Coupling Satellite Rainfall Estimates and Machine Learning Techniques for Flow Forecast: Application to a Large Catchment in Southern Africa

José P. Matos
Ph.D. student, Laboratoire de Constructions Hydrauliques (LCH) École Polytechnique Fédérale de Lausanne (EPFL), Switzerland; Instituto Superior Técnico (IST), TechnicalUniversity of Lisbon, Portugal. Email: jose.matos@epfl.ch

Théodora Cohen Liechti
Ph.D. student, Laboratoire de Constructions Hydrauliques (LCH) École Polytechnique Fédérale de Lausanne (EPFL),Switzerland. Email: theodora.cohen@epfl.ch

Maria M. Portela
Professor, Instituto Superior Técnico (IST), TechnicalUniversityofLisbon, Portugal.
Email: mps@civil.ist.utl.pt

Anton J. Schleiss
Full Professor, Laboratoire de Constructions Hydrauliques (LCH) École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Email: anton.schleiss@epfl.ch

ABSTRACT: Accurate river flow forecasting is an important asset for stream and reservoir management, being often translated into substantial social, economic and ecological gains. This contribution aims at coupling satellite rainfall estimates and machine learning techniques for daily flow forecast. Two lead times, of 30 and 60 days, were tested for flows at Victoria Falls, in Southern Africa. Six distinct machine learning models were compared with optimized ARMA models and benchmarked against a Fourierseries approximation. Results show that the addition of rainfall data generally enhanced the performances of machine learning models at 30 days but did not improve forecasts at 60 days. Also, it was shown that traditional ARMA models do not make use of the rainfall information. Regarding a lead time of 60 days, the machine learning models appear to bear great advantages compared to ARMA models which, for such a lead time have shown practically no forecast capabilities.

KEY WORDS: Artificial Neural Networks,Flow forecast, Kariba, Support-Vector Regression, Zambezi.

1 INTRODUCTION

Accurate river flow forecasting is an important asset for stream and reservoir management, being often translated into substantial social, economic and ecological gains. In the past, considerable advances have been accomplished on this subject, ranging from the development of physical distributed and lumped conceptual approaches to data-driven models.

In the last decades, a wealth of alternative data-driven models has been proposed among which autoregressive moving-average (ARMA)(e.g. Anderson, 1977; Mohammadi et al., 2006; Valipour et al., 2013), one of the most popular times series models for reservoir design and operation (Wang et al., 2009), autoregressive integrated moving-average (ARIMA)(e.g. Carlson et al., 1970), which is a non-static generalization of the ARMA model (Valipour et al., 2013), artificial neural networks (ANN)(Abrahart and See, 2000; Cigizoglu, 2005; Shamseldin and O'Connor, 2001), genetic programming (GP)(Londhe and Charhate, 2010; Wang et al., 2009), support vector regression (SVR)(e.g. Lin et al., 2006; Wang et al., 2009), and k-nearest neighbors (KNN) (e.g. Sivakumar et al., 2002)are examples. The performances reported in literature vary greatly(Wu and Chau, 2010).

Developed in the framework of the African Dams Project (ADAPT)(Mertens et al., 2013), a flow forecast system is proposed and evaluated for the Zambezi River at Victoria Falls. This section drains the over 360 000 km^2 UpperZambezi catchment. Not far downstream that section laysKariba, the World's

largest artificial reservoir by volume, whose operation greatly influences the Zambezi basin's economy and ecosystem. Characterized by its large size, pronounced seasonal changes in its hydrological response, covering a range of climates from tropical to semi-arid, and being influenced by large and complex wetlands, the UpperZambezi catchment presents an interesting challenge with regard to flow forecasting. The location and main features of the catchment are presented in **Figure 1**.

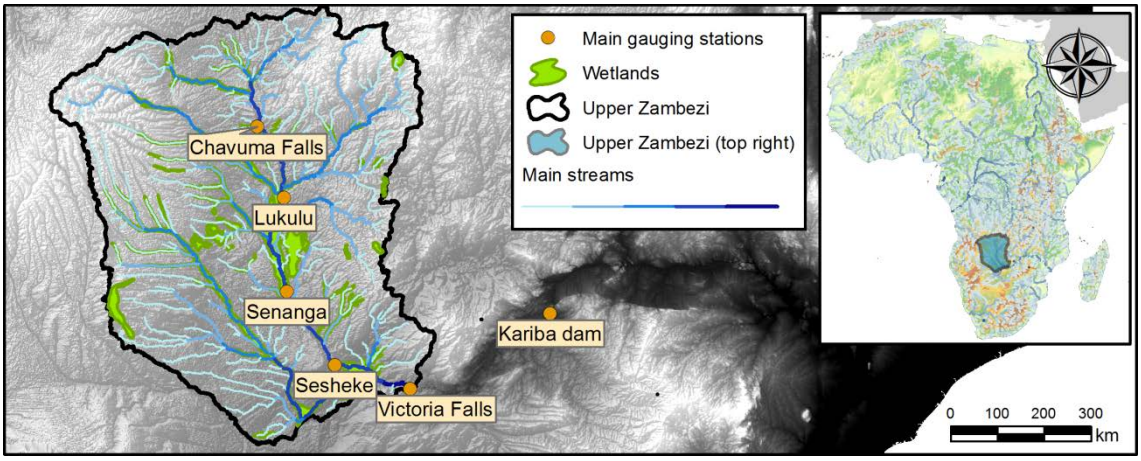

Figure 1 Location and detail of the Upper Zambezi River basin and main discharge gauging stations

In particular, in the present contribution the flow of the Zambezi River is forecast for lead times of 30 and 60 days based on gauge measurements and satellite rainfall estimates. Two distinct pre-processing procedures for the satellite rainfall data aggregation using flow length maps are evaluated. Alternative data-driven forecasting models are compared, namely ARMA and machine learning models such as multilayer perceptron ANNs (MLP), radial basis function networks (RBFN), nonlinear autoregressive with exogenous inputs ANNs (NARX), layer-recurrent ANNs (LR), least squares support vector regression (LSSVR), and LSSVR NARX.

In section 2 the discharge and rainfall data employed in the study are presented. An overview of the applied methods is given in section 3. Section 4 is dedicated to the presentation and discussion of the main results of the study. Finally, conclusions are drawn in section 5.

2 DATA

The historical discharge data for Victoria Falls, provided by the Department of Water Affairs of Zambia,spans a period from 1958-11-7 to 2010-3-6. The series is characterized by a mean discharge of 1100 m^3/s and substantial decadal average flow variations (Matos et al., 2010). Data of additional gauging stations upstream of Victoria Falls were also analyzed: Chavuma Falls, Lukulu, Senanga and Shesheke. These data were provided by the Department of Water Affairs of Zambia, the Zambezi River Authority and the Global Runoff Data Centre, 56068 Koblenz, Germany. The series are represented in **Figure 2**.

As aforementioned, the Upper Zambezi is a large basin which holds wetlands with a great influence on the Zambezi's flow, the most important being the Barotse plains, which spreads between Lukulu and Senanga, and the Chobe-Zambezi confluence, between Sesheke and Victoria Falls (**Figure 1**).In **Table 1** the average daily lags which yield the highest cross-correlations between discharge series at each pair of gauging stations are presented. It can be seen that the Barotse plains delay the hydrograph by approximately 18 days, whereas the Chobe-Zambezi confluence has an impact of 10 days. From Chavuma Falls to Sesheke, and especially to Victoria Falls, the lags amount to nearly one month.

The rainfall data was derived from the Tropical Rainfall Measuring Mission (TRMM), namely in the form of the TRMM Multi-satellite Precipitation Analysis (TMPA) 3B42 version 7a product (Huffman et al., 1995; Huffman et al., 2007) which produces merged microwave and infrared precipitation at three-hourly and 0.25° resolutions.

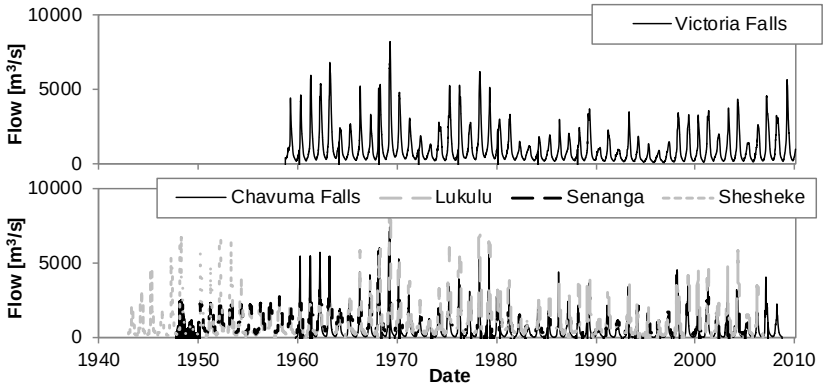

Figure 2 Historical series of daily records from the main discharge gauging stations in the Upper Zambezi River

Table 1 Average lag (in days) between the main discharge gauging stations in the Upper Zambezi River

Chavuma Falls	Lukulu	Senanga	Shesheke	Victoria Falls
-1				
-17	-18			
-22	-18	-1		
-28	-25	-9	-10	

Reliable daily or sub-daily satellite rainfall estimates have only been made available for the Upper Zambezi area from 1998 onwards – the period covered by TMPA 3B42 data. In order to meet the goals of the present contribution a period of overlapping daily rainfall and river discharge was sought. Accordingly, the focus of the remainder of the text will be on the period spanning from 1998-1-1 to 2010-3-6. In **Figure 3** the average yearly rainfall is depicted for the Upper Zambezi area and the period of interest. As can be seen, average annual rainfall values vary from nearly 2000 mm in the northern part of the catchment to 400 mm in the south.

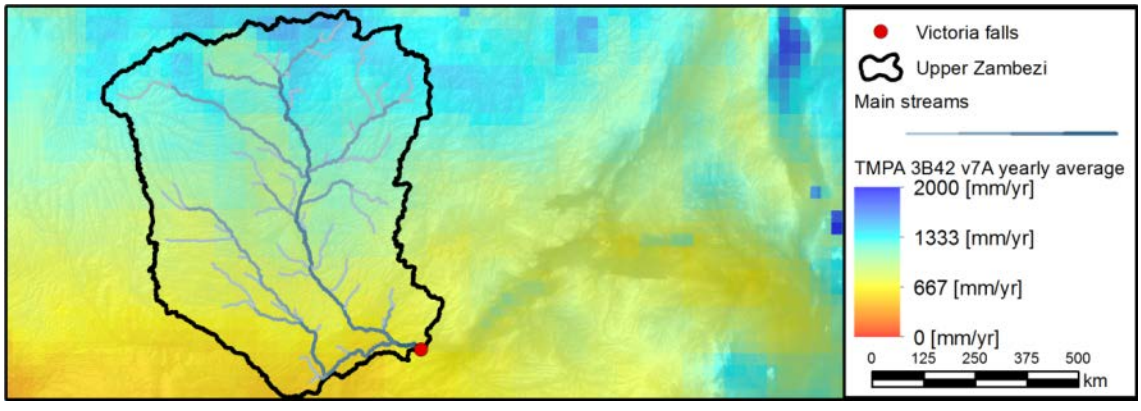

Figure 3 Average yearly rainfall estimated by the TMP 3B42 version 7a product from 1998-1-1 to 2010-3-6

In order to fit and evaluate the performance of the different models a cross-validation procedure was implemented by partitioning the data into training (60% of the data), validation (20%) and test (20%) periods. In **Figure 4** the three periods are illustrated along with the mean daily rainfall estimates over the catchment.

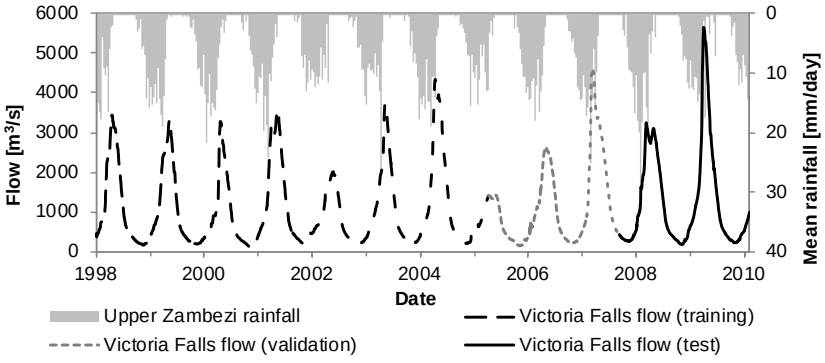

Figure 4 Studied discharge series at Victoria Falls showing training, validation and test periods. Mean daily rainfall over the Upper Zambezi displayed on top and the right axis.

3METHODS

3.1Applied models

3.1.1 ARMA

ARMA are mathematical models of the persistence, or autocorrelation, in a time series (Wu and Chau, 2010). Widely used in hydrology, anARMA(p,q) model is characterized by p, the autoregressive order, and q, the moving-average order. In the present case, given the evident seasonal behavior of the flood, and being that the model is applied to daily data, an external input in the form of a Fourier series approximation was combined with the ARMA model.

In order to select the adequate model for each of the tested conditions, the values of p and q were made to vary from 0 to 10 in a stepwise procedure. The chosen model was the one with the minimum Akaike Information Criterion value.

3.1.2 Multilayer perceptron

The multilayer Perceptron (MLP) is the most common feedforward ANN. Numerous successful applications of this type of model have been reported in hydrology and, particularly, in attempts to forecast discharges (e.g. Baratti et al., 2003; Kumar et al., 2004). Being thoroughly described in the literature (e.g. Haykin, 1994), the MLP is a non-parametric static network commonly used for regressions tasks. In the application of the MLP there are a number of choices to be made, namely of the number of hidden layers and the number of nodes amongst them, the activation functions and the training algorithms. This, adding to the intrinsic stochastic behavior of the training procedure, renders the output of MLP models prone to substantial variations.

In order to achieve some significance of the results, each MLP architecture was run 1000 times for each scenario, being the presented performances (always referring to the test period) the average of the best 100 validation performances in terms of Nash-Sutcliffe Efficiency Coefficient. Architectures of 5 and 10 sigmoid nodes within one single hidden layer, followed by a linear output node, were tested. The Levenberg-Marquardt backpropagation algorithm(Hagan, 1994) was chosen for training.

3.1.3 Nonlinear autoregressive with exogenous inputs ANN

The nonlinear autoregressive with exogenous inputs ANNs (NARX) differ from the MLP due to a delayed feedback connection from the output layer to the input layer. In practice, this results in the capacity of the network to exhibit a dynamic behavior. Due to the presence of the feedback connection, the gradients that serve as a basis for the adjustment of the network to the training data must be approximated,which hinders training.Although obviously interesting in view of the task at hand, where each estimate is partially correlated to the preceding, the NARX model's potential benefits must be

weighed against the likely loss of training performance.

A single architecture of one hidden layer with 5 sigmoid nodes and a linear output node was tested. As before, the Levenberg-Marquardt backpropagation algorithm was chosen. Being much slower to train than the MLPs, for each scenario 20 models were fitted, being the presented test performances averages from the 20% networks that evidenced higher Nash-Sutcliffe coefficients for the validation period.

3.1.4 Layer recurrent ANN

The layer recurrent ANN (LR), or recurrent multiplayer perceptron (Haykin, 1994) is quite similar to the NARX ANN. In this type of network, however, it is not the output that is fed back, but the state of each node on the hidden layers. With an increased number of feedback connections the LR ANN can approximate sophisticated dynamic behaviors. This is, however, accomplished at the expense of additional training difficulties.

The Bayesian regularization backpropagation algorithm (MacKay, 1992) was used to train the LR models. Attending to the specificities of this training algorithm, a relatively larger architecture was adopted, including 20 hidden sigmoid nodes in the hidden layer. Due to the highly demanding computational training requirements, 12 networks were trained for each scenario over 200 iterations, being the compared performances the average resulting from the four with the highest training Nash-Sutcliffe coefficient.

3.1.5 Radial basis function networks

The radial basis functions networks (RBFNs)(Broomhead and Lowe, 1988) are a type of static ANN fundamentally different from the aforementioned models. Typically comprised of two main layers, the RBFN projects the input space into a high dimensionality space of generalized distances to training points using kernels. Once the coordinatesin this high dimensionality space are known, the RBFN performs a multiple linear regression in order to compute its outputs. The Gaussian kernel was used in the present work. Consequently, the network has one free parameter, σ, which controls the width of the kernel and ultimately tunes be behavior of the RBFN.

The correct definition of the free parameter $\sigma$is essential to attain good performances. In the present contribution, this value has been optimized using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm(Auger and Hansen, 2005; García et al., 2009; Hansen, 2006; Hansen, 2010), which is a random search evolutionary algorithm adequate to real-parameter optimization of non-linear, non-convex functions, in which the candidate solutions are sampled according to a multivariate normal distribution.

3.1.6. Least squares support vector regression

The least squares support vector regression (LSSVR)(Suykens et al., 2002; Suykens and Vandewalle, 1999) is a variation of the more widespread support vector regression concept(Drucker et al., 1996), being the practical difference between the two models the fact that the latter minimizes a function based on the mean absolute error and LSSVRs minimize the mean squared error. Like RBFNs, LSSVRs apply kernels to project the inputs into a high dimensionality space. In the case of LSSVRs, however, the regression is subject to a principle of structural risk minimization.Like the RBFNs, LSSVR models with a Gaussian kernel have a free parameter, σ, which should be optimized. Like before, the CMA-ES algorithm was used.

LSSVR has the potential to improve the promising results of support vector regression found in similar studies (e.g. Lin et al., 2006) as, by minimizing the mean squared error instead of the mean absolute error, the integral of its forecasts converges to the true river runoff volume, which is not generically true for SVRs. Additionally to the static LSSVR, a LSSVR NARX implementation was attempted(Suykens and Vandewalle, 2000).

3.2Tested forecasting approaches and error metrics

The different models were applied using four alternative forecasting procedures. The first, schematized in **Figure 5**, is usual in ARMA implementations. It performs a one-step-ahead forecast and,

as lead times increase, estimates are fed into the model as inputs until the desired lead time is reached.

In **Figure 5**, $\hat{y}_i$represents the desired estimate; $\boldsymbol{x}^*$ a generalized vector of inputs comprised by estimates, $\hat{y}$ (or observations y) and other ancillary variables,$\boldsymbol{x}$; the index i indicates the time frame of interest; and $\boldsymbol{W}$ is a matrix characterizing the parameters of the model.

A purely feedforward technique was applied to the MLP, RBFN and LSSVR models. Using the same notation, this technique is illustrated in **Figure 6**. The main difference from the prior formulation is that earlier estimates are never used by the model, being that the forecast is made directly for a predetermined lead time (lag) based solely on ancillary variables and lagged observations. Consequently, no estimates are explicitly interdependent.

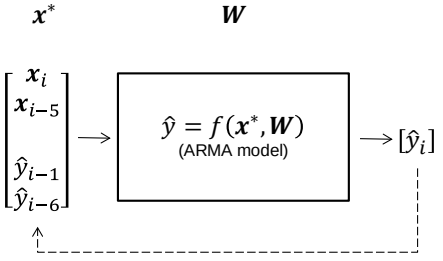

Figure 5 One-step-ahead forecast model's scheme

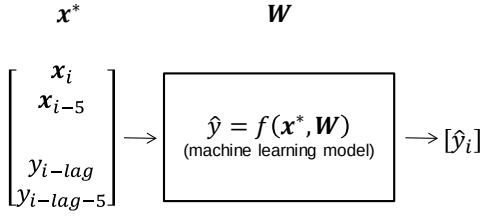

Figure 6 Feedforward forecast models' scheme

Figure 7 shows the implementation of the NARX and LSSVR NARX models. While this implementation directly computes the estimate associated with a predetermined lead time, information from earlier estimates is also passed on to the model. A potential advantage of such a formulation when compared to the feedforward approach is the possibility of attaining added stability of the forecasts. A forth alternative, characteristic of the LR model, is illustrated in **Figure 8**. Being similar to the implementation adopted for the NARX models, instead of using past estimates, this method employs information from past states of the estimator as inputs ($\boldsymbol{\varphi}$).

After initial trials it was remarked that all the models performed better once a Fourier series approximation of the annual discharge cycle was introduced as an ancillary variable. This approximation followed the form of equation (1) with $K = 4$, and $T = 365$.In order to compare the alternatives, flow forecast performances are evaluated for different lead times in terms of Nash-Sutcliffe efficiency coefficient (Nash and Sutcliffe, 1970), Pearson correlation coefficient, Spearman correlation coefficient, relative accumulated runoff deviation, root mean squared error (RMSE), and mean absolute error (MAE).

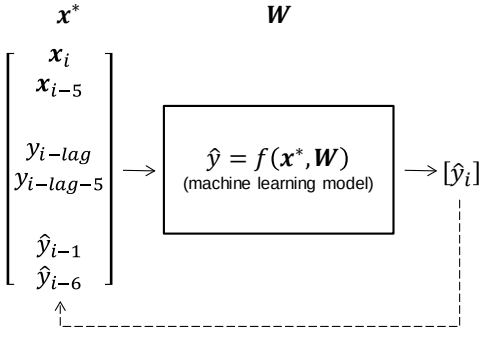

Figure 7 NARX forecast model's scheme

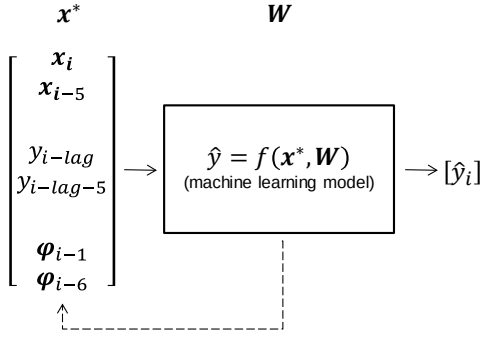

Figure 8 LR forecast model's scheme

$$x_i = \left[\frac{a_0}{2} + \sum_{k=1}^{K}\left[a_k \cos\left(\frac{2\pi k}{T}\cdot t(i)\right) + b_k \sin\left(\frac{2\pi k}{T}\cdot t(i)\right)\right]\right] \quad (1)$$

The Nash-Sutcliffe efficiency coefficient (NS) attains a value of 1 for a perfect fit, a value of 0 for a

constant estimate equal to the average of the true values and can reach minus infinity for worse fits. The Pearson correlation coefficient is 1 for a perfect linear correspondence between estimates and true values, -1 for a perfect inverse linear correspondence and 0 for the absence of a linear relationship. The relative accumulated runoff deviation is the relative bias of the whole forecasted series. In the absence of bias it equals 0 whereas positive values indicate an overestimation and, conversely, negative values reflect an underestimation. The RMSE, as the MAE, are computed through sums of errors and depart from 0 as estimates and true values diverge. Their difference resides in the fact that the former is based on a sum of squares and the later on a sum of absolute values. As such, the RMSE is more sensitive to large deviations than the MAE.Being widely described in literature, the equations of these metrics will not be transcribed.

The Spearman's rank correlation coefficient, ρ, is a non-linear metric of correlation which evaluates the degree to which estimates and true values can be related by any monotonic function. Varying between -1 and 1 it can be interpreted similarly to the Pearson correlation coefficient.

3.3 Rainfall as an input

The substantial time flows take to travel along the Upper Zambezi added to the regularization effect of the Barotse Plains and the Chobe-Zambezi confluence potentiate the accuracy of more or less distant forecasts. In this work, satellite rainfall estimates are input to the different models in order to test whether they improve forecast performances.The applied TMPA 3B42 version 7a data was pre-processed in two main steps: the aggregation of daily accumulation into 6 equidistant distance bands obtained by division of the longest flow path (**Figure 9**); and the smoothing of this data through the use of a moving average with a window of 20 days.

Based on the 6 distance bands, the models were also trained using only rainfall data from the 3 most upstream bands, being these are the ones that exert most influence on the hydrographs at Victoria Falls when lead times larger than 30 days are considered.

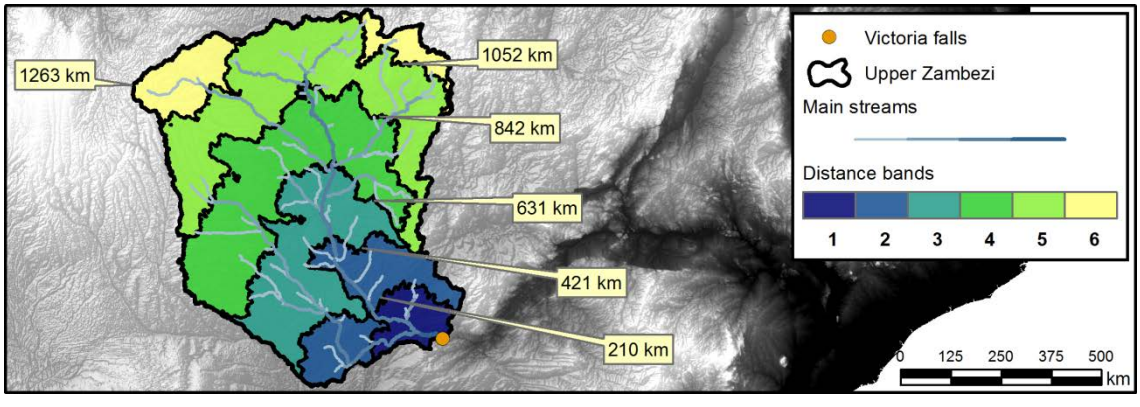

Figure 9 Illustration of the six rainfall bands applied in the study

4 RESULTS AND DISCUSSION

4.1 Using observed discharges as inputs

The first results pertain to a lead time of 30 days obtain using solely past observed discharges and the Fourier approximation of the discharge as inputs.In **Table 2** performances are quantified for each metric and model. For each metric, the bars in the background are proportional to the relative differences between models. The best model for each metric is highlighted in bold red font.

The Fourier series approximation is independent of preceding flows and does not make forecasts but rather depicts the average annual hydrograph.As such, itprovides a good benchmark for the predictive capabilities of the other models regardless of lead time. Analyzing **Table 2** it can be observed that most models are able topropagate information of past flows 30 days into the future. Both MLP models perform similarly to the fitted ARMA(6,2). The NARX model however, seems to outperform all others in the

majority of the metrics.The LR_{20} model proved to be of difficult calibration and should be considered an outlier. This, however, might be related to a high number of hidden nodes and the training algorithm that was used (Bayesian regularization).

When a lead forecast of 60 days is endeavored (**Table 3**), the comparative results change appreciably. The ARMA(6,2) model seems to collapse into the Fourier series approximation, indicating that this linear model fails to use information from recent observations. Most of the other models show performances close to those attained for a lead forecast of 30 days. The best models seem to have been the MLP_5 and LR_{20}. As for the LSSVR results it can be remarked that, for both lead times, they lead to the smallest relative deviation, which is an important parameter for reservoir management.

Table 2 Results for a lead time of 30 days using observed discharges and 6 rainfall bands as inputs

Model	Nash Sutcliffe	Pearson	Spearman	Relative Deviation	RMSE	MAE
Fourier	0.736	0.917	0.982	-0.243	21.730	304.587
ARMA(6,2)	0.833	0.919	0.983	-0.109	17.268	231.763
MLP_5	0.826	0.929	0.978	-0.077	16.423	**217.448**
MLP_{10}	0.831	0.921	0.980	-0.063	16.818	231.845
LSSVR	0.807	0.910	0.980	**0.020**	18.834	268.028
$NARX_5$	**0.870**	**0.939**	**0.984**	-0.082	**15.438**	221.255
LR_{20}	0.169	0.730	0.874	1.257	37.235	628.372
RBFN	0.844	0.930	0.972	-0.087	16.921	219.901
LSSVR NARX	0.785	0.897	0.971	-0.024	19.840	333.881

Table 3 Results for a lead time of 60 days using observed discharges and 6 rainfall bands as inputs

Model	Nash Sutcliffe	Pearson	Spearman	Relative Deviation	RMSE	MAE
Fourier	0.736	0.917	**0.982**	-0.243	21.730	304.587
ARMA(6,2)	0.749	0.892	0.979	-0.186	21.162	296.437
MLP_5	0.843	**0.931**	0.967	-0.113	16.811	**270.424**
MLP_{10}	0.810	0.911	0.970	-0.082	18.307	291.359
LSSVR	0.843	0.923	0.967	**-0.054**	17.076	297.893
$NARX_5$	0.828	0.928	0.969	-0.146	17.803	277.856
LR_{20}	**0.847**	0.926	0.958	-0.039	**16.796**	290.925
RBFN	0.762	0.889	0.933	-0.062	21.006	370.401
LSSVR NARX	-0.478	0.615	0.842	-0.031	52.346	666.453

4.2Usingobserveddischarges and 6 rainfall bands as inputs

To evaluate the potential benefits of adding rainfall information as input to the forecast, two modelswere dropped from the analysis: RBFN and LSSVR NARX. This decision was made dueto their poor performances and long and computationally demanding training and optimization.As can be seen in **Table 4**, when data from the accumulated rainfall in the 6 distance bands was added to the input vector results have improved marginally for the 30 days lead forecast, particularly those of the MLP and LSSVR. The ARMA(3,2) model shows little changes when compared to the earlier ARMA(6,2), hinting that the linear model has difficulties to make use of the rainfall series. Finally, the best model, $NARX_5$, has actually displayed a loss in performance.

Table 4 Results for a lead time of 30 days using observed discharges and 6 rainfall bands as inputs

Model	Nash Sutcliffe	Pearson	Spearman	Relative Deviation	RMSE	MAE
Fourier	0.736	0.917	0.982	-0.243	21.730	304.587
ARMA(3,2)	0.833	0.922	**0.984**	-0.125	17.496	233.902
MLP_5	0.849	0.925	0.972	-0.063	16.591	**229.760**
MLP_{10}	0.824	0.912	0.970	-0.055	17.846	246.935
LSSVR	0.854	0.927	0.972	**0.003**	16.361	233.578
$NARX_5$	**0.856**	**0.933**	0.970	-0.106	**16.360**	243.059
LR_{20}	0.404	0.696	0.859	-0.262	31.398	480.762

Regarding the 60 days lead forecast (**Table 5**), a general degradation of the model's performances was remarked with the addition of the 6 series of accumulated rainfall. Indeed, it is possible that the relevant information that is eventually contained in these data does not outweigh the instability it induces in the model's responses and training procedures.

Table 5 Results for a lead time of 60 days using observed discharges and 6 rainfall bands as inputs

Model	Nash Sutcliffe	Pearson	Spearman	Relative Deviation	RMSE	MAE
Fourier	0.736	**0.917**	**0.982**	-0.243	21.730	304.587
ARMA(3,5)	0.749	0.891	0.976	-0.184	21.564	308.172
MLP_5	0.747	0.885	0.940	-0.152	21.573	327.539
MLP_{10}	**0.756**	0.885	0.939	**-0.130**	**21.142**	333.881
LSSVR	0.744	0.889	0.974	-0.150	21.826	**284.675**
$NARX_5$	0.727	0.881	0.916	-0.147	22.736	335.041
LR_{20}	0.505	0.760	0.903	-0.176	29.444	474.970

4.3 Usingobserved discharges and 3 upstream rainfall bands as inputs

Due to the hypothesis that the full 6 bands mightadd too much noise to the input vector, an additional effort using only the 3 most upstream bands was undertaken. This was aimed at removing the data of the downstream bands close to Victoria Falls, which based on the lagged correlation evaluation presented in **Table 1** should not have much impact in the flows 30 or more days into the future.

Results for a lead time of 30 days, presented in **Table 6**, show that a general increase in the models' performances was attained. Relevant exceptions were the ARMA models, which, again, do not seem sensitive to rainfall data, and $NARX_5$, which evidenced a minor degradation in nearly all metrics.

Table 6 Results for a lead time of 30 days using observed discharges and 3 upstream rainfall bands as inputs

Model	Nash Sutcliffe	Pearson	Spearman	Relative Deviation	RMSE	MAE
Fourier	0.736	0.917	0.982	-0.243	21.730	304.587
ARMA(3,2)	0.833	0.922	**0.984**	-0.126	17.495	233.741
MLP_5	0.871	0.937	0.979	-0.046	15.343	**217.494**
MLP_{10}	0.849	0.931	0.974	-0.029	16.174	230.922
LSSVR	**0.874**	**0.938**	0.979	**0.020**	**15.232**	235.462
$NARX_5$	0.860	0.930	0.977	-0.047	16.158	226.149
LR_{20}	0.258	0.634	0.743	-0.255	35.678	536.823

Focusing on the60 days lead time performances,presented in **Table 7**, it can be remarked that the addition of the 3 upstream rainfall bands does not contribute to improved forecast and, although the estimates are better than those resulting from the use of the full set of 6 rainfall bands, for most models there is still a disadvantage in adding this type of rainfall aggregation to the inputs. One exception is the ARMA(7,2) model, which improved marginally compared with earlier results.

Table 7 Results for a lead time of 60 days using observed discharges and 3 upstream rainfall bands as inputs

Model	Nash Sutcliffe	Pearson	Spearman	Relative Deviation	RMSE	MAE
Fourier	0.736	0.917	0.982	-0.243	21.730	304.587
ARMA(7,2)	0.785	**0.922**	**0.983**	-0.201	19.838	281.355
MLP_5	0.786	0.904	0.959	-0.130	19.826	303.393
MLP_{10}	**0.813**	0.912	0.963	**-0.115**	**18.569**	296.512
LSSVR	0.745	0.898	0.976	-0.173	21.759	**276.687**
$NARX_5$	0.761	0.898	0.938	-0.178	21.292	331.040
LR_{20}	0.685	0.834	0.954	-0.138	22.322	349.399

Having the error metrics for the 6 scenarios and all the models been presented, the discussion should be interpreted in light of some limitations inherent to the study. In this and similar problems, longer series of data are always desirable and could, all too often, contribute to more concluding analyses. Due to the

recent inception of satellite areal rainfall estimations, training, validation and test periods have been drawn from a sample of 4419 daily records, leaving a period of roughly two years for model testing, which could be considered short.Notwithstanding, the test period is both demanding on the forecasts and enlightening of the different model's strengths and limitations, for it contains, by far, the highest flood on record (5651 m^3/s, in 2009, comparable with a training peak of 4340 m^3/s in 2004) and, consequently, enables comparisons where models are most uncertain: when extrapolations are made.

The tested models have different geneses and, despite the best efforts of the modeler, can seldom be compared in truly fair circumstances. Moreover, a daunting number of variables can be tuned in order to boost model's performances.Above, many such variables were adopted by convention, empirical rules, and brief exploration of alternatives, this meaning, in practical terms, that the presented results could certainly be improved and (improbably, albeit not impossibly) tend to different comparisons.

Finally, the rainfall data was pre-processed by being aggregated in 6 distance bands and averaged according to a moving window of 20 days. Slightly different variations of this scheme were tested, but surely there are more advantageous procedures and/or additional steps that might enhance results.

4.4 Proposed models

From the analyses of the earlier results, it is evident that substantial differences can be expected in the forecasts from the various types of model. One common denominator was that all non-linear models have proven quite sensitive to noise in the input set, from which stems that rainfall data should only be added when there is good evidence that it leads to better forecasts. Additionally, more complex models tended to display higher variability among results, occasionally with cases of miss-training and test performances far below those of the Fourier series approximation.

Overall, the addition of rainfall data was only relevant for the 30 days lead time forecasts butthis information was only harnessed by the non-linear models. The way in which the data are pre-processed is relevant in terms of performance. The dynamical recurrent models have proven to be more unstable than static ones. Although the NARX network with 5 hidden nodes has displayed promising performances, the LR and LSSVR NARX models generally failed to attain reliable responses. For their simplicity and reliability, it is considered that static models emerge as best and, among these, the simpler ones, namely MLP_5, combine good performances with stable responses.The ARMA models, although surpassed in terms of performance in both lead times, have very interesting features of reliability and quantification of uncertainty and, based on the results, should not be dismissed for the proposed tasks.

The LSSVR model, however inadequate for the 60 days lead time, resulted in the best performances for the 30 days lead time and was consistently the best estimator of flow volumes. Being a type of kernel regressor, a LSSVR model loosely "recalls" training events in order to produce estimates and, due to the structural risk minimization principle that it incorporates, when no similar events are available to be recalled, gradually tends to a "non-compromising" responsewhich may lead to misleading forecasts.

From these considerations, the proposed models emerging from this study are the MLP and the ARMA. In **Figure 11** characteristic forecast hydrographs are shown for a 30 days lead time. Also, histograms of forecast deviations are illustrated.

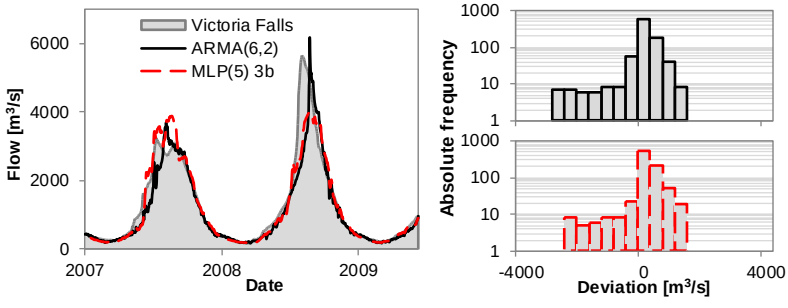

Figure 11 Example of 30 days' forecasts (left) and deviations (right) from the ARMA(6,2) and MLP_5 3b (3 rainfall bands) models

In this case the MLP_5 model with 3 rainfall bands added to the inputs performed better than the best ARMA alternative. Although this can be intuited when the deviation histograms are compared, the forecasted hydrographs can be misleading. In fact, the ARMA model appears to estimate the peak flood much better than the MLP; this, however, is not accurate. Having in mind that the hydrographs are estimated 30 days into the future, the ARMA model is actually estimating the peak after it was observed, having missed it when it occurred. The MLP's forecast also misses the peak (although by a smaller difference), but forecasted the recessionmore accurately.

A similar effect can be seen for the 60 days lead time (**Figure 12**), where the MLP_5 model is visibly better. Here, after missing the 2008 flood's peak altogether, the ARMA model starts overestimating the recession flows after the peak was observed, an effect that also influences the MLP's estimates to a smaller degree.

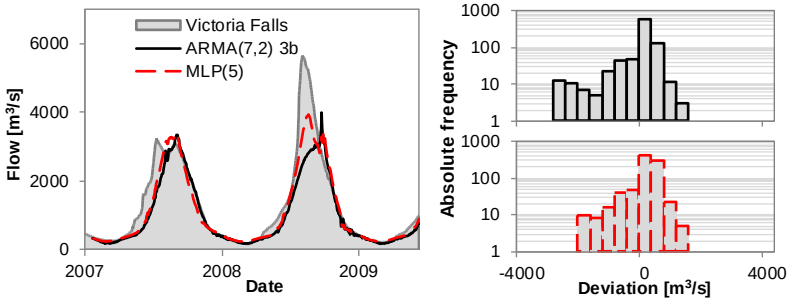

Figure 12 Example of 60 days' forecasts (left) and deviations (right) from the ARMA(7,2) 3b (3 rainfall bands) and MLP_5 models

5 CONCLUSIONS

The present contribution aimed at coupling satellite rainfall estimates and machine learning techniques for daily flow forecast. Two lead times were tested: 30 and 60 days. The machine learning models were compared with a fitted ARMA model and benchmarked against a Fourier series approximation of the annual hydrographs. Results are promising and clearly show that, even considering lead times of 60 days, forecasts can be made with surprising accuracy when non-linear models are used.

It was observed that satellite rainfall could enhance forecasts for the 30 days lead time, but improvements were only registered for the non-linear models. The additional noise conveyed by too detailed or unrelated rainfall input data appears to hinder the training and fitting processes.

For the 60 days lead time, forecasts by the non-linear models clearly outperform those obtained using ARMA equations, the latter approaching the Fourier series approximation benchmark or, in other words, almost displaying no forecast capabilities.

Among all the tested models, small MLP networks (30 and 60 days) and ARMA (30 days) models are recommended. However, dynamic NARX networks and static LSSVR alternatives are interesting and can justify further work that aims at resolving issues with training stability and the choice of parameters that avoid misleading forecasts for large flood events.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Department of Water Affairs of Zambia, the Zambezi River Authority and the Global Runoff Data Centre which kindly provided the discharge data applied in this contribution. The authors also thank Artur Silva for his valuable insights regarding the calculations performed using the R software. This work was mainly supported by the Portuguese Foundation for Science and Technology (Fundaçãopara a Ciência e a Tecnologia, FTC) under the scholarship SFRH/BD/33669/2009 with additional contribution from the ÉcolePolytechniqueFédérale de Lausanne.

References

Abrahart, R.J., See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrological Processes, 14(11-12): 2157-2172.

Anderson, O.D., 1977. Time series analysis and forecasting: another look at the Box-Jenkins approach. Journal of the Royal Statistical Society. Series D (The Statistician), 26(4): 285-303.
Auger, A., Hansen, N., 2005. A restart CMA evolution strategy with increasing population size, IEEE Congress on Evolutionary Computation. IEEE, pp. 1769-1776 Vol. 2.
Baratti, R. *et al.*, 2003. River flow forecast for reservoir management through neural networks. Neurocomputing, 55(3): 421-437.
Broomhead, D., Lowe, D., 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks, Royal signals and radar establishment Malvern, England.
Carlson, R.F., MacCormick, A.J.A., Watts, D.G., 1970. Application of linear random models to four annual streamflow series. Water Resources Research, 6(4): 1070-1078.
Cigizoglu, H.K., 2005. Generalized regression neural network in monthly flow forecasting. Civil Engineering and Environmental Systems, 22(2): 71-81.
Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. Advances in neural information processing systems: 155-161.
García, S., Molina, D., Lozano, M., Herrera, F., 2009. A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 special session on real parameter optimization. Journal of Heuristics, 15(6): 617-644.
Hagan, M., 1994. Training feedforward networks with the Marquardt algorithm. IEEE transactions on neural networks, 5(6): 989-993.
Hansen, N., 2006. The CMA Evolution Strategy: A Comparing Review. In: Lozano, J., Larrañaga, P., Inza, I., Bengoetxea, E. (Eds.), Towards a new evolutionary computation. Studies in Fuzziness and Soft Computing. Springer Berlin / Heidelberg, pp. 75-102.
Hansen, N., 2010. The CMA evolution strategy: A tutorial. http://www.lri.fr/~hansen/cmatutorial.pdf.
Haykin, S., 1994. Neural networks: a comprehensive foundation. Prentice Hall PTR Upper Saddle River, NJ, USA.
Huffman, G.J., Adler, R.F., Rudolf, B., Schneider, U., Keehn, P.R., 1995. Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. Journal of Climate, 8(5): 1284-1295.
Huffman, G.J. *et al.*, 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. Journal of Hydrometeorology, 8(1): 38-55.
Kumar, D.N., Raju, K.S., Sathish, T., 2004. River flow forecasting using recurrent neural networks. Water resources management, 18(2): 143-161.
Lin, J.-Y., Cheng, C.-T., Chau, K.-W., 2006. Using support vector machines for long-term discharge prediction. Hydrological Sciences Journal, 51(4): 599-612.
Londhe, S., Charhate, S., 2010. Comparison of data-driven modelling techniques for river flow forecasting. Hydrological Sciences Journal, 55(7): 1163-1174.
MacKay, D.J., 1992. A practical Bayesian framework for backpropagation networks. Neural computation, 4(3): 448-472.
Matos, J.P., Cohen, T., Boillat, J.-L., Schleiss, A.J., Portela, M.M., 2010. Analysis of flow regime changes due to operation of large reservoirs on the Zambezi River, 6th Internationational Symposyium on Environmental Hydraulics, Athens, Greece.
Mertens, J. *et al.*, 2013. Adapted reservoir management in the Zambezi river basin to meet environmental needs. Hydropower & Dams, 20(2): 80-84.
Mohammadi, K., Eslami, H.R., Kahawita, R., 2006. Parameter estimation of an ARMA model for river flow forecasting using goal programming. Journal of Hydrology, 331(1): 293-299.
Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology, 10(3): 282-290.
Shamseldin, A.Y., O'Connor, K.M., 2001. A non-linear neural network technique for updating of river flow forecasts. Hydrology and Earth System Sciences, 5(4): 577-597.
Sivakumar, B., Jayawardena, A., Fernando, T., 2002. River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. Journal of Hydrology, 265(1): 225-245.
Suykens, J.A.K., Gestel, T.V., Brabanter, J.D., Moor, B.D., Vandewalle, J., 2002. Least squares support vector machines. World Scientific Pub Co Inc.
Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural processing letters, 9(3): 293-300.
Suykens, J.A.K., Vandewalle, J., 2000. Recurrent least squares support vector machines. IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, 47(7): 1109-1114.
Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476(1): 433-441.
Wang, W.-C., Chau, K.-W., Cheng, C.-T., Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of Hydrology, 374(3): 294-306.
Wu, C.L., Chau, K.W., 2010. Data-driven models for monthly streamflow time series prediction. Engineering Applications of Artificial Intelligence, 23(8): 1350-1367.