Range Mgmt. & Agroforestry 35 (1) : 43-50, 2014
ISSN 0971-2070

Usefulness of *in-situ* hyperspectral data to develop prediction model of dual purpose Sorghum (*Sorghum bicolor* (L.) Moench) grown under semi-arid condition of India

D. Deb, J. P. Singh* and S. K. Rai
Indian Grassland and Fodder Research Institute, Jhansi-284003 (UP), India
*Corresponding author email: jpsingh.igfri@gmail.com
Received: 2nd December, 2013 Accepted: 12th February, 2014

Abstract

Crop yield forecasts are extremely useful in formulation of policies regarding stock, distribution and supply of agricultural produce. In this paper, an effort was made to develop regression models for biomass and seed yield production of dual purpose sorghum (*Sorghum bicolor*) incorporating canopy spectral reflectance indices collected during three different growth stages of the crop in 2011 and 2012. The results showed that the correlation between the reflectance at each wavelength and the aboveground biomass has the maximum negative correlation coefficient (r= - 0.21, p< 0.05) at 690nm. The r values changed sharply from 690nm to 750nm. The maximum positive correlation coefficient (r=0.61, p<0.05) comes around 800nm in the Near Infra-Red (NIR) portion. In order to determine the plant stages more appropriate for yield forecasting, the indices more sensitive to yield variations, pearson correlation coefficient was calculated between the indices and grain, fodder yield for each sampling date. It was found that at flowering stage, almost all the indices were highly sensitive to both grain and fodder yield variations. Different narrowband vegetative indices like Normalized Difference Vegetation Index (NDVI) - based VIs ($NDVI_1$, $NDVI_2$, $NDVI_3$ and $NDVI_4$) and Ratio-based VIs (RVI_1 and RVI_2) were calculated to estimate the above ground biomass. Applying linear and non-linear regression approach it was found out that for both dry fodder yield and grain yield could be well estimated using non-linear function $Y = a * X^b$ using the ratio indices *i.e.,* X=RVI_2(735,706) and the root mean square error (RMSE) is minimum and R^2 is maximum for this function.

Keywords: Canopy reflectance, Dual purpose Sorghum, Fodder, Regression, Vegetative indices, Yield forecast.

Abbreviations: DAS: Days after sowing; **GPS:** Global positioning system; **LAI:** Leaf area index); **NDVI:** Normalized difference vegetation index; **NIR:** Near infra-red; **RMS:** Root mean square; **RMSE:** Root mean square error; **RVI:** Ratio-based vegetative index.

Introduction

Sorghum (*Sorghum bicolor L. Moench*) is fast-growing, warm weather annual crop that can provide plenty of feed in mid-summer during lean period. Amongst different variants of this crop, dual purpose one is extensively grown at semi-arid part of India for both grain and fodder purpose. So, reliable and timely forecast of both grain and fodder yield of the crop is important for proper, foresighted and informed planning to overcome several uncertainties associated with agriculture. For breeding programme also early prediction of crop yield can be an important tool for identifying promising genotypes. The large scale determination of grain and fodder yield using plant characters requires collection of data from farmer's fields on characters which are not easy to measure without involving much expertise, cost and sophisticated instruments. Some characters contributing significantly towards yield may not find place in the model due to these limitations (Agarwal, 2006). So it requires including other variables like hyperspectral indices in the model to take care of such variables indirectly. Using hyperspectral reflectance indices to forecast crop yield is a fast, cheap and accurate technique. This process of crop yield prediction mainly depends on crop spectral characteristics and their growth and yield (Shu *et al.*, 2006).

A few reports are available on the hyperspectral reflectance of sorghum to explore the relation between canopy reflectance and important physiological characteristics of the crop. Richardson and Wiegand (1992) performed multisite analyses of spectral-biophysical data for sorghum and Duli *et al.* (2005) made a study on Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum, but till date no statistical model for predicting yield of the crop using hyperspectral reflectance parameter is available. Keeping this in view, an attempt has been made in the present study to find out the most suitable hyperspectral Indices to develop regression models for prediction of dual purpose Sorghum grain and fodder yield.

Materials and Methods

Experiment was carried out at Central Research Farm of Indian Grassland and Fodder Research Institute, Jhansi in 2011 and 2012 on dual purpose Sorghum genotypes grown in All India Coordinated Sorghum Improvement Project. The experiment comprised a Randomised Block Design with 3 replicates and included 20 dual purpose Sorghum genotypes. The plots were of 18.25 m^2 (6.75 m x 2.7 m) (6 rows 45 cm apart), the GPS location of the experiment site was between 25.52^{0}N 78.55^{0}E, 25.52^{0}N 78.55^{0}E, N25.52^{0}N ^{0} 78.55^{0}E, N25.52^{0}N 78.55^{0}E; altitude 212.75m. The standard agronomic practises of sowing date and amount of fertilizer applied to the experiment were followed. Canopy spectral reflectance on flowering, filling and ripening stages were measured by using portable Reflectance recording device, Field Spec® (Analytical Spectral Devices, Inc., Boulder CO, USA) which measures radiance in 751 contiguous bands from 325 to 1075 nm wavelengths. The field of view is 25^0. Measurements were taken on clear, sunny days at midday (11:00h to 14:00h). The sensor was held at a height of 1m above the ground to take readings from nadir position. Three sampling points (each being the average of 25 scans) in each test area were recorded, and the average value was used as hyper-spectral reflectance of this treatment. The calibration was performed with a white Spectral on reference panel during measurement. There after these narrow bands spectral reflectance data are used to calculate several Vegetative Indices (VI). The goal of the research was to find out the optimal hyperspectral narrow band VIs that best help model grain and dry fodder yield of dual purpose sorghum. In the past also researchers have used reflectance from single narrow bands (Mariotti *et al*., 1996), derivatives of reflectance spectra (Elvidge and Chen, 1995; Curran *et al.* 1991), various ratio indices (Aoki *et al*., 1981; Carter, 1994; Lichtenthaler, 1987) or a combinations of these (Thenkabail *et. al.* 2000) and linear mixture modelling approach (McGwire *et al.* 2000). Since studies revealed that these VIs tend to asymptotically saturate in response to high aboveground biomass (Thenkabail *et al.* 2000; Mutanga and Skidmore, 2004), we can infer that linear or non-linear relationship exists between VIs and the biomass for the study site. Linear and non-linear regression was carried out between grain and fodder yield and each of the calculated VIs. The regression analysis was performed using SAS 9.3 software.

Results and Discussion

The typical temporal change of sorghum canopy reflectance spectral during three stages of crop season are shown in Fig.1. These stages were 50 DAS, Physiological maturity and before maturity, 120 DAS. Reflectance characteristics of the crop were almost similar to those of general green plant. Liu *et al.* (2006) observed green peak of 550nm and red light low valley of 680nm in visible light region of 400 to 700nm as well as plateau area of 780-1100nm in near infrared region. At 50 DAS, the reflectance at 555 and 675nm was lower comparatively to other stages because of the greenness of the plant chlorophyll content. As the crops grew, particularly at 20/25 DAS before harvesting the green region showed maximum reflectance near 554 nm. However, when crop was almost dry, due to non-availability of chlorophyll content reflectance was higher in red region (670-690nm). Correlation between the reflectance at each wavelength and the aboveground biomass was worked out using Pearson correlation coefficients at each growth stage from flowering to grain maturity stage.

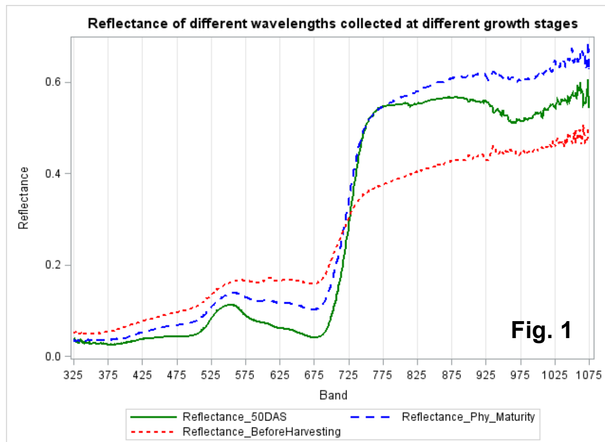


Fig. 1

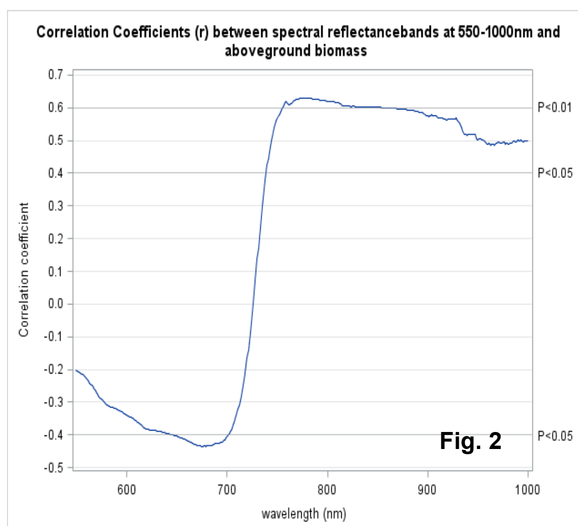


Fig. 2

Table1. Definition of vegetation indices used in this study

Vegetative Index	Equation	Advantages	References
Normalized Difference Vegetative Index 1	$\frac{P_{800} - P_{670}}{P_{800} + P_{670}}$	Enhance between vegetation and soil while minimizing illumination effects	Rouse *et al.* (1974)
Normalized Difference Vegetative Index 2	$\frac{P_{755} - P_{746}}{P_{755} + P_{746}}$	Best correlation with high biomass	Mutanga and Skidmore (2004)
Normalized Difference Vegetative Index 3	$\frac{P_{918} - P_{682}}{P_{918} + P_{682}}$	Best correlation with biomass	Thenkabail *et al.* (2000)
Normalized Difference Vegetative Index 4	$\frac{P_{708} - P_{565}}{P_{708} + P_{565}}$	Best correlation with biomass	Hansen and Schjoerring (2003)
Ratio Vegetation Index 1	$\frac{P_{833}}{P_{614}}$	Enhance between vegetation and soil while minimizing illumination effects	Pearson and Miller (1972)
Ratio Vegetation Index 2	$\frac{P_{755}}{P_{706}}$	Best correlation with high biomass	Mutanga and Skidmore (2004)

Table 2. Correlation between canopy reflectance hyper-spectral parameters

Crop stages	Hyperspectral Indices	Pearson correlation coefficient	
		Grain yield (Kg/ha)	Dry fodder yield (Kg/ha)
Boot stage (50 DAS)	$NDVI_1$(800,670)	0.12863	0.26291
	$NDVI_2$(755,746)	0.05758	0.19554
	$NDVI_3$(918,682)	0.1273	0.25859
	$NDVI_4$(708,565)	-0.24545	-0.34168*
	RVI_1(833,679)	0.11383	0.25089
	RVI_2(735,706)	0.04601	0.18016
Flowering stage (80 DAS)	$NDVI_1$(800,670)	0.71156**	0.67924**
	$NDVI_2$(755,746)	0.69193**	0.63342*
	$NDVI_3$(918,682)	0.69071**	0.65617**
	$NDVI_4$(708,565)	-0.56031*	-0.60774*
	RVI_1(833,679)	0.72596**	0.67465**
	RVI_2(735,706)	0.73906**	0.68711**
Grain maturity stage (120 DAS)	$NDVI_1$(800,670)	-0.07939	-0.03091
	$NDVI_2$(755,746)	-0.11771	-0.07957
	$NDVI_3$(918,682)	-0.09901	-0.05324
	$NDVI_4$(708,565)	-0.07762	-0.11513
	RVI_1(833,679)	-0.08451	-0.03648
	RVI_2(735,706)	-0.15108	-0.09554

* $p< 0.05$, ** $p < 0.01$

The maximum negative correlation coefficient ($r= - 0.21$, $p< 0.05$) appears around 690nm. The r values change sharply from 690nm to 750nm (Fig. 2). The maximum positive correlation coefficient ($r=0.61$, $p<0.05$) comes around 800nm in the NIR portion. Thereafter, the spectral reflectance indices were calculated (Table 1). These vegetation indices based on narrow spectral bands are reported to be well correlated with variety of vegetation parameters such as LAI, Biomass, chlorophyll concentration and photosynthetic activity. Table 1 summarizes the narrow band vegetation indices used to estimate the aboveground biomass. The selected VIs can be grouped into 2 categories: NDVI- based VIs ($NDVI_1$, $NDVI_2$, $NDVI_3$ and $NDVI_4$ and Ratio-based Vis (RVI_1 and RVI_2).

As several studies suggest that many VIs tend to asymptotically saturate in response to high above ground biomass (Thenkabail *et al.,* 2000; Mutanga and Skidmore, 2004), we can infer that there might be a linear or nonlinear relationship between the VIs and biomass in the present study. Although linear model equation is the most popular regression equation used in predicting various biological parameters, the biological systems never follow linearity hence has little use towards prediction. Therefore, keeping the same thing in mind some curves are also tried using non-linear regression analysis between the above ground biomass and each of the VIs listed in table 1 at each growth stage. In order to determine the plant stages more appropriate for yield forecasting, the indices more sensitive to yield variations, Pearson Correlation Coefficient was calculated between the indices and grain, fodder yield for each sampling date (Table 2).

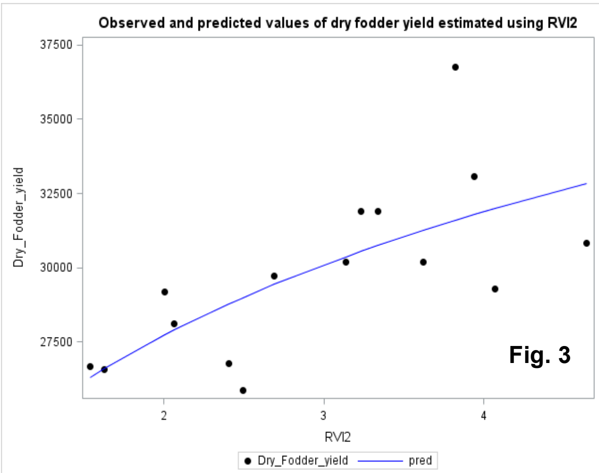


Fig. 3

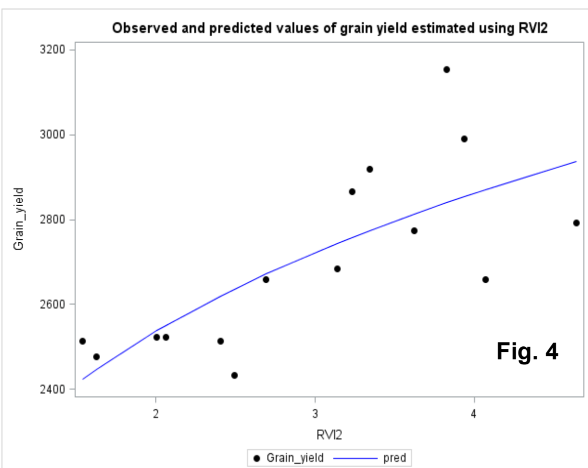


Fig. 4

As the above table suggests the relationships between each spectral index and grain/fodder yield, we found a little association between them at both boot stage (50 DAS) and grain maturity stage (120 DAS). At flowering stage, almost all of the indices were found highly sensitive to both grain and fodder yield variations. So, usefulness of spectral reflectance indices to forecast sorghum grain or fodder yield depends on the sampling date. It is well known that non-green components contribute to the canopy spectral reflectance; hence vegetation indices have been reported to vary due to the presence of non-green vegetation (Van Leeuwen and Huete, 1996). Liu *et al.* (2004) and Feng *et al.* (2007) found that correlations between characteristic spectral indices and wheat yield from tilling to maturity are high and yield prediction models are established accordingly. Table 2 shows that flowering stage (80 DAS) was the most appropriate developmental stage for yield assessment. So, the data collected in this stage is used to develop prediction models. Six non-linear functions were tried using all these VIs collected at this developmental stage as explanatory variables separately for above ground biomass. Non-linear regression procedure of SAS 9.3 was applied and the parameters were estimated using non-linear least square method following modified Gauss-newton method (Draper and Smith, 1981). Normality of the explanatory variables was checked by normality test prescribed by Shapiro and Wilk (Rao *et al.,* 1985). As listed in table 3, all these models were tried using the indices, which were highly sensitive to yield and fodder variations and the root mean square error (RMSE) and R^2 are also given.

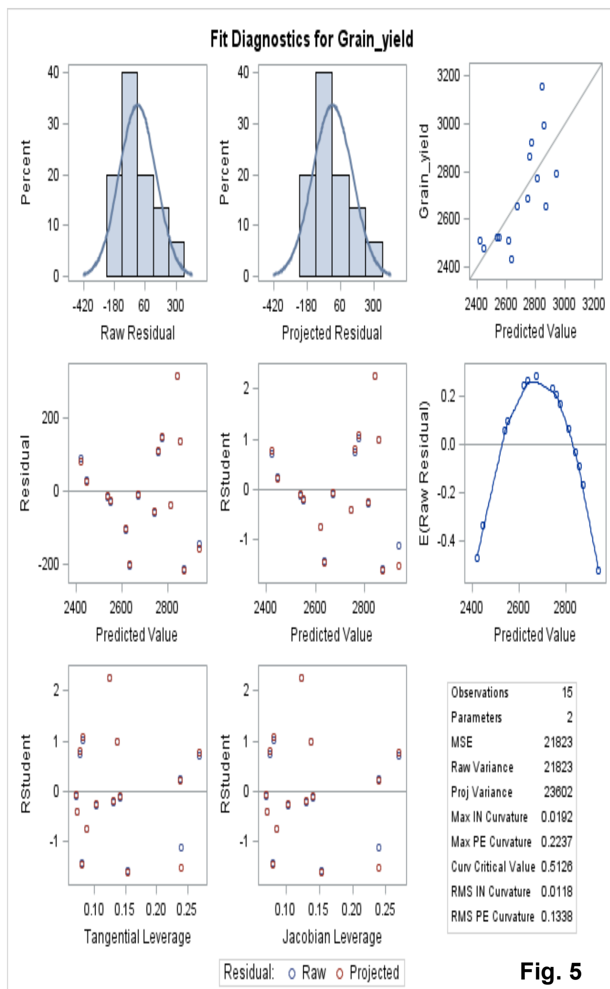


Fig. 5

Table 3. Different models tried between the VIs and grain/fodder yield at flowering stage

		Dry fodder yield (kg/ha)		Grain yield (Kg/ha)	
Model	**Hyper-spectral indices**	**R^2**	**RMSE**	**R^2**	**RMSE**
Linear (Y = a + b * X)	$NDVI_1$(800,670)	0.46	2207.79	0.50	155.50
	$NDVI_2$(755,746)	0.40	2327.80	0.47	159.78
	$NDVI_3$(918,682)	0.43	2270.05	0.47	160.04
	$NDVI_4$(708,565)	0.36	2388.96	0.50	155.50
	RVI_1(833,679)	0.47	2185.64	0.52	152.20
	RVI_2(735,706)	0.47	2185.64	0.54	149.09
$Y = a * X^b$	$NDVI_1$(800,670)	0.44	2234.19	0.48	159.05
	$NDVI_2$(755,746)	0.42	2290.72	0.48	158.93
	$NDVI_3$(918,682)	0.42	2285.61	0.46	162.36
	$NDVI_4$(708,565)	0.41	2294.10	0.34	178.57
	RVI_1(833,679)	0.47	2185.284	0.53	150.98
	RVI_2(735,706)	**0.48**	**2154.78**	**0.55**	**147.72**
$Y - a + X^b$	$NDVI_1$(800,670)				
	$NDVI_2$(755,746)	ð -	ð -	ð -	ð -
	$NDVI_3$(918,682)				
	$NDVI_4$(708,565)	0.44	2233.61	0.36	175.77
	RVI_1(833,679)	0.28	2551.37	0.42	168.14
	RVI_2(735,706)	0.20	2680.42	0.35	177.42
$Y = a + b * X^c$	$NDVI_1$(800,670)	ð -	ð -	ð -	ð -
	$NDVI_2$(755,746)	0.42	2283.95	0.48	164.63
	$NDVI_3$(918,682)	ð -	ð -	ð -	ð -
	$NDVI_4$(708,565)				
	RVI_1(833,679)	0.47	2274.51	0.30	191.82
	RVI_2(735,706)	0.48	2242.01	ð -	ð -
$Y = a - b * C^x$	$NDVI_1$(800,670)	0.49	2232.47	0.56	151.77
	$NDVI_2$(755,746)	ð -	ð -	0.49	163.15
	$NDVI_3$(918,682)	0.44	2333.85	0.51	161.15
	$NDVI_4$(708,565)	ð -	ð -	ð -	ð -
	RVI_1(833,679)	ð -	ð -	ð -	ð -
	RVI_2(735,706)	ð -	ð -	ð -	ð -
$Y = \frac{X}{a + b * X}$	$NDVI_1$(800,670)	0.43	2264.02		
	$NDVI_2$(755,746)	ð -	ð -		
	$NDVI_3$(918,682)	0.41	2303.41	ð -	ð -
	$NDVI_4$(708,565)	0.44	2243.03		
	RVI_1(833,679)	0.46	2215.71	0.49	156.61
	RVI_2(735,706)	0.48	2165.79	ð -	ð -

ð -: does not meet the convergence criterion
* = Multipilication

Table 4. Global nonlinearity measures

	In case of grain yield	In case of dry fodder yield
Max Intrinsic Curvature	0.0192	0.0255
RMS Intrinsic Curvature	0.0118	0.0156
Curvature Critical Value	0.5126	0.5126
Max Parameter-Effects Curvature	0.2237	0.1801
RMS Parameter-Effects Curvature	0.1338	0.5126

It can be observed from the table that for both dry fodder yield and grain yield estimation, RMSE is minimum and R^2 is maximum for function $Y = a * X^b$. The following two graphs are showing the data points for observed and predicted values using the nonlinear model $Y = 2248.2 * X^{0.1743}$ and $Y = 24132.4 * X^{0.2010}$ for grain yield and dry fodder yield respectively.

If we study the post convergence diagnostics for the proposed model as revealed by the SAS procedure (SAS, 2011) from the table below (Table 4) we find that the maximum and RMS intrinsic curvatures compared to the critical curvature value suggests that intrinsic curvature property of the model is not a highly non-linear one. As such performing diagnostics with the raw residuals can be suggestive towards the adequacy of the model.

The partial results from this SAS procedure run are shown in panel charts (Figs 5 and 6) made on both the dependent variables which support the previously mentioned expectations, the low correlation between raw residuals with the predicted values, non-significant difference between tangential and Jacobin leverages and projected residuals overcome some of the shortcomings of the raw residuals.

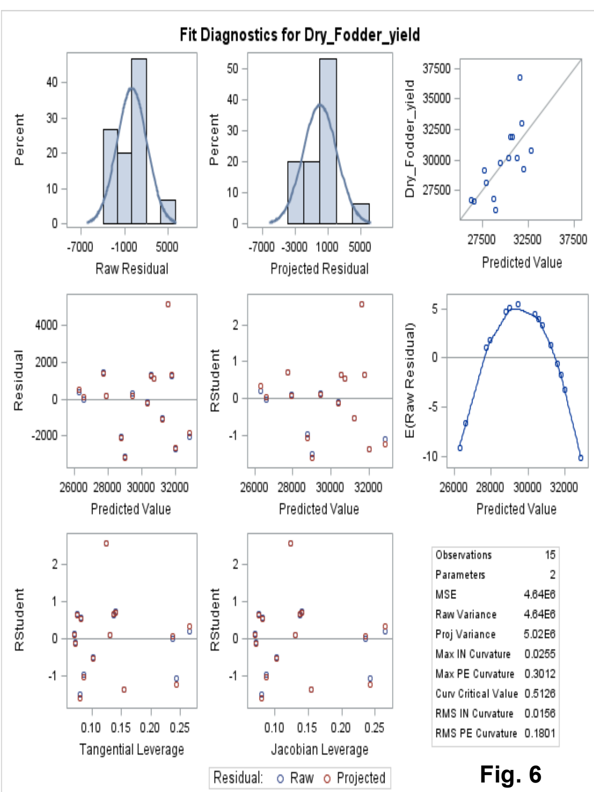


Fig. 6

To check the model adequacy, validity of assumptions of regression analysis has been checked here using some diagnostic methods, as gross violations of the assumptions may yield an unstable model in the sense that a different sample could lead to a totally different model with opposite conclusions and we usually cannot detect these departures from underlying assumptions by merely examining the standard summary statistics like *t* or *F* statistics or R^2 (Montgomery *et al.,* 2003). These diagnostics methods are primarily based on the study of model residuals.

To validate that the residuals have constant variance, the residuals are plotted against the estimate (Figs 7 and 8). The residuals plotted against the predicted values show no trends or patterns. If there are any patterns like ñconeò or ñsphereò shapes, that will indicate lack of model fit and unequal variances.

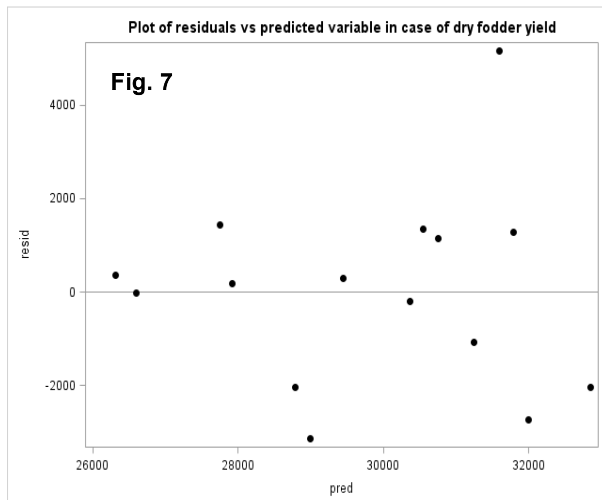


Fig. 7

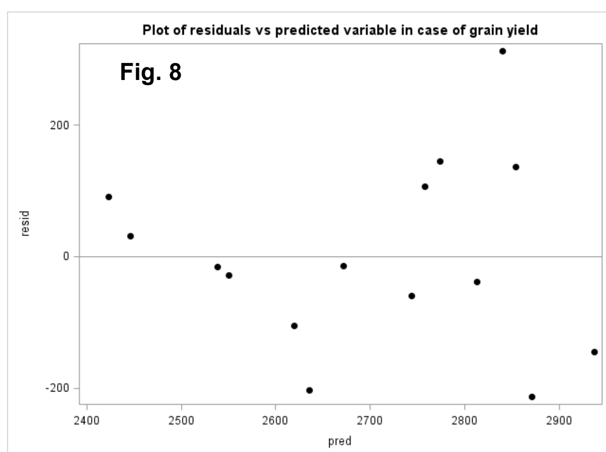
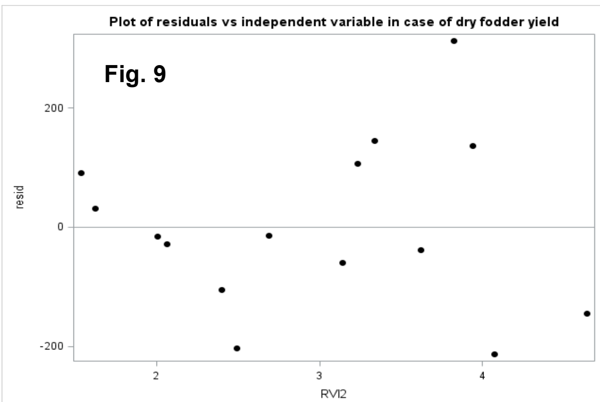
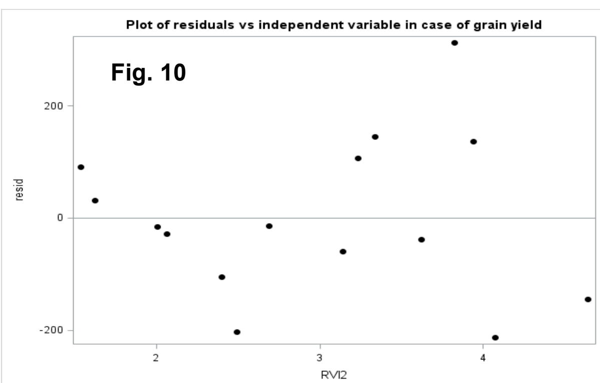
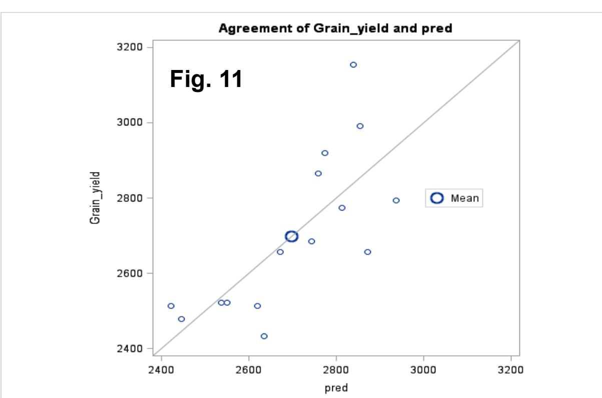


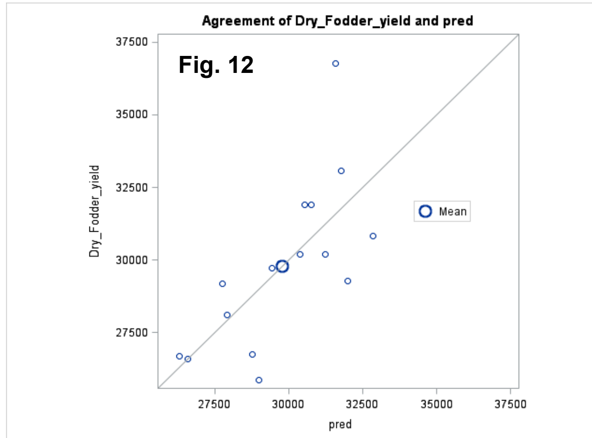
Fig. 8

To confirm that the residuals are not continuously being over/under estimated, the residuals are again plotted against the independent or explanatory variable (Figs 9 and 10) and Paired t-test on the observed and predicted values (Figs 11 and 12) was applied to find out the agreement between the observed grain yield and dry fodder yield with that of the predicted one (Ajit and Parsad, 2011). It is clear from above agreement plots of predicted and observed values that almost all predicted values are in close agreement with the observed ones. Moreover, the t-value of the difference between predicted and observed values is almost zero with non-significant differences again implies that predicted values are very close to observed values.









Conclusion

The capacity of hyperspectral reflectance indices to forecast grain and fodder yield of dual purpose sorghum is of much importance for researchers in the field of agriculture and production and maintenance of livestock as well as for the policy makers to have prior information about the produce. Our results showed that, in spite of some indices being sensitive to yield variations in previous growth stages, flowering stage (80 DAS) was the most appropriate stage for yield assessment. Hyperspectral reflectance measured at this stage and particularly the Vegetative Index RVI_2 (735,706) has been proved to be a very useful tool for dual purpose sorghum grain and dry fodder yield estimation. So here attempts were made to develop prediction model for both the yield parameters using RVI_2 (735,706). The different results showed that predicted values of the proposed model were in preferred conformity with actual observed values and residual diagnostics established that the model can be reliable to predict dual purpose sorghum actual grain and fodder yield.

References

Agarwal, Ranjana. 2006. *Forecasting techniques in crops*, Indian Agricultural Statistical Research Institute, New Delhi.

Ajit and Parsad. 2011. Tree growth and biomass modeling: An Illustration using SAS. web.iasri.res.in/sscnars/Agro/2-Tree-Growth-Modelling-SAS-illustration.pdf.

Aoki, M., K. Yabuki and T. Totsuka. 1981. An evaluation of chlorophyll content of leaves based on the spectral reflectivity in several plants. *Res. Rep. Natl. Inst. Environ. Stud. Jpn.* 66: 125-130.

Carter, G. A. 1994. Ratios of leaf reflectance in narrow wavebands as indicators of plant stress. *Int. J. Remote Sens.* 15: 697-703.

Curran, P. J., J. L. Dungan, B. A. Macler and S. E. Plummer. 1991. The effect of a red leaf pigment on the relationship between red-edge and chlorophyll concentration. *Remote Sensing of Environment*. 35: 69-75.

Draper, N. R. and H. Smith. 1981. *Applied Regression Analysis*, John wiley & Sons, New York 709.

Duli, Z., K. R. Reddy, V. G. Kakania and V. R. Reddy. 2005. Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum. *Eur. J. Agron.* 22: 391-403

Elvidge, C. D. and Z. Chen. 1995. Comparison of broad-band and narrow-band red and near infrared vegetation indices. *Remote Sensing of Environment*. 54: 38-48.

Feng, W., Y. Zhu, Y. C. Tian, X. Yao, T. C. Guo and W. X. Cao. 2007. Model for predicting grain yield with canopy hyper-spectral remote sensing in wheat. *Journal of Triticeae Crops.* 27 (6): 1076-84.

Hansen, P. M. and J. K. Schjoerring. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment.* 86: 542-553.

Lichtenthaler, H. K. 1987. Chlorophylls and carotenoids: pigments of photosynthetic membranes. *Methods in Enzymology.* 148: 350-382.

Liu, L. Y., J. H. Wang, W. J. Huang, CH. J. Zhao, B. Zhang and Q. X. Tong. 2004. Improving winter wheat yield prediction by novel spectral index. *Transactions of the CSAE.* 20(1): 172-175.

Liu, Z. Y., J. F. Huang, X. H. Wu, Y. P. Dong, F. M. Wang and P. T. Liu. 2006. Hyper-spectral remote sensing estimation models onvegetation coverage of natural grassland. Chinese Journal of Applied Ecology. 17(6): 997-1002.

Mariotti, M., L. Ercoli and A. Masoni. 1996. Spectral properties of iron-deficient corn and sunflower leaves. *Remote Sensing of Environment*. 58: 282-288.

McGwire, K., T. Minor and L. Fenstermaker. 2000. Hyperspectral Mixture Modelling for Quantifying Sparse Vegetation Cover in Arid Environments. *Remote Sensing of Environment*. 72(3): 360-374.

Montgomery, D. C., E. A. Peck and G. G. Vining. 2003. *Introduction to linear regression analysis*, Third edition, John Wiley & Sons.

Mutanga, O. and A. K. Skidmore. 2004. Narrow band vegetation indices solve the saturation problem in biomass estimation. *Int. J. Remote Sens*. 25: 1-16.

Pearson, R. L. and L. D. Miller. 1972. *Remote mapping of Standing Crop Biomass for Estimation of the Productivity of the Short-grass Prairie, Pawnee National Grasslands, Colorado* (Ann Arbor, MI: ERIM).

Rao, C. R., S. K. Mitra, A. Maithal and K. G. Ramamurthy. 1985. *Formula and Tables for Statistical Works*. Statistical Publishing Society, Calcutta.

Richardson, A. J. and C. L. Wiegand. 1992. Multisite analyses of spectral-biophysical data for Sorghum. *Remote Sensing of Environment.* 44: 71-82.

Rouse, J. W., R. H. Haas, J. A. Schell, D. W. Deering and J. C. Harlan. 1974. Monitoring the vernal advancements and retro gradation of natural vegetation. NASA/GSFC, Final Report, Greenbelt, MD, USA.

SAS Institute, 2011 SAS/STAT Userþs guide and SAS 9.3 Help and Documentation (Cary, North Carolina: SAS Institute Inc.).

Shu, J., Q. Wang and J. Sun. 2006. Applications of hyper-spectral remote sensing. *Journal of East China normal university.* 4: 1-10.

Thenkabail, P. S., R. B. Smith and E. DePauw. 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*. 71: 158-182.

Van Leeuwen, W. J. D. and A. R. Huete. 1996. Effects of standing litter on the biophysical interpretation of plant canopies with spectral indices. *Remote sensing of Environment*. 55: 123-138.