ISBN 978-93-84422-69-1
Proceedings of 2016 International Conference on Biological and Environmental Science (BIOES-16)
New Jersey (Near NEW YORK), USA, June 8-9, 2016 pp.1-7

Non Linear Bilinear Type Models to Forecast Critical Dry Spell Lengths in Anamaduwa, Sri Lanka

S C Mathugama[1] and T S G Peiris[2]

[1]Institute of Technology, University of Moratuwa

[2]Department of Mathematics, Faculty of Engineering, University of Moratuwa

Abstract: *Unexpected longer dry spells are a recurrent feature of the climate of Sri Lanka. Such events caused many negative impacts to the economic status of the country and in particular for agriculture sector. For achieving maximum benefits from dry land agriculture the knowledge of distribution of dry spells (length and start date) within a year is useful. This information can be used for various decisions making in diverse fields. A major challenge of drought research is therefore to develop suitable methods for forecasting the onset and termination points of droughts and length of dry spell [4],[5]..In this study a novel type of model is developed to predict the length of critical dry spells in Anamaduwa, Sri Lanka using critical dry spell series during1950 to 2005. Different types of non linear bilinear type (NBLX) models with exogenous variables were tried and found that the best fitted model to the critical dry spell series to Anamaduwa is NBLX(2,2,0,1,1,2). The both statistical and non statistical validity of the model were checked using various diagnostics statistics and thus fitted model can be recommended to predict critical dry spell length in Anamaduwa. The best exogenous variable was obtained based on the dry spell series. Thus it also recommended to explore the possibility of such nonlinear bilinear type model (NBLX) to predict the length of critical dry spells in other parts of Sri Lanka as well.*

Keywords: Bilinear Models, Dry Spells, Non Linear Bilinear Rainfall

1. Introduction

Importance of the prior knowledge of the length and time of occurrence of dry spells for efficient planning has been highlighted by many authors ([4],[6]. A major challenge of climate research is to develop suitable methods for forecasting the onset and termination points and length of droughts. Furthermore, they have highlighted that less attention has been given to forecast dry spells in climate change studies, particularly in Sri Lanka. When the dry spell length is known in advance, it is helpful for farmers to select a particular crop or variety and they can select breeding varieties of various maturity durations. Similarly irrigation engineers can make their decisions on irrigation water demand.

Anamaduwa (7.52N, 80.0E, and 259ft) is located in the Dry Intermediate zone in Sri Lanka. An average of four long dry spells per year was observed in Anamaduwa [4]. As the length of dry spells and numbers of dry spells have been highly varied, their research was limited to four longest dry spells within a year. The four longest dry spells in a year were named, as the four critical dry spells. Of the four critical dry spells, the one which first occurred is defined as the ‘first critical dry spell’. Similarly ‘second’, ‘third’, and ‘fourth’ critical dry spells were defined according to the order of occurrence. The four length series of the four critical dry spells were combined to make one dry spell series and forecasting techniques were explored. In a previous study by the authors [4] developed ARIMAX(2,1,1,1) model with exogenous input variable to forecast dry spell length. They have discussed the advantages and disadvantages in such models and stressed the importance of non linear models. Thus nonlinear type models with inclusion of exogenous input variable/s are investigated in this study to develop a better model to forecast the lengths of critical dry spell series in Anamaduwa.

2. Materials and Methods

2.1. Data

Fifty six years of daily rainfall (1950 – 2005) in Anamaduwa location were used. A single dry spell series was formed based on four critical dry spells as explained in [5].

2.2. Non Linear Bilinear Type Model (NBLX)

In recent years there has been a growing interest in non linear modeling of time series data by many authors ([10].[9], [7],[3] and [2]) One of these nonlinear models, proposed by [8]) is called a Bilinear model, denoted by BN(p,q,m,k) and given by equation (1).

$$Y_t + \sum_{i=1}^{p} a_i Y_{t-i} = \sum_{j=0}^{q} d_j e_{t-j} + \sum_{i=}^{m} \sum_{j=1}^{k} b_{ij} Y_{t-i} e_{t-j} \quad (1)$$

where $\{e_t\}$ is a sequence of iid random variables, p= order of MA term, q=order of AR term, m=order of bilinear term (Y) and k= order of bilinear term (e). When the last part is removed from the equation (1) with $m = k$, it becomes an autoregressive moving average model with orders p and q, that is ARMA (p, q) model.For our studyan exogenous input variable is included and thus inclusion of an exogenous input variable X and the corresponding interaction terms into model (1) becomes $BLX(p, q, r, s, k, u, v, w$ with the form of

$$Y_t = \sum_{j=0}^{q} \lambda_j e_{t-j} - \sum_{i=1}^{p} \delta_i Y_{t-i} + \sum_{i=1}^{r} \sum_{j=1}^{s} \mu_{ij} Y_{t-i} e_{t-j} + \sum_{l=1}^{k} \gamma_l X_{t-l} + \sum_{j=1}^{u} \sum_{l=1}^{v} \theta_{jl} e_{t-j} X_{t-l} + \sum_{i=1}^{w} \sum_{l=1}^{x} \eta_{il} Y_{t-i} X_{t-l} \quad (2)$$

However, as the model (2) consists of lot of parameters and makes the model more complex, the models with bilinear terms of same order (that is, $r = s, u = v \; and \; w$ were considered. Then equation (2) becomes (3) as follows.

$$Y_t = \sum_{j=0}^{q} \lambda_j e_{t-j} - \sum_{i=1}^{p} \delta_i Y_{t-i} + \sum_{i=1}^{r} \sum_{j=1}^{r} \mu_{ij} Y_{t-i} e_{t-j} + \sum_{l=1}^{k} \gamma_l X_{t+1-l} + \sum_{j=1}^{u} \sum_{l=1}^{u} \theta_{jl} e_{t-j} X_{t-l} + \sum_{i=1}^{w} \sum_{l=1}^{w} \eta_{il} Y_{t-i} X_{t-l}$$
(3)

The type of modelin the form of (3) was named as, Nonlinear Bilinear type model with input variable X and denoted as NBLX(p,r,k,u). Furthermore, models were limitedto third order, with $r = u = w$. When there were more than one input variable, (say X1 and X2), the model becomes NBLX1X2(p,q,r,k1,k2,u). The inclusion of input variables was also limited to maximum of three input variables.

2.3. Selecting Subset Models

In some cases all the terms upto full order were not significant and only a subset of terms were significant. For an example, suppose only AR (1) and AR (b) terms were significant when AR part AR (p) model was fitted. In such cases the AR order was written as which means AR(p1,b) subset terms were only fitted. Now the model has the form of NBLX($p_{1,b}$,q,r,k,u)..The same terminology applies for other terms as well.

2.4. Identification of Exogeneous Variables

Identifying a suitable exogenous input variable is a difficult task. However, the possibility of using average length of previous dry spell as covariate/s can be considered. Thus three customer specific exogenous variables were defined and denoted by CDM1, CDM2 and CDM3 respectively as described below.

2.4.1. Customer specific conditional mean 1 (CDM1)

Conditional mean 1 (CDM1) for first, second, third and fourth dry spell length series is considered as the average length of critical dry spelled occurred previously. If the corresponding lengths of the four critical dry spells in a year are denoted by LDS_1, LDS_2, LDS_3 and LDS_4 separately then the first covariate (CDM1) for the year (1950 + n), where n = 1,2…......55are defined as:

$$\mathrm{CDM1\_i}(1950+n) = \frac{\mathrm{LCDS_i}(1950) + \dots\dots\dots + \mathrm{LCDS_i}(1950+n-1)}{n}, \quad i = 1,2,3,4.$$

2.4.2. Customer Specific Conditional Mean 2 (CDM2)

Conditional mean 2 for first, second, third and fourth dry spell length series is considered as the average of four previously occurred dry spells lengths irrespective of the order. Thus for a given year (say, 195) CDM2_1 to CDM2_4 can be denoted as the average of four previously occurred dry spell lengths.

$$\mathrm{CDM2\_1}(1951) = \frac{\mathrm{LCDS_1}(1950) + \dots\dots\dots + \mathrm{LCDS_4}(1950)}{4}$$

$$\mathrm{CDM2\_2}(1951) = \frac{\mathrm{LCDS_2}(1950) + \dots\dots\dots + \mathrm{LCDS_1}(1951)}{4}$$

$$\mathrm{CDM2\_3}(1951) = \frac{\mathrm{LCDS_3}(1950) + \dots\dots\dots + \mathrm{LCDS_2}(1951)}{4}$$

$$\mathrm{CDM2\_4}(1951) = \frac{\mathrm{LCDS_4}(1950) + \dots\dots\dots + \mathrm{LCDS_3}(1951)}{4}$$

2.4.3. Customer specific conditional mean 3 (CDM3)

Third conditional mean was calculated considering the month of starting date of the critical dry spell. A sample calculation is shown in Table 1.

The length of dry spell series and starting date (month) of the corresponding dry spells are tabulated as shown above. If the dry spell occurred in a particular month, mean length of dry spells previously occurred in the same month was considered as the conditional mean 3. For example CDM3 of L6 was L5 because both L5 and L6 occurred in month 5 and the only previously occurred spell in month 5 is L5. Similarly L9 was also occurred in month 5, then the conditional mean of L9 was calculated as (L6+L5)/2 where this is the average of lengths occurred in month 5 prior to L9. Since L12 occurred in month 12 and no prior spell occurred in the same month length occurred lastly in previous month (i.e month 11) was considered as the conditional mean. This method of calculating CDM3 has not been used in any previous study. In Table 1, the continuous CDM3 series started at L8 and CDM3 values prior to L8 has to be neglected.

TABLE I: Calculation of CDM3

Critical Dry Spell			CDM3
Length	Starting Date	Month	
L1	9	1	
L2	119	4	
L3	175	6	
L4	243	8	
L5	121	5	
L6	151	5	L5
L7	205	7	
L8	324	11	
L9	114	5	(L5+L6)/2
L10	163	6	L3
L11	308	11	L8
L12	362	12	L11
L13	23	1	L1
L14	116	4	L2
L15	158	6	(L10+L3)/2

2.5. Model Fitting Procedure

- First, fit a higher order AR model, say AR(7) .

- Then select significant AR terms at 5% significance level and determine the order of autoregressive terms p.
- Adjust model for MA terms by adding MA terms one by one up to maximum order MA (7) and select significant MA terms at 5% significance level. Determine the order of moving average terms q.
- Add bilinear terms $Y_{t-i}e_{t-j}$ for $i = j$ where $i_{max} = j_{max} = 3$.
- Include significant input variable X_t and significant terms of X_{t-l} with $l = 1, ..., k$.
- Add bilinear terms $X_{t-l}e_{t-j}$ for $l = j$ where $l_{max} = j_{max} = 3$. If single terms with $l = j$ =2 (X_{t-2} and e_{t-2}) are significant then first add the bilinear term $X_{t-2}e_{t-2}$ and test for the significance of the term, then adjust the model for other bilinear terms $X_{t-1}e_{t-1}$ and $X_{t-3}e_{t-3}$.
- Follow the same procedure when adding the next bilinear term $Y_{t-i}X_{t-l}$for $i = l$ where $i_{max} = l_{max} = 3$.
- Select the best set of models based on significance of parameters, minimum RMSE values and white noise residuals with homoscedasticity of residual variance.
- Among the best set of models the most suitable model was selected based on the percentage of prediction errors of the fitted models.

2.6. Model Diagnostics for Constant variance, Randomness and Normality of errors

The constant variance, normality and homoscedasticity of residuals were confirmed using White test, Breusch–Godfrey Lagrange Multiplier test and ShapiroWilk test respectively.

3. Results and Discussion

3.1. Non linear bilinear type models with input variable (NBLX) for Anamaduwa

In order to find the impact of the customer specfic variables on the length series cross-correlations were checked with the variable length and three conditional means CDM1, CDM2 and CDM3.The cross-correlations with the variable length for Anamaduwa were significant with CDM2. As the term was significant, input variable with MA and AR terms were tried. However a best model with significant parameter estimates could not be found. Cross-correlations were checked and found significant (5%) with CDM3 and the parameter estimates were found.NBLX models suggested are given in Table 2.Of these 9 models, models with parameter estimates significant at 10% were selected at the first stage of screening and details of the selected four models are presented in Table 3.

Of these four models, model (12), NBLX(2,2,0,1) has minimum root mean square error (RMSE=19.02) with white noise residuals and constant variance can be considered as the best model. Also model (8), NBLX(2,2,0,3) has slight higher RMSE than that in model (12), it also satisfies the randomness and the contant variance of the errors. Thus of the four models those two can be considered as most suiatble models. In order to decide the best one out of those two, critical dry spell lengths were predicted for the period 1950 to 1999. The predicted values were computed and the difference between observed and predicted was compared. The percentage values for the difference falling between [±0 - ±5], [±6 - ±10], [±11 - ±15], [±16 - ±20], [±21 - ±25], > +25 and< -25 days are shown in Table 3.

Among these two best models, prediction errors were low in NBLX(2,2,0,1) compared to NBLX(2,2,0,3). Approximately 70% of the predicted values were within ±15 days of the observed values. Hence the model NBLX(2,2,0,1) can be considered as the most adequate model for the lengths of critical dry spells in Anamaduwa.

4. Conclusions

As the dry spell length series is not equally spaced non linear time series techniques were used in this study. The developed non linear bilinear type time series modeles are suitablein forecastingthe length of a given critical dry spell for Anamaduwa location. The best fitted non linear bilinear type model recomended is NBLX (2,2,0,1,1,2). The length of dry spells could be used for deciding a particular crop or variety in a given location, and for breeding varieties of various maturity durations. The fitted model therefore can be effectively used for planner for various applications. Such types of models can be used to forecast critical dry spell lengths in other locations in Sri Lanka. One disadvantage of such model is lag length and identification of covariates. Thus it is suggested to improve this type of models by increasing the lag length.

TABLE II: Fitted NBLX Models for Anamaduwa

Model	
$Y_t = C + \sum_{i=1}^{2} \delta_i Y_{t-i} + \gamma_1 X_t + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j} + \mu_{22} Y_{t-2} e_{t-2}$	(4)
$Y_t = C + \sum_{i=1}^{3} \delta_i Y_{t-i} + \gamma_1 X_t + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j}$	(5)
$Y_t = C + \sum_{i=1}^{2} \delta_i Y_{t-i} + \sum_{k=1}^{3} \gamma_k X_{t+1-k} + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j}$	(6)
$Y_t = + \sum_{i=1}^{2} \delta_i Y_{t-i} + \sum_{k=1}^{3} \gamma_k X_{t+1-k} + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j}$	(7)
$Y_t = \sum_{i=1}^{2} \delta_i Y_{t-i} + \sum_{k=1}^{3} \gamma_k X_{t+1-k} + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j}$	(8)
$Y_t = \sum_{i=1}^{2} \delta_i Y_{t-i} + \sum_{k=1}^{2} \gamma_k X_{t+1-k} + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j} + \eta_{11} Y_{t-1} X_{t-1}$	(9)
$Y_t = \sum_{i=1}^{2} \delta_i Y_{t-i} + \gamma_1 X_t + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j} + \eta_{11} Y_{t-1} X_{t-1}$	(10)
$Y_t = \sum_{i=1}^{2} \delta_i Y_{t-i} + \gamma_1 X_t + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j} + \eta_{11} Y_{t-1} X_{t-1} + \theta_{11} X_{t-1} e_{t-1}$	(11)
$Y_t = \sum_{i=1}^{2} \delta_i Y_{t-i} + \gamma_1 X_t + e_t - \sum_{j=1}^{2} \lambda_j e_{t-j} + \sum_{i=1}^{2} \sum_{j=1}^{2} \eta_{ij} Y_{t-i} X_{t-j} + \theta_{11} X_{t-1} e_{t-1}$	(12)

TABLE III: Parameter Estimates of Selected Models

Model	Parameter	Estimate	SE	P value	Diagnostics
Model (8)	δ_1	1.8276	0.0470	<0.0001	RMSE=21.82
	δ_2	-0.8419	0.0464	<0.0001	WT=15.59, p=0.74
NBLX (2,2,0,3,0,0)	λ_1	2.0661	0.0376	<0.0001	(constant variance)
	λ_2	-1.0916	0.0381	<0.0001	G_1=0.00, p=0.99
	γ_1	0.9892	0.0901	<0.0001	(white noise)
	γ_2	-1.8617	0.1836	<0.0001	SW=0.90, p=0.00
	γ_3	0.8869	0.1009	<0.0001	(non normal)
Model (10)	δ_1	1.2575	0.2358	<0.0001	RMSE=22.73
	δ_2	-0.4699	0.0970	<0.0001	WT=18.35, p=0.65
NBLX (2,2,0,1,0,1)	λ_1	0.8070	0.1424	<0.0001	(constant variance)
	λ_2	-0.4750	0.1156	<0.0001	G_1=8.74, p=0.001
	γ_1	0.8037	0.1025	<0.0001	(not white noise)
	η_{11}	-0.0136	0.0033	<0.0001	SW=0.90, p=0.00
					(non normal)
Model (11)	δ_1	1.3506	0.1856	<0.0001	RMSE=22.16
	δ_2	-0.4649	0.0903	<0.0001	WT=12.42, p=0.79
	λ_1	1.6247	0.2014	<0.0001	(constant variance)
NBLX (2,2,0,1,1,1)	λ_2	-0.5013	0.1099	<0.0001	G_1=5.44, p=0.01
	γ_1	0.7794	0.1022	<0.0001	(not white noise)
	η_{11}	-0.0150	0.0026	<0.0001	SW=0.90, p=0.00
	θ_{11}	-0.0164	0.0042	<0.0001	(non normal)
Model (12)	δ_1	1.7749	0.1648	<0.0001	RMSE=19.02
	δ_2	-0.9568	0.1933	<0.0001	WT=16.2, p=0.70
NBLX (2,2,0,1,1,2)	λ_1	1.7720	0.1485	<0.0001	(constant variance)
	λ_2	-0.6871	0.0861	<0.0001	G_1=0.00, p=0.96
	γ_1	0.7397	0.0820	<0.0001	(white noise)
	η_{11}	-0.0206	0.0029	<0.0001	SW=0.90, p=0.00
	η_{22}	0.0080	0.0033	0.0159	(non normal)
	θ_{11}	-0.0156	0.0033	<0.0001	

TABLE IV: Comparison of distribution of errors (in days) between Model(8) and model (12)

Class interval of errors (in days)	**% of** points wrt to total points	
	$NBLX(2,2,0,3,0,0)$ **model (8)**	$NBLX(2,2,0,1,1,2)$ **model (12)**
[±0 - ±5]	26.5	25.6
[±6 - ±10]	14.8	23.9
[±11 - ±15]	14.8	22.1
[±16 - ±20]	9.7	7.9
[±21 - ±25]	16.2	10.1
> +25 and< -25	18.0	10.4

5. References

[1] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 4, pp. 570-578, July 1993.

http://dx.doi.org/10.1109/72.238312

[2] K.S. Chan, and H.Tong, "A note on invertibility of nonlinear ARMA models," *Journal of Statistical Planning and Inference*, vol. 140, pp. 3709-3714, 2009.

http://dx.doi.org/10.1016/j.jspi.2010.04.036

[3] C.W.J. Granger, and A.P. Anderon, *An Introduction to Bilinear Time Series Model*s. Vandenhoeck and Ruprecht, Gottingen, 1978.

[4] S.C. Mathugama, and T.S.G. Peiris, "Critical Evaluation of Dry Spell Research," *International Journal of Basic and Applied Sciences*, vol. 11(6), pp. 153-160, 2011.

[5] S.C. Mathugama, and T.S.G. Peiris, "Spatial and Temporal Analysis of Critical Dry Spells in Sri Lanka," *International Journal of Climate Change: Impacts and Response*, vol. 3(3), pp.71-88, 2012.

[6] S.C. Mathugama, and T.S.G. Peiris, "Use of ESACF and SCAN methods to develop ARIMAX(p,d,q,b) to model critical lengths of dry spells in Anamaduwa, Sri Lanka, Sri Lankan Journal of Applied Statistics., in press., 2016

[7] T.S. Rao, "On the theory of Bilinear time series models," *Journal of Royal Statistics Society B*, vol. 43, pp. 244-255, 1981.

[8] T.S. Rao, and M. M. Gabr, *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture notes in Statistics 24. Springer, Verlag, 1984.

http://dx.doi.org/10.1007/978-1-4684-6318-7

[9] H. Tong, *Non-linear Time Series: A Dynamical System Approach*. Oxford: Claredon Press, 1990.

[10] R.S. Tsay, *Analysis of Financial Time Series*. 2nd ed., Wiley India, 2009.