Mysore J. Agric. Sci., 52 (1) : 24-28, 2018

Fitting of Statistical Models for Paddy Crop Area of Selected Districts of Karnataka after BIAS Correction

SUNIL GAYAKAWAD, G. B. MALLIKARJUNA, K. N. KRISHNAMURTHY AND A. SATHISH
Department of Agricultural Statistics, Applied Mathematics and Computer Science
College of Agriculture, UAS, GKVK, Bangalore - 560 065
E-mail : gbm267@gmail.com

ABSTRACT

THE BIAS correction is done based on 17 years (1998-2015) data on area and production for Paddy crop at selected districts of Karnataka such as Bellary, Davangere and Raichur. The study revealed that MDM recorded least value of normalized root mean square error (NRMSE) for paddy crop area in all the selected districts compared to DM method. Auto correlation was found to be absent in corrected paddy crop area of all the selected districts. Thus, Model fitting was done using linear and nonlinear models and the results showed that cubic model was the best fit with high R^2, Adj. R^2 and least root mean square error (RMSE) value for all the districts.

Keywords : Statistical Models, difference method, modified difference method

ONE of the important subjects in agriculture is crop yield forecasting. Their use includes monitoring of agricultural production changes, planning of agricultural interventions, development of projects, development of early warning systems and preparation of macro-economic accounts. Poor agricultural data can lead to misallocation of scarce resources and policy formulations that fail to resolve critical development problems. The advance estimates of crop production are needed much before the actual harvest of the crops for making various decisions such as pricing, distribution, export and import *etc.* However, the final estimates of crop production which are based on area through complete enumeration and yield rate through Crop Cutting Experiments are made available much after the harvest of the crop. Therefore, there is great need for developing suitable and reliable models using information from different sources like agricultural inputs, meteorological data and remote sensing data for providing the reliable and timely forecast of crop Area/Production. Accurately estimating crop yields is never easy and is even more of a challenge in the context of farming systems that are characterized by small area holder farms that produce a wide range of diverse crops. Challenges that may occur include information on land use, intercropping, non-uniform plots in a wide range of sizes, not all planted area is harvested and significant post-harvest losses.

Crop Cutting Experimental methods that have greater precision at small areas, become invalidate at country level. Currently, the agriculture department officials visit the village or tahsil where they enquire about crop acreage and expected yield. Based on these types of sampling, the results are projected to acquire the acreage and yield information. This methodology, though prevalent for a long time is neither very accurate nor very scientific. It is having other limitations such as extremely tedious, time-consuming, costly, inconsistent and labor-intensive.

Alternatively, Remote sensing data has been used for forecasting purpose. It does not require close contact between the sensing organs and the external objects. It deals with remote sensing data attained through earth observation satellite. Remote sensing-based methods have already been proven as an effective alternative for mapping crop area and forecasting crop production. The benefits of remote sensing technology include: (i) spatial coverage over a large geographic area; (ii) availability during all seasons; (iii) relatively low cost, since some optical images are freely available although radar data are usually a bit costly; (iv) efficient analysis; (v) they provide information on a timely manner; and (vi) they are capable of delineating detailed spatial distributions of areas under crop cultivation. Problems that limit the current usefulness of remote sensing for developing countries include cloud coverage, the need for expensive ground truthing, the need for specialist knowledge, and the need of expensive image processing software (Reynolds *et al.* 2000). Under this situation precise estimate will be done only by smoothing the data generated for minimizing the variation. Smoothing of data has to be done by having appropriate bias correction to the data before having the proper prediction model. Gallego (2006) indicated that crop area estimation from satellite imagery is typically calculated using the product of the resolution of an image and the area of an agricultural feature delineated with a spectral classifier. They also revealed that, Co-location inaccuracies and considerable overlap between spectral categories can induce further error. Graham *et. al.* (2007) and Weiland *et. al.* (2010) have used delta method for the bias correction.

Material and Methods

The present study was based on the secondary data on Paddy crop area of selected districts of Karnataka such as Bellary, Davanagere and Raichur district. The data over a period of 17 years (1998-2015) was collected from the Directorate of Economics and Statistics (DES), Government of Karnataka and Karnataka State Remote Sensing Application Centre (KSRSAC), Bangaluru. The data obtained from the Directorate of Economics and Statistics (DES) is an observed data which is based on Crop Cutting Experiment. Remote estimates which are obtained from the Karnataka State Remote Sensing Application Center is a modeled data. Here, observed data is normally accurate compared to modeled data. But, because of limitations of area coverage, timely availability and so on, remote estimates of KSRSAC have been considered. Since the data generated by KSRSAC is having long resolution and pixels, it might have not been so accurate compared to Crop Cutting Experiment estimates i.e. bias might have been noticed. In this study two bias correction methods are used to bring modeled data (satellite estimates) close to observed data (crop cutting experiment estimates). Further, appropriate prediction models were evaluated for the bias corrected data by following the procedures of model fitting.

BIAS Corrections

Following two methods are applied to bring the modeled (remote estimates) data close to the observed ones. Each value is converted with the correction methods.

1. Difference method

In this method, averaged yearly difference (Δx) of observed and modeled values of cropped area is taken. The term Δx was considered as correction factor, which was added to the modeled uncorrected value (x $model_{uncor}$) to correct it (x $model_{cor}$) so that the values approach the observed ones.

$$Model_{cor} = Model_{uncor} + (\Delta x)$$

where Δx - Averaged difference of observed and modeled values of cropped area.

2. Modified difference method

The modified difference method (MDM) is similar to the difference method (DM); however, some statistical parameters were added to improve the correction function. For example, in area correction, μ and σ are added which aimed at shifting and scaling to adjust the μ and σ^2 (Leander and Buishand, 2007).

$$Model_{cor} = (Model_{uncor} + (\Delta x)) \times \left(\frac{\pm Area_{obs}}{\pm Area_{\mathrm{mod}}} \right)$$

where Δx - Averaged difference of observed and modeled values of a parameter

Validation of bias corrective measures

The correction capability of these measures were tested by coefficient of variation (CV%) expressed as Normalized Root Mean Square Error (NRMSE).

$$NRMSE = \frac{\left[\sum_{i=1}^{n} \frac{(P_i - O_i)^2}{n} \right]^{0.5}}{\bar{O}} \times 100$$

Where,

P_i = Predicted value; O_i = Observed value; $\bar{O}$ = Mean of observed value; n = Number of observations ranging from 1 to n

Model fitting for bias corrected model data

As the data collected is a time series data; Durbin Watson test for autocorrelation was performed to know the absence or presence of autocorrelation to the bias corrected data. Growth models (linear/ non-linear) such as Linear, Quadratic, Cubic, Exponential, MMF, Rational, Sinusoidal and Logisitic models or AR/MA/ARIMA models were considered depending on the outcome of Durbin-Watson test.

The best fit models for paddy crop area were assessed based on R^2 (Coefficient of determination), Adj. R^2 and RMSE values. The model with the highest R^2, Adj R^2 and the lowest RMSE value is considered as the best model.

Diagnostic checking : Different models obtained for various combinations of AR and MA individually and collectively are tested using the diagnostics checking such as Plot of residual ACF (plotting the ACF of residuals of the fitted model) and Non-significance of auto correlations of residuals via Portmonteau tests (Q-tests based on Chi-square statistics)-Box-Pierce or Ljung-Box texts

Box-Pierce statistic (a function of auto correlations of residuals) whose approximate distribution is Chi-square and is computed as follows:

$$Q = n\sum_{j=1}^{k} r_{(j)}^{\ 2}$$

where n is the number of observations in the series, r (j) is the estimated autocorrelation at lag j; k can be any positive integer and is usually around 20. Q follows Chi-square with (k-m-1) degrees of freedom where m-1 is the number of parameters estimated in the model. A modified Q statistic is the Ljung-box statistic which is given by

$$q = n(n+2)\sum_{j=1}^{k} \frac{r_{(j)}^{\ 2}}{(n-j)}$$

The modeled data (Remote sense data) was subjected to bias correction using 2 methods *viz.*, Difference method (DM) and Modified difference method (MDM) for data on Paddy crop area. To identify suitable methodology to smoothen the modeled data, NRMSE for each (Model uncorrected, Model corrected by DM and MDM methods) was worked out.

Results and Discussion

The Table I showed that calculated NRMSE values for Paddy crop area is least for MDM in all the Districts. This indicated that MDM was a better bias correction method for getting smoothening data compared to DM. Kim *et al.* (2016) indicated that, raw satellite-based rainfall estimates require a post processing of bias correction before data can be useful for forecasting and impact studies. To address this issue, they suggested several bias correction methods.

Table I

NRMSE values for paddy crop area (ha) for selected districts of Karnataka

District	Model uncorrected	Model corrected	
		DM	MDM
Bellary	8.12	7.72	7.36
Davanagere	7.81	7.69	7.50
Raichur	6.20	5.85	5.28

Bias corrected time series data of Paddy crop area has been checked for the Auto correlation. Results of Autocorrelation test made with the Durbin-Watson test are presented in Table II.

Table II

Durbin-Watson values of paddy crop area

District	DW Value
Bellary	1.83
Davanagere	2.01
Raichur	1.78

From Table II, it could be noticed that Durbin-Watson value for area of all the Districts shows absence of autocorrelation (1.5 < DW< 2.5). This leads to fitting of linear and nonlinear models for the Paddy crop area.

Results of Linear and nonlinear models *viz.*, linear, quadratic, cubic, exponential, MMF, rational,

TABLE III

Linear and non-linear models of paddy crop area in Bellary

Model	Parameters				Criteria		
	A	B	C	D	R^2	Adj.R^2	RMSE
Linear	67.3991 **	3.4601 **			0.5804 **	0.5524	13.67
Quadratic	56.2286 **	6.9876 **	-0.1959		0.6158 **	0.5609	13.79
Cubic	82.1557 **	-8.1744	1.8509 **	-0.0758 **	0.7110 **	0.6443	11.95
Exponential	1.8348 **	0.0162 **			0.5729 **	0.5444	14.10
MMF	10.429	0.866	101.014 **	83.808 **	0.1980	-0.0693	22.78
Rational	3350795	-1E-07	-133765	2580.31	0.5690 **	0.4253	16.71
Sinusoidal	98.642 **	6.104	4.028 **	-7.356	0.3800	0.1733	24.95
Logistic	1.198 **	1.981 **	2.962 **		0.5800 **	0.4831	17.50

Fitted model: $\hat{Y}_{Area} = 82.1557 - 8.1744X - 1.8509X^2 - 0.0758X^3$

The value of the criterion for a model with bold numerals shows that the model is better than the other models with respect to that criterion.

**indicate significant at 1% level of probability.

sinusodial and logistic fitted to the corrected data on area of Paddy crop for Bellary, Davanagere and Raichur districts are presented in Tables III, IV and V.

The results presented in Table III, IV and V revealed higher R^2, adj.R^2 least RMSE for the Cubic model pertaining to Bellary, Davangere and Raichur districts. This indicated that, cubic model can be chosen for forecasts of Paddy crop area in all the Districts. Hassan *et. al*. (2011) revealed that among the entire models, cubic model was selected as the best fitted model. They claimed that by using the cubic growth model, coarse rice prices can be forecasted.

From the above outcomes it could be inferred that, Crop cutting experiments have its own limitations mainly in the coverage of cropped area in a limited period and unable to provide the projections. Alternatively, remote sensing data covers larger area.

TABLE IV

Linear and non-linear models of paddy crop area in Davanagere

Model	Parameters				Criteria		
	A	B	C	D	R^2	Adj.R^2	RMSE
Linear	97.6595	1.6876			0.1279	0.0698	21.58
Quadratic	120.4187	-5.4994	0.3992		0.2640	0.1589	19.82
Cubic	158.023 **	-27.4903**	3.3680 **	-0.1099 **	0.4720 **	0.3502	17.14
Exponential	4.9787 **	0.007			0.1099	0.0506	23.15
Rational	-120000000	-170000000	-2268916	56804.22	0.3180	0.0907	21.82
Logistic	112.291 **	-4133.81**	-1238.2		0.3812	0.2369	20.41
Sinusoidal	112.929 **	-6.195	3.353 **	4.505	0.3241	0.0933	26.00

Fitted model: $\hat{Y}_{Area} = 158.023 - 27.4903X + 3.3680X^2 - 0.1099X^3$

The value of the criterion for a model with bold numerals shows that the model is better than the other models with respect to that criterion.

** indicate significant at 1% level of probability.

TABLE V

Linear and non-linear models of paddy crop area in Raichur

Model	Parameters				Criteria		
	A	B	C	D	R^2	Adj.R^2	RMSE
Linear	116.1896 **	3.0417 **			0.2461 **	0.1958	26.0777
Quadratic	118.3712 **	2.3527	0.3894		0.2468	0.1392	24.0649
Cubic	164.426 **	-24.5799	3.6741 **	-0.1346 **	**0.4118 ****	**0.2761**	**23.0333**
Exponential	2.05792 **	0.009826 **			0.2341 **	0.183	28.35
Logistic	143.565 **	1181.198 **	907.625 **		0.3025	0.1415	28.55
Sinusoidal	143.209 **	-10.821	2.876 **	-1.258	0.2812	0.0416	30.37

Fitted model: $\hat{Y}_{Area} = 158.023 - 27.4903X + 3.3680X^2 - 0.1099X^3$

The value of the criterion for a model with bold numerals shows that the model is better than the other models with respect to that criterion.

** indicate significant at 1% level of probability.

But, in view of its large resolution, bias may be observed in the data collected. The study revealed that MDM recorded least value of NRMSE for Paddy crop area collected by the Remote sensing in all the selected districts. This indicated MDM can be used for smoothening. Autocorrelation was found to be absent in corrected paddy crop area of all the selected districts. Thus, Model fitting was done using linear and nonlinear models and the results showed that Cubic model was the best fit with high R^2 value for all the districts.

REFERENCES

GALLEGO, F. J., 2006, Review of the main remote sensing methods for crop area estimates agriculture unit, *Proc. Remote Sensing Support to Crop Yield Forecast and Area Estimates*, ISPRS Archives XXXVI-8/W48, pp. 65-70.

GRAHAM, L. P., ANDRÉASSON, J. AND CARLSSON, B., 2007, Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods - A case study on the lule river basin. *Clim. Change*, **81** : 293 - 307.

HASSAN, M. F., ISLAM, M. A., IMAM, M. F. AND SAYEM, S. M., 2011, Forecasting coarse rice prices in Bangladesh. *Prog. Agric.*, **22** (1&2) : 193 - 201.

KIM, K. B., BRAY, M. AND HAN, D., 2016, An improved bias correction scheme based on comparative precipitation characteristics. *Hydrological Processes*, **29** (9) : 2258 - 2266.

LEANDER, R. AND BUISHAND, T. A., 2007, Resampling of regional climate model output for the simulation of extreme river flows. *J. Hydrology*, **332** (3) : 487 - 496.

REYNOLDS, C. A., YITAYEW, M., SLACK, D. C., HUTCHISON, C. F., HUELE, A. AND PETERSEN, M. S., 2000, Estimating crop yields and production by integrating FAO crop specific water balance model with real-time satellite data and ground-based auxiliary data. *Int. J. Remote Sensing*, **21** : 3487 - 3508.

WEILAND, S. F. C., BEEK, V. L. P. H., KWADIJK, J. C. J. AND BIERKENS, M. F. P., 2010, The ability of a GCM-forced hydrological model to reproduce global discharge variability, *Hydrol. Earth Sys. Sci.*, **14** (8) : 1595 - 1621.

(*Received :* January, 2018 *Accepted :* March, 2018)