THE MODELLING AND FORECASTING OF CULTIVATED LAND IN KHYBER PAKHTUNKHWA USING ARIMIA MODELS

Mukhtiar Ahmad^a, Farhad Ali^b and Alamgir^c

Department of Statistics University of Peshawar, Peshawar Pakistan ^{a, c} Institute of Numerical Sciences, Kohat University of Sicence and Technology Kohat Pakistan ^b,

ate of Humerican Sciences, Ronat Oniversity of Science and Teenhology Ronat Fakista

Emails: mukhtiar_ahmad41@yahoo.com_farhadmardan@gmail.com,

ABSTRACT Agriculture is one of the most important sectors, which significantly contribute to the socio-economic development of Pakistan. However, this sector is suffering due to lack of evidence-based policymaking and proper planning. This study tries to examine the relationship between the culturable waste and cultivated land in the KPK province and presents the forecast analysis – by using the Auto-Regressive Integrated Moving Average (ARIMA) models. The data set was obtained from the agriculture department of KPK. The estimation results show that the cultivated area and culturable waste forecasts for the year 2025 would be around 1.8621 and 1.39 million hectors respectively. Further, it was found that the cultivated area have a negligible increase, whereas the culturable waste have an upward trend in the near future. The key findings and implications of this study are quite significant to the concerned policymakers and government agencies for proper land management and improved crop production in the province.

Keywords: Forecasting, Culturable Waste, Cultivated Land, ARIMA Model.

(1): INTRODUCTION:

Agriculture play a significant role for the economy of any country in the world. Therefore, it should be developed and planned. Moreover, it is very necessary to manage the land use for various purposes and enhance the production for a country. In today world, modern methods are using for land use of agriculture planning. Hence the motivation behind this study is that to use the data available and applied the modern statistical research technique to forecast for future management and planning of land use for Khyber Pakhtunkhwa (KPK) province and hence for Pakistan. For management and planning of land for this country, no such work had done before it. So, it would also be beneficial for other researchers and planners who desire to work in this area. This forecast can also be used for agricultural planning in other countries of the world having the same land and environment like the province KPK Pakistan.

Effective land use planning plays a key role in the sustainable production capacity of agriculture sector. Land would be degraded over the years, if not properly planned and optimally managed. One of the burning issues that the world faces today is the continuous degradation of lands which ultimately lead to under productivity and the occurrences of disasters. Several natural or manmade factors are involved in land degradation such as, wind and water erosion, water logging and salinity, deforestation and desertification etc. In order to prevent land degradation, countries should adopt short and long term measures to control and better manage lands for socio-economic activities. Effective efforts, therefore, are needed both at national and international levels to formulate methodologies and devise appropriate policies and programmes to combat land degradation and improve proper utilization of lands.

The land is defined as that part of earth which is not covered by oceans or other bodies of water. In other words, land is the solid surface of the Earth that is not permanently covered by water. The area which sown during the year at least once or during past year will be referred as Cultivated Area. Further, Cultivated Area=Net Area sown + Current Fallow. While Culturable Waste is that uncultivated farm which is not cropped during the whole year even not in the previous year, but it is suitable for cultivation [1].

The lands which are used primarily for production of food and fibre or related goods is referred as agriculture land. Where the classification of the farm area in accordance with its use is called land use statistics. Province KPK consists of around 22 million people. The large part of the population (82%) is living in rural areas leading a big pressure on land areas. This province includes 10.17 million hectares of land of which the 2.75 million hectare is the cultivable area, of which only 1.8 million hectares are cultivated land and the remaining 1.08 million hectares are culturable waste. Mostly cultivated land depends on rain that include 49.5% of the cultivated area [1].

The main objective of this paper is to analyse the time series data for agriculture land use of KPK Pakistan for two variables (culturable waste, cultivated area), make modelling and forecast for next 9 years. Therefore we developed models for these two variables which are found to be *ARIMA* (1, 2, 1) (0, 1, 1) respectively. The Box – Jenkins methodology are used for analyses. Usual thinking is that the cultivated land increases rapidly and the calturable waste decreases, but our prediction shows that the cultivated land increase negligibly as compare to the culturable waste.

(2): Literature Review

Saleem A. *at el* [15], conducted a study about the future estimates of maize production and area for Khyber Pakhtunkhwa. They analysed the data using time series [3,5] analysis using ARIMA model popularized by Box – Jenkins Methodology (1976). Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD) and Mean Squared Deviation (MSD) were used as model selection criteria to forecast errors. He found that there is a decreasing trend both in maize area and production.

Mehmood and Ahmad [6, 7], conducted study to examine the growth of area of mangoes in Pakistan. They used time series data for 49 years from 1961 to 2009. They found ARIMA (0, 1, 0) is a suitable model for the series to forecast. This research concluded that in 2025 the area for mangoes in Pakistan would be 318.5 thousand hectares. Further, the area for mangoes would be increase 87% in 2025 as compared to today, which is more than 6% per year. For more literature see Suleiman N. and Sarong S., Mustafa K., Burhan A. and Nugroho, A. Simanjuntak, H. B. [8,10] etc.

Prabakaran [11], used annual data of cultivated land for the years 1950-51 to 2011-12 in India. He fitted Auto Regressive Integrated Moving Average (ARIMA) (1, 1, 0) and ARIMA (1, 1, 1) models for cultivated land and for the production of wheat, from which he made forecast for the next four years. He found an increase in wheat cultivated areas and production.

Suleiman and Sarpong [16], calculated that milled rice production expected an upward trend for coming ten years. Data was collected from 1960 to 2010 in Ghana state. The data was model using ARIMA (Box – Jenkins methodology). They fitted ARIMA (2, 1, 0) model, in which, they concluded that the milled rice production have an increasing trend in next ten years.

Badmus and Ariyo [2], conducted a study to forecast cultivated areas and maize production. They used last 36 years data from 1970 to 2005 in Nigeria. Box and Jenkins (1976) univiriate time series model was used for modelling the data. They fitted ARIMA model and found that the maximum maize cultivation area was 9244000 hectares in 2007 and minimum was 425000 hectares in 1979 in Nigeria. They found that cultivation land and production indicated an increasing trend. They concluded that government should increase fund for agriculture sector and make a link between the farmers and research institute. By doing this, the yield will be increased in the future, specially, for cropped areas and for the production of maize.

Najeeb *at el*,[9], conducted a study about the future of wheat area and production in Pakistan to year 2022. They projected that the wheat production and area both are expected to be increased. They applied the Box – Jenkins methodology and selected ARIMA (1, 1, 1) and ARIMA (2, 1, 2) as best models. The results were shown by forecasting that the wheat area and production for the year 2022 would be 8475.1 hectares and 29774.8 thousand tons respectively.

Sahu, [13], worked forecasting production of major food crops in four major SAARC countries. Sajid, Nouman and Hina [14] tried to forecast production and yield of sugarcane and cotton crops in Pakistan for the period 2013-2030.

(3): Research Methodology

This study utilizes secondary data, which has been taken from the Directorate of Crop Reporting Services Agriculture Department, Khyber Pakhtunkhwa (www.crs.kp.gov.pk) [4].

The annual data of the two variables, cultivated area and culturable waste for the period from 1971-72 to 2016-17 is used to model and forecast for these variables.

To select a best model to produce more accurate forecast based on past pattern in the historical data and in which way to determine the best model orders. Statisticians George Box and Gwilym Jenkins improved a practical method to developed ARIMA model that better fit to a time series data satisfy the principle of parsimony. These concepts have very basic role in time series analysis and forecasting.

There are three steps in Box-Jenkins methodology: they are, identification of model, estimation of parameters and diagnostic tests to find the best model among all of ARIMA models [12]. It is recurring many times until in the end a suitable model is selected. In last the selected model may be used to forecast of the series (Adhikari & Agrawal)

An Autoregressive Integrated Moving Average Process model is a way of explaining that in which way a time series observation is attached to its own past value. The general form of the *ARIMA* (p, d, a) model is:

The general form of the ARIMA
$$(p, d, q)$$
 model
 $\varphi_p(B) \quad (1-B)^d X_t = \theta_q(B) Z_t$

$$\varphi_p(D)$$
 (1)

Where, θ_q and φ_p are the parameters, B is back shift operator, p is the order of AR component, q is MA model order and d is the differencing order.

When the original time series is stationary, then d = 0 and the ARIMA models convert to ARMA models.

Box-Jenkins process deals with mixed ARIMA models for a data set. The aim of ARIMA model is to find out the stochastic process and make forecast in a precise manner. In building models for discrete time series and dynamic systems this procedure is used and beneficial in various situations [17].

ARIMA models were study by George Box and Gwilym Jenkins in 1968 that is why their names have categorically been used those days for ARIMA models applied for time series analyses and forecasting. While to make optimal forecast stochastic models were used. Where stochastic process is both stationary and non-stationary. As most time series models are non- stationary and the ARIMA models deals with stationary time series, therefore, Box-Jenkins used order differences process to convert non-stationary series to a stationary series.

The stages in Box-Jenkins Methodology are:

Step 1: Identification: Appropriate values of p, d and q are found first. The tools used for identification are the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF) and the resulting correlogram and partial correlogram.

Step 2: Estimation: Having identified p and q values estimation of parameters of the autoregressive and moving average terms are estimated using simple least squares.

Step 3: Diagnostic checking: Before forecasting a given series, first is to verify it by applying one simple test Box-Ljung statistic through which residuals analysed [13].

Step 4: Forecasting: On contrary traditional econometric modelling, ARIMA modelling is more reliable to make forecast. ARIMA process is a good way to used its own lag values to forecast. That is why an ARIMA model is applied to fit the best weighted average forecasts for a single time series [18].

(4): RESULTS & DISCUSSION

For analyses R i386 3.1.2 are used for this study.

(4.1): Culturable Waste: Figure 1 represent that time series observation on culturable waste is non stationary. Since, taking first difference to make the series stationary.

Figure 2 shows that even the first difference has not made it stationary. Therefore, the 2^{nd} difference is calculated and



Figure 2: Time Plot for 1st Differences of Culturable Waste The time series of first differences appears that it is not stationary in mean and variance, so, we take again a second difference.



Figure 3: Time Plot for 2nd Differences of Culturable Waste

Figure 3 shows that the data is stationary in variance and mean. Thus, the trend component is removed and probably

an ARIMA (p, 2, q) model is suitable to this series despite some spikes which is shown in the Figure.

Unit Root Test for Culturable Waste:

Table 1: Augmented I	Dickey-Fuller Te	st for Culturable V	Vaste
	•/		

Test	Dickey-Fuller	P-value
Original Series	-3.0	0.17
1 st difference of series	-2.9	0.21
2 nd difference of series	-4.15	0.01

In Table 1 Augmented Dickey-Fuller Test verify that on 2nd differences of culturable waste is converted to stationary series, having P value 0.01 which mean that the series is significant.

To find the values of p and q in order to find an appropriate ARIMA model, partial and auto correlogram for the stationary time series are examined.

We see from the correlogram that the first ACF exceed the significant bound, where, all other tail off to zero. This suggests a tentative value of q=1.



Next we examined the PACF to select the appropriate model, which is given in 5.



Figure 5: Partial Auto-correlation for 2nd Difference on Culturable Waste

We see that first PACF exceed the lower bound of the confidence limits, all other partial correlogram are within the limits. Since both correlogram and partial correlogram are tail off after lag 1. So, this gives rise to ARIMA (1, 2, 1).

But to reach a best final model, it is a common practice to fit various tentative models lying in the neighbourhood of ARIMA (1, 2, 1) as well.

For the best model we construct other models in Table 2.

Table 2: Best ARIMA Models for Culturable Waste					
Models	Log-Likelihood	AIC	BIC		
ARIMA(2, 2, 0)	47.35	-88.69	-83.34		
ARIMA(2, 2, 2)	50.2	-90.41	-81.49		
ARIMA(1, 2, 1)	49.43	-92.86	-87.51		
ARIMA(1, 2, 2)	50	-92.01	-84.87		
ARIMA(2, 2, 1)	47.42	-86.85	-79.71		

175

By principle of parsimony *ARIMA* (1, 2, 1) is the fitted and best model. Which is estimated in the following equation: $X_t = 1.9896X_{t-1} - .9792X_{t-2}$ -0.0104 $X_{t-3} + Z_t + Z_{t-1}$ X_t Is stationary time series, Z_t is a white noise having zero mean and constant variance. ϕ and θ are the parameters having values -0.0104 and -1.00 respectively.

Diagnostic checking:

After selecting the model is to check whether the model adequately represent the data and whether it satisfies the assumption the important step in Box–Jenkins methodology Some diagnostic checks are performed for this purpose including plots of the residual and Ljung-Box test for serial correlation. Ljung Box test gave the value for P 0.80 which is concluded that there is weak evidence for auto correlations to be zero. Therefore, ARIMA(1, 2, 1) model is suitable to represent this time series data.

Moreover, Figure 6 explicitly reveal that there is no autocorrelation among the lag observations, except some are by chance, which are irrelevant to the data.



Figure 6: Ljung - Box Statistic for Culturable Waste



Figure 7: Residuals for Culturable Waste

In Figure 7 residuals are normally distributed with constant variance and near to zero mean with one or two spikes by chance.

Using the estimated model ARIMA (1, 2, 1) we make forecast for the given series along with 95% and 80% prediction intervals for the next 9 years.

Forecasting:

Table	3: Fore	cast	for	Culturable	Waste	for	next 9 years
				**!		~ -	

Point	Forecast	Lo 80	Hi 80	Lo 95 Hi 95
2017	1.345419	1.247811	1.443027	1.1961398 1.494698
2018	1.351822	1.213013	1.490630	1.1395326 1.564111
2019	1.358214	1.186703	1.529725	1.0959104 1.620518
2020	1.364607	1.164681	1.564532	1.0588465 1.670367
2021	1.370999	1.145321	1.596677	1.0258547 1.716143
2022	1.377392	1.127799	1.626984	0.9956724 1.759111
2023	1.383784	1.111629	1.655940	0.9675584 1.800010
2024	1.390177	1.096499	1.683854	0.9410356 1.839318
2025	1.396569	1.082196	1.710942	0.9157771 1.877361

There are 46 observations of original time series for culturable waste. Forecast are made for the next 9 years (2016—2025) as well as 80% and 95% prediction intervals. The last 46th observation for the years 2016-17 is 1.34

million hectare, and the ARIMA model gives the approximated forecast value 1.39 million hectare for 2025, means that much increase in the culturable waste is expected. Which is clearly shows in Figure 8.





If these errors are normally distributed with constant variance and zero mean, we make a histogram and time series plot for the forecasts errors. The time plot in figure 9 explicitly reveal that the variance for the forecast errors are look to be constant with time except one or two spikes which occur by chance.



Figure 9: Time Plot for Forecast Errors of Culturable Waste



Figure 10: Histogram of Forecast Errors for Culturable Waste

The histogram for this time series reveal that these errors are distributed and mean close to zero. It means forecast errors are distributed normally with constant variance and mean zero. Therefore, it is apparently valid. Therefore, consecutive errors do not look to be related and it is distributed normally constant variance and zero mean, the *ARIMA* (1, 2, 1) does seem to be a suitable predictive

model for area of culturable waste of KPK Pakistan agriculture land.

(4.2): Cultivated Area:

The first step in box – Jenkins methodology is to plot the data to see whether the series is stationary. Figure 11 shows that the data on cultivated area is non stationary. Since, difference is to be taken to convert it into stationary data.



Figure 11: Time Plot for Cultivated Area

By differencing the series once, one see that it becomes stationary and taking the following shape.



Figure 12: 1st Differences for Cultivated Area

Figure 12 shows that the series is stationary. Thus, the trend component is removed and probably an ARIMA (p, 1, q) model is fit for this series despite a spike which is clear from the Figure.

Unit Root Test for Cultivated Area:

Table 4: Dicke	y-Fuller Test for Cultivated Area
----------------	-----------------------------------

Test	Dickey-Fuller	P-value
Original series	-4.07	0.15
1 st difference of series	-3.61	0.04

From the table 4 the Augmented Dickey-Fuller Test verify that on 1st differences of cultivated area is stationary having P value 0.04 which reveals that the series is significant.

To find the values of p and q in order to find an appropriate ARIMA model, correlogram and partial correlogram for the stationary series are examined.

We see from the correlogram that the first ACF exceed the significant bound, where, all other tail off to zero. This suggests a tentative value of q=1.



Figure 13: Auto- Correlation Function for Cultivated Area

In next figure PACF is examine to select the appropriate model, which is given in 14.



Figure 14: Partial Auto- Correlation Function for Cultivated Area

As all PACFs are within the limits. Since one correlogram is tail off after lag 1 and all are in the limits, where, PACF are within the limits. So, this gives rise to ARIMA (0, 1, 1). But, for the selection of the best model it is important to fit others models than ARIMA (1, 1, 1). Which are computed in table 5.

Table 5: Best ARIMA Models for Cultivated Area					
Models	Log-Likelihood	AIC	BIC		
ARIMA(1, 0, 1)	70.53	-123.07	-125.75		
ARIMA(0, 1, 1)	68.68	-133.35	-129.74		
ARIMA(1, 1, 1)	69.43	-132,86	-127.44		
ARIMA(2, 1, 2)	69.44	-128.69	-119.66		
ARIMA(1, 1, 0)	68.78	-123.55	-124.94		

By principle of parsimony on the basis of used criteria ARIMA (0 1 1) is the best model. The estimated model will be written as:

 $X_t = X_{t-1} + z_{t+0.2183} z_{t-1}$

 X_t is the stationary time series, Z_t is a white noise with mean close to zero and constant variance. Theta is the estimated parameter. The estimated value for theta is -0.2183 in case of ARIMA (0, 1, 1) model for cultivated area of agriculture land.

Diagnostic Test:

The important step in Box - Jenkins methodology after selecting the model is to check whether the model adequately represent the data and whether it satisfies the assumption. Some diagnostic checks are performed for this purpose including plots of the residual and Ljung-Box test for serially interdepended. Ljung Box test gave the P value is 0.50, so, as a result we observe that there is a little evidence for auto correlations to be non-zero. Therefore, ARIMA (0, 1, 1) model is suitable to represent this time series data.



Figure 15: Ljung - Box Statistic for Cultivated Area

Above Figure represent that P value is significant, and therefore, there is no auto-correlation among all the past (lag) values, except one or two are violated.



Figure 16: Residuals for Cultivated Area

In Figure 16 it is shown that the residuals are distributed normally having mean near to zero and permanent variance. On the basis of estimated model *ARIMA* (0, 1, 1) we make forecast for the given series. Two prediction intervals are also calculated, i.e 80% and 95% for the next 9 years.

Forecasting:

Table 6: Forecast of Cultivated Area for Next 9 Years						
Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95	
2017	1.862331	1.796523	1.928138	1.761687	1.962975	
2018	1.860055	1.771565	1.948546	1.724721	1.995390	
2019	1.862276	1.752093	1.972460	1.693765	2.030787	
2020	1.860108	1.734974	1.985243	1.668732	2.051485	
2021	1.862225	1.720985	2.003464	1.646217	2.078232	
2022	1.860159	1.706913	2.013405	1.625789	2.094528	
2023	1.862176	1.695576	2.028776	1.607383	2.116968	
2024	1.860207	1.683266	2.037147	1.589599	2.130814	
2025	1.862129	1.673552	2.050706	1.573725	2.150533	





The original time series for cultivated area includes 46 observation. The forecast for next 9 years (2017—2025) along with 80% and 95% prediction intervals are computed in Table 6. The last 46th observation for the year 2016-17 of cultivated area is 1.8601 million hectors and the ARIMA model gives the forecasted value for the year 2025 is 1.8621 million hectors, it means that there would be very

negligible improvement in cultivated land is expected. Which is shows in figure 17.

To find whether the forecast errors are normally distributed with constant variance and mean near to zero, we make a histogram and time plot of the forecasts errors.

The time plot in figure 18 reveal that the variance for these errors are look to be same over time except one or two spikes which occur by chance.



The histogram reveal that these forecast errors are help distributed normally having mean near to zero with plan constant variance. Since, it is apparently appropriate for The

this series and valid. Therefore consecutive forecast errors does not seem to be inter-related and it is distributed normally with constant variance and mean close to zero, the *ARIMA* (0, 1, 1) does look to be an appropriate model for cultivated area of KPK Pakistan agriculture land.

(5): CONCLUSION & FUTURE DIRECTIONS:

Land use statistics play a vital role in planning for the future of a country. Pakistan is an agriculture country and it is more necessary to make record of land for sustainable planning. KPK Pakistan mainly depends on agriculture and its area is also less than the others provinces. Agriculture land use data were collected from the Crop Reporting Services of the Agriculture Department KPK Pakistan. Statistical modelling and forecasting approaches were used to predict that what would be the position of different variables for the next 9 years. Box-Jenkins methodology was used for these variables that resulted in valid models.

It is concluded that increase in cultivated land is very slow and there is link between culturable waste area and cultivated land. Culturable waste area would be converted to cultivated area in near future to enhance the economy of the country. Increase in cultivated land would increase in agriculture yields in KPK Pakistan. These projections will help the government to make policies regarding land use planning and to enhance agriculture production in future.

The 44% rural population of KPK Pakistan is living below the poverty line. This is because the low economic growth due to declining in jobs, lack of education, mismanagement of agricultural land and a range of natural resources problems. For sustainable economic growth, particularly in rural area like KPK it is necessary for the GOVT to develop agriculture and livestock sectors along with exploring other natural resources. In KPK and throughout the country (Being an agricultural country) special attention should be made to the towards the agriculture sector to cut down the poverty line. The mean concern of the agricultural land are, the rain fed cultivated land, culturable waste, crumble and uneconomical land holding, weak coordination between farmers and agricultural department, coordination trend in farming community, weak insufficient utilization of water resources, non availability of quality seed and fertilizers and absence of agricultural based processing units in rural areas like KPK. To improve the economic condition and control the poverty level, the GOVT should take positive steps of providing agricultural and high productivity livestock friendly environment to its citizen. Reduce the burden of taxes and prices of fertilizers. Provide good seeds, fertilizers and new methods of farming by introduce new machinery to the agricultural sector. The new machinery will increase cultivated land and reduce the calturable waste. These measures will certainly boast the economics condition of the country and cut down the poverty level.

(6): REFERENCES:

- S. K. Qureshi, Annual General Meeting of Pakistan Society of Development Economists 2006, Vol 1, 166-167.
- [2] A. M. Badmus and S. O. Ariyo, Asian Journal of Agricultural Science 2011, 3 (3), 171-176.
- [3] C. Chatfield, Chapman & Hall/CRC: Bath, UK, **1995**; fifth edition.
- [4] Government of Khyber Pakhtunkhwa. 2010-11. Crops Statistics of Khyber Pakhtunkhwa, Crops Reporting Services, Agriculture, Livestock Cooperative Department, Peshawar.
- [5] D. Jonathan and S. C. Kung, Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA, 2008; second edition.
- [6] S. Mehmood and Z. Ahmad, *Academy of Contemporary Research Journal* **2013**, Volume 2(1), pp. 10-15.
- [7] S. Mehmood, Journal of Contemporary Issues in Business Research 2012, Volume 1(3), pp.96-110.
- [8] K. Mustafa and A.Burhan, *Iinternational Journal of Agriculture and Biology* 2006, volume 8(4), pp. 459-462.

- [9] I. B. Najeeb, M. Khuda, Asif and S.A. Abid, Journal of Agricultural and Social Sciences 2005, 1(2): 120-122.
- [10] A. Nugroho and H. B. Simanjuntak, International Journal of Computer Science Issues 2014, 11 (3).
- [11] K. Prabakaran and C. Sivapragasam, *Golden Research Thoughts* **2013**, *3* (3).
- [12] N. M. F. Rahman, Journal of the Bangladesh Agriculture University 2010, Volume 8(1), pp. 103-112.
- [13] P. K. Sahu, International Journal of Statistical Sciences 2010, Volume 10, pp. 71-92.
- [14] A. Sajid, B. Nouman, F. Hina, Sarhad Journal of Agriculture 2015, Volume 31.
- [15] A. Saleem et al. *Pakistan European Academic Research* 2014, Volume 2, Issue 4.
- [16] N. Suleiman, and S. Sarong, International Journal of Agricultural Management & Development 2012, 2(2): 79-84.
- [17] C. W. J. Granger P. Newbold, Journal of Econometrics, 1974 2(2), 111-120.
- [18] M. D. Rahulamin and M. A. Razzaque, M.A. Journal of Financial Management and Analysis, 2000, 13(1): 74-80.