ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Physics and Chemistry of the Earth xxx (2007) xxx–xxx

PHYSICS and CHEMISTRY of the EARTH

www.elsevier.com/locate/pce

Monthly streamflow prediction in the Volta Basin of West Africa: A SISO NARMAX polynomial modelling

B.A. Amisigo [a,b,*], N. van de Giesen [c], C. Rogers [a], W.E.I. Andah [b], J. Friesen [c]

[a] *Centre for Development Research (ZEF), Bonn University, Germany*
[b] *CSIR-Water Research Institute (CSIR-WRI), Ghana*
[c] *Delft University of Technology, The Netherlands*

Received 14 April 2006

Abstract

Single-input-single-output (SISO) non-linear system identification techniques were employed to model monthly catchment runoff at selected gauging sites in the Volta Basin of West Africa. NARMAX (Non-linear Autoregressive Moving Average with eXogenous Input) polynomial models were fitted to basin monthly rainfall and gauging station runoff data for each of the selected sites and used to predict monthly runoff at the sites. An error reduction ratio (ERR) algorithm was used to order regressors for various combinations of input, output and noise lags (various model structures) and the significant regressors for each model selected by applying an Akaike Information Criterion (AIC) to independent rainfall-runoff validation series.

Model parameters were estimated from the Matlab REGRESS function (an orthogonal least squares method). In each case, the sub-model without noise terms was fitted first followed by a fitting of the noise model. The coefficient of determination (*R*-squared), the Nash-Sutcliffe Efficiency criterion (NSE) and the *F* statistic for the estimation (training) series were used to evaluate the significance of fit of each model to this series while model selection from the range of models fitted for each gauging site was done by examining the NSEs and the AICs of the validation series.

Monthly runoff predictions from the selected models were very good, and the polynomial models appeared to have captured a good part of the rainfall-runoff non-linearity. The results indicate that the NARMAX modelling framework is suitable for monthly river runoff prediction in the Volta Basin. The several good models made available by the NARMAX modelling framework could be useful in the selection of model structures that also provide insights into the physical behaviour of the catchment rainfall-runoff system.

Keywords: NARMAX; NARX; Rainfall-runoff modelling; Non-linear models; Polynomial models; Dynamic models; Structure selection; Systems identification

1. Introduction

A proper basin scale water resources development and management scheme requires both existing and a good prediction of future stream flow series. Flood management and sizing of on-stream reservoirs require knowledge of both the magnitude and frequency of high flows while drought management, water withdrawals from streams and waste load carrying capacity of streams are determined from the magnitude and frequency of low flows (Tabrizi et al., 1998). An accurate simulation of river runoff series is, therefore, an important input to a comprehensive water resources management at the basin scale.

In the semi-arid Volta Basin of West Africa, existing monthly stream flow series for almost all the gauging stations are short and incomplete. In general these flow series are 20% incomplete with several stations having as high as 80% gaps (Taylor, 2004). Thus, stream flow simulation in this basin is essential not only as a means of filling in some of these gaps but also for the extension of these series in

* Corresponding author. Address: Centre for Development Research (ZEF), Bonn University, Germany.
E-mail address: barnyy2002@yahoo.co.uk (B.A. Amisigo).

1474-7065/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pce.2007.04.019

Please cite this article in press as: Amisigo, B.A. et al., Monthly streamflow prediction in the Volta Basin of West Africa: ..., J. Phys. Chem. Earth (2007), doi:10.1016/j.pce.2007.04.019

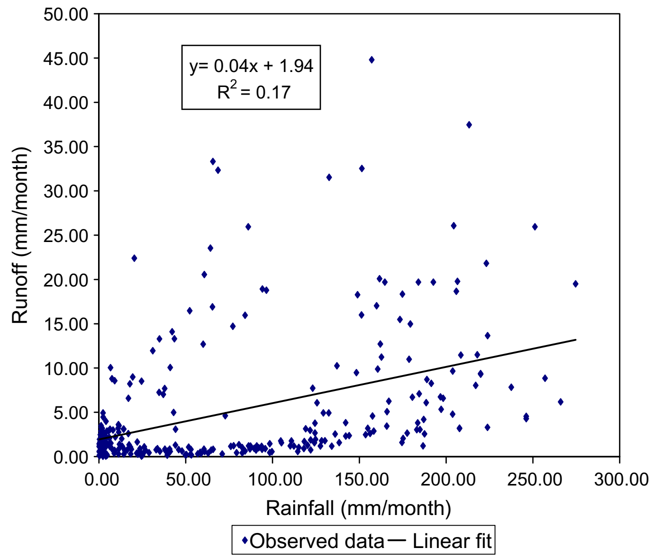

Fig. 1. Catchment runoff vs. rainfall for Bamboi on the Black Volta River.

order to provide adequate information for the water resources management of the basin.

The rainfall-runoff relationship in the Volta Basin has been recognised as highly non-linear (Andreini et al., 2000). This high non-linearity is clearly illustrated in Fig. 1 for one of the gauging stations used in this study. The figure is a plot of monthly catchment runoff vs rainfall and shows the high non-linearity in the relationship, at least at the monthly scale, with a very poor linear fit R-squared value of 0.17. Therefore, modelling the monthly rainfall-runoff system in this basin is a non-linear estimation problem.

2. The NARMAX polynomial model and system identification

A non-linear input–output model widely used in system engineering and found suitable for systems identification of a wide variety of non-linear systems including environmental systems is the NARMAX (Non-linear Autoregressive Moving Average with eXogenous Input) polynomial model (Billings and Leontaritis, 1982; Leontaritis and Billings, 1985b; Chen and Billings, 1989). This model is usually expressed as a non-linear polynomial function expansion of lagged input, output and noise terms, and, for the single-input-single-output (SISO) model, is represented as:

$$y(t) = f^d\begin{pmatrix} y(t-1), y(t-2), \ldots, y(t-n_y), u(t-n_k), \ldots, u(t-n_k-n_u), \\ e(t-1), \ldots, e(t-n_e) \end{pmatrix} + e(t) \tag{1}$$

where f^d is the polynomial of degree $d(d > 1)$; $y(t), u(t), e(t)$ are the output, input and white noise signals respectively at time t; n_y, n_u, n_e are the maximum output, input and noise lags, respectively; $n_{k(>0)}$ is the input signal time delay (measured in sampling intervals)

In non-linear systems identification in general, n_k is usually taken as at least 1. However, since in this study $u(t)$ is monthly rainfall, $n_k = 0$, i.e. $y(t)$ also a function of current input, would be considered. Thus for $n_y = n_u = n_e = 1, d = 2$ and $n_k = 0$, for example, the polynomial expansion in (1) for $y(t)$ is:

$$y(t) = \begin{pmatrix} \theta_1 + \theta_2 y(t-1) + \theta_3 u(t) + \theta_4 u(t-1) + \\ \theta_5 y(t-1)^2 + \theta_6 u(t)^2 + \theta_7 u(t)u(t-1) + \theta_8 u(t-1)^2 + \\ \theta_9 y(t-1)u(t) + \theta_{10} y(t-1)u(t-1) + \\ \theta_{11} y(t-1)e(t-1) + \theta_{12} u(t)e(t-1) + \theta_{13} u(t-1)e(t-1) + \\ \theta_{14} e(t-1) + \theta_{15} e(t-1)^2 \end{pmatrix} + e(t) \tag{2}$$

where $\theta_1, \theta_2, \ldots, \theta_{15}$ are parameters of the model.

A non-linear polynomial model of the form of (1), but without the noise terms, i.e,

$$y(t) = f^d[y(t-1), y(t-2), \ldots, y(t-n_y), u(t-n_k), \ldots, u(t-n_k-n_u)] + e(t) \tag{3}$$

is the NARX (Non-linear Autoregressive with eXogenous Input) model. This model is also general and can describe any non-linear system well (Stenman, 2002). In addition, it is not recursive as the regressors are independent of previous model outputs whereas in the NARMAX representation the noise terms can only be derived from previous model outputs. It is therefore more convenient to work with. However, the absence of a noise model in the structure means that a large number of regressor terms have to be included in order for the model to adequately represent both the system and noise dynamics (Stenman, 2002). Due to this limitation and also to avoid bias in the estimated parameters (Chiras et al., 2000), the full NARMAX model is preferred. However, to simplify the model selection process, the NARX part of the NARMAX model is fitted first in this study and then the noise terms fitted afterwards to obtain the full model.

2.1. Formulation of the model

For a given input–output series and any set of n_y, n_u, n_e, n_k and d, the polynomial represented in (1) above can generally be expressed as:

$$y(t) = \sum_{m=1}^{np} P_m \theta_m + e(t) \tag{4a}$$

where np is the number of terms in the polynomial expansion, P_m is the mth regressor term with $P_1 = 1$, and θ_m is the regression parameter for term m. The regressor terms are formed, as in (2), by various combinations of lagged values of the output and noise terms and both lagged and current (when $u_k = 0$) values of the input term.

In matrix form (4a) becomes:

$$y = P\theta + \varepsilon \tag{4b}$$

Here P is n by np and y n by 1 regressor and output matrices, respectively, Θ is np by 1 and ε is n by 1 parameter and white noise vectors, respectively, with n being the number of samples in the input–output series.

Please cite this article in press as: Amisigo, B.A. et al., Monthly streamflow prediction in the Volta Basin of West Africa: ..., J. Phys. Chem. Earth (2007), doi:10.1016/j.pce.2007.04.019

Eq. (4) is linear in the parameters and so its parameters can be estimated by the use of well-established parameter estimation techniques developed for linear systems identification, such as orthogonal least squares methods, although recursion is required to estimate the noise terms.

The system identification problem in modelling the output y using the NARMAX polynomial representation then consists of:

(i) Determination of the input, output and noise lags and input time delay (i.e., model order).
(ii) Selection of the polynomial degree.
(iii) Estimation of the parameters of the model represented as in (4) 0.

Procedures in (i) and (ii) involve model structure selection while procedure (iii) is the parameter estimation.

2.2. Parsimonious model selection and error reduction ratio (ERR) algorithm

A major difficulty in systems identification using NARMAX models is obtaining a model that is parsimonious in the number of parameters and represents the dynamics of the system adequately. This is because of the enormous number of parameters that is often involved in such models. For example, a model with $n_y = n_u = n_e = d = 3$ and $nk = 0$ results in 286 parameters. To assist in the selection of the most significant terms to consider in a NARMAX model, the error reduction ratio (ERR) algorithm (Billings and Leontaritis, 1982; Chen et al., 1989; Korenberg et al., 1988; Billings and Jones, 1992), which is derived from the orthogonal least square algorithm used for solving equations such as (4), is often used. The ERR is used to order regressors according to the levels to which each regressor reduces the mean square model error (MSE). The regressor with the largest reduction in the MSE is ranked first and is the first to be considered in a forward regression solution to Eq. (4). The strength of the procedure lies in the fact that it does not require the estimation of the full model in order to rank the regressors. The number of terms to include in the final model is determined by the application of information criteria (Billings and Leontaritis, 1982; Chiras et al., 2000) such as final prediction error, FPE (Akaike, 1974), Akaike information criterion, AIC (Akaike, 1974) and Bayesian information criterion, BIC (Kashyap, 1977) on validation series. In this study, the AIC is used. The goodness-of-fit of the model to the data is usually determined from measures such as the coefficient of determination, R-squared, the F statistic and the Nash-Sutcliffe Efficiency criterion, NSE (Nash and Sutcliffe, 1970).

2.2.1. Derivation of the error reduction ratio (ERR) algorithm

The QR decomposition method can be used in the formulation of the ERR using the QR function available in Matlab (Mathworks, 2002). A QR orthogonal triangular decomposition can be performed on the n by np matrix P in (4b) to obtain

$$P = QR \tag{5}$$

where Q is an n by np orthogonal matrix such that $Q^TQ = I$, the np by np identity matrix, and R an np by np upper triangular matrix.

Let

$$g = Q^T y \tag{6}$$

an np by 1 vector, so that for the np by 1 parameter vector, Θ,

$$R\Theta = g \tag{7}$$

Then

$$y = P\Theta + \varepsilon = (PR^{-1})(R\Theta) + \varepsilon = Qg + \varepsilon \tag{8}$$

Thus, the sum of squares of the observed output samples is

$$y^T y = \sum_{i=1}^{np} g_i^2 (q_i^T q_i) + \varepsilon^T \varepsilon = \sum_{1}^{np} g_i^2 + \varepsilon^T \varepsilon \tag{9}$$

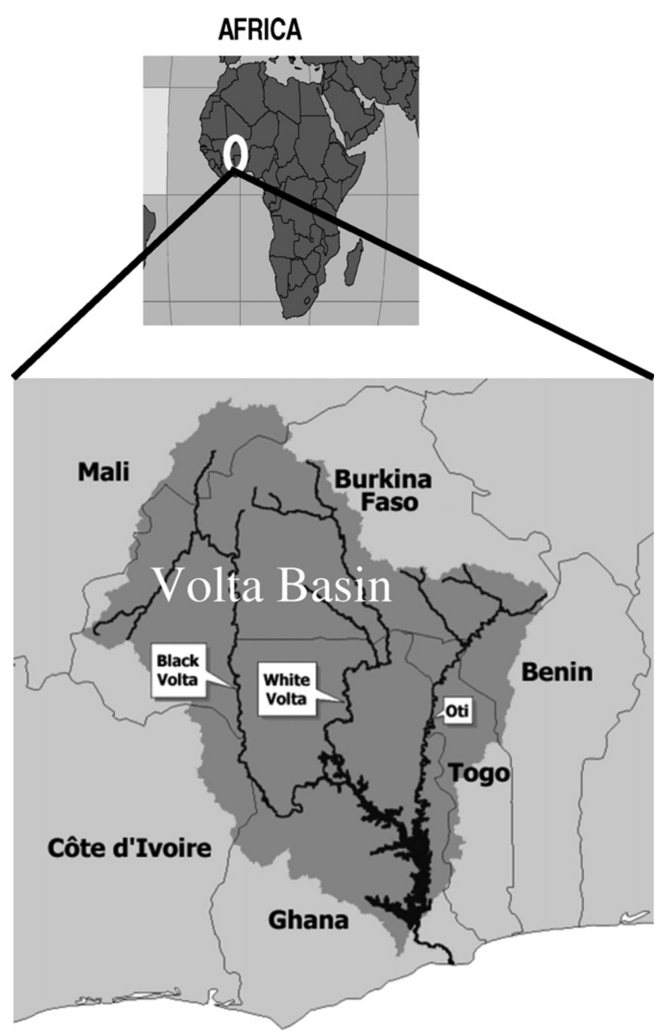

Fig. 2. Map of the Volta Basin showing the three main river systems.

where g_i is element i of vector g, q_i column i of matrix Q and the orthogonality of Q and the mean of $\varepsilon = 0$ hold.

The error reduction ratio ERR_i due to the ith regressor term is defined as the proportion of the observed output variance explained by that term (Chen et al., 1989), i.e.:

$$\text{ERR}_i = \frac{g_i^2}{y^T y} \tag{10}$$

Eq. (10) is thus used to order the regressors in each model according to their ERR_i values and the forward regression procedure used to select significant regressors and estimate parameter values.

In this study, rainfall-runoff modelling of monthly river flows of selected catchments in the Volta Basin of West Africa (Fig. 2) was undertaken using NARMAX polynomial models. The study was carried out as part of a more comprehensive riverflow modelling in the basin and was used as a first step to understanding the nature of the rainfall-runoff non-linearity in the basin.

3. The river catchments and application of the model

The river catchments used in this study are all located in the Volta Basin of West Africa. Selected characteristics of these stations are indicated in Table 1. The flow periods shown in the table are those for which runoff data without gaps are available. The 400,000 km^2 Volta Basin consists of three main sub-basins – the White Volta, Black Volta and Oti sub-basins – and occupies various areas of the six West African states of Mali, Burkina Faso, Benin, Togo, La Cote d'Ivoire and Ghana (Fig. 2). Mean annual rainfall ranges from about 150–500 mm in the extreme northern parts of the basin to about 1500–2000 mm in the south (MWH, 1998a). Annually, the basin contributes a total of 2.85×10^{10} m^3 (MWH, 1998b) to Lake Volta, the largest water body in the basin that is the source of the hydroelectric power that provides much of Ghana's electrical energy.

Monthly rainfall(u)–runoff(y) series for the selected river catchments in the basin were used to fit and evaluate SISO NARMAX polynomial models of various structures. A zero input time delay and maximum polynomial degree of 3 (i.e., $n_k = 0, d = 3$) were used. The ERR algorithm was employed to order the regressors involved in each case and the most significant regressors selected by the forward regression method and the application of the AIC on independent validation series. The parameters of the model at each forward regression stage were obtained with Matlab REGRESS function and used to compute the predicted values of the validation series. The predicted validation series at each stage of the forward regression were used with model evaluation criteria to determine when to stop regressing.

The NARX part of the model was fitted first using Eq. (10) and the forward regression procedure to select the significant regressors. The equation was used again to order the noise terms from the residuals generated by the fitted NARX sub-model and the forward regression procedure used again to select the significant noise terms. The general steps adopted in the modelling procedure for each of the river catchments considered were as follows:

1. Select a model structure (i.e., pick n_a, n_b, and n_c; $n_k = 0$ and $d = 3$ for all structures in this study).
2. Form the regressors for the NARX sub-model and rank them using their ERR values.
3. Perform forward regression and select the best sub-model (with optimum number of parameters) with validation series using both the AIC and NSE criteria.
4. Form noise terms using the residuals from the model selected in 2 above and rank these using their ERR values.
5. Perform forward regression adding terms from the noise terms to the selected sub-model. Select full model (with optimum number of parameters) with validation series using both the AIC and NSE criteria.

Table 1
Selected characteristics of gauging stations used in the study (data source: Taylor, 2004)

Station	River	Co-ordinates (decimal degrees)		Drainage area (km^2)	Period of available complete runoff series
		Longitude	Latitude		
Black Volta sub-Basin					
1.Lawra	Black Volta	2.90 W	10.60 N	96,000	1951–1973
2. Dapola	Black Volta	2.90 W	10.57 N	96,437	1951–1990
3. Bui	Black Volta	2.10 W	8.20 N	111,853	1954–1971
4. Bamboi	Black Volta	1.90 W	8.15 N	134,200	1951–1975
White Volta sub-Basin					
5. Wiasi	Sissili	1.30 W	10.33 N	12,105	1962–1973
6. Yagaba	Kulpawn	1.2 W	10.10 N	9,100	1958–1972
7. Nasia	Nasia	0.75 W	10.10 N	6,070	1969–1975
8. Nabogo	Nabogo	0.80 W	9.70 N	3,040	1963–1974
Oti sub-Basin					
9. Porga	Oti	0.90 E	11.05 N	27,197	1952–1984
10. Mango	Oti	0.40 E	10.30 N	36,287	1953–1973
11. Koumangou	Koumangou	0.40 E	10.20 N	6,070	1959–1973
12. Sabari	Oti	0.20 E	9.28 N	72,775	1960–1973

Please cite this article in press as: Amisigo, B.A. et al., Monthly streamflow prediction in the Volta Basin of West Africa: ..., J. Phys. Chem. Earth (2007), doi:10.1016/j.pce.2007.04.019

6. Repeat for other model structures and select the best model(s) for the catchment from the values of the AIC and NSE of the validation series.

The following forms of the AIC (Ljung, 1999) and NSE criteria were used:

$$\mathrm{AIC} = \log(V) + \frac{2np_s}{n} \tag{11a}$$

$$\mathrm{NSE} = 100\left(1 - \frac{var(\varepsilon)}{var(y))}\right) \tag{11b}$$

where $V = \varepsilon^T\varepsilon$, np_s is the number of selected regressors and n is the sample size of the monthly runoff series (estimation or validation) being considered.

4. Results and discussion

Fig. 3 is a plot of total monthly catchment rainfall and corresponding total monthly gauging station runoff for Mango on the Oti River. The figure illustrates clearly the non-linearity in the rainfall-runoff process in the catchment. There is much more variation in the runoff than in the rainfall for the station. This high difference in the variations of the rainfall and runoff is typical of the Volta Basin as a whole (Andreini et al., 2000). The partitioning of the observed series into estimation and validation series, as indicated on the plot, was done in such a way as to obtain estimation series long enough to be able to account for the variations in the validation series as much as possible.

Tables 2 and 3 summarize the results of the model selection process for Bamboi on the Black Volta River and Mango on the Oti River for various model structures considered in the study. Similar model selection procedures were undertaken for the rest of the stations studied. The tables indicate that of the 23 model structures considered for each station only 8 of them in each case included significant terms from the noise model from the selection process. This suggests that for the remaining models enough NARX terms were fitted to account for both the system dynamics and the noise input. Considering both the NSE and AIC criteria for the validation series (columns 6, 10, 12 and 16), Table 2 shows that only 3 of the model structures for Bamboi are important. These are models 1 ([$na\,nb\,nc$] = [1 1 1]), 5 ([$na\,nb\,nc$] = [1 3 3]) and 7 ([$na\,nb\,nc$] = [1 4 2]). However, considering parameter parsimony, model 1 is the best with its 13 parameters (column 13). Models 5 and 7 with 22 and 23 parameters, respectively, are likely to be overparameterized.

A similar analysis of Table 3 shows that 5 good models can be identified, i.e., models 3 ([$na\,nb\,nc$] = [1 3 1]), 9 ([$na\,nb\,nc$] = [1 5 1]), 10 ([$na\,nb\,nc$] = [1 5 2]), 11 ([$na\,nb\,nc$] = [1 5 3]) and 23 ([$na\,nb\,nc$] = [5 1 1]). Here, models 9 and 11 are superior considering their lower AIC values (column 16). However, model 9, with 3 parameters less than and with an F value for the estimation series (the fitted series) higher than that of 11 (indicating the fit to 9 is more significant than for 11), should be considered the better of the two.

Fig. 4a is a plot of AIC and NSE values against number of parameters included in the best NARMAX model ([$na\,nb\,nc$] = [1 5 1]) for Mango on the Oti River. Plots like this were used to select the best (optimum) model in each model class (of the same structure) such as listed in Tables 2 and 3. For each class, the best model was the one with minimum AIC for the validation series. The NSE value shown in the figure is the value for the validation series at the optimum number of parameters. The plot shows that there is a continuous fall in the AIC and rise in NSE with increasing number of parameters for the estimation series. Since this series is the fitting series and is known to the model, an increase in the number of parameters always leads to a better fit to the series. Not so for the validation series. The model does not know the output component of this series and so cannot adjust to fit it. Thus, insignificant parameters soon show as an increase in the AIC. Therefore, the application of the AIC on validation data can be very effective in determining when to stop adding parameters to a model. However, the use of this criterion alone would not always result in a parsimonious model, as clearly shown in Tables 2 and 3, where the number of parameters for some of the models is rather very high. Other goodness-of-fit criteria, such as, as the F statistic for the estimation series, need also be considered to arrive at an optimum model.

Visual evaluation of the selected model in each class was made through plots such as the one shown in Fig. 4b for the best models for Mango. The figure shows a plot of observed and predicted total monthly runoff for the NARMAX formulation. The plot and the validation NSE

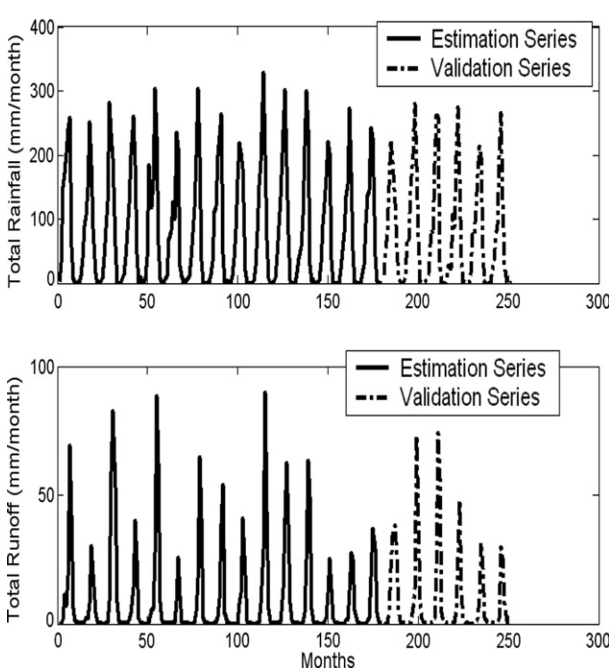

Fig. 3. Observed monthly rainfall and runoff series at Mango on the Oti River.

Table 2
Model structure selection layout for Bamboi on the Black Volta River (shaded rows highlight best model structures)

				NARX sub-model						Full NARMAX model					
Model no.	*na*	*nb*	*nc*	NSE (%)		Opt. np	*R*-squared	*F*	AIC (Validation)	NSE %		Opt. np	*R*-squared	*F*	AIC (Validation)
				Estimation series	Validation series		Estimation series			Estimation series	Validation series		Estimation series		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
1	1	1	1	94.34	88.96	12	0.94	235.63	6.54	94.44	89.93	13	0.94	218.67	6.46
2	1	2	1	95.14	89.26	21	0.95	143.04	6.64	–	–	–	–	–	–
3	1	3	1	94.33	88.01	16	0.94	167.35	6.67	95.45	88.89	18	0.96	183.46	6.63
4	1	3	2	″	″	″	″	″	″	95.45	88.89	18	0.96	183.46	6.63
5	1	3	3							95.70	91.05	22	0.96	153.49	6.47
6	1	4	1	94.42	89.42	19	0.95	138.47	6.59	95.08	90.41	20	0.95	148.98	6.51
7	1	4	2							95.35	91.31	23	0.95	133.60	6.45
8	1	4	3	″	″	″	″	″	″	95.08	90.41	20	0.95	148.98	6.51
9	1	5	1	93.38	86.46	18	0.94	123.12	6.81	–	–	–	–	–	–
10	1	5	2	″	″	″	″	″	″	–	–	–	–	–	–
11	1	5	3	″	″	″	″	″	″	–	–	–	–	–	–
12	1	6	1	92.63	88.11	15	0.93	136.00	6.63	–	–	–	–	–	–
13	2	1	1	95.05	89.03	17	0.95	180.09	6.60	–	–	–	–	–	–
14	2	1	2	″	″	″	″	″	″	–	–	–			
15	2	2	1	94.88	87.89	17	0.95	173.70	6.70	94.96	88.81	18	0.95	165.14	6.64
16	2	2	2	″	″	″	″	″	″	–	–	–	–	–	–
17	2	3	1	93.62	87.27	17	0.94	137.34	6.75	–	–	–	–	–	–
18	2	4	1	93.73	88.57	19	0.94	122.50	6.67	–	–	–	–	–	–
19	3	1	1	94.90	88.33	17	0.95	174.19	6.66	–	–	–	–	–	–
20	3	2	1	94.70	87.67	18	0.95	156.26	6.73	–	–	–	–	–	–
21	3	3	1	93.66	86.70	17	0.94	138.28	6.79	–	–	–	–	–	–
22	4	1	1	94.70	88.37	17	0.95	167.02	6.65	–	–	–	–	–	–
23	5	1	1	93.60	86.31	15	0.94	158.20	6.78	–	–	–	–	–	–

Please cite this article in press as: Amisigo, B.A. et al., Monthly streamflow prediction in the Volta Basin of West Africa: ..., J. Phys. Chem. Earth (2007), doi:10.1016/j.pce.2007.04.019

Table 3
Model structure selection layout for Mango on the Oti River (shaded rows highlight best model structures)

				NARX						NARMAX					
Model no.	*na*	*nb*	*nc*	NSE (%)		Opt. np	*R*-squared	*F*	AIC (Validation)	NSE (%)		Opt. np	*R*-squared	*F*	AIC (Validation)
				Estimation series	Validation series		Estimation series			Estimation series	Validation series		Estimation series		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
1	1	1	1	94.12	87.18	9	0.94	341.63	8.11	–	–	–	–	–	–
2	1	2	1	94.10	86.68	12	0.94	242.74	8.22	94.12	87.66	13	0.94	221.98	8.18
3	1	3	1	93.48	89.34	14	0.94	182.04	8.05	93.50	90.13	15	0.94	168.74	8.00
4	1	3	2	″	″	″	″	″	″	–	–	–	–	–	–
5	1	3	3	″	″	″	″	″	″	–	–	–	–	–	–
6	1	4	1	93.26	88.69	12	0.93	210.20	8.04	–	–	–	–	–	–
7	1	4	2	″	″	″	″	″	″	–	–	–	–	–	–
8	1	4	3	″	″	″	″	″	″	–	–	–	–	–	–
9	1	5	1	93.60	88.86	12	0.94	222.43	8.02	93.70	91.22	13	0.94	206.06	7.81
10	1	5	2	″	″	″	″	″	″	94.34	90.02	14	0.94	211.70	7.97
11	1	5	3	″	″	″	″	″	″	94.46	92.28	16	0.95	185.42	7.77
12	1	6	1	93.66	89.62	12	0.94	224.08	7.94	–	–	–	–	–	–
13	2	1	1	91.26	85.30	10	0.91	196.56	8.27	91.51	86.34	11	0.92	181.62	8.22
14	2	1	2	″	″	″	″	″	″	91.51	86.34	11	1.92	181.62	8.22
15	2	2	1	93.41	84.96	13	0.93	196.52	8.37	–	–	–	–	–	–
16	2	2	2	″	″	″	″	″	″	–	–	–	–	–	–
17	2	3	1	93.51	87.67	15	0.94	168.81	8.22	–	–	–	–	–	–
18	2	4	1	93.66	89.18	17	0.94	149.56	8.14						
19	3	1	1	94.24	86.48	11	0.94	274.95	8.20	–	–	–	–	–	–
20	3	2	1	95.01	86.35	19	0.95	169.18	8.44	95.13	87.09	20	0.95	163.37	
21	3	3	1	90.98	84.29	9	0.91	214.70	8.29	–	–	–	–	–	–
22	4	1	1	94.97	87.90	14	0.95	239.86	8.17	–	–	–	–	–	–
23	5	1	1	95.35	90.48	17	0.95	207.68	8.01	–	–	–	–	–	–

Please cite this article in press as: Amisigo, B.A. et al., Monthly streamflow prediction in the Volta Basin of West Africa: ..., J. Phys. Chem. Earth (2007), doi:10.1016/j.pce.2007.04.019

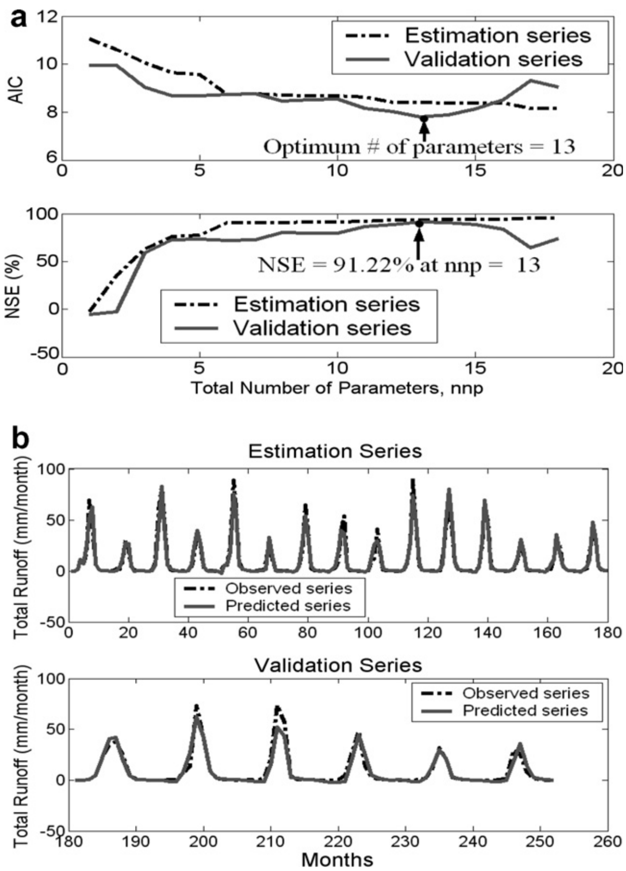

Fig. 4. AIC, NSE, observed and predicted runoff plots for Mango on the Oti River – $[nanbnc] = [1\,5\,1]$ AIC and NSE vs. number of NARMAX parameters (nnp) Observed and predicted monthly runoff for both estimation and validation series.

shown in Fig. 4a show very good fit and predictions from the model. The number of fitted parameters, the estimation and validation NSEs for the best models for all the stations used in this study are presented in Table 4. The validation NSEs shown in the table indicate fairly good fits and predictions from the selected models.

In Table 5, the selected regressors and the full regressor set for both the NARX and NARMAX formulations are presented for the best model for Bamboi on the Black Volta River. The rankings of the regressors as indicated in the table are obtained from the application of the ERR algorithm. The NSE values for the validation series as each regressor term is included in the model up to one term after the optimum number of terms are also indicated in the table. It is important to note that while the full regressor terms number 35, only 13 were found to be significant for the full NARMAX case for this model structure.

5. Conclusions

A SISO NARMAX polynomial model was successfully formulated and applied to monthly rainfall-runoff series for runoff prediction at selected river gauging stations in the Volta Basin of West Africa. Several model structures for each station were evaluated using the AIC, NSE (both for the validation series), R-squared, and F (for the estimation series) goodness-of-fit criteria. The combination of these criteria enabled the identification of the ''best'' model from all the models tested for each river catchment. The AIC criterion was particularly useful in selecting the optimum models from the several available for each model structure. However, the other goodness-of-fit criteria were required to successfully identify parsimonious models in some of the cases.

Monthly runoff predictions from the selected models were very good, and the polynomial models appeared to have captured a good part of the rainfall-runoff non-linearity, even though some peak and/or low flows were not adequately simulated in the cases investigated. The use of multiple goodness-of-fit criteria in which different criteria are employed for fitting simulations to different parts of

Table 4
Number of parameters and the NSE values for the best NARMAX models for all river catchments used in the study

Station	River	Number of parameters in best model	NSE (%)	
			Estimation series	Validation series
Black Volta Basin				
Lawra	Black Volta	9	87.5	88.4
Dapola	Black Volta	14	88.8	68.2
Bui	Black Volta	10	95.2	91.3
Bamboi	Black Volta	13	94.4	89.9
White Volta Basin				
Wiasi	Sisilli	11	87.9	84.9
Yagaba	Kulpawn	13	86.9	72.4
Nasia	Nasia	10	94.3	69.6
Nabogo	Nabogo	10	80.1	70.0
Oti Basin				
Porga	Oti	11	93. 8	82.4
Mango	Oti	13	93.7	91.2
Koumangou	Koumangou	13	88.4	87.1
Sabari	Oti	14	93.0	92.1

Table 5
Regressors for model $[na\,nb\,nc] = [1\,1\,1]$ for Bamboi on the Black Volta

Optimum regressor terms			Validation series NSE (%)*	Full regressor terms		
Regressor #	Regressor rank	Regressor (P)		Regressor #	Regressor rank	Regressor (P)
NARX sub-model				NARX sub-model terms		
1	1	1	−5.35	1	1	1
2	2	y_{t-1}	−5.36	2	2	y_{t-1}
3	3	u_t	68.71	3	3	u_t
4	4	u_{t-1}	67.25	4	4	u_{t-1}
5	5	$(y_{t-1})^2$	67.56	5	5	$(y_{t-1})^2$
6	6	$(y_{t-1})^*\ u_t$	79.85	6	6	$(y_{t-1})^*\ u_t$
7	7	$(y_{t-1})^*(u_{t-1})$	80.39	7	7	$(y_{t-1})^*(u_{t-1})$
9	8	$u_t^*(u_{t-1})$	86.55	9	8	$u_t^*(u_{t-1})$
10	9	$(u_{t-1})^2$	87.72	10	9	$(u_{t-1})^2$
8	10	u_t^2	88.42	8	10	u_t^2
13	11	$(y_{t-1})^{2*}(u_{t-1})$	88.42	13	11	$(y_{t-1})^{2*}(u_{t-1})$
14	12	$(y_{t-1})^* u_t^2$	**88.96**	14	12	$(y_{t-1})^*\ u_t^2$
	13		88.82	15	13	$(y_{t-1})^* u_t^*(u_{t-1})$
NARMAX model				16	14	$(y_{t-1})^*(u_{t-1})$^2
1	1	1	−5.35	11	15	(y_{t-1})^3
2	2	y_{t-1}	−5.36	12	16	$(y_{t-1})^{2*}u_t$
3	3	u_t	68.71	17	17	ut^3
4	4	u_{t-1}	67.25	18	18	$ut^{2*}(u_{t-1})$
5	5	$(y_{t-1})^2$	67.56	19	19	$u_t^*(u_{t-1})$^2
6	6	$(y_{t-1})^*\ u_t$	79.85	20	20	(u_{t-1})^3
7	7	$(y_{t-1})^*(u_{t-1})$	80.39			
9	8	$u_t^*(u_{t-1})$	86.55	Noise sub-model terms		
10	9	$(u_{t-1})^2$	87.72	29	1	$(e_{t-1})^*\ u_t^2$
8	10	u_t^2	88.42	25	2	$(e_{t-1})^2$
13	11	$(y_{t-1})^{2*}(u_{t-1})$	88.42	33	3	$(e_{t-1})^{2*}u_t$
14	12	$(y_{t-1})^*\ u_t^2$	88.96	34	4	$(e_{t-1})^{2*}(u_{t-1})$
29	13	$(e_{t-1})^* u_t^2$	**89.93**	26	5	$(e_{t-1})^*(y_{t-1})^2$
	14		87.51	30	6	$(e_{t-1})^*\ u_t^*(u_{t-1})$
				31	7	$(e_{t-1})^*(u_{t-1})^2$
				32	8	$(e_{t-1})^{2*}(y_{t-1})$
				35	9	$(e_{t-1})^3$
				23	10	$(e_{t-1})^* u_t$
				27	11	$(e_{t-1})^*(y_{t-1})^* u_t$
				21	12	e_{t-1}
				22	13	$(e_{t-1})^*(y_{t-1})$
				24	14	$(e_{t-1})^*(u_{t-1})$
				28	15	$(e_{t-1})^*(y_{t-1})^*(u_{t-1})$

*The NSE value refers to the value for the model with number of regressors up to the current regressor.

the hydrograph, especially the peak and low flow sections, could result in improvements in the predicted hydrographs.

Nevertheless, the results show the appropriateness of non-linear representation of the rainfall-flow process. They also indicate that the NARMAX modelling framework is suitable for monthly river runoff prediction in the Volta Basin. The drawback of the method, as applied here, is its inability to provide physically interpretable results. However, the several good models made available by the NARMAX modelling framework may be useful in the selection of model structures that also provide insight into the physical behavior of the catchment rainfall-runoff system.

Acknowledgements

Support of the GLOWA Volta Project, under Grant 07GWK01 provided by the German Federal Ministry of Education and Research and the State of North-Rhine Westphalia is gratefully acknowledged. We are also very grateful to the two anonymous reviewers whose constructive criticisms and useful suggestions have contributed to improving the quality of the paper.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automatic. Control 19 (6), 716–723.

Andreini, M., van de Giesen, N.C., van Edig, A., Fosu, M., Andah, W. 2000. Volta Basin Water Valance; ZEF-Discussion Papers on Development Policy, Number 21. Bonn, Germany. Center for Development Research.

Billings, S.A., Leontaritis, I.J., 1982. Parameter estimation techniques for non-linear systems. In: 6th IFAC Symposium on System Identification and Parameter Estimation, Washington D.C., USA, pp. 505–510.

Billings, S.A., Jones, G.N., 1992. Orthogonal least-squares parameter estimation algorithms for non-linear stochastic systems. Int. J. Syst. Sci. 23 (7), 1019–1032.

Chen, S., Billings, S.A., 1989. Representations of non-linear systems: the NARMAX model. Int. J. Control 49 (3), 1013–1032.

Chen, S., Billings, S.A., Luo, W., 1989. Orthogonal least squares methods and their application to non-linear system identification. Int. J. Control 50 (5), 1873–1896.

Chiras, N., Evans, C., Rees, D., 2000. Non-linear gas turbine computer modeling using NARMAX structures, In: Proceedings of the IEEE Instrumentation and Measurement Technology Conference IMTC2000, Baltimore, vol. 3, pp. 1278–1284.

Kashyap, R.L., 1977. A bayesian comparison of different classes of dynamical models using empirical data. IEEE Trans. Automatic Control 22 (5), 715–727.

Korenberg, M.J., Billings, S.A., Liu, Y.P., Mollroy, P.J., 1988. Orthogonal parameter estimation algorithm for non-linear stochastic systems. Int. J. Control 48 (1), 193–210.

Leontaritis, I.J., Billings, S.A., 1985b. Input–output parametric models for non-linear systems. Part II: stochastic non-linear systems. Int. J. Control 41 (2), 329–344.

Ljung, Lennart, 1999. System Identification: Theory for the User, second ed. Prentice-Hall, Upper Saddle River, New Jersey, ISBN 0-13-656695-2.

Mathworks (2002). Matlab Release 13. The MathWorks, Inc., Natick, MA 01760-2098.

MWH 1998a. Ghana water resources management study – International waters ‘‘building block’’ study. Main report (final). Ministry of Works and Housing, Accra, Ghana.

MWH, 1998b. Ghana water resources management study – Information ‘‘building block’’ study. Part III. Executive summary. Vol. 1: Summary of information in the three basin systems (final report). Ministry of Works and Housing, Accra, Ghana.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. J. Hydrol. 10, 282–290.

Stenman, A., 2002. On Model Structure Selection for Nonparametric Prediction Methods. http://www.control.isy.liu.se.

Tabrizi, M.H.N., Said, S.E., Badr, A.W., Mashor, Y., Billings, S.A., 1998. Non-linear modeling and prediction of a river flow system. Int. J. AWRA 34 (8), 1333–1339.

Taylor, J., 2004. Watershed management in the Volta River Basin, West Africa: Providing information through hydrological modelling. M.Sc. Thesis, Cornell University.

Please cite this article in press as: Amisigo, B.A. et al., Monthly streamflow prediction in the Volta Basin of West Africa: ..., J. Phys. Chem. Earth (2007), doi:10.1016/j.pce.2007.04.019