09 - FORECASTING RAINFALL AND FLOOD EVENTS AT DODDER BASIN USING A COMBINATION OF SUPPORT VECTOR REGRESSION AND RAINFALL-RUNOFF MODELS

Bidroha Basu[1], Maria Nogal[2] and Alan O'Connor[3]

[1,2,3] *Department of Civil, Structural and Environmental Engineering, Trinity College Dublin*

Abstract

Over the past few decades there have been scientific evidences that global climate has changed significantly due to several anthropogenic influences. One of the major concerns due to this climate change is the change in rainfall processes and how the changes affect the flooding pattern in a river basin. Hydrologists are focused on assessing the impact of climate change on a watershed scale at several river basins around the world, in order to devise strategies for sustainable management of water resources in those basins. For this purpose, future forecasting of rainfall events are necessary to obtain flood forecast. This study develops a support vector regression (SVR) based downscaling model by formulating regression equations between the rainfall and set of predictors, which are large scale atmospheric variables (LSAVs) obtained from European reanalysis data ERA-40. Subsequently, the developed SVR model is used in arriving at future projections of rainfall using the seasonal climate forecast obtained from the national centers for environmental prediction coupled forecast system model version 2 (CFSv2). The rainfall projections in the study are at monthly scale and hence they were subsequently disaggregated to a daily scale by using a k-nearest neighbour disaggregation technique. Once the daily projections of rainfall are obtained, they can be used as an input in a physically based rainfall-runoff model to obtain future forecast of streamflow data. The forecasted flows can then be used to perform a real-time flood forecasting. The approaches described in this paper were implemented to obtain future flood forecasts at the tributary of River Liffey near Dublin, Ireland. It has been observed from the analysis that the frequency and intensity of flood events are expected to increase in future due to climate change. Flood risk assessment can be performed using the forecasted flood events.

1. INTRODUCTION

Streamflow forecasts in river basin are needed for effective management of water resources. Streamflow forecast can be obtained by using two types of models, namely data driven models and physically based models. The data driven model develops a mathematical/statistical relationship between the streamflow and variables (predictors) affecting streamflow. On the other hand, the physically based model utilizes the physics of inter-relationships among various hydrological processes in the basin contributing flow to the target location. The major advantage of data driven models are that they are simple and easily implementable. However, those models cannot consider the physics of hydrological processes and hence require very large datasets to develop an effective model. The physically based models are complex in nature but are more accurate in flood forecasting. Some examples of data driven models are Artificial Neural Networks (ANN), fuzzy logic based modelling, time series analysis and support vector machine, of which ANN have been extensively used by several hydrologists (e.g. ElShafie *et al.*, 2006; Shu and Ouarda, 2008, Campolo *et al.*, 2010, He *et al.*, 2014). Examples of some of the physically based models are Soil Water Assessment Tool (SWAT), TOPMODEL, Hydrologic Engineering Center's River Analysis System (HEC-RAS), and Variable Infiltration Capacity (VIC) model.

The physically based models can be broadly classified as lumped or distributed models based on the modelling approach. In lumped models, the entire catchment is considered as a single unit having same hydrological process. Distributed models discretize the catchment into a large number of elements/grids and develop separate equations of hydrological processes corresponding to every element/grid. Parameter values representing the hydrological process in each grid can be different. In a catchment, areas with the same soil type and land use form a Hydrologic Response Unit (HRU). HRU can be considered as a basic computational unit which is assumed to be homogeneous in hydrologic response.

The main source of hydrological input to a catchment is rainfall. To obtain forecast of streamflow, first, forecast of rainfall needs to be obtained by considering the climate change scenario at river basin scale. The approach to obtain rainfall forecast is by using the downscaling approaches. Downscaling approaches can be broadly classified as the dynamic downscaling and statistical downscaling. In the dynamic downscaling approach, a regional climate model is embedded into a general circulation model (GCM) [or climate system forecast (CFS) model], with its boundary conditions specified by the GCM/CFS. Statistical downscaling approach involves developing statistical relationships between the watershed-scale meteorological variables (rainfall) and GCM/CFS simulated climate variables available at a relatively coarser scale. This study considered a statistical downscaling model to obtain the forecasted rainfall due to its simplicity in application.

2. PROPOSED METHODOLOGY

This section presents the SVR approach to obtain future projections of rainfall. Following this, theoretical background of the SVR approach and SWAT model to obtain the streamflow projections from rainfall are provided.

2.1. SVR approach to downscaling of rainfall

The SVR model is used to develop a relationship between the large scale atmospheric predicted variables (LSAPVs) and the historical rainfall. The developed relationship is used to obtain the future projections of rainfall considering climate change scenario in the study area. The methodology is given in the following steps.

1. Identification of predictor variables from the large scale atmospheric variables (LSAVs) available in both observed/reanalysis data and CFS simulations, such that they are reasonably well correlated with the rainfall (predictand) at the target location in the study area. The identified predictor variables are henceforth referred to as large scale atmospheric predictor variables (LSAPVs). Due to unavailability of observed values of the LSAVs, reanalysis data are considered for analysis. Further, CFS simulation data are considered at monthly scale, as those simulations at finer time scale (daily/sub-daily) are not considered to be reliable (Prudhomme et al., 2002, Brown et al., 2008). In situations where the spatial resolutions of CFS and reanalysis data are different, the CFS data should be spatially interpolated to the resolution of reanalysis data by using software such as Grid Analysis and Display System (GrADS; Doty and Kinter, 1993). In the present study, European reanalysis data ERA-40 (Interim) data at 2.5°×2.5° were considered at four pressure levels (850mb, 700mb, 500mb and 200mb) in which the CFS data are also available. The reanalysis data were downloaded from "http://apps.ecmwf.int/datasets/data/interim-full-moda/", and the CFSv2 data were downloaded from "https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/climate-forecast-system-version2-cfsv2#CFSv2 Operational Forecasts". The spatial resolution of the CFSv2 data along longitudinal direction were 2.5°, however the resolution along latitudinal direction varied non-uniformly (ranging from 2.4619° to 2.4827°), and hence the data were interpolated to 2.5°×2.5° grid resolution, which matched with the ERA reanalysis data.

2. Before developing the downscaling model, the data (LSAPVs and historical observations of rainfall) were divided in two sets, namely the calibration set and the validation set. The model is developed considering the LSAPVs and rainfall data from the calibration set and then the performance of the model is evaluated based on the data from validation set. The time period

corresponding to calibration set is termed as baseline period. The reanalysis data corresponding to each of the LSAPVs for calibration period are standardized by subtracting its respective mean and dividing by the standard deviation. Standardization of the LSAPV is necessary to nullify the effect of differences in magnitude, range and variance of values corresponding to the LSAPVs.

3. The standardized LSAPVs is provided as an input to the least squares support vector regression (LS-SVR) model to develop regression relationships between the LSAPVs and the contemporaneous observed monthly values of rainfall data at the target site. To develop the regression model, values of the two SVR parameters γ and C are needed. As the optimal values of those parameters are not known *a priori*, a grid search procedure (Gestel *et al.*, 2004) has been used in the study. The optimal parameters were identified as those for which output of the model produce low error in terms of Root Mean Square Error (RMSE), and high value of Nash–Sutcliffe Efficiency (NSE).

4. Performance of the developed model was then validated using the data from the validation set. For this purpose, the LSAPVs corresponding to the validation set were first standardized by subtracting the mean and then dividing by the standard deviation. The standardized LSAPVs were then considered as an input to the developed SVR model to obtain rainfall projections. The projected rainfalls were compared with the observed rainfall and the errors are noted in terms of performance measure (RMSE and NSE). If the model performance is satisfactory, the model can be used for future projections of rainfall. Otherwise, it is necessary to scrutinize the chosen sets of predictor variables to assess whether they are adequate, and whether any additional LSAPVs are necessary to develop an effective downscaling model.

5. The future projections of rainfall were then estimated based on the validated SVR downscaling relationship. To obtain the future projection, each LSAPV corresponding to the CFS data was bias-corrected using the SVR approach. For this purpose, SVR relationships for bias-correction were developed between the LSAPV values obtained using the reanalysis data and the CFS data for baseline period. The developed SVR relationships were then used for bias correction of the future CFS data. Subsequently, the bias-corrected LSAPVs for CFS data were provided as input to the downscaled SVR relation to obtain future projected rainfall at monthly scale.

6. To obtain future streamflow, the rainfall data were needed at daily scale. For this purpose, the monthly rainfall is disaggregated to daily scale using the k-nearest neighbour disaggregation methodology (Srinivas *et al.*, 2014). The methodology involves finding a nearest neighbour to the downscaled monthly values of predictand (rainfall). A nearest neighbour refers to an observed monthly value of rainfall that is deemed closer to the downscaled monthly value of the rainfall. Historical relationship between the nearest neighbour and its corresponding daily values of rainfall for the site is used to disaggregate the downscaled monthly values to daily values. The daily values of future projected rainfall can be considered as an input to a physically based rainfall-runoff model to obtain future projections of streamflow for site in the study area. It can be noted that the percentage increase (decrease) in daily rainfall averaged over a month is assumed to be equal to the percentage increase (decrease) in monthly forecasted rainfall value corresponding to that month.

2.2. Support Vector Regression

This section describes the procedure to develop SVR relationship (Vapnik, 1995) in terms of input vector (predictor) $\boldsymbol{x}_t = \left[x_{t,1}, \ldots, x_{t,m} \right] \in \Re^m$ where m denote the number of predictors, and output (predictand) $y_t \in \Re$ corresponding to time $t = \left\{1, 2, \ldots, N\right\}$.

The relationship between $\boldsymbol{x}_t$ and y_t can be expressed as,

$$y_t = f\left(\boldsymbol{x}_t\right) + \varepsilon_t \tag{1}$$

where $\left\{f\left(\cdot\right);\Re^m \rightarrow \Re\right\}$ is a nonlinear transformation function, and $\boldsymbol{\varepsilon}_t$ is white noise whose expected value $E\left[\boldsymbol{\varepsilon}_t\right]$ is zero.

Let us define a function $\phi(\cdot)$ to map $\boldsymbol{x}_t$ at a higher p-dimensional space, where linear relationship exists between $\phi\left(\boldsymbol{x}_t\right)$ and y_t, which can be expressed as,

$$\left[y_t\right]_{1\times1} = \left[\phi\left(\boldsymbol{x}_t\right)\right]_{1\times p}\left[\boldsymbol{w}\right]_{p\times1} + \left[b\right]_{1\times1} \tag{2}$$

where $$\left[\boldsymbol{w}\right]_{p\times1} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} \tag{3}$$

The parameters $w_1, w_2, \ldots, w_p$ and b can be estimated by optimizing the following objective function.

$$\mathcal{L}\left(\boldsymbol{w},\boldsymbol{e}\right) = \min_{\boldsymbol{w},b,\boldsymbol{e}}\left[\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\frac{1}{2}\sum_{t=1}^{n} e_t^2\right] \tag{4}$$

where $$e_t = y_t - \phi\left(\boldsymbol{x}_t\right)\boldsymbol{w} - b \tag{5}$$

and $\left[\boldsymbol{e}\right]_{N\times1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$. C is a SVR parameter considered to be the weight associated to the sum of squares of error e_t at a given time. Choosing higher value of the parameter reduces (increases) the error in model prediction by overfitting (under fitting) the model.

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.

Introducing Lagrange multiplier, the optimization problem becomes,

$$\mathcal{L}\left(\boldsymbol{w},b,\boldsymbol{e};\boldsymbol{\alpha}\right) = \min_{\boldsymbol{w},b,\boldsymbol{e}} \mathcal{L}\left(\boldsymbol{w},\boldsymbol{e}\right) - \sum_{t=1}^{N}\alpha_t\left\{\phi\left(\boldsymbol{x}_t\right)\boldsymbol{w} + b + e_t - y_t\right\} \tag{6}$$

where α_t are the Lagrange multipliers and $[\boldsymbol{\alpha}]_{N\times1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}$. The optimal solution can be obtained

from the conditions
$$\begin{aligned} &\frac{\partial\mathcal{L}}{\partial \boldsymbol{w}} = 0; \quad \Rightarrow \boldsymbol{w} = \sum_{t=1}^{N} \alpha_t \phi(\boldsymbol{x}_t) \\ &\frac{\partial\mathcal{L}}{\partial b} = 0; \quad \Rightarrow \sum_{t=1}^{N} \alpha_t = 0 \\ &\frac{\partial\mathcal{L}}{\partial e_t} = 0; \quad \Rightarrow \alpha_t = C \times e_t \, \forall t \\ &\frac{\partial\mathcal{L}}{\partial \alpha_t} = 0; \quad \Rightarrow \boldsymbol{w}^T \phi(\boldsymbol{x}_t) + b + e_t - y_t = 0 \, \forall t \end{aligned} \tag{7}$$

Based on those conditions in equation (7), and eliminating $\boldsymbol{w}$ and $\boldsymbol{e}$ we get,

$$\begin{aligned} &\sum_{t=1}^{N} \alpha_t = 0 \\ &\sum_{t'=1}^{N} \alpha_{t'} \phi(\boldsymbol{x}_{t'})^T \cdot \phi(\boldsymbol{x}_t) + b + \frac{\alpha_t}{C} = y_t; \quad \text{for } \forall t = 1,\ldots,N \end{aligned} \tag{8}$$

where $\phi(\boldsymbol{x}_t)^T \cdot \phi(\boldsymbol{x}_t)$ is the dot product of the input vector in the high-dimensional transformed space. The dot product can be expressed by a Kernel function (Vapnik, 1995) given as,

$$\phi(\boldsymbol{x}_i)^T \cdot \phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{9}$$

This study considered Gaussian radial basis kernel function for analysis, which can be expressed as

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}, \quad \gamma > 0 \tag{10}$$

For chosen values of γ and C, values of $\{\alpha_t, t = 1,\ldots,N\}$ and b were estimated by solving equation (8). Subsequently, future projections of the predictand y_{t_f} can be obtained as,

$$y_{t_f} = \sum_{t=1}^{N} \alpha_t \phi(\boldsymbol{x}_t)^T \cdot \phi(\boldsymbol{x}_{t_f}) + b = \sum_{t=1}^{N} \alpha_t K(\boldsymbol{x}_t, \boldsymbol{x}_{t_f}) + b \tag{11}$$

where $\boldsymbol{x}_{t_f}$ is the future projection of predictor vector corresponding to time t_f.

2.3. Soil Water Assessment Tool (SWAT)

The SWAT model uses the Digital Elevation Model (DEM) to obtain stream network in the target catchment. The land-use/land-cover and soil information is merged with the DEM information to form sub-catchments in the study area. Subsequently, the Hydrologic Response Unit (HRU) were prepared for each of the sub-catchments. The identified HRUs are then convoluted with the forecasted rainfall at daily scale to obtain future forecast of daily streamflow.

3. RESULTS AND DISCUSSION

In the present study daily streamflow value observed at stream gauge located at Waldron's bridge (station no. 09010) on tributary of River Liffey (River Dodder) were considered for analysis. The data were obtained from "https://waterlevel.ie/". The data ranged from 1st February, 1986 to 30th April, 2016. Catchment of the stream gauge was delineated by using the "spatial analyst toolbar" in ArcGIS framework by processing SRTM DEM data having 90 m resolution. The DEM were extracted from the website "http://srtm.csi.cgiar.org/".

The rainfall data considered for the study is observed at Merrion Square, Dublin (Rain ID 3923). The rain gauge had been selected as it is the nearest rain gauge from the stream gauge (approx. 3.62 km). The rainfall data were downloaded from "http://www.met.ie/climate/daily-data.asp" and data ranged from 1st January 1948 to 30th April, 2016. It can be noted that in practice, one should use rainfall data available at multiple rain gauges in a river basin to obtain weighted average rainfall by using techniques such as the Thiessen polygon or the Isohyetal maps, for better estimation of rainfall time series over a river catchment. In this study, rainfall data corresponding to a single rain gauge is used for illustration, even though rainfall data for several gauges are available for the basin.
However, since streamflow data at the target location were available from 1st February, 1986, rainfall data from this date were considered for analysis. The daily average rainfall at the rain gauge for the chosen period were found to be 1.97 mm. The time series of daily rainfall is shown in Figure 1(a) and the monthly mean of daily rainfall are plotted in Figure 1(b). It can be noted from the Figure 1(a) that the number of extreme rainfall days was low from 1986 to 1992, and number of extreme rainfall days started to increase from the year 1993 till 2003. There were again low rainfall events from 2004 to 2007 and extreme rainy days increased from early 2008. From Figure 1(b), it can be noted that the monthly average rainfall is low in the winter and summer months (January to July). The monthly mean rainfall in the study area increases from the month of August and the rainfall is, in general, high in the months of October and November.

To develop the SVR model, 10 LSAPVs at six reanalysis grid points (in latitudinal direction: 52.5°N and 55°N; longitudinal direction: 5°W, 7.5°W and 10°W) surrounding the study area were chosen. The selected LSAPVs are found to be reasonable well correlated with the rainfall data. The LSAPVs are: Air Temperature at 500mb, Geopotential Height at 700mb and 200mb, U-wind at 850mb and 200mb, V-wind at 700mb and 200mb, W-wind at 500mb, Precipitable water and Mean Sea Level pressure. Data ranging from February 1986 to December 2005 were considered as the baseline period and data ranging from January 2006 to April 2016 were considered for the validation period.

To identify optimal values of the SVR parameters γ and C, values of both γ and C were varied from 1 to 200 with a unit interval. The optimal values of the SVR parameters were found to be $\gamma = 7$ and $C = 27$. The performance of the developed SVR model were quantified in terms of RMSE and NSE. The error measures are shown if Table 1.

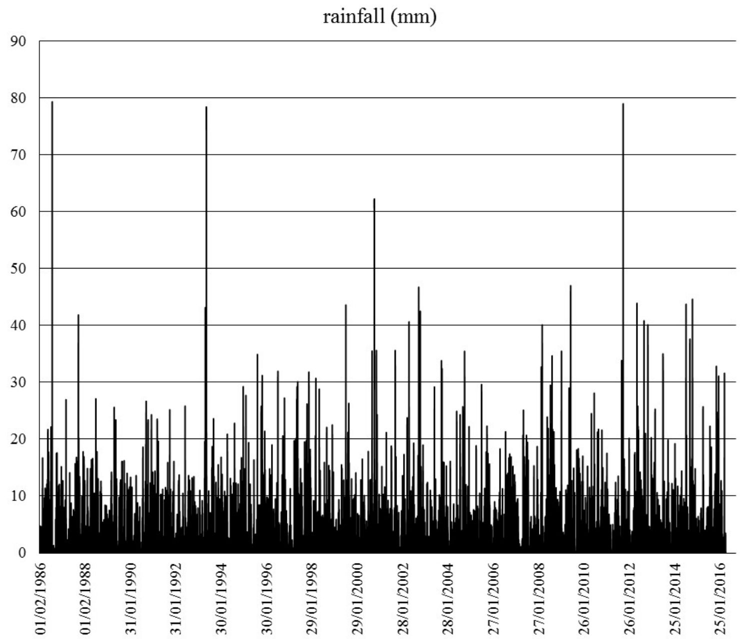

Figure 1a: *Time series of observed daily rainfall at Merrion Square, Dublin.*

To obtain the future projections of rainfall, LSAPVs for CFSv2 simulated data for the period September 2016 to May 2017 were selected. The simulated LSAPV data were first bias corrected using SVR approach (described in step 5 in section 2.1), and the bias-corrected LSAPVs were standardized by subtracting the mean and dividing by the standard deviation. The standardized values of LSAPVs were then considered as an input to the developed SVR downscaling model to obtain monthly forecast of rainfall data at the target site. The rainfall data were then disintegrated from monthly scale to daily scale using k-nearest neighbor disaggregation methodology. The forecasted rainfall values are shown in the Figure 2. It can be noted from the figure that the rainfall intensity is expected to increase in October to November, whereas the rainfall is expected to reduce from January to May. It has been noted that the forecasted rainfall is high in September and October and reduces from beginning of November. The rainfall is expected to be lower in January to May when compared to historical observations. Since the weather in Ireland is mainly influenced by Atlantic Ocean (http://www.met.ie/climate-ireland/climate-of-ireland.asp), change in the sea surface temperature (SST) in the ocean plays a major role in the rainfall pattern over the country. More investigation relating the large scale atmospheric variables and the oceanic variables is needed to attribute the change in rainfall pattern in the study area.

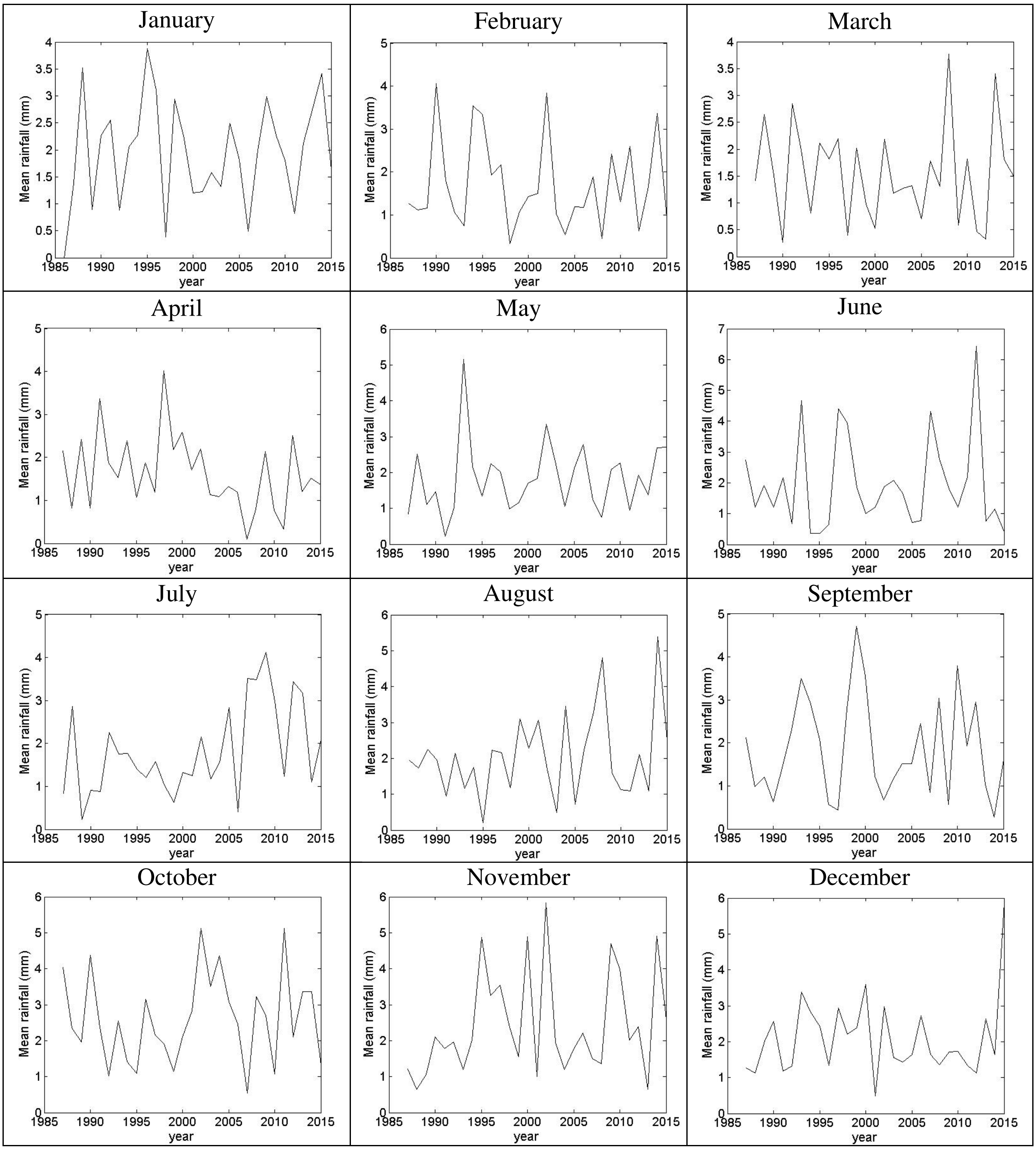

Figure 1b: *Monthly mean of daily mean rainfall from 1986-2015.*

Table 1: *Performance measure of developed SVR model.*

Error measure	Calibration	Validation
RMSE	0.4096	0.8468
NSE	0.8519	0.4784

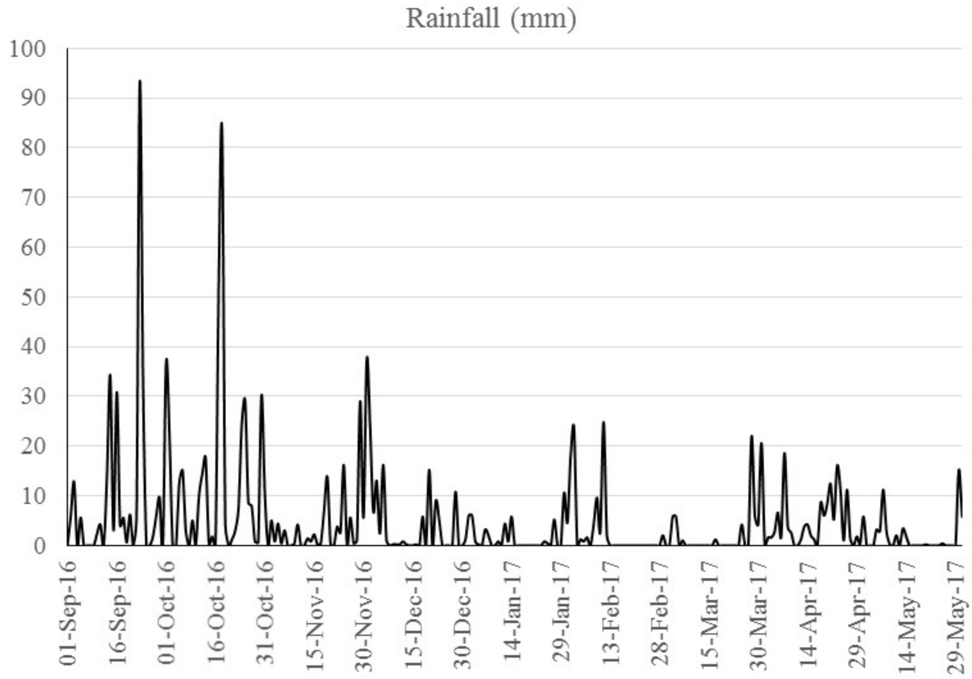

***Figure 2:** Forecast of daily rainfall at Merrion Square, Dublin.*

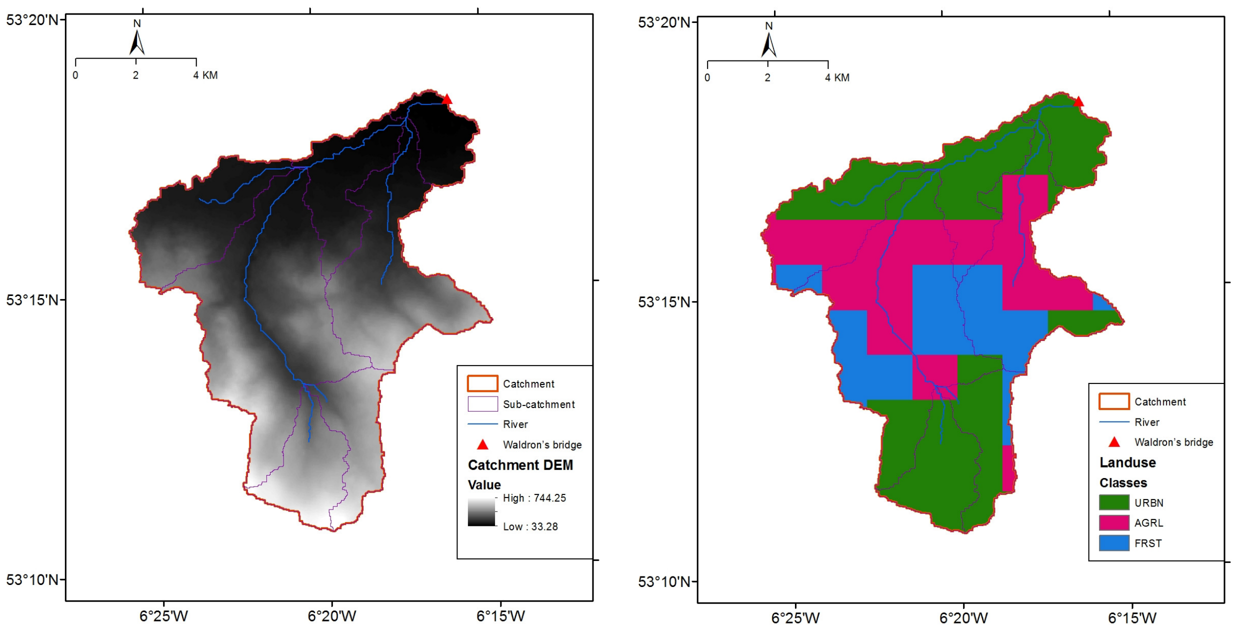

***Figure 3:** DEM, river network, catchment, sub-catchment and land use details (agriculture: AGRL, forest: FRST, urban: URBN, and wetland: WETL).*

To obtain streamflow forecast, SRTM DEM of 90m resolution is used in the study. The river network had been obtained by analyzing DEM in ArcView interface of the SWAT model (see Figure 3). The soil information was obtained from the general soil map of Ireland (Gardiner and Radford, 1980) and the land use information were obtained from the CORINE land use map (CORINE, 1989) which classifies the land according to the dominant use (agriculture: AGRL, forest: FRST, urban: URBN, and wetland: WETL). The SWAT analysis generated 34 HRUs in the study area. The HRUs are then convoluted with the forecasted rainfall to obtain the forecasted streamflow from September 1st, 2016 to May 31st, 2017 (see Figure 4).

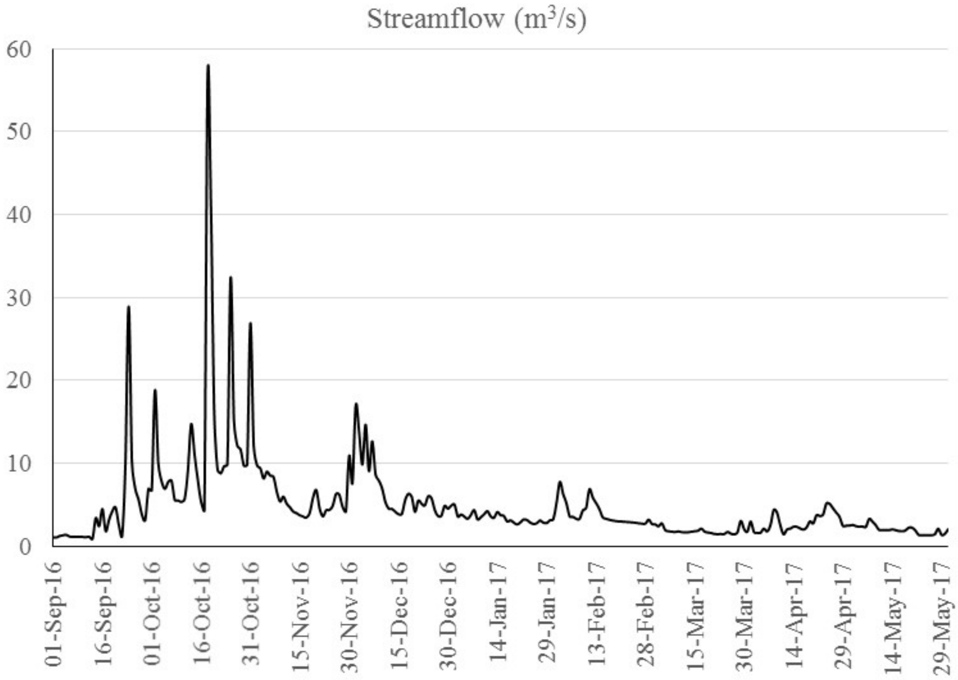

Figure 4: *Forecast of daily streamflow at Waldron's bridge, Dublin.*

It can be observed from the figure that the streamflow is expected to increase in the autumn season (September to December) and is expected to decrease substantially from January to May. The maximum discharge noted at the location from the historical data was 45.7m^3/s. It can be noted that the forecasted peak flow is approximately 30% higher than that observed in the historical data. Further, the mean value of the forecasted streamflow from months September to May was found to be 3.94m^3/s, whereas the historical mean during this period (September to May, excluding June-August) was estimated as 2.51m^3/s. The increase in mean is due to forecasted high flows in October and November. It can be noted from Figure 2 that for the months September to November the forecasted values of rainfall are high. Also, the number of high rainfall events are expected to increase for those months when compared to that observed in the historical data. The combined effect of expected increase in rainfall intensity and the number of high rainy events results in significant increase in mean and peak flow values of streamflow at the outlet of the river basin.

4. CONCLUSION

The paper develops a support vector regression (SVR) based downscaling approach to obtain future projections of rainfall in Dublin based on a set of large scale atmospheric predictor variables (LSAPVs) available from Climate Forecast System. For this purpose, a set of LSAPVs were identified for the study area. A Soil Water Assessment Tool based physical rainfall-runoff model were developed to obtain future forecasts of streamflow at a river basin in Dodder river, Ireland. The catchment considered in the study is a small catchment with steep slope. It can be noted that it is a challenging problem to forecast hydro-meteorological variables (rainfall and streamflow) at smaller basins having higher slope (due to orographic precipitation and flash floods) at daily/sub-daily time scale. The analysis indicated that the rainfall and streamflow is expected to increase in the autumn season and reduce in the winter and summer seasons. The forecasted streamflows are found to be higher than the historically observed streamflow in the study area, which can be attributed to the increase in forecasted rainfall intensity as well as the number of high intensity rainfall events. A detailed analysis is underway to attribute the changes in rainfall and streamflow pattern in Ireland.

5. REFERENCES

Brown, C., Greene, A., Block, P., and Giannini, A. (2008) Review of downscaling methodologies for Africa climate applications. IRI Technical Report 08-05, International Research Institute for Climate and Society, Columbia University, USA.

Campolo, M., Andreussi, P., and Soldati A. (1999) River flood forecasting with a neural network model. Water Resources Research, vol. 35, no. 4, pp. 1191–1197.

CORINE, (1989) Land Cover Project, The Nomenclature. Commission of the European Communities, Directorate-General Environment Nuclear Safety and Civil Protection, Rue de la Loi 200, B-1049 Brussels, Belgium.

Doty, B., and Kinter III J. L. (1993) The Grid Analysis and Display System (GrADS): a desktop tool for earth science visualization, American Geophysical Union 1993 Fall Meeting, San Fransico, CA, 6-10 December.

El-Shafie, A., Taha, M.R., and Noureldin, A. (2006) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. Water Resources Management 21, 533–556.

Gardiner, M.J., and Radford, T. (1980) Soil Associations of Ireland and their land use potential (Explanatory Bulletin to Soil Map of Ireland 1980). Published by An Foras Taluntais, 19 Sandymount Avenue, Dublin 4, Ireland.

Gestel, T. V., Suykens, J. A. K., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., Moor, B. D., and Vandewalle J. (2004) Benchmarking least squares support vector machine classifiers, Mach. Learn., 54 (1), 5–32.

He, Z., Wen, X., Liu, H., and Du, J. (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology 509: 379–386.

Prudhomme, C., Reynard, N., and Crooks S. (2002) Downscaling of global climate models for flood frequency analysis: where are we now? Hydrological Processes, 16: 1137–1150.

Shu, C., and Ouarda, T.B.M.J. (2008) Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. Journal of Hydrology 349, 31–43.

Srinivas, V. V., Basu, B., Nagesh Kumar, D. and Jain, S. K. (2014) Multi-site downscaling of maximum and minimum daily temperature using support vector machine. International Journal of Climatology, Wiley & Royal Meteorological Society, Vol. 34, Issue 5, pp.1538-1560, April 2014, doi: 10.1002/joc.3782

Vapnik, V (1995) The nature of statistical learning theory. Springer-Verlag, New York, 1995.