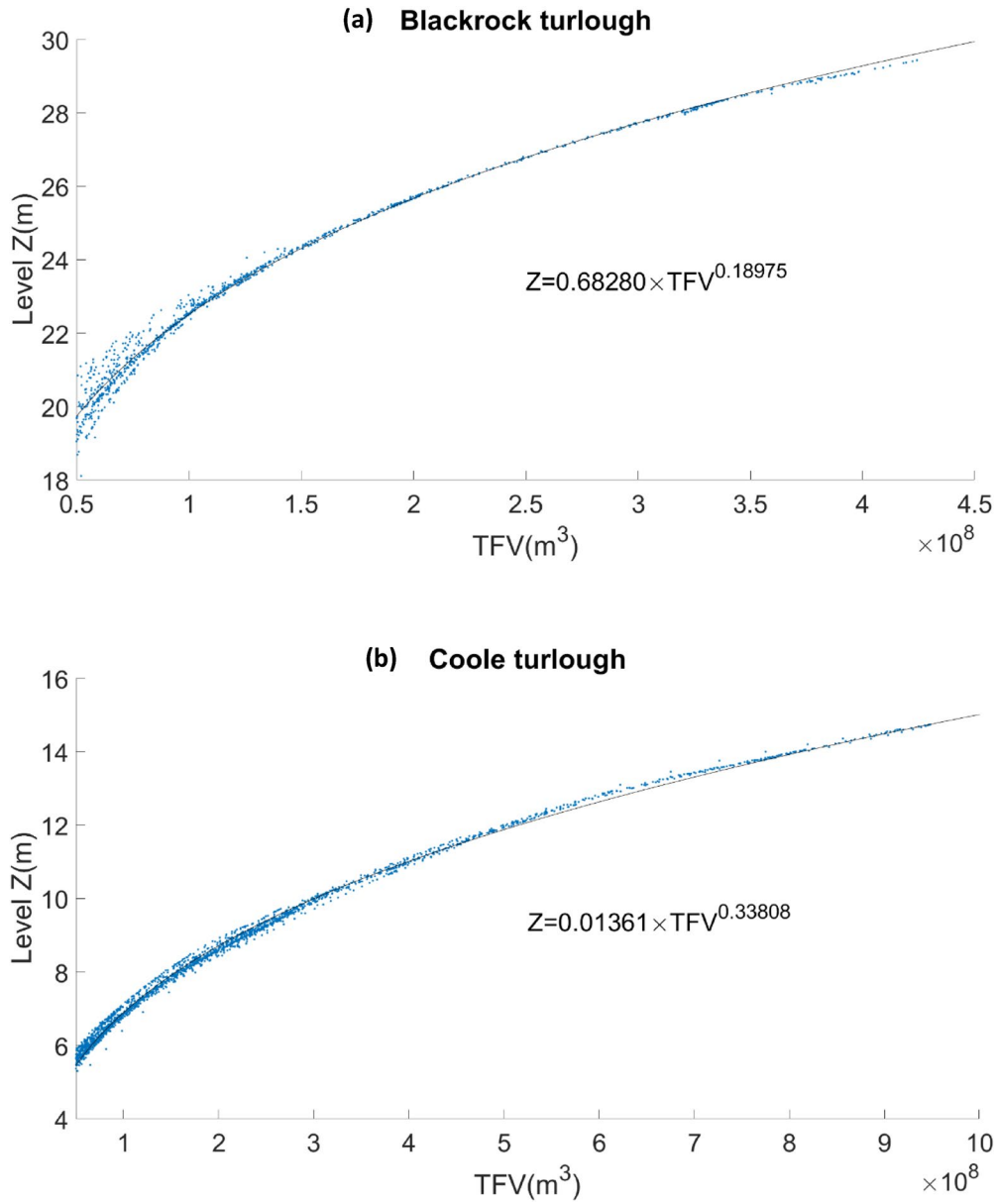

Figure 9. Total flooded volume versus water level (Z) relationship for (a) blackrock and (b) coole turloughs.

telemetry every 15 min. Hence, the water level (stage) of Blackrock turlough located in the upper network and Coole turlough located in the lower part of the network have been related to the total flooded volume of the upper and lower altitude turloughs respectively. Since the goal is to focus on peak volumes and the corresponding water level, a TFV greater than 50×10^6 m^3 was used to develop the water level (Z) versus TFV relationship for both continuously monitored turloughs (Figure 9). The model can therefore be used to estimate the water level rise from the TFV into the future given different levels of future predicted rainfall as an early warning forecast system for the region. These rainfall projections can be made according to short-term weather forecasts. In situations where the forecasted water level is expected to rise above a certain threshold, necessary precautions can be taken. For example, if the water level reaches a critical threshold level at Blackrock and/or Coole turloughs on a given day, one can estimate the TFV for the past 2–4 days using the TFV versus water level relationship, while the past 4 days rainfall and tidal amplitude data would also have been recorded. Based on the information, the amount of rainfall in the upcoming week that will create flooding by exceeding critical water levels in those turloughs can be simulated. In situations where the observed rainfall in this period is above such simulated rainfall values, a flood

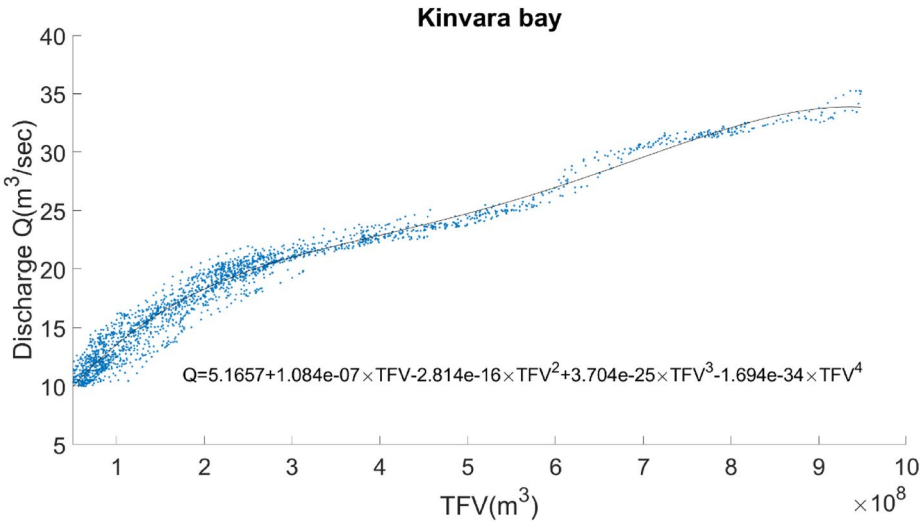

Figure 10. Relationship between the total flooded volume from the three lower altitude turloughs and the discharge at Kinvara bay to the ocean.

warning can be issued. A recent flood study of the area (Morrissey et al., 2020) has defined the target maximum flood levels at Blackrock to be 27.1 mAOD and at Coole to be 13.35 mAOD. Hence, for an early warning system, it would be recommended that the model starts to be consulted on a daily basis when the levels get to 24.5 mAOD at Blackrock and/or 11.5 mAOD at Coole, and different forward projections can be applied.

In addition, the model can be used to quantify freshwater discharge into the Atlantic Ocean at Kinvara bay, which is not possible to measure directly as it is an inter-tidal spring. To do this the spring discharge (as simulated by the Infoworks model) was related with the TFV from the lower altitude set of turloughs, as shown in Figure 10. Again, TFV values greater than 50×10^6 m^3 were considered for the development of the relationship between the discharge Q (m^3/sec) into Kinvara bay and the TFV (m^3) from the three lower altitude turloughs (Coole, Garryland and Caherglassaun). Hence, for any value of existing or future predicted TFV, an equivalent spring discharge into Kinvara bay can be easily calculated according to that polynomial relationship. This could be of interest to mariculture farmers in the bay (mussels are harvested there) as well as more broadly for studies trying to assess, for example, the impact of terrestrial nutrients into the bay. Finally, in the future this type of approach could also be linked to regularly updated earth observation data, for example, using SAR satellite data to map the spatial extent of flooding, which can then be combined with the DEM to predict depth and volume.

6. Conclusions

This study has demonstrated the application of different nonlinear modeling approaches being able to capture the dynamics of a complex natural lowland karst system with frequent and widespread groundwater/surface water interactions and their development into an effective groundwater flood forecasting tool. For this karst system, a NARX model was deemed to provide the most reliable predictions up to 60 days into the future, beyond which the model's performances start to deteriorate, although it is still able to identify the timing of peak floods successfully up to 90 days in the future. The optimum NARX model requires inputs of the past 5 days' flood volume, and both daily rainfall data and tidal amplitude data across the past 4 days. Existing real-time telemetric monitoring of water level data, one situated in the upper part of the catchment, the other in the lower part of the catchment can be fed into the model to allow it to function as an early flood warning tool. The model also predicts freshwater discharge from the inter-tidal catchment outlet spring into the Atlantic Ocean.

The development of a NARX model to understand such complex and nonlinear natural systems is a challenge and needs to be explored in the future. However, this study has shown a potential application of this methodology which can be extended to other locations, for example, modeling and prediction of spring discharges from karst systems or more broadly into other hydrological systems. The NARX model can be a viable alternative to other semi-distributed groundwater models, to make the relationships between key input variables easier to

characterize. The NARX model is a statistical model generated solely based on the available historical data and hence the nonlinear framework can be applied to other natural systems (other than groundwater) as well. This approach also enables one to explore and identify the most important variables that influence the flood generation mechanism in a catchment, thereby helping to refine conceptual models of the natural systems.

Data Availability Statement

The data sets that were used to develop the models for this research can be found in Gill, L. (2021), "Lowland karst groundwater flooding data for forecast model," Mendeley Data, V1, (doi:10.17632/s7jb2snd2j.1).

Acknowledgments
The field data that these models have been developed upon in this karst catchment have been collected over several years from many studies funded by different sources including the National Parks and Wildlife Service, AXA Research Fund, Geological Survey Ireland and Galway County Council (as per the cited references in the paper). Open access funding provided by IReL.

References

Abusaada, M., & Sauter, M. (2013). Studying the flow dynamics of a karst aquifer system with an equivalent porous medium model. *Ground Water*, *51*(4), 641–650.

Al-Fugara, A., Pourghasemi, H. R., Al-Shabeeb, A. R., Habib, M., Al-Adamat, R., Al-Amoush, H., & Collins, A. L. (2020). A comparison of machine learning models for the mapping of groundwater spring potential. *Environmental Earth Sciences*, *79*, 206. https://doi.org/10.1007/s12665-020-08944-1

Amorocho, J. (1967). The nonlinear prediction problem in the study of the runoff cycle. *Water Resources Research*, *3*(3), 861–880. https://doi.org/10.1029/wr003i003p00861

Amorocho, J., & Brandstetter, A. (1971). Determination of nonlinear functional response functions in rainfall-runoff processes. *Water Resources Research*, *7*(5), 1087–1101. https://doi.org/10.1029/wr007i005p01087

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000a). Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*, *5*(2), 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. (2000b). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering*, *5*(2), 124–137.

Bailly-Comte, V., Jourde, H., Roesch, A., Pistre, S., & Batiot-Guilhe, C. (2008). Time series analyses for karst/river interactions assessment: Case of the Coulazou River (southern France). *Journal of Hydrology*, *349*, 98–114. https://doi.org/10.1016/j.jhydrol.2007.10.028

Basu, B. & Srinivas, S. S. (2013). Formulation of a mathematical approach to regional frequency analysis. *Water Resources Research*, *49*(10), 6810–6833. https://doi.org/10.1002/wrcr.20540

Beven, K. J. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Billings, S. A. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons.

Billings, S. A., Chen, S., & Korenberg, M. J. (1989). Identification of MIMO non-linear systems using a forward-regression orthogonal estimator. *International Journal of Control*, *49*(6), 2157–2189. https://doi.org/10.1080/00207178908559767

Billings, S. A., & Fadzil, M. B. (1985). The practical identification of systems with nonlinearities. *IFAC Proceedings*, *18*(5), 155–160. https://doi.org/10.1016/s1474-6670(17)60551-2

Billings, S. A., Korenberg, M. J., & Chen, S. (1988). Identification of non-linear output-affine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, *19*(8), 1559–1568. https://doi.org/10.1080/00207728808964057

Billings, S. A., & Leontaritis, I. J. (1981). Identification of nonlinear systems using parameter estimation techniques. In *IEE conference control and its applications, Warwick* (pp. 183–187).

Billings, S. A., & Voon, W. S. F. (1984). *Least squares parameter estimation algorithms for non-linear systems*. ASCE Research Report 225. Department of Control Engineering, University of Sheffield.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. a. P., Parajka, J., Merz, B., et al. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*, 108–111.

Bonacci, O. (2014). Ecohydrology of karst poljes and their vulnerability. In P. Sackl, R. Durst, D. Kotrošan, & B. Stumberger (Eds.), *Dinaric karst poljes - floods for life*. (pp. 25–38).

Borghi, A., Renard, P., & Cornaton, F. (2016). Can one identify karst conduit networks geometry and properties from hydraulic and tracer test data? *Advances in Water Resources*, *333*(90), 99–115. https://doi.org/10.1016/j.advwatres.2016.02.009

Brockwell, P. J., Davis, R. A., & Fienberg, S. E. (1991). *Time series: Theory and methods*. Springer Science & Business Media.

Chang, F. J., Chen, P. A., Lu, Y. R., Huang, E., & Chang, K. Y. (2014). Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, *517*, 836–846. https://doi.org/10.1016/j.jhydrol.2014.06.013

Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, *50*(5), 1873–1896. https://doi.org/10.1080/00207178908953472

Chen, S., Cowan, C. F., & Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks*, *2*(2), 302–309. https://doi.org/10.1109/72.80341

Chen, Z. & Goldscheider, N. (2014). Modeling spatially and temporally varied hydraulic behavior of a folded karst system with dominant conduit drainage at catchment scale, Hochifen–Gottesacker, Alps. *Journal of Hydrology*, *514*, 41–52. https://doi.org/10.1016/j.jhydrol.2014.04.005

Cobby, D., Morris, S. E., Parkes, A., & Robinson, V. (2009). Groundwater flood risk management: Advances towards meeting the requirements of the EU floods directive. *Journal of Flood Risk Management*, *2*(2), 111–119. https://doi.org/10.1111/j.1753-318x.2009.01025.x

Dreiss, S. J. (1982). Linear kernels for karst aquifers. *Water Resources Research*, *38*(10), 865–876. https://doi.org/10.1029/wr018i004p00865

Drew, D. (2008). Hydrogeology of lowland karst in Ireland. *The Quarterly Journal of Engineering Geology and Hydrogeology*, *41*, 61–72. https://doi.org/10.1144/1470-9236/07-027

Drew, D. (2018). *Karst of Ireland*. Geological Survey Ireland.

Ghasemizadeh, R. F., Hellweger, C., Butscher, I., Padilla, D., Vesper, M., Field, M., & Alshawabkeh, A. (2012). Review: Groundwater flow and transport modeling of karst aquifers, with particular reference to the North Coast Limestone aquifer system of Puerto Rico. *Hydrogeology Journal*, *20*(8), 1441–1461. https://doi.org/10.1007/s10040-012-0897-4

Gill, L. W., Babechuk, M. G., Kamber, B. S., McCormack, T., & Murphy, C. (2018). Use of trace and rare earth elements to quantify autogenic and allogenic inputs within a lowland karst network. *Applied Geochemistry*, *90*, 101–114. https://doi.org/10.1016/j.apgeochem.2018.01.001

Gill, L. W., Naughton, O., & Johnston, P. M. (2013). Modeling a network of turloughs in lowland karst. *Water Resources Research 49*(6), 3487–3503. https://doi.org/10.1002/wrcr.20299

Gill, L. W., Naughton, O., Johnston, P. M., Basu, B., & Ghosh, B. (2013). Characterisation of hydrogeological connections in a lowland karst network using time series analysis of water levels in ephemeral groundwater-fed lakes (turloughs). *Journal of Hydrology*, *499*, 289–302. https://doi.org/10.1016/j.jhydrol.2013.07.002

Gill, L. W., Schuler, P., Duran, L., Johnston, P. M., & Morrissey, P. (2020). An evaluation of semi-distributed-pipe-network and distributed-finite-difference models to simulate karst systems. *Hydrogeology Journal*, *29*, 259–279. https://doi.org/10.1007/s10040-020-02241-8

Guzman, S. M., Paz, J. O., & Tagert, M. L. M. (2017). The use of NARX neural networks to forecast daily groundwater levels. *Water Resources Management*, *31*(5), 1591–1603. https://doi.org/10.1007/s11269-017-1598-5

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics*, *52*(3), 218–242. https://doi.org/10.1002/2013rg000443

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press.

Hu, C., Hao, Y., Yeh, T. J., Pang, B., & Wu, Z. (2008). Simulation of springflows from a karst aquifer with an artificial neural network. *Hydrological Processes*, *22*, 596–604. https://doi.org/10.1002/hyp.6625

Hughes, A. G., Vounaki, T., Peach, D. W., Ireson, A. M., Jackson, C. R., Butler, A. P., et al. (2011). Flood risk from groundwater: Examples from a chalk catchment in southern England. *Journal of Flood Risk Management*, *4*(3), 143–155. https://doi.org/10.1111/j.1753-318x.2011.01095.x

Jacoby, S. L. S. (1966). A mathematical model for nonlinear hydrologic systems. *Journal of Geophysical Research*, *71*(20), 4811–4824. https://doi.org/10.1029/jz071i020p04811

Jayawardena, A. W., & Lai, F. (1994). Analysis and prediction of chaos in rainfall and stream flow time series. *Journal of Hydrology*, *153*(1–4), 23–52. https://doi.org/10.1016/0022-1694(94)90185-6

Jukić, D., & Denić-Jukić, V. (2008). Estimating parameters of groundwater recharge model infrequency domain: Karst springs Jadro and Žrnovnica. *Hydrological Processes*, *22*, 4532–4542.

Kavvas, M. L. (2003). Nonlinear hydrologic processes: Conservation equations for determining their means and probability distributions. *Journal of Hydrologic Engineering*, *8*(2), 44–53. https://doi.org/10.1061/(asce)1084-0699(2003)8:2(44)

Kovacic, G., & Ravbar, N. (2010). Extreme hydrological events in karst areas of Slovenia, the case of the Unica River basin. *Geodinamica Acta*, *23*(1–3), 89–100. https://doi.org/10.3166/ga.23.89-100

Kovács, A., & Sauter, M. (2007). Modelling karst hydrodynamics. In N. Goldscheider, & D. Drew (Eds.), *Methods in karst hydrogeology, IAH international contributions to hydrogeology* (Vol. 26, p. 264).

Kurtulus, B., & Razack, M. (2007). Evaluation of the ability of an artificial neural network model to simulate the input-output responses of a large karstic aquifer: The La Rochefoucauld aquifer (Charente, France). *Hydrogeology Journal*, *15*, 241–254. https://doi.org/10.1007/s10040-006-0077-5

Labat, D., Ababou, R., & Mangin, A. (2000). Rainfall-runoff relations for karstic springs. Part I: Convolution and spectral analyses. *Journal of Hydrology*, *238*, 123–148. https://doi.org/10.1016/s0022-1694(00)00321-8

Labat, D., Ababou, R., & Mangin, A. (2001). Introduction to wavelet analyses to rainfall/runoff relationship for a karstic basin: The case of Licq-Athery karstic system (France). *Ground Water*, *39*(4), 605–615. https://doi.org/10.1111/j.1745-6584.2001.tb02348.x

Larocque, M., Mangin, A., Razack, M., & Banton, O. (1998). Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France). *Journal of Hydrology*, *205*, 217–231. https://doi.org/10.1016/s0022-1694(97)00155-8

Lee, W. K., & Tuan Resdi, T. A. (2016). Simultaneous hydrological prediction at multiple gauging stations using the NARX network for Kemaman catchment, Terengganu, Malaysia. *Hydrological Sciences Journal*, *61*(16), 2930–2945. https://doi.org/10.1080/02626667.2016.1174333

Mayaud, C., Gabrovšek, F., Blatnik, M., Kogovšek, B., Petrič, M., & Ravbar, N. (2019). Understanding flooding in poljes: A modelling perspective. *Journal of Hydrology*, *575*, 874–889. https://doi.org/10.1016/j.jhydrol.2019.04.092

McCormack, T., Naughton, O., Johnston, P. M., & Gill, L. W. (2016). Quantifying the influence of surface water-groundwater interaction on nutrient flux in a lowland karst catchment. *Hydrology and Earth System Sciences*, *20*, 2119–2133. https://doi.org/10.5194/hess-20-2119-2016

Morrissey, P., McCormack, T., Naughton, O., Johnston, P. M., & Gill, L. W. (2020). Modelling groundwater flooding in a lowland karst catchment. *Journal of Hydrology*, *580*, 124361. https://doi.org/10.1016/j.jhydrol.2019.124361

Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, *31*(9), 2761–2775. https://doi.org/10.1007/s11269-017-1660-3

Naughton, O., Johnston, P. M., & Gill, L. (2012). Groundwater flooding in Irish karst: The hydrological characterisation of ephemeral lakes (turloughs). *Journal of Hydrology*, *470–471*, 82–97. https://doi.org/10.1016/j.jhydrol.2012.08.012

Naughton, O., Johnston, P. M., McCormack, T., & Gill, L. W. (2017). Groundwater flood risk mapping and management: Examples from a lowland karst catchment in Ireland. *Journal of Flood Risk Management*, *10*, 53–64. https://doi.org/10.1111/jfr3.12145

Naughton, O., McCormack, T., Drew, D., Gill, L. W., Johnston, P. M., Regan, S., et al. (2018). The hydrology and hydrogeology of the Gort Lowlands. *Irish Journal of Earth Sciences*, *36*, 1–20. https://doi.org/10.3318/IJES.2018.36.3

Neshat, N., Hadian, H., & Behzad, M. (2018). Nonlinear ARIMAX model for long–term sectoral demand forecasting. *Management Science Letters*, *8*(6), 581–592. https://doi.org/10.5267/j.msl.2018.4.032

Noone, S., Broderick, C., Duffy, C., Matthews, T., Wilby, R. L., & Murphy, C. (2017). A 250-year drought catalogue for the island of Ireland (1765–2015). *International Journal of Climatology*, *37*, 239–254. https://doi.org/10.1002/joc.4999

Pinault, J. L., Amraoui, N., & Golaz, C. (2005). Groundwater-induced flooding in macropore-dominated hydrological system in the context of climate changes. *Water Resources Research*, *41*, W05001. https://doi.org/10.1029/2004wr003169

Porst, G., Naughton, O., Gill, L., Johnston, P., & Irvine, K. (2012). Adaptation, phenology and disturbance of macroinvertebrates in temporary water bodies. *Hydrobiologia*, *696*, 47–62. https://doi.org/10.1007/s10750-012-1181-2

Ristic, D. M. (1976). Water regime of flooded poljes. In V. Yevjevich (Ed.), *Karst hydrology and water resources*. *Karst Hydrology* (Vol. 1, pp. 301–318). Water Resources Publications.

Schuler, P., Cantoni, E., Duran, L., Johnston, P., & Gill, L. W. (2020). Using wavelet coherence to characterize surface water infiltration into a low-lying karst aquifer. *Ground Water*, *58*(4), 71–79. https://doi.org/10.1111/gwat.13012

Shen, H. Y., & Chang, L. C. (2013). Online multistep-ahead inundation depth forecasts by recurrent NARX networks. *Hydrology and Earth System Sciences*, *17*(3), 935–945. https://doi.org/10.5194/hess-17-935-2013

Sivakumar, B., & Singh, V. P. (2012). Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. *Hydrology and Earth System Sciences*, *16*(11), 4119–4131. https://doi.org/10.5194/hess-16-4119-2012

Swain, A. K., & Billings, S. A. (2001). Generalized frequency response function matrix for MIMO non-linear systems. *International Journal of Control*, *74*(8), 829–844. https://doi.org/10.1080/00207170010030144

Te Chow, V. (2010). *Applied hydrology*. Tata McGraw-Hill Education.

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer Science & Business Media.

Waldren, S., Allott, N., Coxon, C., Cunha Periera, H., Gill, L., Gonzalez, A., et al. (2015). *Turlough hydrology, ecology and conservation*. Unpublished Report, National Parks & Wildlife Services. Department of Arts, Heritage and the Gaeltacht.

Wei, H. L., Billings, S. A., & Liu, J. (2004). Term and variable selection for non-linear system identification. *International Journal of Control*, *77*(1), 86–110. https://doi.org/10.1080/00207170310001639640

White, W. B., & White, E. L. (2005). Ground water flux distribution between matrix, fractures, and conduits: Constraints on modeling. *Speleogenesis And Evolution Of Karst Aquifers*, *3*, 1–6.

Wunsch, A., Liesch, T., & Broda, S. (2018). Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). *Journal of Hydrology*, *567*, 743–758. https://doi.org/10.1016/j.jhydrol.2018.01.045

Wunsch, A., Liesch, T., & Broda, S. (2021). Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX). *Hydrology and Earth System Sciences*, *25*, 1671–1687. https://doi.org/10.5194/hess-25-1671-2021

Ye, L., Hanson, L. S., Ding, P., Wang, D., & Vogel, R. M. (2018). The probability distribution of daily precipitation at the point and catchment scales in the United States. *Hydrology and Earth System Sciences*, *22*(12), 6519–6531. https://doi.org/10.5194/hess-22-6519-2018