Peer-reviewed version available at *Water* **2018**, *10*, 1038; doi:10.3390/w10081038

Article

Simple and Low-Cost Model for Monthly and Yearly Streamflow Forecasts during the Current Hydrological Year

Fernando Delgado-Ramos [1], Carmen Hervás-Gámez [2]

[1] Dr. Eng. at Institute of Water Research, University of Granada, Spain; Scientific Advisor of *Fundación Centro Andaluz de Investigaciones del Agua (CETAQUA)* fdelgado@ugr.es;
[2] MEng. at Department of Structural Mechanics and Hydraulic Engineering, University of Granada, Spain; karmina@correo.ugr.es

Abstract: Forecasting streamflow accurately is essential to achieve an efficient integrated water resources management strategy and provide consistent support to water decision-makers. We present a simple, low-cost and robust approach for forecasting monthly and yearly streamflows during the hydrological year in course, applicable to headwater catchments. It combines the use of regression analysis techniques, the two-parameter Gamma continuous cumulative probability distribution function and the Monte Carlo method. It is based on a probabilistic comparison of the progression of the current hydrological year with the historic observed series. The methodology has been successfully applied to two headwater reservoirs within the Guadalquivir River Basin in southern Spain. The root-mean-square error and correlation coefficient were used to measure the accuracy of the model and the results showed good levels of reliability. The outputs are the probabilistic monthly and yearly streamflows and 80% confidence interval. Further reductions in prediction errors may be achieved from increasing the number of observed years. These risk-based predictions are of great value, especially, before the intensive irrigation campaign starts (usually in April), when Water Authorities are to take responsible management decisions about the best allocation of the available water volume between the different water users and environmental needs.

Keywords: Integrated water resources management; support to decision-making process, streamflow forecast; simple and low-cost forecasting model; Guadalquivir River Basin; Genil River; Canales reservoir; Quéntar reservoir.

1. Introduction

Nowadays, water authorities and decision-makers are facing considerable challenges to achieve a sustainable and integrated water resources management system, especially, in water-stressed areas. They are to take responsible management decisions about the optimum allocation of the available water volume from a wide range of possible sources (regulated or non-regulated rivers, groundwater resources, water re-use schemes, desalination plants, etc.) between, usually, multisectoral demanding water users (domestic and municipal, agriculture, industry, tourism, energy, etc.).

These decisions are to consider not only the satisfaction of the water demands but the protection of this natural and finite resource, the environmental needs, the environmental impact, the compliance with the increasing water quality legal requirements, the social equity, the costs, and the promotion of economic growth. Additionally, the effects of the climate change on the spatial distribution and temporal climate variability coupled with the increasing population are altering the traditional approach to water resources planning, management and decision processes.

To cope with this situation, a wide variety of conservation policies from supply augmentation (i.e. new infrastructure such as reservoirs, desalination plants, rainwater harvesting, grey and black

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

water reuse schemes, water transfers, groundwater recharge) or demand reduction (i.e. water use efficiency, water restrictions, pricing policies, governance) at the basin scale can be adopted [1].

Therefore, advanced hydrological information and the provision of streamflow forecasts accurately is one of the key aspects to provide consistent support to decision-makers. Short-term forecasting such as hourly or daily forecasting is crucial for flood warning and sediment control. Medium-term forecasting based on monthly, seasonal or annual time scales is fundamental to decide on critical aspects of the current hydrological year such as the reservoir outflows planning, scheduling irrigation releases, allocating water to downstream users, drought mitigation and managing river treaties or implementing compact compliance. [2]. Long term forecasting is key for planning new strategic water infrastructure, such as reservoirs, water transfers between catchments, etc. and to inform the preparation of the River Basin Hydrological Plans.

In Mediterranean countries (for example Spain), where the agriculture plays an important socio-economic role, the critical decision point is just before the irrigation campaigns commence, usually in March-April. While the typical annual streamflow hydrograph peaks occur generally during the months of January-February-March (assuming there are no others hydrological processes such as snow melting or subterranean inflows), the peak agricultural water demands occur during the most water-stressed months (June, July and August). Therefore, supply infrastructure such as reservoirs provide a reliable source of water storage during the winter months and are the essential infrastructure to deal efficiently with the spatial and temporal climate irregularity distinctive of the Mediterranean area.

At the seasonal level, a skillful streamflow forecast may allow more efficient water allocation and predictable trade-offs between flows for energy, irrigation, municipalities, environmental services, etc. Such forecasts often provide the ability to prepare for anticipated conditions and not simply react to existing conditions, potentially reducing climate-related risks and offering opportunities ([3], [4]).

However, the level of accuracy achieved by the seasonal climate forecasts provided by the Spanish Meteorological Agency, AEMET (*Agencia Estatal de Metereología*) [5] for this year 2017-2018 has not been as accurate as expected. For example, AEMET predicted in February 2018 for the south-eastern quarter of Spain, that March-April-May would be much drier than average, however, March 2018 has been the wettest month registered in historic records.

The importance of achieving an adequate level of accuracy when predicting streamflows has been highlighted by many authors ([6], [7], [8], [9]). There is a wide variety of methods that have been used to build streamflow prediction models. Streamflow forecasting models fall into two general categories: process-driven and data-driven ([2], [9], [10]). Shalamu [2] and Yu X [10] provided a very detailed description and review of the different models, limitations and applications. Conceptual hydrologic models replicating observed historic data sets have been traditionally used for predicting the future. However, some of their shortcomings might come from the quality, accuracy and completeness of the input data, complexity due to the number of variables required and the difficulty to capture some hydrological phenomena such as the snow storage and melting processes, as well as the calibration and parameter optimization processes. Gragne [11] proposed to improve those difficulties by complementing conceptual models with simple error models and their results presented more accurate inflow forecasts into hydropower reservoirs several hours ahead. Data-driven models use the support vector machine, genetic programming and seasonal autoregressive techniques, such as for example the work done by Wang ([12]) and applied to the Three Gorges Reservoir.

This research work contributes towards the development of a novel, simple, low cost and robust methodology to forecast streamflows within the hydrological year in course. This is applicable to headwater systems and can provide support to strategic water management decisions. The methodology has been successfully applied to two headwater reservoirs located at the upper area of the Genil River (Guadalquivir River Basin) namely, Quéntar and Canales.

The model was first put into operation in October 2016 and the forecasts achieved satisfactory results with a relative error varying from 5% (Canales) up to 20% (Quéntar). These risk-based forecasts are useful not only to water reservoir operators and authorities but all the Stakeholders involved in the planning, management and decision-making processes. The model has also the option to incorporate seasonal climate predictions and climate change effects.

This paper is organised as follows: Section 2 states the aim and model performance objectives, Section 3 describes the methodology, Section 4 presents the results for two cases of study including the description, findings, model performance tests using reliability metrics and discussion, Section 5 presents the discussion on the predictive results obtained and Section 6 provides the concluding remarks.

2. Aim and objectives

The aim of this work was to develop a simple and low cost statistical model (data-driven model) to forecast monthly and annual streamflows during the current hydrologic year. The model was created to achieve the following objectives:

- To minimise the cost, the model is to use free hydrological data sources available in the public domain. Where possible, data should be downloaded in an easy, free and quick manner from a reliable official online resource;
- To simplify the model structure, the number of hydrological variables is to be minimised (i.e. the selection of important hydrological variables as the predictors is key);
- The running time of the model to obtain the results is to be minimum (instantaneous if possible);
- To ensure future performance, operability and flexibility of the model, new observed data from gauging stations is to be easily and regularly integrated and updated in the model (where possible, this should be an automatic process);
- The results obtained from the model should be sufficiently consistent and robust to support strategic and management decisions associated with annual or quarterly water cycles. The model results are to delimit the solution within the 80% confidence interval;
- The model is to be flexible and allows the integration of seasonal climate predictions, climate change effects or any other variation that might be needed in the future to improve the model;
- The model presents an intuitive, clear and easy-to-use interface with the final user (who does not necessarily need to be a programmer or specialist on this field of research);
- The model allows the variation or integration with other type of hydrological models (conceptual or distributed models) in the future if needed.

2. Materials and Methods

2.1. Introduction

The proposed methodology is to predict the monthly and annual streamflows during the ongoing hydrological year. It combines the use of: i) regression analysis techniques between relevant hydrological variables, ii) the two-parameter Gamma continuous and cumulative probability distribution function, and iii) the Monte Carlo simulation. It is based on a probabilistic comparison of the progression of the current hydrological year (in terms of the cumulative monthly rainfall or cumulative monthly streamflow) with the historical data series.

Figure 1 below shows the flowchart of this methodology (described in detail in the following Sections and applied to two cases of study detailed in Section 3):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at *Water* **2018**, *10*, 1038; doi:10.3390/w10081038

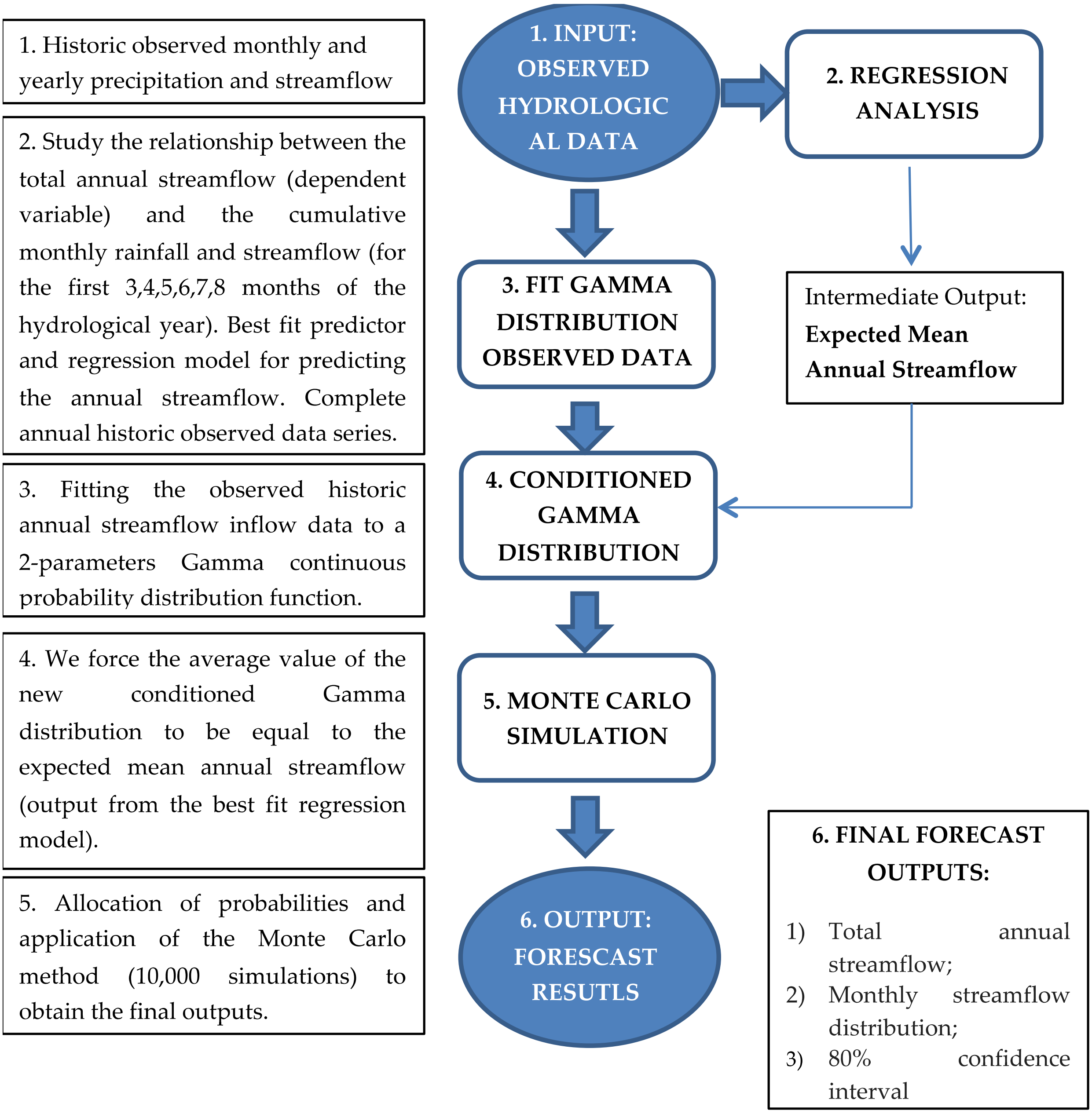


Figure 1. Flowchart of the methodology

3.2 Application and Limitations

The total annual streamflow was chosen as the prediction target (instead of the precipitation) to avoid the need of creating an additional model to transform the precipitation into streamflow.

However, it is important to note that if there are rapid and significant land use changes within the catchment of study, these might alter the observed hydrological cycle within the basin and might generate misleading streamflow forecasts. For example, a recent and rapid increase in impermeable area would generate more volume of surface water flows and quick streamflow peaks than the historic observed catchment response. In this situation, the precipitation as the prediction target (instead of the streamflow) might be more appropriate and an additional model to transform the precipitation into streamflow contribution might be required.

The proposed methodology does not take into account the contributing inflows from upstream systems (for example, regulated outflow releases from other reservoirs upstream). Therefore this methodology is applicable to headwater systems located at the upper catchment areas.

There is no limitation in catchment size however those catchments with considerable climatic irregularity or very rare seasonal sporadic rainfall events are difficult to predict and should be

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

carefully studied prior to progressing the modelling works (since this methodology is based on a statistical treatment of the known past to predict the future).

It is recommended to apply the methodology when there is good quality input data and at least, 30 years of observed data (although 40 years would be desirable to increase the forecasts accuracy). The minimum number of observed years should be greater in those catchments with greater standard deviation.

3.3 Model Time step

A monthly time step has been considered adequate for this work since the results are to support strategic and management decisions associated with annual or quarterly cycles. Therefore, the predictions from the model might not be applicable to other reservoir operations (for example, hydropower plants or flood forecasting) with lower time scale data requirements (hours, days).

It is also important to understand the phenomena that will be picked up within the selected time step (snow storage and melting processes, subterranean inflows, etc.) although the distribution or time of the year may vary.

3.4 Data sources and treatment

The aim was to use free hydrological data sources available in the public domain. The data should ideally be downloaded in an easy, free and fast manner from a reliable official online resource.

To carry out the analysis and methodology proposed in this paper, the data requirements are the monthly and yearly observed data sets of precipitation and streamflow. From this information, the cumulative precipitation and streamflow monthly time series are calculated as the sum of the monthly data from the start of the hydrological year (i.e. in Spain from October up to the month of study).

For example, the cumulative monthly rainfall up to the 1st April will be the result of adding the monthly precipitation of October, November, December, January, February and March (i.e. six months of cumulative rainfall from the 1st October).

Based on the above, the cumulative precipitation and streamflow monthly time series for the first 3, 4, 5, 6, 7 and 8 observed months are obtained (which will be used for the regression analysis). To achieve sufficiently good correlation coefficient values, we recommend starting the regression analysis using the cumulative values corresponding with the three first months of the hydrological year (i.e., October, November and December). Hence, the model is able to provide monthly and yearly streamflow forecasts from the 1st January.

3.5 Simple regression analysis and best fit predictor

Regression analysis is a well-known statistical technique used to study the relationship between a dependent variable (target or prediction) and a number of independent variables (or predictors). This technique has been traditionally used for forecasting and establishing relationships among different input variables.

For the proposed methodology, we use simple (linear and potential) regression analysis techniques. The analysis is based on the assumption that the hydrological cycle for a specific basin is in balance at the end of the hydrological year. Therefore, a greater correlation between the total annual streamflow (target or prediction) and the cumulative time sets is expected as the number of observed months increases.

We seek to predict the total annual streamflow (thereinafter referred to as, $\boldsymbol{A}_{annual}$) from relevant hydrological descriptors. The variables that we have selected as best predictors are the cumulative monthly precipitation (thereinafter referred to as, $\boldsymbol{P}_{cum}$) and the cumulative monthly streamflow (thereinafter referred to as, $\boldsymbol{A}_{cum}$).

We therefore used the regression analysis (linear and potential models) to investigate the relationship (type and strength) between the total annual streamflow (dependent variable or prediction) with, on the one hand, the cumulative monthly rainfall (independent variable or

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

predictor) and, on the other hand, the cumulative monthly streamflow (independent variable or predictor).

Subsequently, the results from the linear and potential models using both, the $\boldsymbol{P_{cum}}$ and $\boldsymbol{A_{cum}}$ to predict $\boldsymbol{A_{annual}}$ were compared using the R-squared values for each regression type and strength achieved. The results for the two cases of study are presented below.

It is important to note that the best fit regression model and best fit predictor does not necessarily need to be same for all the observed months of study. It might vary depending on the predominant hydrological process at each specific time or season of the year (snow storage and melting processes, subterranean inflows, seasonal extreme and sporadic rainfall events). In this situation, a combined regression model should be used to achieve the best correlation and predictive results.

In a nutshell, from the previous regression analysis, we identify the best fit regression model (linear or potential model) and best fit hydrological predictor (cumulative rainfall or cumulative streamflow) for each month of study. The expected total annual streamflow (based on the best fit regression model and predictor for each month) will be used an intermediary output as shown in Figure 1 above.

3.6 Two-parameter Gamma cumulative probability distribution function

The next step is fitting the historic observed annual streamflow data series to a two-parameters Gamma (α, β) cumulative probability distribution function.

There is an extensive literature describing the properties, parameters estimation and applications of the two-parameters Gamma (α, β) distribution function [13]. This distribution has shown to fit well to rainfall and streamflow data sets and has been widely applied to hydrological data-driven models, such as described in the studies of Buishand [14], Stephenson [15], Wang and Nathan [16], Chen [17], Chowdhury [18].

The Gamma distribution is a continuous probability distribution with two parameters α and β, known as the shape and scale parameters. This distribution is used to model exponentially distributed random variables. A random variable, X, is said to have the two parameter Gamma continuous probability density function if its distribution is given by:

$$f_X(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, \qquad x\geq 0,\ \alpha>0 \text{ and } \beta>0 \tag{1}$$

$$0, \qquad x<0$$

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \tag{2}$$

Where:

- x: total annual streamflow
- Γ: Gamma function.
- α: shape parameter
- β: scale parameter

The cumulative Gamma distribution is given by:

$$F_X(x) = \frac{1}{\Gamma(\alpha)} \int_0^x \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha}}, \ x\geq 0,\ \alpha>0 \text{ and } \beta>0 \tag{3}$$

For a random variable following a two-parameters Gamma distribution, the mean and variance are given the following functions:

$$E(x) = \alpha\beta \tag{4}$$

$$Var(x) = \alpha\beta^2 \quad (5)$$

Where:

- E(x) = Mean of the distribution function
- Var (x) = Variance of the distribution function

From the above equations and the observed annual streamflow series, we estimate the mean and variance of the observed series and the parameters α and β, as follows:

$$\beta = \text{Var (x) / Mean (x)} \quad (6)$$

$$\alpha = \text{Mean (x)} / \beta \quad (7)$$

This distribution provides the non-exceedance (or maximum) probability that the total annual streamflow will take a value less than or equal to a determined value as well as the maximum total annual streamflow value associated to a specific probability.

It is important to check how well the Gamma distribution fits with the observed data series of total annual streamflows for each case of study. As shown for the two cases of study, the Gamma distribution was found to provide the best fit to the total annual cumulative streamflow distribution.

3.7 Conditioned Two-parameter Gamma cumulative probability distribution function (using the output from the regression models)

The next step is to determine the influence of the estimated total annual streamflow (‘A_{annual}’, intermediary output from the previously obtained regression models) on the Gamma cumulative distribution function (fitted to the historic observed data) in order to obtain the Conditioned Gamma distribution function.

We impose the mean value of the new Conditioned Gamma distribution to be equal to the result obtained from the regression model (‘A_{annual}’ or expected total annual streamflow). The scale parameter β is kept the same as per the already obtained Gamma cumulative distribution function fitted to the observed historic total annual streamflow data sets. However, the shape parameter α will be re-calculated to obtain the Conditioned Gamma distribution applying Equation (8) above, this is;

$$\alpha \text{ (Conditioned Gamma)= Mean (Conditioned Gamma)} / \beta = \quad (8)$$

$$= \text{Expected total annual streamflow (from the regression analysis)} / \beta$$

We then estimate the Conditioned Gamma cumulative probability distribution function using (3) above and the two-parameters (α_{cond}, β).

It is important to highlight that while the Gamma cumulative distribution is derived from the observed and complete hydrological year series (from October to September, both inclusive) and it is a fixed curve, the Conditioned Gamma distribution uses the prediction (expected total annual streamflow) given by the regression models (based on the progression of the current ongoing hydrological year and different for each month of study). Therefore, the Conditioned Gamma distribution curve will vary month to month depending on the output from the regression models.

If the Conditioned Gamma distribution falls to the right of the Gamma distribution means that the prediction is a wetter hydrological year than the observed mean hydrological year, while this will be a drier year if the conditioned Gamma distribution falls to the left of the Gamma distribution. If the Conditioned Gamma falls close to the observed Gamma Distribution means that the ongoing hydrological year is similar to the mean observed hydrological year.

The Conditioned Gamma distribution function (α_{cond}, β) allow us to assign a probability to each year of the historical series based on their total annual streamflow. In doing so, we obtain greater probabilities for those years whose annual streamflow is similar to the estimated annual streamflow (‘A_{annual}’) and lower probabilities for those years whose annual streamflow differs from the estimated annual streamflow. For each year of the historical series we will therefore have its probability (p) given by:

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

$$P(x) = \frac{1}{\beta^{\alpha_{cond}}\,\Gamma(\alpha_{cond})} x^{\alpha_{cond}-1} e^{-\frac{x}{\beta}} \tag{9}$$

Where:

- x = the observed annual streamflow of the historic records (hm^3)
- α_{cond} = the shape parameter of the conditioned Gamma (-).
- β = the scale parameter of the Gamma distribution (the same as per the non-conditioned Gamma distribution) (-).

3.8. Monte Carlo method

The Monte Carlo Method is a numerical statistical method that allows the replication of random behavior of real non-dynamic systems through the generation of random numbers to which an event is assigned to based on their probability distribution.

The model generates 10,000 random input numbers in the interval [0,1]. Each of these random numbers is assigned to one year of the observed historical series of streamflows according to the cumulative probability obtained from the Conditioned Gamma distribution function (which best represent the intermediary output from the regression model).

This simulation will allow us to obtain 10,000 years and their monthly distribution according to their probability of success (which correspond to the observed years of the historical series).

The statistical analysis of the aforementioned series of streamflows will allow us to obtain the values of the expected monthly streamflows distribution (based on the average value of the 10,000 simulations) as well as the 80% confidence interval (10^{th} and 90^{th} percentiles of the 10,000 simulations, i.e., 80% of the values simulations will be found within these intervals).

3.9. Model Running Time, Test and Validation

The running time of the model is minimum, no longer than one minute. To verify that the model performed as expected and to measure its predictive accuracy and identify potential limitations, the root-mean-square error and correlation coefficient were used to measure the accuracy of the model and the results showed good levels of reliability. The results are shown and discussed in detail below for the case study.

4. Application to Canales and Quéntar Reservoirs (Upper Guadalquivir River Basin, Spain)

4.1 Study Area Description

The Guadalquivir River is the main river in southern Spain that serves water to a total population of over four million people and over eight hundred thousand hectares for irrigation. This system is formed by an interconnected system of currently operating 64 dams [19]. Although there are alternative water resources from aquifers, springs and water re-use schemes, nowadays reservoirs are the essential infrastructure to deal efficiently with the spatial and temporal climate irregularity distinctive of this catchment area.

The Guadalquivir River has a total contributing catchment area of 57,527 km^2 and delimited by Sierra Morena to the north, the Betic mountain to the south and the Atlantic Ocean. The altitude at the mountainous borders varies between 1,000mAOD and 3,480mAOD, which contrasts with the low altitude of the Guadalquivir river valley. The climate is Mediterranean defined by the warm temperatures (16.8ºC annual average) and by the irregularity of the precipitations (550 l / m^2 annual average). The rains frequently are torrential and occur after long periods of drought and high temperatures, with a marked susceptibility to erosion [19]. Figure 2 below shows the location of the Guadalquivir River Basin in Spain and the area of study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

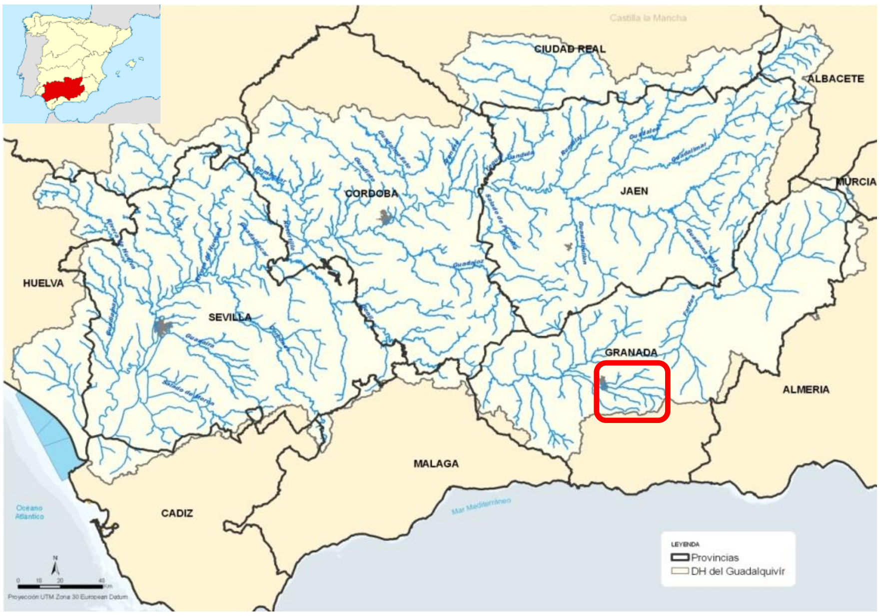


Figure 2: Location of the Guadalquivir River Basin [1] in Spain and area of study

The Guadalquivir River Basin Authority (RBA) is responsible for the elaboration of the Guadalquivir River Basin Hydrological Plan, as well as the administration and control of the hydraulic public domain [19]. The RBA has vast knowledge and experience managing water resources in this area, especially during drought and water scarcity periods when critical decisions and actions are to be taken. Last year, the RBA published the Draft version of the Guadalquivir River Basin Special Drought Management Plan (2017). This document establishes the general management principles and course of action for different drought and water scarcity threshold scenarios and each sub-catchment area within the basin [20].

The methodology has been applied to two headwater reservoirs of study, namely, Quéntar, and Canales. These are located at the upper area of the Genil River within Granada administrative area and to the south-east of the Guadalquivir River Basin (refer to Figure 3 below).

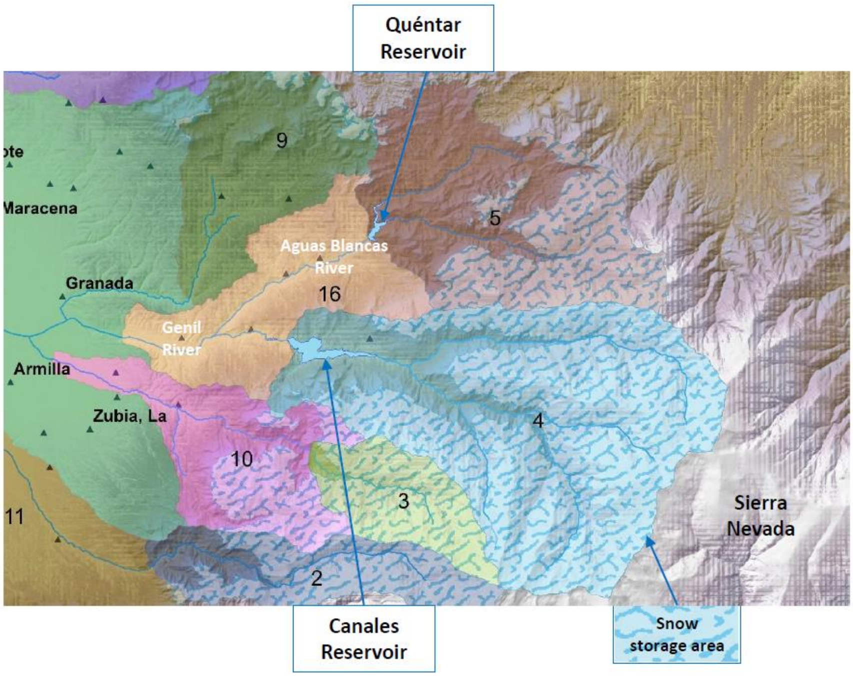


Figure 3. Location of Quéntar and Canales reservoirs and snow storage areas within the basin

Canales reservoir is located at the upstream area of the Genil River (close to Sierra Nevada), in the town of Güejar Sierra with a total contributing catchment area of 176.5km^2. The dam was built in 1989 with a total capacity of 70 hm^3 and the average streamflow is 80.42 hm^3 / year. This catchment area is affected by snow storage and melting processes in Sierra Nevada.

Quéntar reservoir is located on the Aguas Blancas River (tributary of the Genil River) with a total contributing catchment area of 101.2km^2. This dam was built in 1975 with a total volume capacity of 13.5 hm^3. The average streamflow is 28.84 hm^3 / year. This catchment area is affected by subterranean inflows due to the aquifer and lithology present.

These two reservoirs form the main infrastructure which serve water mainly for urban and irrigation purposes. The urban water users are formed by Granada city and fourteen towns of its metropolitan area with up to 300,000 inhabitants. The Vega Alta del Río Genil traditional irrigations cover over 4,000 hectares and are fed by an extensive irrigation channel system which diverts water from the Genil River (at the downstream area of the Canales and Quéntar reservoirs).

This system is supplemented by a network of currently operating thirteen underground water wells located in the upper area of the Vega de Granada aquifer (south-east of the city of Granada, on both banks of the Monachil River and the A-395 motorway). These were built by the Guadalquivir RBA after the serious social, economic and environmental consequences suffered during and after the 1992-1995 drought (in which the Canales and Quéntar reservoirs were exhausted).

The objective of these works was to supplement the existing surface water resources (provided by the Canales-Quéntar reservoirs system) with the extraction of groundwater to serve the urban water supply of Granada and metropolitan area. To guarantee its correct management, operation and maintenance, the Guadalquivir RBA delivered this infrastructure in 1995 to Emasagra (Local Water and Sewage Company) who, since then, is responsible for its management. The Guadalquivir RBA is responsible for the supervision, reservoir management and compliance of the water allocation.

In years with average precipitation and snow, the system water demands can be satisfied. However, in years with below-normal precipitation, typically, require a higher underground water volume extraction and/or reduction in per right allocation. During prolonged periods of drought and water scarcity issue, the urban water supply has theoretically priority over the agricultural demand.

Strategic decisions made by the Guadalquivir RBA on the controlled released outflows from the Quéntar-Canales reservoirs system are critical to ensure the most resource-efficient and sustainable allocation of the available water resources. These decisions are especially relevant before the intensive irrigation campaign starts (usually in April), when the Water Authorities are to take responsible management decisions about the best allocation of the available water volume between the different water users and environmental needs for the rest of the hydrological year. However, despite their vital importance, decisions made by the RBA are mainly based on current water storage in reservoirs and not on reservoir inflow predictive models.

4.2 Data sources and treatment

For the Guadalquivir River Basin, the information has been downloaded from a freely available official data-sharing online portal known as 'Automatic Hydrological Information System (SAIH)'. This is a public service provided and maintained by the Guadalquivir River Basin Authority [21].

Information such as, streamflow, outflow, rainfall, reservoir level, temperature, etc. is available for up to 57 reservoirs, 52 non-regulated rivers, 20 canals, 10 hydro power plants as well as rain, snow and temperatures gauging stations across the basin. The available temporal data sets vary depending on the specific sub-catchment area, but usually these are available from 1989. The time step available is: hour, day and month.

For this particular case of study, the monthly observed data series of precipitation and streamflow for Canales reservoir (observed data from October 1988 to present) and Quéntar reservoirs (observed data from October 1977 to present) were downloaded and used.

From this information, the cumulative precipitation and streamflow monthly time series for the first 3, 4, 5, 6, 7 and 8 observed months were calculated (which will be used for the regression analysis) for each reservoir of study.

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

4.3. Regression analysis, correlation and selection of the best fit estimator and regression model

Using the historic data sets freely available in the public domain (as previously described) and for each reservoir of study (i.e., Canales and Quéntar), a complete statistical study of the correlation between the total annual streamflow (dependent variable or prediction target) with the cumulative monthly rainfall (independent variable or predictor) was carried out and the cumulative monthly streamflow. In all cases, as a minimum, we started using the cumulative data of the three first months of the hydrological year (i.e., October, November and December).

For simplicity and avoid repetition, the application of the methodology will be detailed for Quéntar reservoir. For Canales, only the results will be presented.

a) Quéntar reservoir

Figure 4 below shows the linear and potential regression models obtained for the first 3,4,5,6,7 and 8 observed months using the cumulative monthly reservoir inflow (as the predictor) for Quéntar reservoir. The same process was also carried out using the cumulative precipitation.

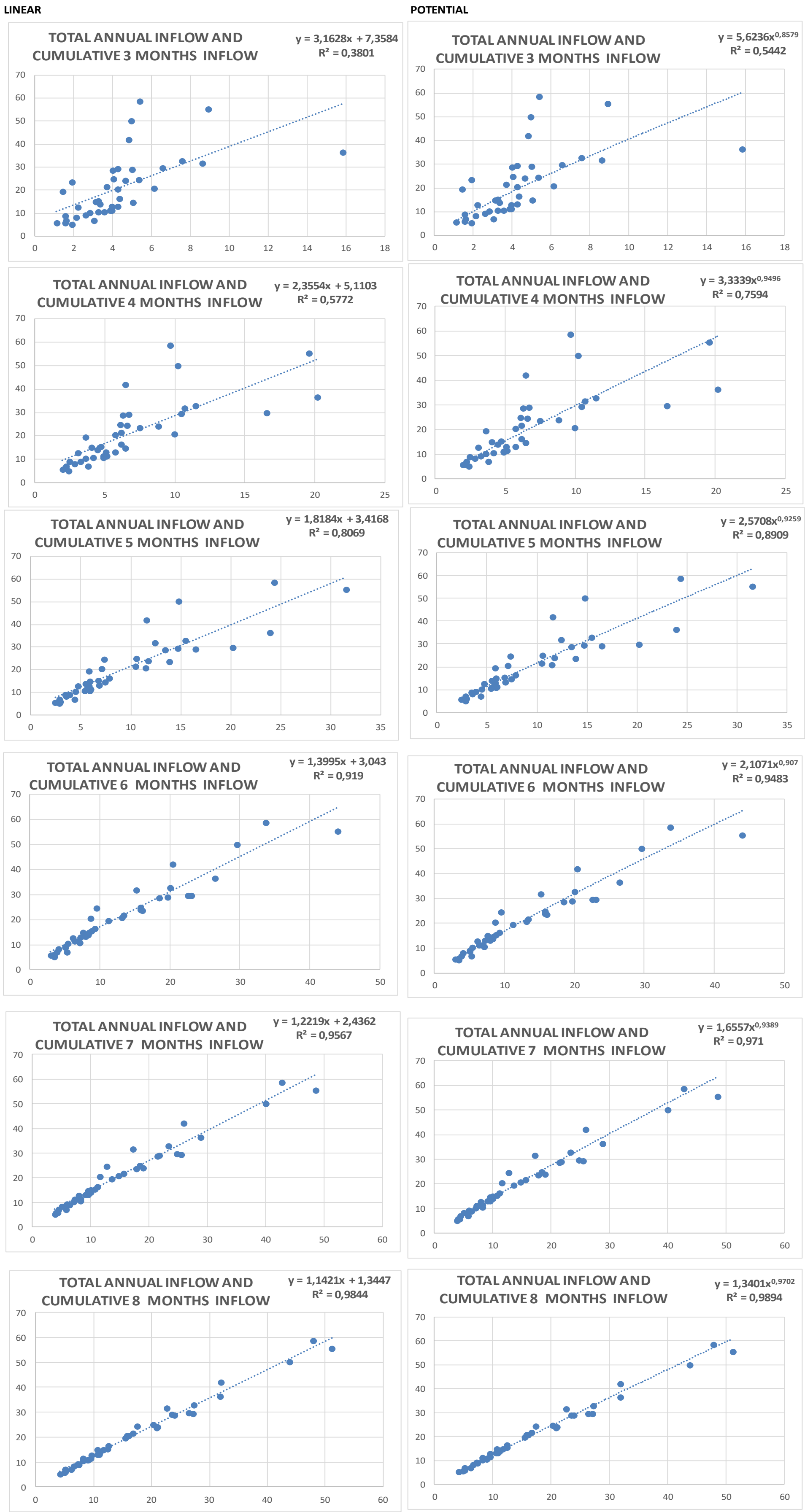


Figure 4*:* Quéntar reservoir: Linear and Potential regression models for the first 3, 4 ,5, 6, 7and 8 months using the monthly cumulative reservoir inflow values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

Subsequently, the results for the linear and potential models using both, the P_{cum} and A_{cum} were compared using the R-squared values for each regression type and strength achieved. Table 1 below shows the results obtained for Quéntar reservoir.

Table 1. Quéntar reservoir: R-squared values for each regression type and strength achieved.

Observed Months	LINEAR R-squared		POTENTIAL R-squared	
	P cum	A cum	P cum	A cum
3	0.3370	0.3801	0.4100	0.5442
4	0.4675	0.5772	0.5653	0.7594
5	0.6428	0.8069	0.6717	0.8909
6	0.7187	0.9190	0.7341	0.9483
7	0.6767	0.9567	0.6724	0.9710
8	0.6479	0.9844	0.6539	0.9894

It can be appreciated from Figure 4 and Table 1 above how the correlation value (R^2) increases with the number of observed months and the potential regression model presents a stronger correlation value.

Based on the above, the potential regression model (using the cumulative reservoir inflow as the best fit predictor) is the best fit model for Quéntar reservoir. Therefore, this model will have the following equation (a, b from the potential regression analysis as shown in Figure 4 above):

$$A_{annual} = a\, A_{cum.}{}^{b} \tag{10}$$

where:

- A_{annual} = Annual reservoir inflow (Hm^3)
- $A_{cum.}$ = Cumulative monthly reservoir inflow since beginning of the hydrological year (Hm^3)
- a = Potential regression model coefficient (-)
- b = Potential regression model coefficient (-)

Figures 5 below show the correlation found for the first 6 months of the hydrological year (i.e. 1st March) with observed data for Quéntar reservoir using the potential regression models.

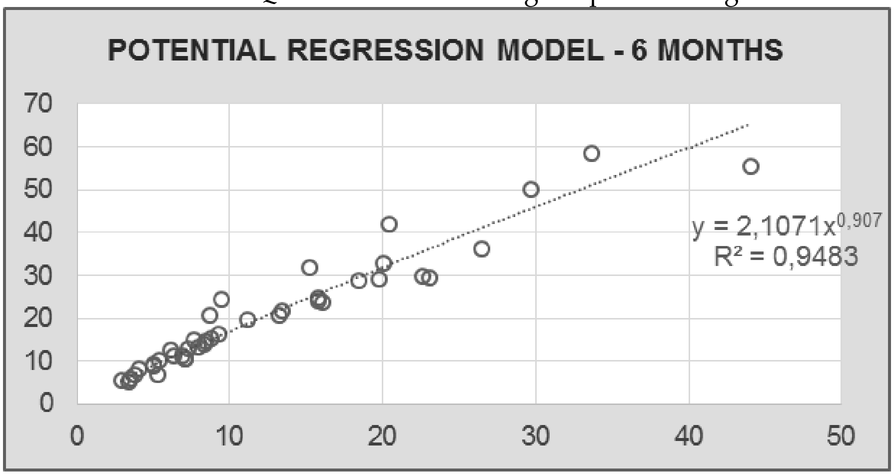


Figure 5. Quéntar reservoir: Potential regression model (observed data Oct-1977 Sept-2017). Total Annual Streamflow (hm^3, axis-y) and first 6 months cumulative monthly streamflow of the hydrological year (hm^3, axis-x)

For the Quéntar reservoir the underground aquifer inflows present in this basin become the most important hydrological process (while the snow storage and melting processes are not relevant in comparison with Canales reservoir), which implies a significant role of the base flow in the total

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at *Water* **2018**, *10*, 1038; doi:10.3390/w10081038

contribution to the reservoir. This is thought to be the reason why the strongest correlation is found between the cumulative streamflow and the annual streamflow.

B) Canales reservoir

A similar process was carried out for Canales reservoir and it was found a stronger relationship between the cumulative rainfall and the annual streamflow than using the cumulative streamflow.

This fact is considered to be due to the influence of snow storage and melting processes distinctive of this basin. Usually the hydrograph of this catchment area has two peaks within the hydrological year, the first is given in January-February due to the rainfall, while the second one is given around April-May which is due to the snow melting processes.

It was found that the potential regression model (using the cumulative precipitation as the best fit predictor) is the best fit model for Canales reservoir. Therefore, this model will have the following equation:

$$A_{annual} = a\, P_{cum.}{}^{b} \tag{11}$$

Where:

- A_{annual} = Annual streamflow (hm³)
- $A_{cum.}$ = Cumulative monthly streamflow since October of the hydrological year (hm³)
- $P_{cum.}$ = Cumulative monthly precipitation since October of the hydrological year (mm)
- a = Potential regression model coefficient (-)
- b = Potential regression model coefficient (-)

The correlation analysis was carried out for the first 3, 4, 5, 6, 7 and 8 months of the hydrological year with observed values for each reservoir. Logically and as expected, a greater correlation coefficient was found as the number of observed months increases.

Figures 6 below show the correlation found for the first 6 months of the hydrological year (i.e. 1st March) with observed data for Canales reservoir using the potential regression models.

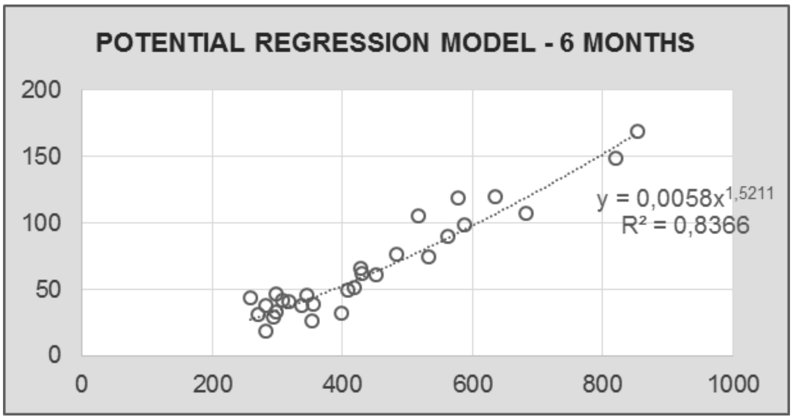


Figure 6. Canales reservoir: Potential regression model (observed data Oct-1988 Sept-2017). Total Annual Streamflow (hm3, axis-y) and first 6 months cumulative monthly precipitation of the hydrological year (mm, axis-x)

The expected annual streamflow ('A_{annual}') can therefore be estimated from the previously obtained potential regression models which are based on complete hydrological years (from October to September, both inclusive) of observed precipitation and streamflow data. This is, if we are to make a prediction during the ongoing hydrological year, the regression models should contain all the historic records up to September last year. It is important to note that this first estimate of the expected annual streamflow is not the final output model but will feed the next phase.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

4.4. Two-parameter Gamma cumulative probability distribution function (observed data) and Conditioned Gamma cumulative probability distribution function (using the output from the regression models)

Figures 7 and 8 below show the 2-parameter Gamma distribution for each reservoir of study. It was found that the Gamma distribution fits very well with the observed series of annual streamflows for the two reservoirs of study. We obtained a correlation coefficient (R) of 0.99 and 0.98 for Quéntar and Canales, respectively, between the observed annual streamflow series and those given by the distribution function.

The conditioned Gamma distribution function will allow us to assign a probability to each year of the historical series based on their annual streamflow. In doing this, we obtain greater probabilities for those years whose annual streamflow is similar to the estimated annual streamflow ('A_{annual}') and lower probabilities for those years whose annual streamflow differs from the the estimated annual streamflow.

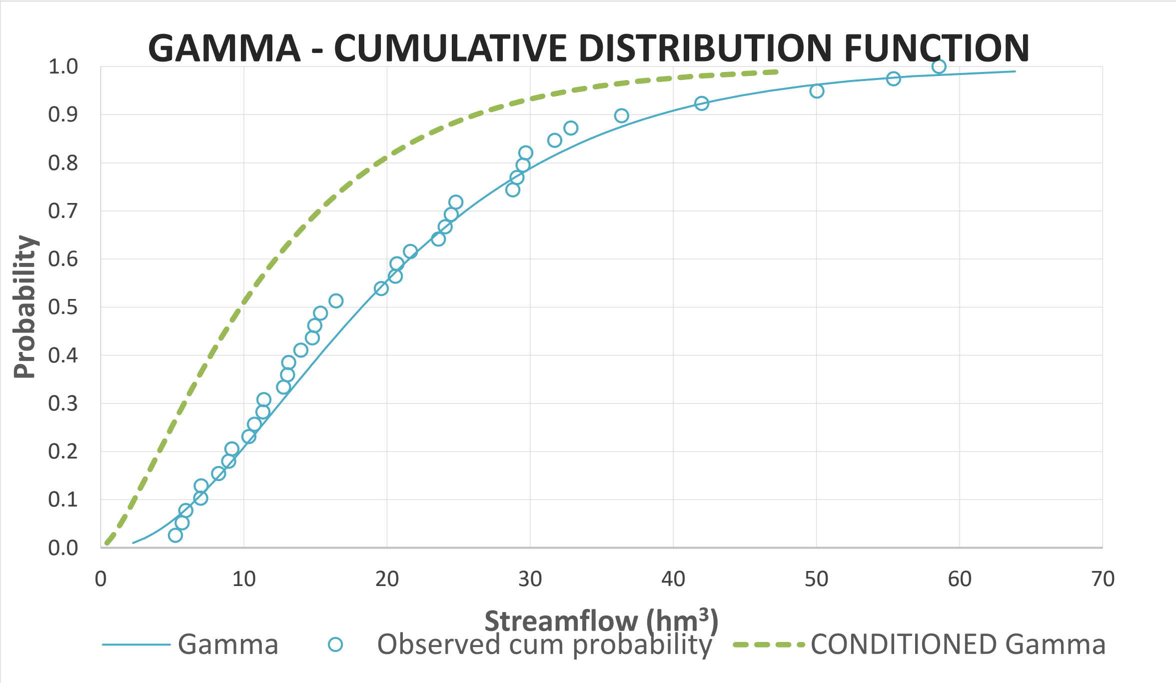


Figure 7: Quéntar reservoir: 2-parameter Gamma cumulative distribution function (observed data Oct-1977 Sept-2016). Conditioned Gamma distribution – 1st April 2017

As example, Table 2 below shows the allocation of probabilities for Quéntar reservoir used to generate Figure 7 above:

Table 2. Quéntar reservoir: assignation of probabilities

GAMMA CUMULATIVE DISTRIBUTION FUNCTION

Number of years	*40*	*E17:E56*
Mean	*20.75*	
Variance	*177.91*	

GAMMA CUMULATIVE DISTRIBUTION FUNCTION		*CONDITIONED GAMMA DISTRIBUTION*	
Alpha	2.42	Alpha	4.42

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at *Water* **2018**, *10*, 1038; doi:10.3390/w10081038

Beta	8.57	Beta	8.57

Initial Year	Final Year	Order number	Annual Res. Inflow	Observed cum probability	P (Gamma)	P (Gamma Conditioned)
1994	1995	1	5.2300	0.0250	0.0649	0.0015
2007	2008	2	5.7000	0.0500	0.0771	0.0021
2006	2007	3	5.9600	0.0750	0.0841	0.0025
2004	2005	4	7.0000	0.1000	0.1145	0.0046
1993	1994	5	7.0300	0.1250	0.1154	0.0047
1999	2000	6	8.2500	0.1500	0.1547	0.0085
2005	2006	7	8.9600	0.1750	0.1789	0.0115
1992	1993	8	9.1800	0.2000	0.1866	0.0125
1991	1992	9	10.3600	0.2250	0.2287	0.0191
2016	2017	10	10.6758	0.2500	0.2401	0.0212
1998	1999	11	10.7500	0.2750	0.2428	0.0217
1982	1983	12	11.3500	0.3000	0.2648	0.0262
2011	2012	13	11.4100	0.3250	0.2670	0.0266
2015	2016	14	12.7935	0.3500	0.3179	0.0389
2001	2002	15	13.0700	0.3750	0.3281	0.0417
1989	1990	16	13.1500	0.4000	0.3310	0.0426
1990	1991	17	14.0000	0.4250	0.3621	0.0520
1980	1981	18	14.8000	0.4500	0.3909	0.0619
1988	1989	19	14.9700	0.4750	0.3970	0.0641
2014	2015	20	15.3688	0.5000	0.4112	0.0695
1987	1988	21	16.4500	0.5250	0.4489	0.0852
2008	2009	22	19.6200	0.5500	0.5518	0.1402
2003	2004	23	20.6000	0.5750	0.5810	0.1597
1981	1982	24	20.7100	0.6000	0.5842	0.1619
1986	1987	25	21.6500	0.6250	0.6108	0.1816
1995	1996	26	23.6100	0.6500	0.6623	0.2249
2002	2003	27	24.0800	0.6750	0.6738	0.2357
1983	1984	28	24.5000	0.7000	0.6839	0.2454
2013	2014	29	24.8300	0.7250	0.6916	0.2532
1985	1986	30	28.8000	0.7500	0.7731	0.3491
1984	1985	31	29.0900	0.7750	0.7783	0.3563
2000	2001	32	29.5200	0.8000	0.7858	0.3668
1996	1997	33	29.7100	0.8250	0.7890	0.3715
1979	1980	34	31.7300	0.8500	0.8209	0.4209
2010	2011	35	32.8600	0.8750	0.8368	0.4482
1997	1998	36	36.4000	0.9000	0.8789	0.5308
1977	1978	37	42.0000	0.9250	0.9257	0.6478

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

Initial Year	Final Year	Order number	Annual Res. Inflow	Observed cum probability	P (Gamma)	P (Gamma Conditioned)
2012	2013	38	50.0500	0.9500	0.9642	0.7794
2009	2010	39	55.4000	0.9750	0.9783	0.8432
1978	1979	40	58.5700	1.0000	0.9840	0.8732

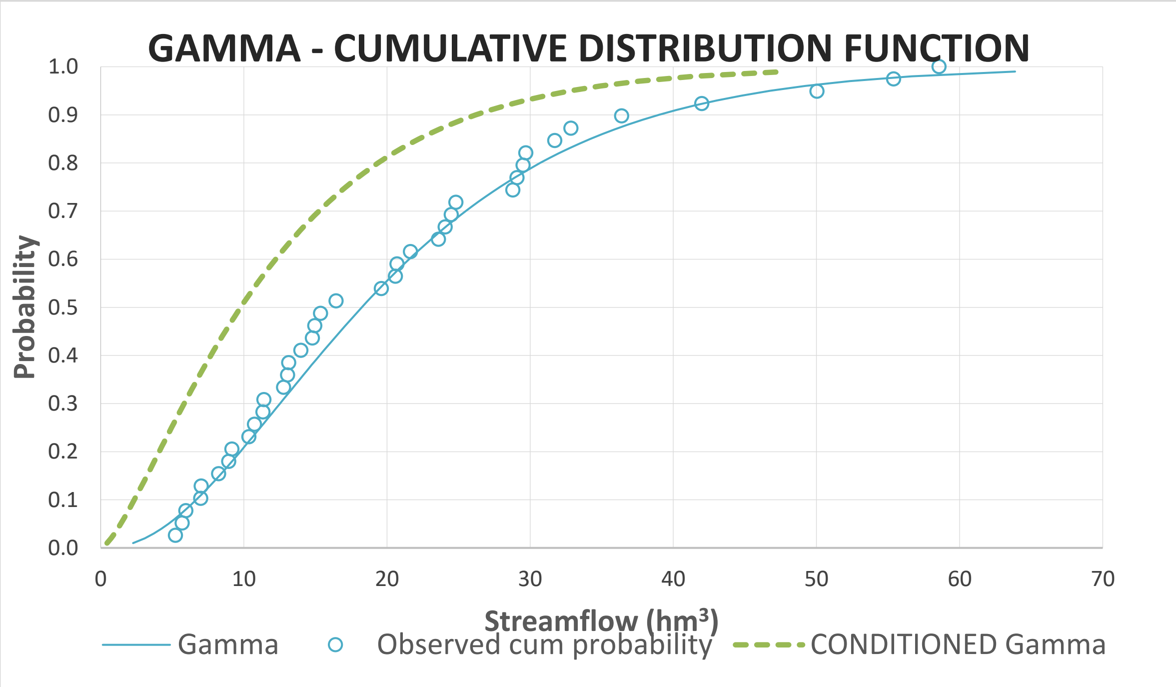


Figure 8: Canales reservoir: 2-parameter Gamma cumulative distribution function (observed data Oct 1988-Sept 2016). Conditioned Gamma distribution – 1st April 2017

It can be seen from Figure 7 and 8 that the prediction given by the model on 1st April 2017 for Canales and Quéntar was for a drier year than the mean historic hydrological year, which is what happened last year. This is explained in further detail in Section 4.7 below.

4.5. Assigning probabilities and Application of Monte carlo method

Table 3 below shows the first 20 Monte Carlo simulations for Quéntar reservoir as well as the results:

Table 3. Quéntar reservoir: assignation of probabilities (1st April 2017)

MONTE CARLO SIMULATION

Month	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Annual
Mean Year	0.92	0.89	1.17	1.44	1.75	2.00	1.52	1.38	0.79	0.57	0.63	0.68	13.75
10% Percentile	0.25	0.37	0.59	0.53	0.45	0.56	0.40	0.33	0.34	0.27	0.20	0.22	5.23
90% Percentile	1.45	1.54	1.61	2.47	4.58	5.36	2.89	3.01	1.28	1.06	1.27	1.55	28.80
Median	0.92	0.60	0.84	0.83	0.81	0.91	1.18	1.05	0.71	0.54	0.35	0.45	10.36

Simulation no.	Random no.	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Annual
1	0.0084	0.92	0.37	0.59	0.53	0.45	0.57	0.40	0.33	0.38	0.27	0.20	0.22	5.23

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

Simulation no.	Random no.	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Annual
2	0.1618	0.92	0.37	0.59	0.53	0.45	0.57	0.40	0.33	0.38	0.27	0.20	0.22	5.23
3	0.7141	1.45	1.26	1.61	1.82	1.73	1.43	1.77	1.38	1.15	0.95	0.88	1.02	16.45
4	0.6652	1.61	1.87	1.57	1.41	0.98	0.96	1.18	1.05	0.57	1.09	1.30	1.21	14.80
5	0.2876	0.27	0.34	0.50	0.84	0.50	0.47	1.04	0.90	0.34	0.19	0.15	0.16	5.70
6	0.7279	1.45	1.26	1.61	1.82	1.73	1.43	1.77	1.38	1.15	0.95	0.88	1.02	16.45
7	0.6409	1.02	1.13	1.17	1.14	1.00	2.84	1.66	1.09	0.85	0.65	0.62	0.83	14.00
8	0.3292	1.05	1.05	0.95	0.73	0.61	0.85	0.55	0.35	0.25	0.18	0.19	0.24	7.00
9	0.2741	0.27	0.34	0.50	0.84	0.50	0.47	1.04	0.90	0.34	0.19	0.15	0.16	5.70
10	0.2904	0.27	0.34	0.50	0.84	0.50	0.47	1.04	0.90	0.34	0.19	0.15	0.16	5.70
11	0.1247	0.92	0.37	0.59	0.53	0.45	0.57	0.40	0.33	0.38	0.27	0.20	0.22	5.23
12	0.5994	0.71	0.75	0.74	0.83	1.64	1.46	1.81	1.62	0.92	0.77	0.77	0.78	12.79
13	0.3368	1.05	1.05	0.95	0.73	0.61	0.85	0.55	0.35	0.25	0.18	0.19	0.24	7.00
14	0.2088	0.92	0.37	0.59	0.53	0.45	0.57	0.40	0.33	0.38	0.27	0.20	0.22	5.23
15	0.1045	0.92	0.37	0.59	0.53	0.45	0.57	0.40	0.33	0.38	0.27	0.20	0.22	5.23
16	0.4328	0.34	0.44	0.78	0.87	0.97	1.64	1.31	0.98	0.26	0.58	0.35	0.44	8.96
17	0.2316	0.92	0.37	0.59	0.53	0.45	0.57	0.40	0.33	0.38	0.27	0.20	0.22	5.23
18	0.6523	1.02	1.13	1.17	1.14	1.00	2.84	1.66	1.09	0.85	0.65	0.62	0.83	14.00
19	0.6638	1.61	1.87	1.57	1.41	0.98	0.96	1.18	1.05	0.57	1.09	1.30	1.21	14.80
20	0.8186	1.08	1.54	1.62	1.48	1.35	1.63	2.89	4.39	1.97	1.05	0.82	0.78	20.60

4.6. Model validation

To verify that the model performed as expected, measure its predictive accuracy and identify potential limitations, a set of different tests and metrics were applied, as described below.

For each year of the observed historic series, the model was used to predict the monthly, yearly and quarterly streamflows using the first 3,4,5,6,7 and 8 months of observed data of that particular hydrological year of study. The year that is being tested is extracted from the historic records, using the remaining years as the model feeding data. This process was carried out for every year of the observed historic series.

Subsequently, to measure the accuracy of the model (predictions and observed data) the root-mean-square error and correlation coefficient were used and the results are listed below in Table 4 and 5 for Canales and Quéntar, respectively:

Table 4. Canales reservoir: Performance test results. RMSE and correlation coefficient values for the monthly, yearly and quarterly series.

Forecast Month	Observed Months	Monthly Series		Yearly Series		Quarterly Series	
		RHO	RMSE	RHO	RMSE	RHO	RMSE
January	3	0.6516	3.3787	0.7846	24.7164	0.8137	8.0667
February	4	0.6927	3.1913	0.8214	23.2191	0.8420	7.4709
March	5	0.7339	2.9995	0.8589	21.3909	0.8682	6.8814
April	6	0.7535	2.9051	0.8920	19.4181	0.8902	6.3542
May	7	0.7655	2.8491	0.9112	18.7671	0.8994	6.1439
June	8	0.7735	2.8067	0.9177	18.2983	0.9043	5.9919

Table 5 Quéntar reservoir: Performance test results. RMSE and correlation coefficient values for the monthly, yearly and quarterly series.

Forecast Month	Observed Months	Monthly Series		Yearly Series		Quarterly Series	
		RHO	RMSE	RHO	RMSE	RHO	RMSE
January	3	0.4023	2.0166	0.4449	11.8933	0.4404	5.2403
February	4	0.4965	1.8322	0.5934	10.0198	0.5562	4.5808
March	5	0.6015	1.6332	0.7558	7.7868	0.7006	3.7373
April	6	0.7120	1.4046	0.9270	4.6128	0.8388	2.8072
May	7	0.7420	1.3414	0.9698	3.2196	0.8725	2.5197
June	8	0.7480	1.3347	0.9832	2.4568	0.8746	2.5003

It can be observed that, as expected, the correlation coefficient increases and the RMSE decreases as the number of observed months increases. It should also be noted that the model performs better forecasting annual and quarterly streamflows than in monthly streamflows, which highlights the difficulty of predicting the monthly distribution and the variability in the behavior of streamflows.

Given that the model has been developed to support management decisions associated with annual or quarterly cycles, we consider that the model is fit for purpose and the results obtained are satisfactory.

Figures 9 and 10 show the total annual streamflow observed and forecasts, as well as the 80% confidence interval.

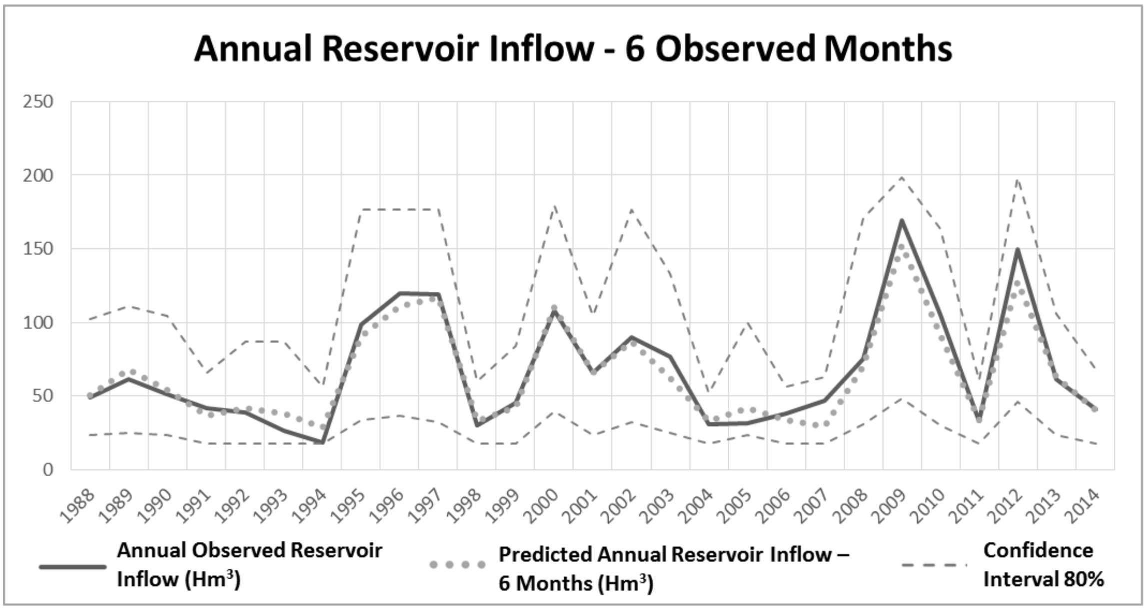


Figure 9. Canales reservoir: Observed annual streamflows and predictions using the model with the first six months of observed data of the hydrological year and 80% confidence interval (hm³)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at *Water* **2018**, *10*, 1038; doi:10.3390/w10081038

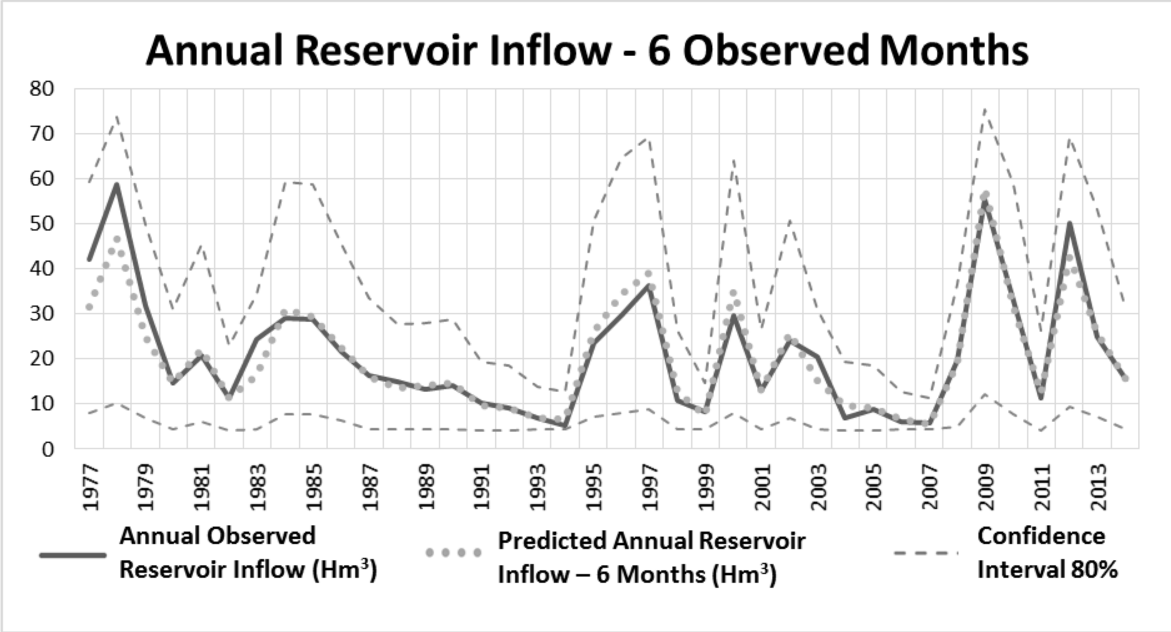
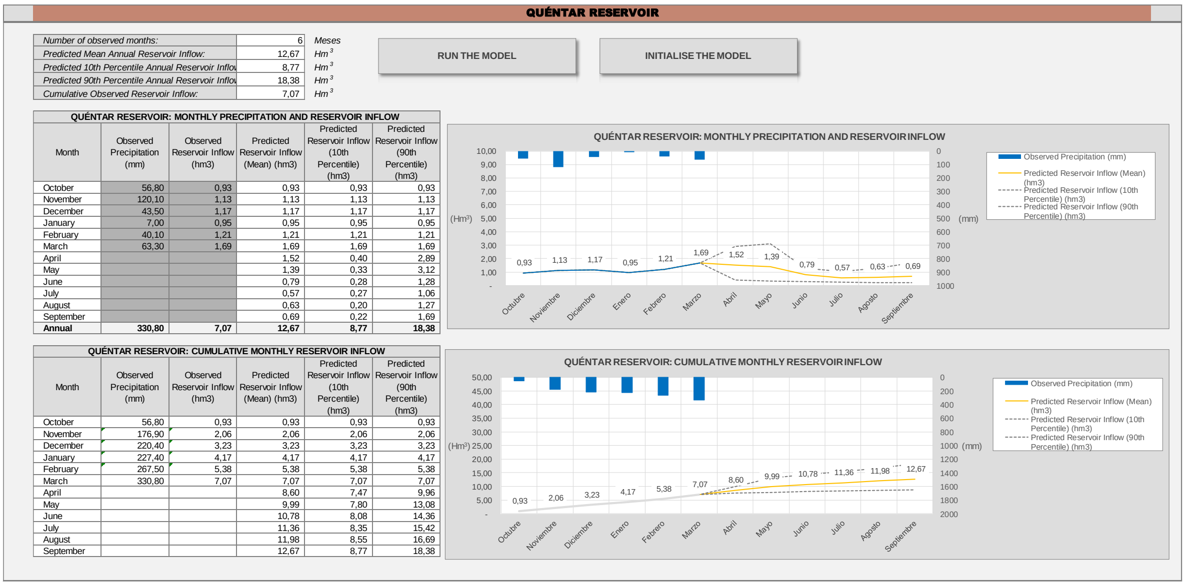


Figure 10. Quéntar reservoir: Observed annual streamflows and predictions using the model with the first six months of observed data of the hydrological year and 80% confidence interval (hm³)

4.7 Model Outputs

The typical model outputs are presented in Figure 11 and Figure 12 below, where the only inputs are the monthly rainfall and streamflow columns (coloured in dark grey):



Number of observed months:	6	*Meses*
Predicted Mean Annual Reservoir Inflow:	12,67	*Hm*³
Predicted 10th Percentile Annual Reservoir Inflo	8,77	*Hm*³
Predicted 90th Percentile Annual Reservoir Inflo	18,38	*Hm*³
Cumulative Observed Reservoir Inflow:	7,07	*Hm*³

QUÉNTAR RESERVOIR: MONTHLY PRECIPITATION AND RESERVOIR INFLOW

Month	Observed Precipitation (mm)	Observed Reservoir Inflow (hm3)	Predicted Reservoir Inflow (Mean) (hm3)	Predicted Reservoir Inflow (10th Percentile) (hm3)	Predicted Reservoir Inflow (90th Percentile) (hm3)
October	56,80	0,93	0,93	0,93	0,93
November	120,10	1,13	1,13	1,13	1,13
December	43,50	1,17	1,17	1,17	1,17
January	7,00	0,95	0,95	0,95	0,95
February	40,10	1,21	1,21	1,21	1,21
March	63,30	1,69	1,69	1,69	1,69
April			1,52	0,40	2,89
May			1,39	0,33	3,12
June			0,79	0,28	1,28
July			0,57	0,27	1,06
August			0,63	0,20	1,27
September			0,69	0,22	1,69
Annual	**330,80**	**7,07**	**12,67**	**8,77**	**18,38**

QUÉNTAR RESERVOIR: CUMULATIVE MONTHLY RESERVOIR INFLOW

Month	Observed Precipitation (mm)	Observed Reservoir Inflow (hm3)	Predicted Reservoir Inflow (Mean) (hm3)	Predicted Reservoir Inflow (10th Percentile) (hm3)	Predicted Reservoir Inflow (90th Percentile) (hm3)
October	56,80	0,93	0,93	0,93	0,93
November	176,90	2,06	2,06	2,06	2,06
December	220,40	3,23	3,23	3,23	3,23
January	227,40	4,17	4,17	4,17	4,17
February	267,50	5,38	5,38	5,38	5,38
March	330,80	7,07	7,07	7,07	7,07
April			8,60	7,47	9,96
May			9,99	7,80	13,08
June			10,78	8,08	14,36
July			11,36	8,35	15,42
August			11,98	8,55	16,69
September			12,67	8,77	18,38

Figure 11. Example of Typical Model Outputs Results

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

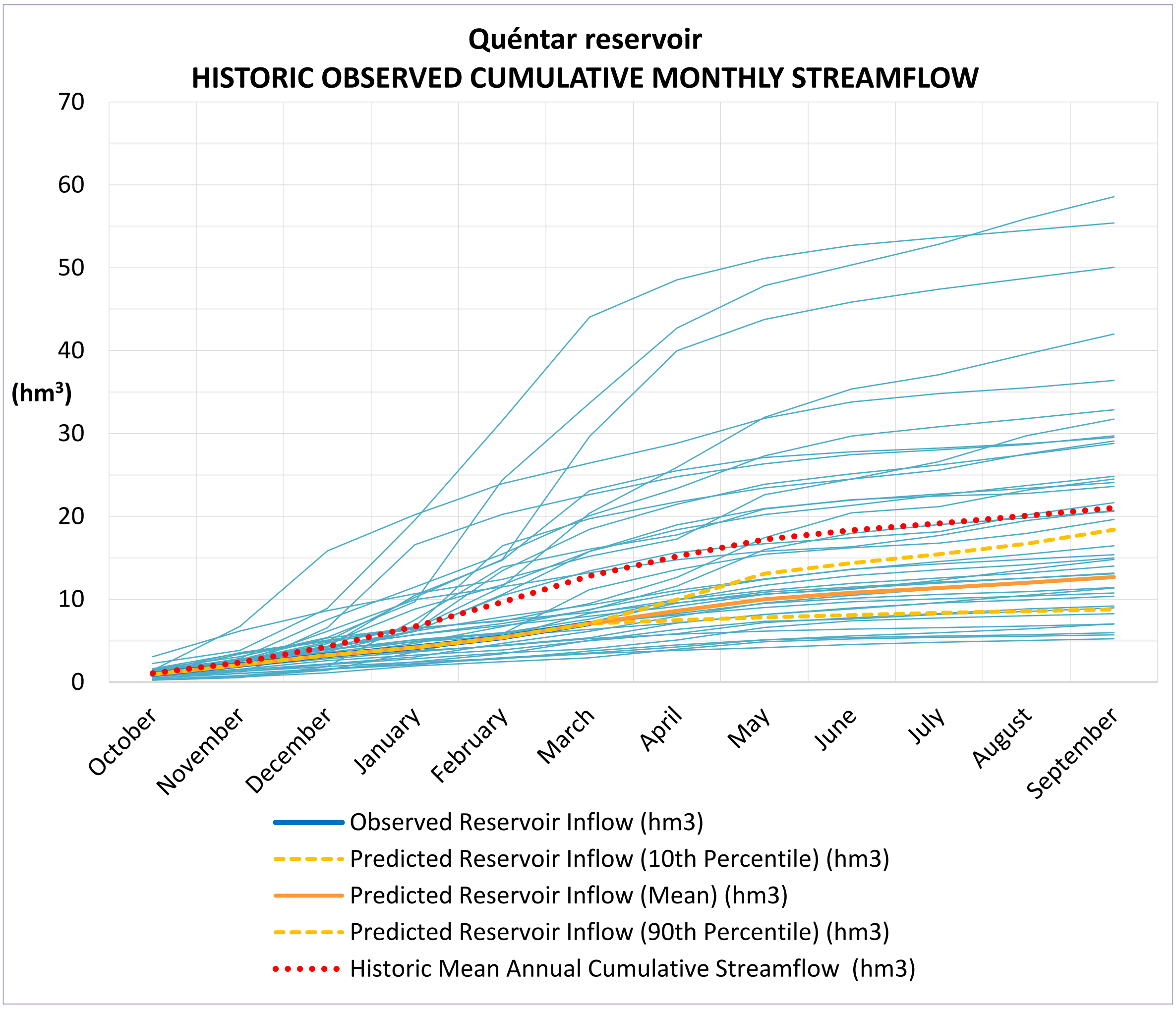


Figure 12. Example of Typical Model Outputs Results

4.7. First Operational Year: Results obtained hydrological year 2016-2017

The model was first put into operation in October 2016 and as an example, Figures 11 and 12 present a comparative analysis of the annual and monthly streamflow estimations made in each month of that hydrological year for Canales and Quéntar reservoirs and the real observed values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 June 2018 doi:10.20944/preprints201806.0164.v1

Peer-reviewed version available at *Water* **2018**, *10*, 1038; doi:10.3390/w10081038

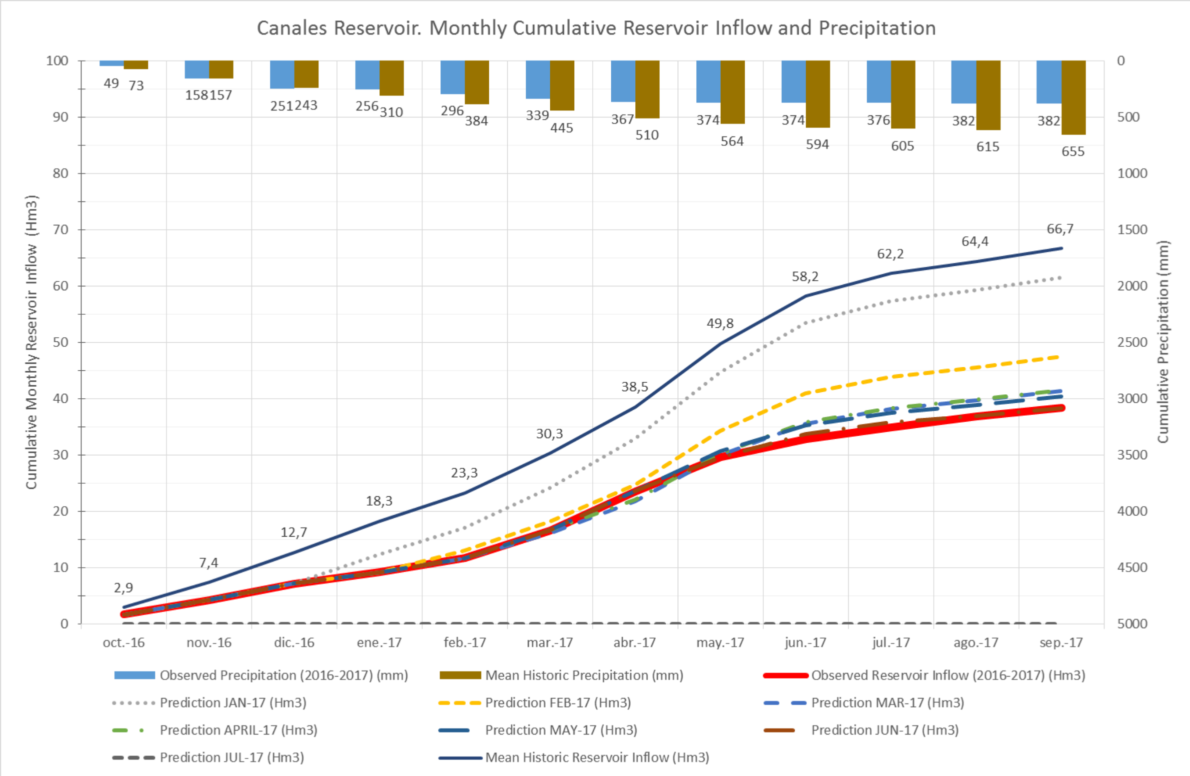


Figure 12. Canales reservoir: Predictions of annual and monthly cumulative streamflow (hm³) and observed annual and monthly cumulative streamflow (hm³)

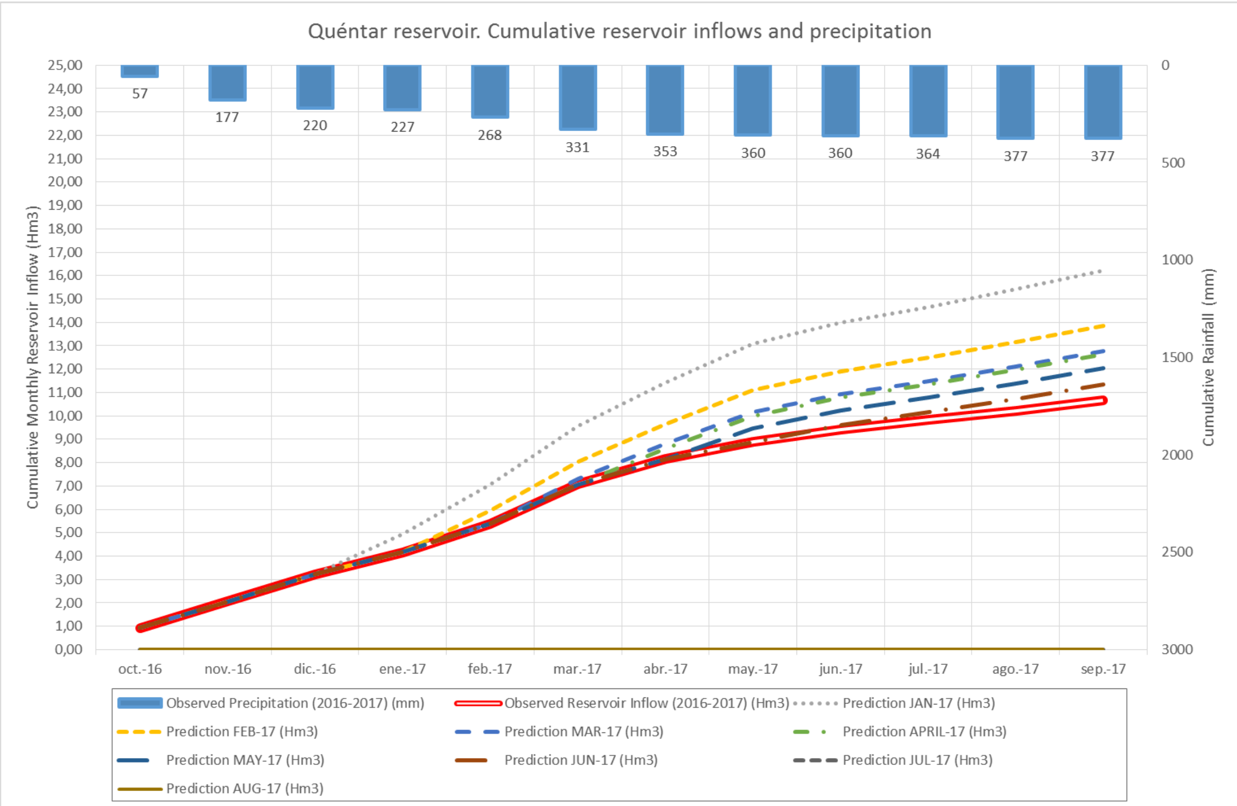


Figure 13. Quéntar reservoir: Predictions of annual and monthly cumulative streamflow (hm³) and observed annual and monthly cumulative streamflow (hm³)

It is observed that for Canales reservoir, the accuracy of the prediction output achieved in March 2017 (for the 2016/2017 total annual streamflow) was very good, approximately 92%. The forecast

Peer-reviewed version available at Water 2018, 10, 1038; doi:10.3390/w10081038

value was 41.52 hm^3 while the observed value was 38.34 hm^3 (compared with the observed historical mean annual streamflow of 80.42 hm^3). The prediction results improved, logically, as the number of observed months increases.

For Quéntar reservoir, the accuracy of the prediction given in March 2017 (for the 2016/2017 total annual streamflow) was slightly lower than for Canales reservoir, an accuracy of 82% was achieve which is good. The total annual streamflow forecast was 12.65 hm^3 while the observed value was 10.68 hm^3 (compared with the observed historical mean annual streamflow of 28.81 hm^3).

We investigated in more detail whether the Quéntar model could be improved. A more detailed evaluation of the behavior and influence of the interannual underground flow in the total annual contribution to the reservoir was carried out. It was found that there is relatively small correlation between the annual streamflow with the previous years. It was also assessed whether the median might be a better estimator instead of the mean value. It was found that, although for the hydrological year 2016-2017 the median was better predictor, in general for a typical wet, normal and dry year, the mean value was the most reliable and accurate predictor. We studied as well the influence of the number of observed years in the results. We concluded that, for this particular, in order to reduce the mean relative error below 12%, the minimum number of observed years should be 40 years.

The model allows the inclusion of the seasonal predictions, for example, those provided by the AEMET in Spain. The seasonal probability prediction values estimated for the next quarter of the ongoing year assigned to each tertile (wet, normal and dry) are applied to the model outputs by multiplying the seasonal predictions to the model outputs (higher, central and lower estimations). Equally, the model allows the integration of the climate change effects by weighing with a higher score to recent years of the historic data sets.

5. Conclusions and future research directions

The protection of the water resource as a vital element to ensure the satisfaction of the present and future human and environmental needs is fundamental. Climate change effects, population growth and increasingly demanding water users are threating the achievement of an integrated water resources management approach.

One of the key steps in achieving an environmentally sustainable and economically rational water resources management approach, apart from applying different supply increase and demand reduction strategies, is through improving the knowledge and experience on the management and decision-making processes. And these are to be supported by reliable and accurate streamflow forecasts models.

This research work contributes towards the development of a novel, simple, low cost and robust methodology to forecasting monthly and annual streamflows within the hydrological year in course. The methodology represents a practical tool, applicable to headwater systems.

In particular, the approach has been successfully applied to two headwater reservoirs located at the upper area of the Genil River (Guadalquivir River Basin) namely, Quéntar and Canales. From the regression analysis, it was shown that the best descriptor for forecasting the total annual streamflow is the cumulative monthly streamflow for Quéntar reservoir (due to the influence of the subterranean flows) and the cumulative monthly rainfall for Canales reservoir (due to the influence of the snow storage and melting processes). It is important to highlight that the best fit regression model and best fit predictor will vary depending on the predominant hydrological process at each specific basin and season of the year (snow storage and melting processes, subterranean inflows, seasonal extreme and sporadic rainfall events). In this situation, a combined regression model should be used to achieve the best correlation and predictive results.

The model has also the option to incorporate seasonal climate predictions and climate change effects. We can conclude that the annual and monthly streamflow forecasts model created, although simple and of low cost, provides satisfactory forecast results within the current hydrological year, with a relatively small error margin.

This model can therefore help on the early detection of critical events such as droughts and water scarcity situations as well as provide support to strategic water management decisions and promote improved resource and cost efficiency when deciding the optimum controlled released outflows time and water allocation, avoiding or minimising the social, economic and environmental consequences of the inappropriate decision.

Supplementary Materials:

Author Contributions: Both authors have contributed equally to the research work.

Funding: This research was funded by *Fundación Centro Andaluz de Investigaciones del Agua* (CETAQUA) and EMASAGRA S.A.

Acknowledgments: We acknowledge B.Eng. Nanton for his collaboration in this work, and CETAQUA and EMASAGRA S.A. for the economic support and technical feed-back.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tanaka, S. K., Zhu, T., Lund, J. R., Howitt, R. E., Jenkins, M. W., Pulido, M. A., Tauber, M., Ritzema, R. S., and Ferreira, I. C. (2006): Climate Warming and Water Management Adaptation for California, Clim. Change, 76, 361–387.
2. Shalamu, A., 2009. Monthly and Seasonal Streamflow Forecasting in the Rio Grande Basin. PhD Thesis. New Mexico State University, U.S.
3. Helmuth, M. E., Moorhead, A., Thomson, M. C., and Williams, J.: Climate Risk Management in Africa (2007): Learning from practice, International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, New York
4. Delorit, J., Gonzalez Ortuya, E. C., and Block, P. (2017): Evaluation of model-based seasonal streamflow and water allocation forecasts. Hydrol. Earth Syst. Sci., 21, 4711–4725, 2017 https://doi.org/10.5194/hess-21-4711-2017
5. Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente, Gobierno de España. Agencia Estatal de meteorología. Available online: http://www.aemet.es/en/serviciosclimaticos/prediccion_estacional (accessed on 01 March 2018).
6. Coulibaly, P., Anctil, F., Bobe´e, B. (2000): Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, Journal of Hydrology 230 (2000) 244–257.
7. Baratti, R., Cannas, B., Fanni, A., Pintus, M., Sechi, G.M., Toreno, N. (2003): River flow forecast for reservoir management through neural networks, Neurocomputing 55 (2003) 421 – 437.
8. Collischonn, W., Morelli Tucci, C.E., Clarke, R.T., Chan Chou, S., Guilhon, L. G., Cataldi, M., Allasia, D. (2007): Medium-range reservoir inflow predictions based on quantitative precipitation forecasts, Journal of Hydrology (2007) 344, 112– 122.
9. Wang, W., 2006. Stochasticity. IOS Press, Nonlinearity and Forecasting of Streamflow Processes.
10. Yu X., Zhang X., Qin H. (2018): A data-driven model based on Fourier transform and support vector regression for monthly reservoir inflow forecasting, Journal of Hydro-environment Research 18 (2018) 12–24 DOI: https://doi.org/10.1016/j.jher.2017.10.005
11. Gragne, A. S., Sharma, A., R. Mehrotra, and Alfredsen, K. (2015): Improving real-time inflow forecasting into hydropower reservoirs through a complementary modelling framework, Hydrol. Earth Syst. Sci., 19, 3695–3714, 2015 www.hydrol-earth-syst-sci.net/19/3695/2015/ doi:10.5194/hess-19-3695-2015
12. Wang Y., Guo S., Chen H., Zhou Y. (2014): Comparative study of monthly inflow prediction methods for the Three Gorges Reservoir, Stoch Environ Res Risk Assess (2014) 28:555–570, DOI 10.1007/s00477-013-0772-4
13. Thom, H. C. S., (1958): A note on the gamma distribution. Mon. Wea. Rev., 86, 117–122.
14. Buishand, T. A. (1978), Some remarks on the use of daily rainfall models, J. Hydrol., 36, 295-308, 1978. https://doi.org/10.1016/0022-1694(78)90150-6
15. Stephenson, D.B., Kumar, K.R., Doblas-Reyes, F.J., Chauvin, E. and Pezzulli, S. (1999). Extreme daily rainfall events and their impact on Ensemble Forecast of the Indian Moonsoon. Monthly Weather Review, 127, 1954-1966

16. Wang, Q. J. and Nathan, R. J (2007).: A Method for Coupling Daily and Monthly Time Scales in Stochastic Generation of Rainfall Series, J. Hydrol., 346, 122–130, https://doi.org/10.1016/j.jhydrol.2007.09.003
17. Chen, J., Brissette, F. P., and Leconte, R. (2010): A Daily Stochastic Weather Generator for Preserving Low-Frequency of Climate Variability, J. Hydrol., 388, 480–490, https://doi.org/10.1016/j.jhydrol.2010.05.032, 2010.
18. Chowdhury, A. F. M. Kamal; Lockart, Natalie; Willgoose, Garry; Kuczera, G.; Kiem, A. S. and Manage, N. P. (2017): Development and evaluation of a stochastic daily rainfall model with long-term variability Hydrol. Earth Syst. Sci., 21, 6541–6558, 2017 https://doi.org/10.5194/hess-21-6541-2017
19. Confederación Hidrográfica del Guadalquivir. Available online: http://www.chguadalquivir.es (accessed on 06 March 2018).
20. Draft Guadalquivir River Basin Special Drought Management Plan (2017) Available online: (accessed on 06 March 2018) https://www.chguadalquivir.es/documents/10182/731658/Memoria+PES+Guadalquivir.pdf
21. Automatic Hydrological Information System (SAIH). Available online: http://www.chguadalquivir.es/saih/ (accessed 2015-2018)