Estimating the Parameters of the Extreme Value Type 1 Distribution for Low Flow Series in Ireland

A. Nasr and M. Bruen
Centre for Water Resources Research

Abstract

In this study two different models have been developed and tested with data from 55 hydrometric stations in the Shannon River Basin in Ireland for estimating the location and the scale parameters of the EV1 distribution of the Minimum, 3-, 7-, 10-, 15-, and 30-days sustained low flow series at ungauged locations. The first is a simple linear model while the other is a fuzzy clustering model. Both models have been calibrated using the unconstrained and the constrained least squares methods. Moreover five different input scenarios including various combinations of some explanatory variables have been investigated with the two models. The results showed that: (i) the simple linear model calibrated by the constrained least squares method was the best model to estimate the EV1 location parameter using catchment area, mean annual rainfall, mean elevation, mean slope, and soil as explanatory variables, and (ii) the fuzzy clustering model calibrated by the unconstrained least squares method was the best for the EV1 scale parameter using only the first four above mentioned explanatory variables.

Keywords: Low flow frequency curve, k-means fuzzy clustering method, EV1 distribution, Catchment characteristics, Model Calibration, Model Validation.

1 Introduction

Estimates of the frequency of drought events or significant periods of low stream flows are vitally important for the planning, design and evaluation phases of any sustainable development project. In particular, such low flow frequency estimates are very often the critical determinant in evaluating the ability of a stream to meet current and future water supply needs and taking account of issues such as water quality, food security, freshwater ecosystem integrity, adaptation to climate change, and governance [1]. This paper describes a new method for estimating the parameters of a frequency analysis model to be used, in Ireland, as a tool for obtaining design values for low flows associated with water quality.

Technically the frequency analysis of low flow data can provide an indication of the adequacy of the natural flow to meet a given demand with a stated probability of experiencing a shortage [2]. Thus the frequency analysis is normally performed on one of a number of possible low flow indices characterising the low flow conditions [3]. Among those indices the sustained low flows over certain periods (e.g. 1, 3, 7, 10 days) are used in most of the published environmental and ecological studies [4] and because of this they are the focus of this paper. For the sustained low flow series of a specified duration, design values can be derived from the low flow frequency curve (LFFC). This curve is constructed, for a chosen statistical distribution, using historical measured flow data at a particular site in the study catchment. It is often extrapolated to predict design values of extreme low flow conditions. This, of course, should only be done if the underlying model is sufficiently robust and future metrological and hydrological conditions remain similar to what have been prevailed. However, this assumption of invariability with time of those conditions may no longer be valid with the expected climate changes and consequently the derived LFFCs need to be adjusted in order to accommodate those changes. Such adjustment is reflected in changes to the parameters of the fitted statistical distribution.

The EV1 distribution, which is widely used in low flow frequency analysis, has two parameters; the location parameter and the scale parameter. At any hydrometric station the two parameters can be calibrated using historical flow measurements which represent the aggregated effects of the past weather conditions and the influence of some important physical characteristics including area, topography, soil, groundwater, and geology on the hydrologic behaviour of the catchment. In this paper a new approach is introduced whereby the two EV1 parameters of a number of low-flow indices including (1) the Minimum or 1-day sustained low flow (Min); (2) 3-day sustained low flow (3-SLF); (3) 7-day sustained low flow (7-SLF); (4) 10-day sustained low flow (10-SLF); (5) 15-day sustained low flow (15-SLF); and (6) 30-day sustained low flow (30-SLF) series are estimated from the mean annual rainfall and catchment characteristics. Two different models representing this approach have been constructed and their performance was tested using data from 55 flow measuring stations in the Shannon River Basin (SRB) in Ireland. The first model is a simple baseline linear relationship relating each of the EV1 parameters with five various combination scenarios of the above-mentioned catchment physical characteristics. The second model is more sophisticated and applies the k-means clustering algorithm [5] to relate the same variables.

2 Characterisation of the Low Flow Frequency Curves (LFFCS)

The primary objective of frequency analysis in a hydrologic context is to infer the probability of some defined event occurring. In low flow analysis the event is a value of a low flow index occurring that is lower than a specified threshold.

This frequency analysis is based on two assumptions [6]: (i) the data in the series form a random representative sample from the distribution of low flow series at the sampling site; and (ii) these series are stationary so that the expected frequency for a particular event does not change over time. The analysis consists of selecting an appropriate frequency distribution, estimating the parameters of the distribution from the data, and then evaluating the distribution function at various points of interest. Some theoretical distributions that have been used in hydrologic frequency analysis are the normal (Gaussian), log normal, exponential, two-parameter gamma, three-parameter gamma, Pearson type III, log-Pearson type III, extreme value (Gumbel) type 1 (EV1), log Gumbel, and Weibull.

The EV1 distribution is used in this study since it is widely adopted in Ireland for both flood frequency estimation and by the Irish Environmental Protection Agency (EPA) for low flow analysis. This distribution has two parameters: (1) the location; and (2) the scale parameters. Values for the two parameters are obtained by one of the recommended fitting techniques [7] and in this study the probability weighted moments (PWMs) method [8] was used.

PWMs are useful in deriving expressions for the parameters of distributions whose inverse form $x = x(F)$ can be explicitly defined [7]. This technique solves two derived relationships for the unknown EV1 parameters. The calculation of the PWMs requires the probability of non-exceedance ($F(x)$) of each value of the low flow index [8]. This can be estimated from an empirical plotting position formula [9] which estimates the non-exceedence probability of each value of the index from its rank within the series. Gumble paper formula [10] was used in this study. In practice $F(x)$ can also be expressed as a return period in years (T) in the LFFC. $F(x)$ and T are related to each other by the equation:

$$F(x) = 1 - \frac{1}{T} \tag{1}$$

The LFFC can be used as a predictive tool to estimate index values associated with specified return periods longer than the record period and vice versa (i.e. estimating the return period corresponding to a specific index value). This is equivalent to calculating a quantile ($x(T)$) from a cumulative function $F(x)$ of the statistical distribution which fits the annual series of low flow indices. Determining the frequency distribution of low flow indices by using this analytical technique has several advantages including: (i) the use of an established procedure for fitting a selected distribution would result in consistent frequency estimates from the same data set by different persons; (ii) error distributions have been developed for some of the theoretical distributions that enable computing the degree of reliability of the frequency estimates; and (iii) it is possible to regionalise the parameter estimates which facilitates making frequency estimates at ungauged locations. In this paper, a new approach is proposed for regionalising the LFFCs through developing a regression model that relates the location and the scale parameters of the EV1 distribution with some physical variables. Two different types of this model have been developed including a simple linear regression model and a more complex k-means fuzzy clustering model. A brief description of the k-means clustering algorithm is given below before discussing the proposed approach.

3 k-means Fuzzy Clustering Algorithm

Clustering is the process of finding groups of objects (or data) such that the objects in a group are similar (or close) to one another and different from (or far from) the objects in other groups. Some defined measure of similarity is required and, in the case of numerical data, Euclidean distance is often used to determine proximity. The k-means clustering algorithm [5] is one of the simplest unsupervised learning algorithms for allocating data to clusters when the number of clusters (k) is specified in advance. It assigns each data point to the cluster whose centroid is nearest. The coordinates of the centroid is the arithmetic mean for each dimension separately of the coordinates of all the points in the cluster.

The k-means clustering algorithm consists of the following steps:

(1) Start with some initial trial choice of positions for the k centroids, one for each cluster. These initial centroids should be chosen carefully because different starting locations generate different results. They should be as far away from each other as possible, given the range of coordinates spanned by the data set.

(2) The next step is to take each point belonging to a given data set and associate it with the nearest centroid.

(3) At this point k new centroids are calculated for each cluster from the points assigned to the cluster in the previous step.

(4) Steps (2) and (3) are repeated until the change in the cluster centroids is insignificant.

In essence the algorithm aims at determining the values of c_j minimising an objective function, in this case a squared error function (J) in the following form:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left\| x_i^{(j)} - c_j \right\|^2 \tag{2}$$

and

$$c_j = \frac{1}{n_j}\sum_{i=1}^{i=n_j} x_i^{(j)} \tag{3}$$

Where $\|x_i^{(j)}-c_j\|$ is a distance measure between the *ith* data point of the *jth* cluster $x_i^{(j)}$ and the cluster centre c_j; n_j is the number of the data points in the *jth* cluster.

4 New model for estimating the EV1 distribution parameters

The traditional method of fitting the EV1 distribution to any flow series at a particular site, as described in section 2 above, requires historical flow data. This in turn limits the frequency analysis to gauged sites where such data is available. To estimate frequencies at other locations modellers use regional frequency analysis [11]. Two main steps are always involved: (i) grouping the available gauged sites into different homogenous groups; and (ii) fitting a distribution to each group and then deriving a relation between the parameters of this distribution and some physical descriptor of that group. In this study, an analogous approach has been followed whereby two different models relating the EV1 distribution parameters (location and scale) with some physical characteristics of a catchment have been developed for a number of low flow indices, i.e. Minimum, 3-SLF, 7-SLF, 10-SLF, 15-SLF, and 30-SLF. The optimum values of the EV1 parameters for the five low flows series in each station were obtained by fitting the EV1distribution to each of the five low flow series using the method of PWM. Six physical characteristics of the catchment including area, rainfall, mean elevation, mean slope, soil, geology, and aquifer type were used.

The first model fitted is a simple baseline linear regression equation relating each of the EV1 parameters of the above-mentioned low flow indices with five different combinations of the six physical features of the catchments. The five combination scenarios include: (Sc1) area and rainfall; (Sc2) area, rainfall, mean elevation, and mean slope; (Sc3) area, rainfall, mean elevation, mean slope, and four separate soil variables representing very well-drained, well-drained, moderate, and poor drainage conditions; (Sc4) area, rainfall, mean elevation, mean slope, and four separate geology variables representing the percentage area in the catchment of limestone, metamorphic, old red sandstone, and sedimentary rocks; and (Sc5) area, rainfall, mean elevation, mean slope, and four separate aquifer variables representing gravel, karstic, poorly productive bedrock, and productive fissured bedrock types. Each of these independent variables is scaled to lie between 0 and 1 before use.

More complex, model was developed in order (i) to relax the implied assumption of homogeneity between all sites in the river basin by allowing for a different model for each group; and (ii) to substitute the linearity assumption of the first model with a non-linearity one. The resulting non-linear model was formulated by combining the k-means fuzzy clustering algorithm with the linear regression equation. In this model the k-means fuzzy clustering method was used to divide the gauged sites into two homogenous groups each of which was represented by a different linear regression equation similar to that in the first model. Each data site is assigned a weight (wj) proportionate with its degree of membership of each cluster, calculated from an exponential membership function that is based on the Euclidean distance of the site from the centroid of the cluster, i.e. between the centroid of the vector of independent characteristics for each cluster and the corresponding characteristics of the site. It has two parameters, the vector giving the location (c_j) of the centriod of the cluster and the scale or the standard deviation of the cluster elements (σ_j), while the vector of physical characteristics of each site ($x_i^{(j)}$) is the input variable. The form of the function is as follows:

$$w_j \quad = \quad f\left(x_i^{(j)}\right) = e^{-\left\|\frac{\left(x_i^{(j)} - c_j\right)^2}{\sigma^2{}_j}\right\|} \tag{4}$$

A separate linear regression equation relating the EV1 parameters to catchment characteristics was then fitted to each group. Finally the predicted values of the EV1 parameters for any individual site can be calculated as a weighted average of the two linear regression models, weighted according to the membership function. The number of homogenous groups was limited in this model to two only in order to avoid any over-parameterisation in the resulting fuzzy model.

Data from 55 stations in the Shannon River Basin in Ireland, described in the next section, was used to test the two models. One set of this data containing 35 stations was used to calibrate the two models while the reaming set of 20 stations was used in the model validation process. The coefficients of the linear regression equation were the only parameters that require calibration in the first model. Whereas the second model has two sets of parameters that must be calibrated and these include: (i) parameters of the exponential membership function which is used to calculate the weights assigned to the sites in each group; and (ii) parameters of the linear regression equation for each group. The first set of parameters was computed from the k-means fuzzy clustering results. Then the unconstrained and the constrained least squares methods were used to calibrate the parameters of the regression equation. The same methods were also used to obtain the parameters of the first model. The first method does not apply any restrictions on the parameters of the linear relationship and there is a danger of producing unrealistic negative values for the EV1 distribution parameters. In contrast, the second method limits the space of the linear relationship parameters and gives only positive values for the EV1 distribution parameters.

5 Study area

The Shannon River Basin (SRB), shown in Figure 1, is the largest in Ireland with an area of more than 18,000 km^2. It includes an extensive area of Central Ireland, from its source in County Cavan to the mouth of the Shannon estuary. The river represents the main source of water for drinking, agriculture, ecology, industry, etc. for the surrounding area which includes a large number of towns and the city of Limerick. Daily flow data for 55 stations in the SRB were used in this study.

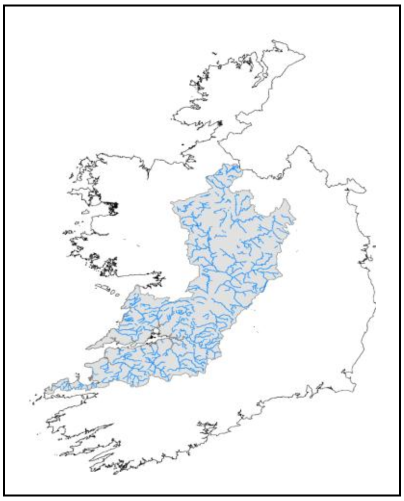

Figure 1: Shannon River Basin District shown shaded in Ireland map.

Six GIS layers representing the catchment boundary, mean annual rainfall, elevation, slope, soil, geology, and aquifer types were constructed for the catchment of each of the 55 hydrometric stations. Then indices representing the six physical features of each catchment to be used in the model were derived from those layers. For example Figure 2 and Figure 3 show the distribution of catchment area and mean annual rainfall for the 55 stations respectively.

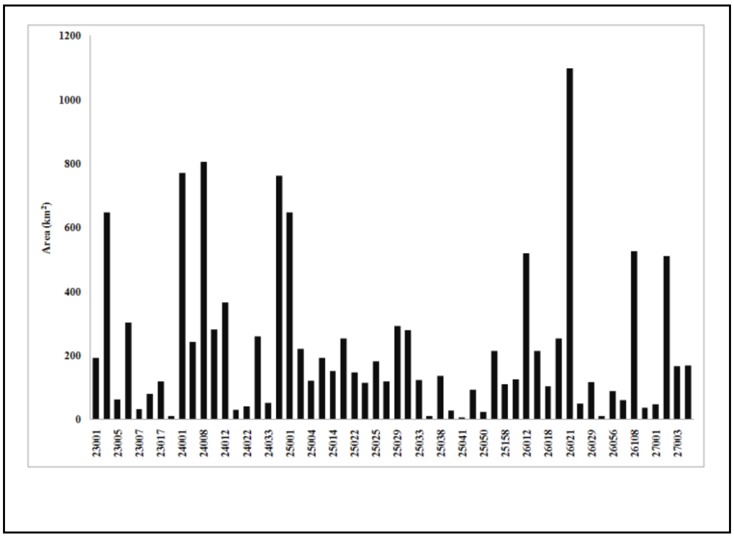

Figure 2: Distribution of the catchment areas of 55 stations in the Shannon River Basin District.

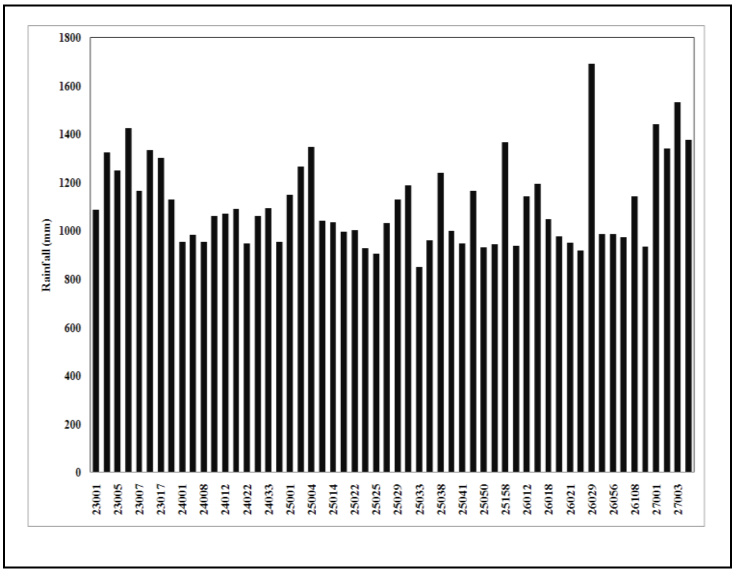

Figure 3: Distribution of the mean annual rainfall (1960-1990) over a catchment for 55 stations in the Shannon River Basin District.

6 Results of model application

The simple linear model calibrated with the unconstrained least-squares method is denoted M1 and that calibrated with the constrained least-squares method is M2. The corresponding fuzzy clustering models are denoted M3, for the unconstrained calibration and M4 for the constrained calibration. The models have been tested using the five different input scenarios (Sc1, Sc2, Sc3, Sc4, and Sc5) as described above.

The data from the 55 hydrometric stations in the SRB used in testing the models has been divided into two sets. The first set, of 35 catchments, was used to calibrate the model parameters while the other set, of 20 catchments, was used in model validation. The Nash-Sutcliffe index (R^2) [12] was calculated from the results of each model for each input scenario case for calibration and validation. To calculate this index the values of the EV1 location and scale parameters produced by the PWM method were taken as the observed data series while the corresponding values produced by any of the models taken as the estimated data series. The index is normally used to assess the model results because it can show the ability of the model to explain the variation in the observed series. The magnitude of this index ranges between 0 and 1 and generally high value for this index approaching 1 indicates a good model. However, when using this index to compare between different models the best model would be the one that produced the highest value of R^2. Table 1 to Table 5 below show the best R^2 value and the model that produced it corresponding to each input scenario for the Min, 3-SLF, 7-SLF, 10-SLF, 15-SLF, and 30-SLF cases.

Table 1. Model results for the Minimum low flow series case

	EV1 Location parameter				**EV1 Scale parameter**			
Scenario	**Calibration**		**Validation**		**Calibration**		**Validation**	
	Best Model	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**
Sc1	M3	0.90	M4	0.53	M3	0.86	M4	0.62
Sc2	M3	0.93	M1	0.52	M3	0.88	M3	0.58
Sc3	M3	0.96	M2	0.60	M3	0.94	M4	0.58
Sc4	M3	0.95	No Model		M3	0.94	M4	0.32
Sc5	M3	0.98	M1	0.49	M3	0.97	M1	0.55

Table 2. Model results for the 3-day sustained low flow series case

	EV1 Location parameter				**EV1 Scale parameter**			
Scenario	**Calibration**		**Validation**		**Calibration**		**Validation**	
	Best Model	**R^2**	**Best model**	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**
Sc1	M3	0.91	M4	0.52	M3	0.87	M4	0.63
Sc2	M3	0.94	M4	0.53	M3	0.88	M3	0.64
Sc3	M3	0.97	M2	0.65	M3	0.94	M1	0.56

Sc4	M3	0.96	No Model		M3	0.95	No Model	
Sc5	M3	0.98	M1	0.49	M3	0.97	M1	0.57

Table 3. Model results for the 7-day sustained low flow series case

Scenario	**EV1 Location parameter**				**EV1 Scale parameter**			
	Calibration		**Validation**		**Calibration**		**Validation**	
	Best Model	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**
Sc1	M3	0.93	M4	0.49	M3	0.88	M1 and M4	0.58
Sc2	M3	0.95	M1 and M4	0.50	M3	0.89	M3	0.69
Sc3	M3	0.98	M2	0.59	M3	0.94	M4	0.62
Sc4	M3	0.97	M2	0.34	M3	0.95	No Model	
Sc5	M3	0.98	M1	0.47	M3	0.97	M4	0.58

Table 4. Model results for the 10-day sustained low flow series case

Scenario	**EV1 Location parameter**				**EV1 Scale parameter**			
	Calibration		**Validation**		**Calibration**		**Validation**	
	Best Model	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**
Sc1	M3	0.94	M4	0.48	M3	0.89	M4	0.61
Sc2	M3	0.96	M1	0.50	M3	0.89	M3	0.71
Sc3	M3	0.98	M2	0.55	M3	0.94	M4	0.65
Sc4	M3	0.98	M2	0.48	M3	0.95	No Model	
Sc5	M3	0.98	M2	0.51	M3	0.97	M4	0.63

Table 5. Model results for the 15-day sustained low flow series case

Scenario	**EV1 Location parameter**				**EV1 Scale parameter**			
	Calibration		**Validation**		**Calibration**		**Validation**	
	Best Model	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**
Sc1	M3	0.94	M2	0.52	M3	0.90	M4	0.72
Sc2	M3	0.97	M2	0.51	M3	0.90	M3	0.79
Sc3	M3	0.98	M4	0.52	3	0.95	M4	0.72
Sc4	M3	0.98	M2	0.56	3	0.96	M2	0.18
Sc5	M3	0.99	M2	0.54	3	0.97	M4	0.73

Table 6. Model results for the 30-day sustained low flow series case

Scenario	**EV1 Location parameter**				**EV1 Scale parameter**			
	Calibration		**Validation**		**Calibration**		**Validation**	
	Best Model	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**	**Best Model**	**R^2**
Sc1	M3	0.93	M2	0.50	M3	0.87	M3	0.59
Sc2	M3	0.97	M4	0.54	M3	0.88	M4	0.53
Sc3	M3	0.99	M2	0.51	M3	0.97	M1	0.17
Sc4	M3	0.99	M2	0.46	M3	0.98	M2	0.56
Sc5	M3	0.99	M2	0.48	M3	0.96	M1	0.07

The first general comment that can be made about the performances of the two models that both models can reproduce reasonably values for the EV1 location and scale parameters in most cases and this is evident from the positive R^2 (>0) values. However, in a small number of cases none of the models was able to achieve acceptable values for R^2 as indicated by "No Model" in the Best Model column in each table. This phenomenon only occurred during the validation of the EV1 location and scale models for the Min, 3-SLF, 7-SLF, 10-SLF, 15-SLF with input scenario Sc4. This could indicate that the data in this scenario is inadequate for estimating the EV1 distribution parameters. The magnitudes of the values in the low series are normally low and far less than the mean annual flows. Such values can be found within the flat part of the recession curve in a hydrograph. The flow in this part of the hydrograph is primarily induced by the soil water storage and the groundwater storage and the effect of geology that underlain both of them is secondary. Consequently the geological variables may not be good predictors in the EV1 distribution parameters models of these low flow series.

It is noticeable that M3 was the best model in all cases during calibration. This of course was expected since M3 has the highest number of parameters and no constraints were imposed on the parameter values due to the use of the unconstrained least squares method of calibration. These two characteristics gave the model sufficient flexibility to fit the data which used in calibration. However, in some cases, this in turn produced a model which overfitted the data for scenario Sc4 and as a result it exhibited a very poor performance during validation.

To compare between the best models for the five input scenarios in each sustained low flow case the validation results are paramount. These show the ability of that model to reproduce measured data set different from the one used in calibration.

For the EV1 location parameter the best validation results were obtained from M2, the simple linear model calibrated by the constrained least squares method, with Sc3 for Min, 3-SLF, 7-SLF, and 10-SLF cases. This indicates that the combination of the catchment area, mean annual rainfall, mean elevation, mean slope, and soil type variables was the best input scenario for the models associated with these low flow series. In the case of 15-SLF M2 was the best model as well but with Sc4 and here the geological variables were better predictors than the soil variables. Anyhow all these variables affect the hydrology of low flow directly and hence they are plausible as explanatory variables for the location parameter of the EV1 distribution. However, it is surprising to find the results showing the simple linear model was better than the fuzzy model in the above cases. This may be attributed to the fact that all stations were in the same River Basin and hence their catchment characteristics do not span a wide range of variations. Thus the allowance provided by the fuzzy clustering model for considering such variations in catchment characteristics may not be required and consequently the resulting model was over-parameterised, degrading the results. In contrast some flexibility was likely required for modelling the location parameter of the EV1 distribution of the 30-SLF series and because of that the best model was found to be M4 with Sc2.

The EV1 scale parameter model of the six sustained low flow series was better estimated by the fuzzy clustering model calibrated by the constrained least squares method in the case of Min low flow series and by the unconstrained least squares method in the other sustained low flow series. In addition Sc1 was the best in the case of Min and 30-SLF sustained low flow series while Sc2 was the best for the other sustained low flow series. This suggests that the scale parameter requires a more complex model as provided by the fuzzy clustering model. The main reason for this could be the fact that the scale parameter defines the distribution of the EV1 curve around the mean value or the location parameter and hence it is a function of the standard deviation of the data which is difficult to be represented by a simple model. The required complexity for the scale model has to be in the form of the model not in the input or the explanatory variables of that model since it was only sufficient to use basic explanatory variables such as Sc2 in order to get the best model.

From the above findings it is clear that M2 and M3 was, in general, the best models for the EV1 location and scale parameters respectively and this means that the method of constrained least squares is suitable for the simple linear model while the method of unconstrained least squares method is suitable for the fuzzy clustering model. The first method imposes some restriction on the parameter spaces to be in the positive domain only and this in turn reduces the degree of freedom of the model. Using the constrained method was necessary to act as a tool preventing the simulation of any unrealistic negative values in the resulting location model. On the other hand the unconstrained least squares method allows of using any values in the parameter spaces in order to achieve the optimum results for the objective function which was used in the calibration. With the fuzzy clustering model it seems there was adequate flexibility acting as a natural barrier which prevents the unconstrained least squares from estimating unrealistic parameters.

Note there is a consistency in the trend of the results for the six sustained low flow series. Most likely this is because the six series in each station have been extracted from the same daily flow data and this means they are subsets of one population and hence that is reflected in the results

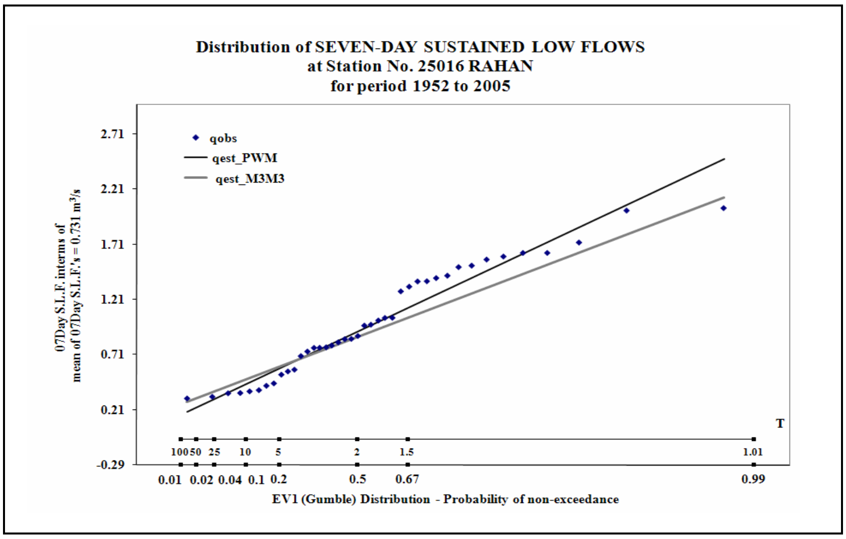

Figure 4: Comparison between EV1 curve obtained from PWM and EV1 curve obtained from the EV1 location and scale parameters models.

The values obtained from M2 for the EV1 Location parameter and from M3 for the EV1 scale parameter were used as an example to demonstrate the results of the newly developed model by plotting the LFFC for the 7-SLF using the EV1 distribution at station 25016 (RAHAN) in the SRB. Figure 4 below shows this curve (qest_M2M3) along with the LFFC obtained from the PWM (qest_PWM). In addition the observed values in the 7-SLF (qobs) are also shown. The qest_PWM is the best line that fits qobs and its parameters have been estimated by the PWM. The other curve qest_M2M3 should be compared with qest_PWM to assess the EV1 location model (M2) and the EV1 location model (M3). The differences at the lower tail of the curve are minimal compared to that at the upper tail. The design values of low flows are normally taken from the lower tail of the curve and corresponding to high return period (e.g. 100 years). It is extremely promising that the newly developed model was able to simulate this part of the curve reasonably well.

7 Conclusions

A new regionalisation approach has been proposed in this study whereby the location and the scale parameters of the EV1 distribution of different sustained low flow series can be estimated from six physical variables including area, rainfall, mean elevation, mean slope, soil, aquifer types, and geology. The resulting LFFC in this case provides a tool for predicting design values of low flows for conditions different than the one is already exist and this makes the tool suitable for use to asses the impact of the expected future climate and land use changes. Two different models have been developed using the proposed approach including (1) a simple linear model, and (2) a fuzzy clustering model. Both models have been tested with data from 55 stations in the SRB in Ireland. The model testing procedure involved an assessment of the two models performances under five different input scenarios with the unconstrained and the constrained least squares methods of calibration. The five input scenarios included (Sc1) area and rainfall; (Sc2) area, rainfall, mean elevation, and mean slope; (Sc3) area, rainfall, mean elevation, mean slope, and four separate soil variables representing the excessive, imperfect, moderate, and poor drainage conditions; (Sc4) area, rainfall, mean elevation, mean slope, and four separate geology variables representing the percentage area in the catchment of limestone, metamorphic, old red sandstone, and sedimentary rocks; and (Sc5) area, rainfall, mean elevation, mean slope, and four separate aquifer variables representing gravel, karstic, poorly productive bedrock, and productive fissured bedrock types. The results indicated that the approach is successful and in general for the LFFC of the sustained low flow series:

(1) The simple linear model with Sc3 and with the unconstrained least squares method was the best model for the EV1 location parameter.
(2) The fuzzy clustering model with Sc2 and with the constrained least squares method was the best for the EV1 scale parameter.

Acknowledgments
The work in this paper was done as part of a STRIVE project funded by the Environmental Protection Agency in Ireland. Many thanks are due Michéal MacCartaigh of the EPA for the hydrometric data used here.

References

[1] UNDP, “Capacity Development: Empowering People and Institutions”, United Nations Development Programme, Annual Report, 2008.
[2] U.S. Army Corps of Engineers, “Hydrologic Frequency Analysis”, Engineer Manual, 1110-2-1415, 1993.
[3] V.Y. Smakhtin. “Low flow Hydrology – a review”, Journal of Hydrology, 240, 147-186, 2001.
[4] A. Gustard, A. Bullock, and J.M. Dixon, “Low flow estimation in the United Kingdom”, Report No. 108, Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OXIO 8BB, United Kingdom, 1992
[5] J.A. Hartigan, and M.A. Wong, “Algorithm AS 139: A k-means clustering algorithm”, Applied Statistics, 28, 100-108, 1979.
[6] M.D. Zaidman, V.K. Ingenieur, A.R. Young, and A.Wall, “Adapting low-flow frequency analysis for use with short-period records”, Water and Environment Journal, 17(2), 74-79, 2003.
[7] C. Cunnane, “Statistical distribution for flood frequency analysis”, Report No. 718, WMO, Geneva, 1989.
[8] J.A. Greenwood, J.M. Landwehr, N.C. Matalas, and J.R. Wallis, “Probability weighted moments – Definition and relation to parameters of distributions expressible in inverse form”, Water Resources Research, 15(5), 1049-1054,1979.
[9] C. Cunnane, “Unbiassed plotting positions – a review”, Journal of Hydrology, 37(3-4), 205-22, 1978.
[10] I.I. Gringorten, “A plotting rule for extreme probability paper”, Journal of Geophysical Research, 68(3), 813-814, 1963.
[11] J.R.M. Hosking, J.R.Wallis, Regional frequency analysis”, Cambridge University Press, 1997.
[12] J.E. Nash, and J.V. Sutcliffe, “River flow forecasting through conceptual models. Part 1. A discussion of principles”, Journal of Hydrology, 10, 282-290, 1970.