www.jard.edu.pl http://dx.doi.org/10.17306/J.JARD.2017.00352

Journal of Agribusiness and Rural Development

pISSN 1899-5241
eISSN 1899-5772

4(46) 2017, 835–841
Accepted for print: 15.10.2017

EFFICIENT PLANNING OF SORGHUM PRODUCTION IN SOUTH AFRICA – APPLICATION OF THE BOX-JENKIN'S METHOD

Rangarirai Roy Shoko✉, Abenet Belete

University of Limpopo, South Africa

Abstract. Estimation and forecasting of crop production are crucial in supporting policy decisions regarding food security and development issues. The present study examines the current status of sorghum production in South Africa. Univariate time series modelling using ARIMA model was developed for forecasting sorghum production. Box and Jenkins linear time series model, which involves autoregression, moving average, and integration, also known as ARIMA (*p*, *d*, *q*) model was applied. The annual production series of sorghum from 1960 to 2014 exhibited a decreasing trend while prediction of sorghum production between 2017 and 2020 showed an increasing trend. The study has shown that the best-fitted model for sorghum production series is ARMA (1, 0, 4). The model revealed a good performance in terms of explaining variability and forecasting power. This study has also shown that sorghum could contribute to the household and national food security because of its drought-tolerant properties.

Keywords: ARIMA, sorghum production, forecasting, South Africa

INTRODUCTION

Sorghum (*Sorghum bicolor*) is one of the most important grain crops in South Africa and is used as a food grain for human consumption, fodder for animal feed and production of biofuels (DAFF, 2010). Traditionally, sorghum was produced largely under subsistence farming, but production has since shifted to commercial farming. Today, Sorghum farming in South Africa is divided into smallholder and commercial farmers (DAFF, 2010) due to the diversities in farm sizes, production systems and marketing techniques. The arid and semi-arid regions of South Africa are well-suited for sorghum cultivation (more than for maize) and its drought and heat-tolerant properties make it the ideal crop. Taylor (2003) revealed that sorghum is increasingly becoming the foundation for successful food and beverage industries in South Africa.

Sorghum is significantly imperative to household food security and is one of the staple foods for the poor and rural communities in the country region. The grain crop has excellent nutritional qualities and studies suggest that there could be prospective health benefits linked with sorghum (Dlamini, 2007). Present epidemiological evidence suggests that sorghum intake reduces the risk of certain types of cancer in humans compared to other cereals (LeBourvellec and Renard, 2012).

The demand for maize products in South Africa is rising rapidly and sometimes outstrips supply. Therefore, sorghum could be a potential substitute and/or compliment for maize and thereby create a balance between demand and supply for grain products. Sorghum is also utilised as a feedstock for biofuel production. The use of maize in the production of biofuels is banned in South Africa amid concerns over food security and fears of

✉MSc, Rangarirai Roy Shoko, Department of Agricultural Economics, University of Limpopo, Private Bag X1106 Sovenga 0727, South Africa, e-mail: rangariraishoko@yahoo.com

price increases (Radebe, 2013). The author also argued that grain sorghum could be used in ethanol production because it yields approximately the same amount of ethanol per hectare as maize. Therefore, growth in sorghum production is crucial to meet the growing demand for grain products in the country.

Reliable forecasts are important for efficient planning of crop production and can possibly assist in making informed policy decisions. Therefore, it is vital to generate forecasts for sorghum production using rigorous statistical modelling techniques that will assist in making informed policy decisions. Univariate time series modelling is convenient in developing forecasting models for crop production. During the past few decades, several statistical forecasting models have been developed due to the advancements in computer technology. One of such models includes the Autoregressive Integrated Moving Average (ARIMA) model.

Using ARIMA model the following authors found robust results. Badmus and Ariyo (2011) determined the ARIMA (2,1,2) to forecast maize production in Nigeria. Awal and Siddique (2011) studied rice production in Bangladesh. The study revealed that production uncertainty of rice could be minimized if production were forecasted with accuracy. Biswas and Bhattacharyya (2013) forecasted production of rice in West Bengal. The study found ARIMA (2,1,1) to be the best fitted model for rice production. Ehab and Frah (2016) modelled sorghum production in Sudan and the study revealed that growth in production is attributed to changes in harvested area land.

Given the significance of sorghum discussed earlier in this section, the main objective of this study is to develop an Autoregressive Integrated Moving Average (ARIMA) model for forecasting sorghum production in South Africa. Estimation and forecasting of crop production are crucial in planning and supporting policy decisions regarding food security and development issues.

MATERIALS AND METHODS

Time series data of sorghum production (from 1960–2014) were extracted from the Abstract of Agricultural statistics (DAFF, 2015) maintained by the Department of Agriculture, Forestry and Fisheries. To model sorghum production the Box and Jenkin (1976) time series method was employed. Autoregressive Integrated Moving Average (ARIMA) is one of the most prominent approaches for forecasting a time series. The main goal of the Box-Jenkin's method is to determine and estimate an econometric model sufficient to be used for forecasting. The ARIMA model is denoted by ARIMA (p, d, q), where 'p' denotes the number of autoregressive terms, 'd' the number of times the series should be differenced before it becomes stationary and 'q' is the number of moving average terms (Gujarati and Porter, 2009). It is vital to note that to use the Box-Jenkin's technique, stationary time series should be available on time series that is stationery after differencing once or several times.

ARIMA MODEL

The development of ARIMA models is based on the methodology quantified in the classic work of Box and Jenkins. The autoregressive integrated moving average (ARIMA) model, represented by ARIMA (p, d, q), is given by:

$$\Delta^d y_t = \delta + \alpha_1 \Delta^d y_{t-1} + \alpha_2 \Delta^d y_{t-2} + \ldots + \alpha_p y_{t-p} + e_t - \varphi_1 e_{t-1} - \ldots - \varphi_q e_{t-q} \tag{1}$$

or equivalently by

$$\omega(\mathrm{B})(\Delta^d y_t - \mu) = \theta(\mathrm{B})\varepsilon_t \tag{2}$$

where: y_t, y_{t-1} signifies the observed sorghum series at time t, e_t, e_{t-1} is a sequence of uncorrelated random variables having zero mean, $\alpha_1, \ldots, \alpha_q, \varphi_1, \ldots, \varphi_q$ are parameters of the model, μ is the mean of $\Delta^d y_t$, $\omega(\mathrm{B})$ is $1 - \omega_1\mathrm{B} - \ldots - \omega_p\mathrm{B}_p$, $\theta(\mathrm{B})$ is $1 - \theta_1\mathrm{B} - \ldots - \theta_q\mathrm{B}^q$ signifies the moving average parameter, Δ and B denote the difference and back-shift operators, respectively, denotes the autoregressive parameter p, q, and d denote the autoregressive, moving average and difference orders of the process, respectively (Awal and Siddique, 2011). The model building procedure involves the following three steps: namely identification, estimation of parameters, and diagnostic checking.

Identification*:* Orders p, d, q of ARIMA models are specified to clarify the number of parameters to estimate. However, the Box-Jenkins ARIMA method can only be applied to time series that are stationary. Thus, the primary step in developing a Box-Jenkins model is to ascertain if the time series data is stationary. Gujarati and Porter (2009) argued that the main reason for requiring stationary data is that any model which is concluded from these data can itself be interpreted as stable or stationary, therefore providing a valid basis for forecasting.

Once stationarity has been addressed, the next phase is to determine the order (p and q) of the autoregressive and moving average terms. The basic tools for accomplishing this are the autocorrelation (ACF) and the partial autocorrelation (PACF) plots.

Estimation: Eviews 8 software package was used to fit the ARIMA model. Akaike information criterion (AIC) and Bayesian information criterion (BIC) values were used for parameter estimation. A model with the smallest values of AIC, BIC and Q-statistics and with high R-square may be considered as an appropriate model for forecasting (Biswas and Bhattacharyya, 2013).

Diagnostic checking: To validate the adequacy of the ARIMA model, relevant diagnostic tests are applied such as the Jargue Bera test for normality and the Ljung-Box Q statistic which provides an overall check of model adequacy. The test statistics Q is specified as:

$$Q_m = n(n+2)\sum_{k-1}^{n} \frac{r_k^2(e)}{n-K}\chi^2_{m-\mathrm{r}} \qquad (3)$$

Where

$r_{k(e)}$ = the residual autocorrelation at lag k

n = sample size

m = lag length available in the test

k = number of parameters estimated in the model

In an application, if computed p-value associated with the Q-Statistics is small (p-value < α), the model is considered unsatisfactory (Gujarati and Porter 2009). Thus, one should repeat the analysis process until a suitable model has been obtained.

Testing for structural change: South Africa experienced a transition to a democratic government in 1994 and this resulted in a change in policies particularly the agricultural marketing policy of 1997 which significantly influenced the South African agriculture sector. The policy involves much less state involvement, regulation and state interference in agricultural markets and product prices (Groenewald, 2000). Lüder et al. (2012) argued that without claiming structural breaks, the estimation of a model is impaired. Therefore, the chow test was used to test the hypothesis that H_0: no structural breaks in 1997, against H_1: there were structural breaks in 1997.

RESULTS AND DISCUSSION

A graphical presentation of the sorghum production series is plotted in Figure 1. A close examination of the graph shows that sorghum production in South Africa reduced dramatically between the late 1980s and 2014. This observation can also be visualized from the plot of standardized production series in Figure 2.

Figure 2 displays a graphical presentation of the standardized production series from 1960 to 2014. A visual inspection of the diagram shows that sorghum production fluctuated above and below average between 1960 and 1997. Between 1998 and 2014 production was predominantly below average and exhibited a decreasing trend. This observation might have been due to notable policy changes that took place in 1994 and 1997.

The results of the chow test evidently confirm that there are structural breaks in the sorghum series that were realised in 1997 as the F-statistic of 22.021 and the associated p-value of 0.0002 is significant at (5%) level, thus we reject the null hypothesis. The findings demonstrate an unexpected shift in sorghum series as a result of policy changes which took place after the democratic dispensation. Therefore, a dummy variable

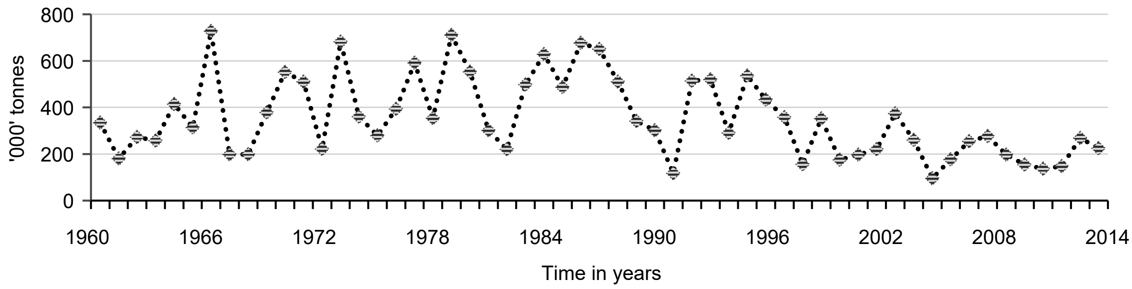

Fig. 1. Sorghum production series
Source: DAFF, 2015.

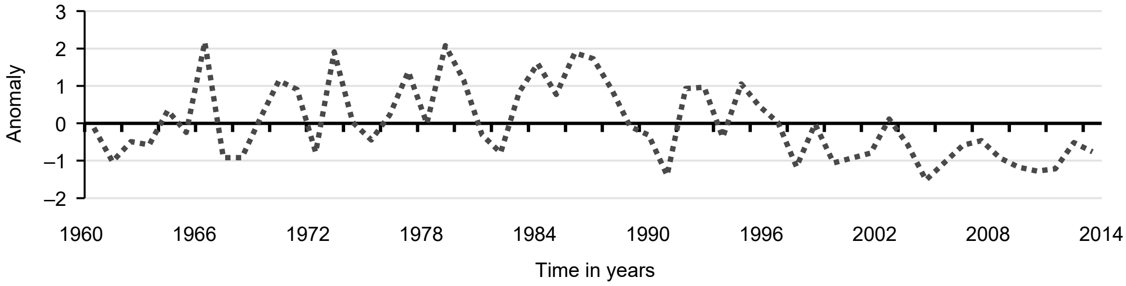

Fig. 2. Standardised sorghum series
Source: DAFF, 2015.

Table 1. Stationarity test results at levels for sorghum production series

Series	ADF test statistic	Critical value	Lag-length	Probability	Conclusion
SP	–4.975	–2.917	10	0.0001**	Stationary

**significant at the 5% level, SP signifies sorghum production.
Source: own elaboration.

was included in the ARMA model to capture the effect of policy changes on sorghum production with years before and after the inception agricultural marketing policy taking values of 1 and 0 respectively.

The Augmented Dickey-Fuller (ADF) test is a popular formal approach of determining stationarity apart from the graphical approach using ADF. Table 1 represents the ADF test at levels for the data series. The optimum lag length was selected based on the Schwarz information Criterion.

The ADF method tests the hypothesis that H_0: $X \sim 1(1)$ is non-stationary against H_0: $X \sim 1(0)$ that is stationary. The critical value for rejection of the null hypothesis of unit root is significant at (5%). The results of the unit roots test show that the sorghum series is stationary at levels. Therefore, it is judged that no differencing of the data series is necessary.

The Box-Jenkins procedure was applied on stationary data series and the corresponding ARMA (p, q) process was identified. The appropriate p and q for the data series were chosen using series correlograms. Figure 3 and 4 display the autocorrelation function and partial autocorrelation function of sorghum series in South Africa. From Figure 3 the series is considered to be

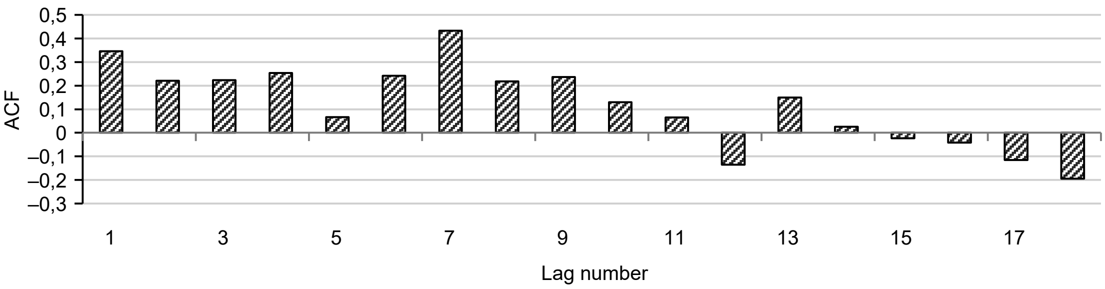

Fig. 3. The ACF of the sorghum series
Source: own elaboration.

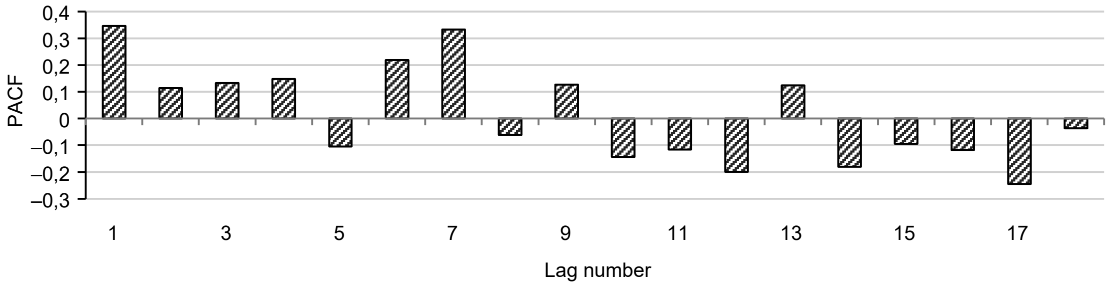

Fig. 4. The PACF of the sorghum series
Source: own elaboration.

adequately stationary because the ACF cuts off, or dies down, fairly quickly. The ACF at lags 1, 4 and 7 seem statistically different from zero, but at all other lags, they are not statistically different from zero. Thus, from the ACF correlogram, the chosen values of the parameter q are 1, 4, and 7. Figure 4 shows that there are two significant spikes for PACF plot, at lags 1 and 7. Thus, to determine the parameter p the PACF correlogram was used and a value of 1 was chosen.

From ACF and PACF plots, several models were estimated in an attempt to reach the right specification. These models are specified as; ARMA (1, 0, 4), ARMA (1, 0, 0) and ARMA (1, 0, 1), ARMA (0, 0, 4), ARMA (4, 0, 3) and ARMA (1, 0, 7). One of these models could be the most appropriate model to use for forecasting. Table 2 shows statistical properties of the six ARMA models.

Based on minimum AIC (Akaike's Information Criterion, BIC (Bayesian Information Criterion) and Hannan-Quinn criteria values and considering the ACF and PACF of the sorghum production series, ARMA (1, 0, 4) model is considered the best-fitted model. The accuracy of the selected model is evaluated by the various tools which includes Root mean square error (RMSE), mean absolute error (MAE) and Theil inequality coefficient (Table 2). The low values of RMSE, MAE and Theil inequality coefficient for ARIMA (1, 0, 4) have shown that the model is accurate in terms of its forecasting ability. The estimation of sorghum series is presented in Table 3. The coefficient of determination, adjusted R-squared is 30%.

Based on the estimated parameters the mathematical model for ARMA (1, 0, 4) is specified as follows:

$$\begin{aligned} Y_t = 5.953 - 0.639\mathrm{DU}M_t + 0.268Y_{t-1} - 0.581e_{t-1} \\ + 0.165e_{t-2} - 0.195e_{t-3} - 0.362e_{t-4} + e \end{aligned} \quad (4)$$

The coefficient of the policy dummy variable is negative and significant at 5% level. This suggests that

Table 2. Comparison of different ARIMA models with model fit statistics for series

ARIMA model	BIC	AIC	HQ	Average	RMSE	MAE	Theil Inequality coefficient
(1, 0, 4)	1.39	1.46	1.43	1.43	223.09	204.25	0.1836
(1, 0, 0)	1.42	1.54	1.46	1.47	232.14	227.68	0.2143
(1, 0, 1)	1.43	1.35	1.77	1.51	237.25	243.71	0.2394
(0, 0, 4)	1.48	1.71	1.56	1.58	245.62	251.25	0.3562
(4, 0, 3)	1.52	1.86	1.65	1.68	287.20	271.36	0.4107
(1, 0, 7)	1.56	1.93	1.87	1.79	307.32	312.54	0.5321

Source: own elaboration.

Table 3. Estimation results of ARMA (1, 0, 4) model

Variable	Coefficient	Std. Error	t-Statistic
Constant	5.953	0.057	9.433**
DUM	–0.639	0.131	–2.471**
AR(1)	0.268	0.088	3.045**
MA(1)	–0.581	0.299	–1.940*
MA(2)	0.165	0.280	0.589
MA(3)	–0.195	0.210	–0.928
MA(4)	–0.362	0.287	–1.258
Adj R^2 = 0.301	DW statistic = 2.060		

**Significant at the 5% level. *Significant at the 10% level, DUM signifies structural break dummy for agriculture marketing policy, the model was estimated in its log form.
Source: own elaboration.

a change in policies that took place after the democratic dispensation could have negatively affected sorghum production. This can also be detected from the chart of the standardised sorghum series (Fig. 2) which shows that sorghum production was predominantly below average after the 1994 period. This observation may have been due to an increase in rural-urban migration which deprived rural areas of labour in farming activities, leading to reduced sorghum production. In South Africa, the level of urbanisation increased by 4.3% between 1996 and 2001 (Kok and Collinson, 2006).

Having chosen ARMA (1, 0, 4) model, and having estimated its parameters, the next step is to see whether the chosen model fits the data reasonably well. The diagnostic checks were done using ACFs of residuals and Ljung and Box Q tests and the results showed that none of the statistical terms were exterior to the confidence intervals. For normality test, the Jargue Bera test was used. The test statistic of 0.306 and the associated p-value of 0.858 shows that the residuals of the model are normally distributed. Based on these findings, the model seems to be satisfactory in terms of its specifications.

Having confirmed the validity of ARMA (1, 0, 4) the model is finally used to forecast the corresponding variables. There are two kinds of forecasts: sample period forecasts and post-sample period forecasts. The sample period forecast was made for the last six years of the dataset (for 2009–2014) to establish confidence about the model and the out of sample forecasts were made up to 2020, to generate forecasts for use in efficient planning and other fundamental purposes. Table 4 shows the generated forecasts and observed values of annual sorghum production of South Africa for the period 2009 to 2020.

Table 4. Forecasting table for sorghum production

Production (000t)		
Year	Observed	Forecasted by ARMA(1,0,4)
2009	197	207.65
2010	155	182.00
2011	137	176.06
2012	147	178.68
2013	268	187.59
2014	225	202.08
2015	**	154.48
2016	**	112.02
2017	**	182.81
2018	**	193.04
2019	**	195.97
2020	**	198.79

Source: own elaboration.

The generated forecasts have shown that sorghum production between 2017 and 2020 exhibit an increasing trend. This outcome suggests an increase in sorghum production in the near future. This observation is consistent with findings obtained by Badmus and Ariyo (2011) who forecasted crop production in Nigeria and observed a future increase in maize production. The increase might be attributed to the fact that sorghum is becoming a potential food security proposition in South Africa given its exceptional drought tolerant and nutritional qualities. The other reason for a surge in sorghum production in the future could be that people are increasingly becoming more health conscious and thereby resorting to healthier diets. Given the health benefits of consuming sorghum discussed earlier in the study, sorghum could contribute in varying degrees to the dietary needs of the growing national population. South Africa's strategy to source all grain needed for biofuel production locally means sorghum output will also rise enormously. Furthermore, in this age of global climate change, sorghum could be the magic bullet South Africa needs to combat poverty and maintain sustainable food security in the long run.

SUMMARY

This study reveals important observations worth sharing. The annual production series of sorghum from 1960 to 2014 exhibited a decreasing trend while prediction of sorghum production between 2017 and 2020 showed an increasing trend. ARIMA (1, 0, 4) was found to be the best-fitted model for the sorghum production series. The model demonstrated a good performance in terms of explaining variability and predicting power. The forecasting of sorghum production is vital as it helps both farmers and policy makers to plan for the future. In drought prone areas, sorghum can provide better household food security than maize. Furthermore, sorghum is also utilised as a feedstock for biofuel production. Therefore, if an increase in the production of sorghum is needed in the future, increase in government support to agriculture, selection of high yielding varieties, increasing agricultural partnerships between farmers and research institutes are critical elements.

REFERENCES

Awal, M. A., Siddique, M. A. B. (2011). Rice production in Bangladesh employing by ARIMA model. Bangl. J. Res., 36(1), 51–62.

Badmus, M. A., Ariyo, O. S. (2011). Forecasting cultivated area and production of Maize in Nigerian using ARIMA Model. Asian J. Agric. Sci., 3(3), 171–176.

Biswas, R., Bhattacharyya, B. (2013). ARIMA modelling to forecast area and production of rice in West Bengal. J. Crop Weed, 9(2), 26–31.

Box, G. E. P., Jenkin, G. M. (1976). Time Series of analysis: Forecasting and control. Jon Wiley & Sons.

DAFF (2010). Sorghum production guidelines. Obtainable from Resource Centre Directorate Agricultural Information Services Private Bag X144 Pretoria 0001.

DAFF (2015). Abstract of Agricultural Statistics. Retrieved August 20th 2016 from: www.daff.gov.za/pdf

Ehab, A., Frah, H. (2016). Sudan Production of Sorghum; Forecasting 2016-2030 Using Autoregressive Integrated Moving Average ARIMA Model. Am. J. Math. Stat., 6(4), 175–181.

Dlamini, N. R. (2007). Effect of sorghum type and processing on the antioxidant properties of sorghum (Sorghum bicolor L. Moench) based foods. PhD Thesis, Texas A&M University, College Station, TX, USA.

Groenewald, J. A. (2000). The Agricultural Marketing Act: A Post – Mortem. South Afr. J. Econ., 68(3), 161–176.

Gujarati, D. N., Porter, D. C. (2009). Basic Econometrics, 5th Edition. New York: McGraw-hill, USA.

Jambhulkar, N. N. (2013). Modelling of rice production in Punjab using ARIMA model. Int. J. Sci. Res., 2(8), 2–3.

Kok, P., Collinson, M. A. (2006). Migration and urbanisation in South Africa. Statistics South Africa; Pretoria. Report 03-04-02.

LeBourvellec, C., Renard, C. M. G. C. (2012). Interactions between Polyphenols and Macromolecules: Quantification Methods and Mechanisms. Crit. Rev. Food Sci. Nutr., 52, 213–248.

Lüder, J., Brauer, A., Jurisch, R. (2012). Breakpoint detection within the time series: modelling approach upon paleoclimatic proxy data. Hist. Soc. Res., 37, 315–325.

Radebe, H. (2013). Biofuels offer sorghum new growth hopes. Bus. Day Live, Jan 4th.

Taylor, J. R. N. (2003). Overview: Importance of sorghum in Africa, in Afripro: Workshop on the Proteins of Sorghum and Millets: Enhancing Nutritional and Functional Properties for Africa, Pretoria, 2–4 April 2003. Retrieved Nov 2013 from: www.afripro.org.uk