Gezira J. of agric. sci. 17(2): 211-229 (2019)

Forecasting rainfed sorghum yield using satellite-derived vegetation indices with limited ground-based information in Gadarif region, eastern Sudan

M.A. Bashir[1], H. Tanakamaru[2], A. Tada[2], A.E. Hassan[3], A.E. Khalid[4] and H.A. Sirelkhatim[1]

[1] Gezira Research Station, Agricultural Research Corporation, Wad Medani, Sudan.

[2] Graduate School of Agricultural Science, Kobe University, Japan.

[3] Gedarif Research Station, Agricultural Research Corporation, Gedarif, Sudan.

[4] Dongla Research Station, Agricultural Research Corporation, Dongola, Sudan.

ABSTRACT

A practical crop growth and yield monitoring system based on satellite data is required and fundamental not only for precision farming, but also very useful for global food security enhancement. This study was performed to determine the optimal vegetation index and also to identify the best time for making a reliable crop yield forecast in one of the major sorghum-growing region (Gedarif State, Sudan). The study was also aimed to develop a simple yield prediction model which was later validated using an official yield data acquired during 2013 and 2014 cropping seasons from the Department of Information System and Statistical Analysis of the State Ministry of Agriculture, Gedarif State. The study used NASA's multi-temporal MODerate resolution Imaging Spectroradiometer (MODIS) land products with limited ground information. Relationship between sorghum yield and crop reflectance indicated that normalized difference vegetation index (NDVI) at the third dekad of September (Sep.III) is the most appropriate to develop sorghum yield prediction model with higher R^2 value of 0.77 ($p<0.05$) compared to other vegetation indices (normalized ratio vegetation index, NRVI and soil-adjusted vegetation index, SAVI). The plotting of estimated yield against actual yield during 2013 and 2014 cropping seasons revealed strong positive and linear correlations (R^2 = 0.64 ($p=0.06$) and 0.74 ($p<0.05$), respectively with average R^2 = 0.71 ($p<0.001$) for both seasons. This study concluded that a good prediction of rainfed sorghum yield could be achieved more than 30 days before harvesting with quick, accurate and cost-effective method compared to traditional field surveys.

INTRODUCTION

Sorghum (*Sorghum bicolor* L.) is a major staple cereal crop in the Sudan, with a growing area of more than 34% of the total staple crops (Elhag and Sulieman, 2017). In the eastern region of the Sudan, sorghum growing is important and its growing acreage and production are over 5% of those of the total main grain crops in this region. Therefore, getting early and accurate information about sorghum yields and production for this region is very important for production management issue as well as for the food supply and demand on the national basis. Traditional sampling methods are based on visual and physical samples taken from the vast cultivated fields during the beginning of the growing season (mid August), after crop maturity and before crop harvest (October) to forecast crop supply and demand for the whole country. It is time consuming, labor intensive and point-representative approach which does not reflect the spatial situation of the whole region. Currently, various methods of yield estimation and prediction are used, including remote sensing-based methods, crop growth models, agro-climate models, and statistical sampling methods (Baez-Gonzalez *et al*., 2002; Bastiaanssen and Ali, 2003; Dadhwal and Ray, 2000; Doraiswamy *et al*., 2003; Zhang and Li, 1999). Remote sensing-based methods are deemed as timely, accurate and less costly compared to other approaches. The estimations are commonly based on related crop parameters (e.g. leaf area index (LAI), normalized difference vegetation index (NDVI), fraction of absorbed photosynthetically active radiation (fPAR), net primary productivity (NPP)…*etc*.) which can be derived from freely available satellite sensors such as MODerate resolution Imaging Spectroradiometer (MODIS), advance very high resolution radiometer (AVHRR), Landsat, SPOT, *etc*.

MODIS provides high frequency observations. However, their spatial resolution is quite coarse. MODIS data are available at 250, 500 and 1000 m spatial resolution depending on the specific product (Justice *et al*., 1998). On the other hand, AVHRR is available at spatial resolutions of 1.1km for local area coverage and 4 km globally (Kidwell, 1998). High spatial resolution data from sensor such as Landsat Thematic Mapper (30m) have also been used for agricultural applications (Rudorff and Batista, 1991; Thenkabail *et al*., 1994). However, the repeat period for Landsat is relatively infrequent (16 days), which presents challenges for agricultural applications that rely on high frequency sampling during critical phases of the crop growth cycle. Recent studies have fused higher spatial resolution map products derived from Landsat with higher frequency of MODIS (Zwart and Bastiaanssen 2004; Bashir *et al*., 2010).

Vegetation indices (VIs) are widely used in crop growth monitoring and yield estimation. Most of the vegetation indices are information-condensed which can reflect terrestrial vegetation cover and growth condition effectively and economically. The use of a simple regression model based on direct correlation between crop yields and satellite-derived vegetation index is the most basic approach to estimate and forecast crop yields on a regional scale (Bolton and friedl, 2013). Becker-Reshef *et al*. (2010) developed a generalized regression-based model for winter wheat yield, which used the seasonal maximum normalized difference vegetation index (NDVI) as the main predictor variable and was able to forecast winter wheat production in Ukraine six weeks prior to harvest. Gnyp *et al*. (2014) concluded that, with suitable band combinations, optimized narrow band RVI and NDVI could significantly improve estimation of rice above ground biomass at different growth stages. Moreover, Ren *et al*. (2008) have tested the suitability of estimating winter wheat yield on a regional scale *via* MODIS-NDVI-based model. Their validated result showed that the relative errors of the predicted yield using MODIS-NDVI were between 4.62% and 5.40% and the whole root mean square error (RMSE) was 214.2 kg/ha. Meanwhile, good yield forecast data of wheat could be achieved 40 days before harvesting time.

This study is an attempt to estimate rainfed sorghum yield based on MODIS derived parameters, field survey and official reports.

MATERIALS AND METHODS

Selection of a crop region

Gedarif region (Fig. 1) is considered as the biggest mechanized rain-fed sorghum cultivation area in the country where the rain-fed agriculture expanded from 500 ha in early 1940s to about 2.3 million ha in 2006 and later to more than 4 million ha in recent years. Nearly 50% of the overall sorghum produced in the Sudan is cultivated in this region. The mechanized rainfed sorghum area covers around 4 million ha and provides about 50% of the overall sorghum production in the Sudan (Hassan and Elasha, 2010). Gedarif State represents the largest mechanized rain fed sorghum planted area (about 2 million ha) with average annual production of 675000 metric tons (Hassan and Elasha, 2010). Accordingly, Gedarif

State was selected for this study due to its status as the main sorghum-growing region in the Sudan. Gedarif State lies in the semi-arid zone (longitudes 33 – 37° E and latitudes 12 – 16° N, and altitude 350 masl with a total area of approximately 78,000 km^2 (Sulieman, 2008). Annual rainfall is during July to September with an average of 450 mm. Temperatures range from a minimum of 21°C in January to a maximum of 37°C in April and May.

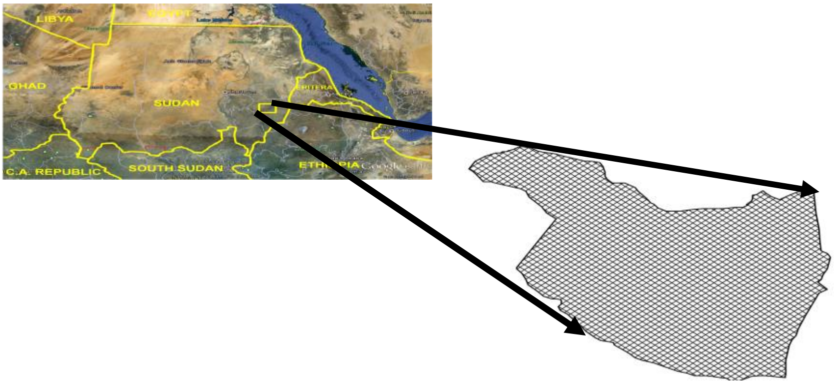

Fig. 1. Study area (Gedarif State)

Data collection and processing

Thirteen MODIS images, relatively cloud-free and at nadir view (refers to the downward-facing viewing geometry of an orbiting satellite) to the study area, were ordered and downloaded from the website of MODIS Data Support (http://daac.gsfc.nasa.gov/) for the period August – December 2010, 2013 and 2014 (Table 1). The images were geographically registered to each other and to the map projection of Universal Transverse Mercator (UTM) using multiple functions of ENVI 4.2 software. The MODIS data were organized on ten-day intervals and used to extract time series values of SAVI, NRVI and NDVI vegetation indices starting after crop establishment stage. The vegetation indices values were computed using the reflectance values in red (RED) and near infrared (NIR) regions obtained from canopy reflectance data. The calculations were based on various equations taken from previous studies (Table 2).

The ground-based information data were provided by Mechanized Agricultural Corporation which included the area, the coordinates of the area, the cultivated land and the yield Table (2).

Table 1. MODIS data and ground-based data used in the study.

MODIS data									
Image dekad	Aug.III	Sep.I	Sep.II	Sep.III	Oct.I	Nov.I	Nov.II	Nov.III	Dec.1	Dec.II
Components of ground-based data										
Region	Name of the scheme	Coordinates of the scheme	Cultivated area (ha)	Actual yield (kg/ha)						

Table 2. Remote sensing indices used in the study.

Remote sensing indices	Definition	Reference
NDVI:Normalized Difference Vegetation	$NDVI = \left(\frac{NIR - RED}{NIR + RED}\right)$	Rouse *et al*. (1973)
NRVI: Normalized Ratio Vegetation Index*	$NRVI = \left(\frac{RVI - 1}{RVI + 1}\right)$	Baret and Guyot (1991)
SAVI:Soil-Adjusted Vegetation Index	$SAVI = \left(\frac{NIR - RED}{NIR + RED + L}\right)(1 + L)$	Huete (1988)

*RVI= Red/NIR.

Conventional determination of sorghum yield in the study area

Many countries, including the Sudan, use the conventional technique of data collection for crop monitoring and yield estimation based on ground reports (Hassan *et al.*, 2010). In Gedarif State, the estimation of sorghum grain yield is usually carried out by the Mechanized Agricultural Corporation by visual estimation visits, however, sometimes the Department of Agricultural Economics in the State Ministry of Agriculture pays some visits for crop estimation using a random crop cutting method for this purpose. Hence, the conventional sorghum yield estimation can be realized through the following two surveys:

1. The initial survey

It is usually carried out in the beginning of the growing season in the period form 1 - 1.5 months after crop sowing in order to estimate sorghum cultivated area per each location by the technical staff who manage these areas. Finally, the estimate of sorghum cultivated area reports will be sent to the Head Quarter of the Mechanized Agricultural Corporation which provides the total cultivated sorghum area for the whole State.

2. The final survey (grain yield estimation)

This is usually carried out after crop maturity and before crop harvest (during October). Within this survey, the crop yield estimation team used to determine sorghum crop yields through an intensive site visits using an eye estimation method. Accordingly, the accuracy of this method depends on the experience of the surveying staff. Sometimes, the experts from the Department of Agricultural Economics used to adjust the eye estimation method for yield determination by using the crop cutting methodology. In both cases (i.e. an eye estimation and crop cutting estimation), the forecasting of sorghum grain yield is taking place late in the season and close to harvesting time which could affect the introduction of the right decision for food security planning. Hence, the following investigation was undertaken to determine sorghum grain yield earlier during the season (before crop maturity) in order to overcome any expected crop failure by provision of the right plan.

RESULTS

Measurement of sorghum canopy reflectance

Temporal changes of sorghum canopy reflectance between Aug.III and Oct.I are shown in Fig. 2. Reflectance values of Aug.III in the visible region were slightly lower than those in the near-infrared region due to sparse crop, however, at Sep.I reflectance values in the infrared were much higher than those in the visible region. Between Sep.II and Oct.I reflectance in the near-infrared region reached the highest value of the season while reflectance in the visible region showed the lowest value due to absorption by chlorophyll and other pigments. Based on the above mentioned results, the latter dates of reflectance measurement (Sep.II to Oct.I) were assumed to be the most accurate for yield forecasting in the mechanized rain-fed sorghum cultivation region, in eastern Sudan.

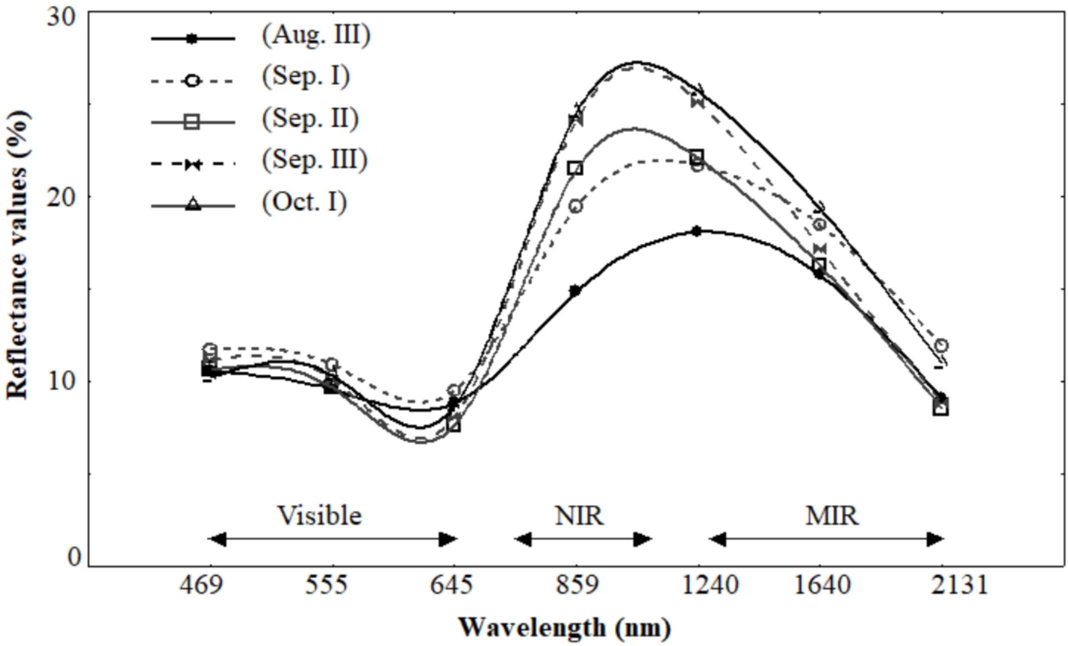

Fig. 2. Time course of sorghum reflectance during the season.

Sorghum prediction model

Figures 3 to 5 show the relationship between sorghum yield and NRVI, SAVI and NDVI values, respectively, measured at Sep.II, Sep.III and Oct.I. The method described by Soper (2018) was used for calculating the P-value of correlation coefficient based on two-tailed probability. Results showed that NDVI and NRVI values during Sep.III seemed to be better correlated with yield compared to other sampling dates along the growing season. While SAVI value during Oct.I, also, was positively correlated with yield compared to September samplings. Relationship between sorghum yield (y) *versus* NRVI, SAVI and NDVI at the three sampling dates are expressed in a linear function. Figures 3 to 5 and Table 3 showed that R^2-value of yield and indices at the above mentioned sampling dates were 0.41 ($p=0.27$), 0.55 ($p=0.13$) and 0.53 ($p=0.14$) for NRVI, 0.34 ($p=0.37$), 0.51 ($p=0.16$) and 0.64 ($p=0.06$) for SAVI and 0.35 ($p=0.36$), 0.77 ($p<0.05$) and 0.49 ($p=0.18$) for NDVI, respectively. Among the nine combinations (indices reflectance X sampling dates), NDVI significantly scored higher R^2-value (0.77) at the third dekad of September (Sep.III), while the lower R^2-value (0.34) scored by SAVI at the second dekad of September (Sep.II).

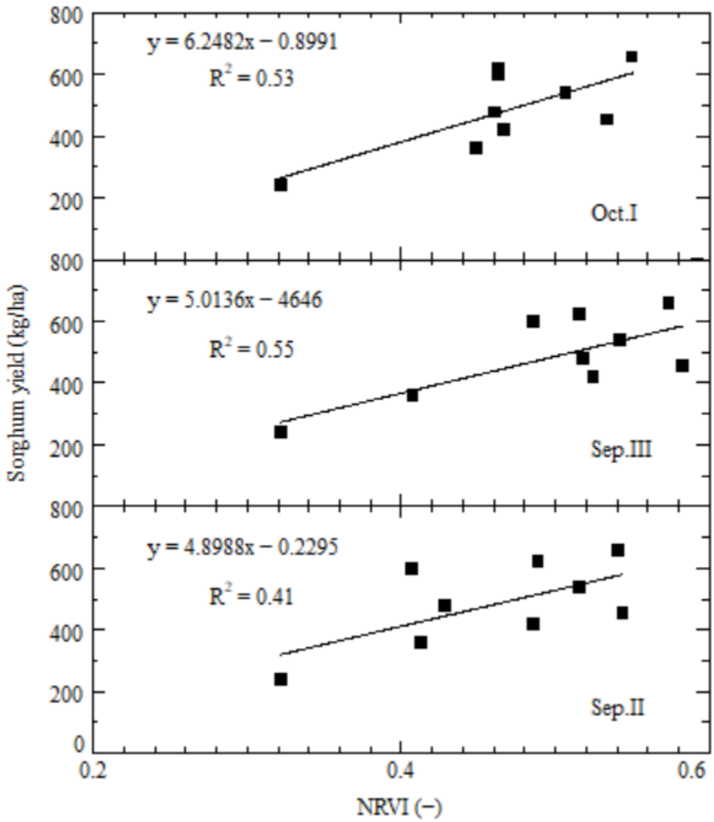

Fig. 3. Relationship between actual yield of sorghum and NRVI from Sep.II through Oct.I in Gedarif area

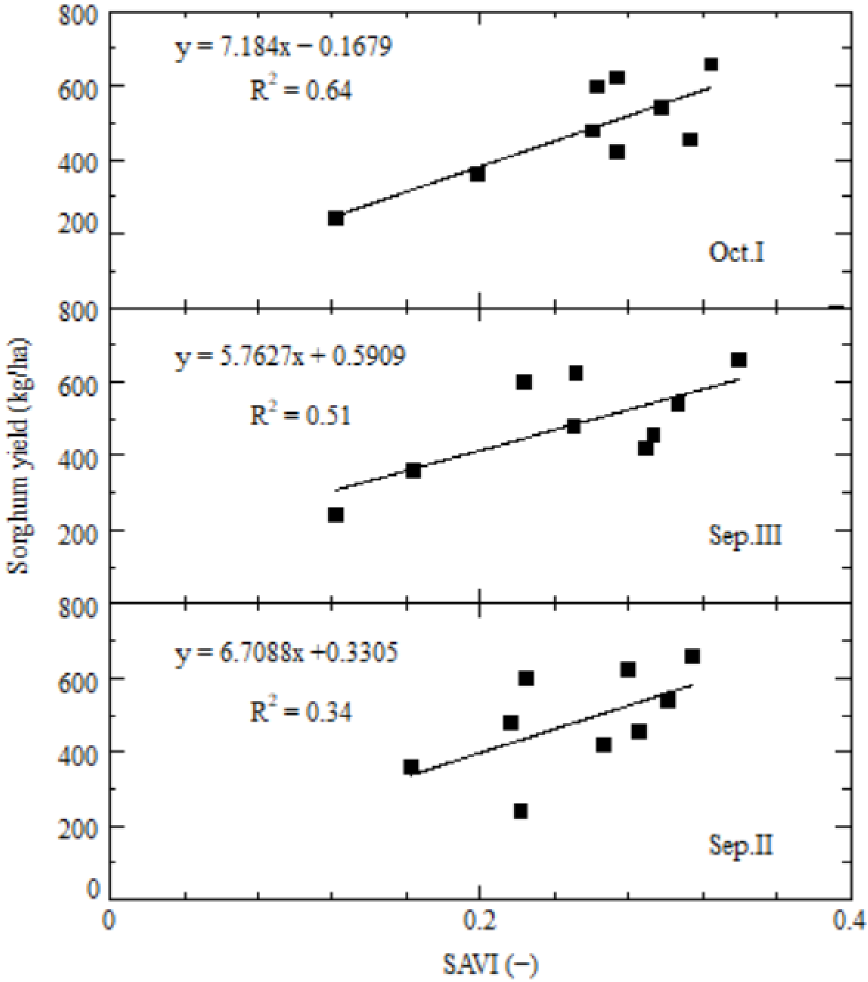

Fig. 4. Relationship between actual yield of sorghum and SAVI from Sep.II through Oct.I in Gedarif area.

The linear regression equations of sorghum yield with the tested indices during the sampling dates were y=4.8988x-0.2295, y=3.0136x-0.4646 and y=6.2482x-0.8991 for NRVI, y=6.7088x+0.3305, y=5.7627x+0.5909 and y=7.184x-0.1679 for SAVI and y=5.3901x-0.3253, y=6.831x-1.0317 and y=5.4621x-0.5358 for NDVI, respectively.

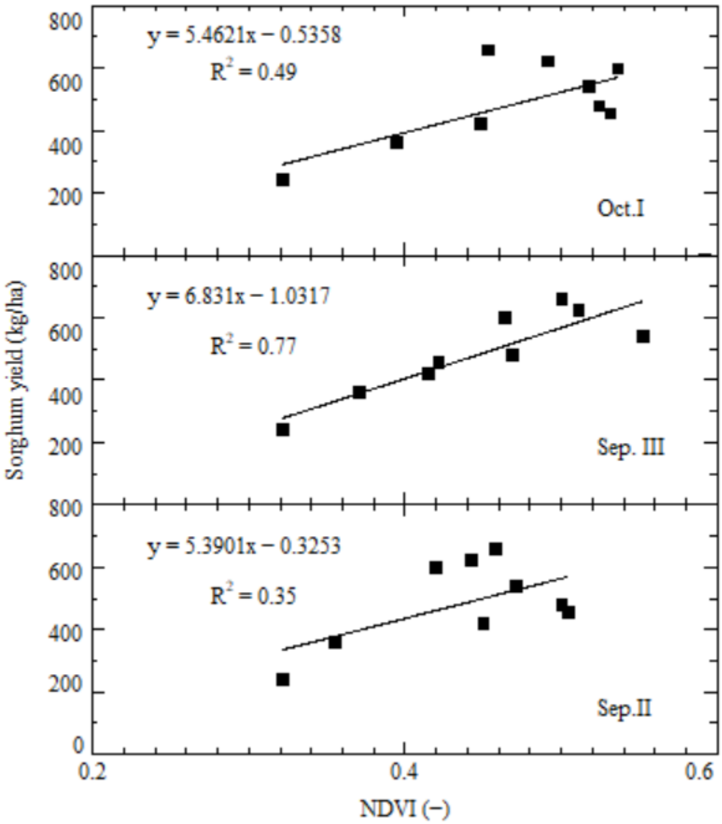

Fig. 5. Relationship between actual yield of sorghum and NDVI from Sep.II through Oct.I in Gedarif area.

Table 3. Models retrieved and corresponding correlation values for different vegetation indices.

Dekad	Vegetation index	Model	R^2-value
Sep.II	NRVI	y = 4.8988x-0.2295	0.41
Sep.III		y = 5.0136x-0.4446	0.55
Oct.I		y = 6.2482x-0.8991	0.53
Sep.II	SAVI	y= 6.7088x+0.3305	0.34
Sep.III		y= 5.7627x+0.5909	0.51
Oct.I		y = 7.1840x-0.1679	0.64
Sep.II	NDVI	y = 5.3901x-0.3253	0.35
Sep.III		y = 6.8310x-1.0317	0.77
Oct.I		y = 5.4621x-0.5358	0.49

Predictive model performance and validation

In order to test the feasibility of the linear model developed for yield prediction, the previous developed equations (Figures 3, 4 and 5) were used to predict yield during 2013 and 2014 cropping seasons. Scatter plots were used for further validation between estimated and actual yield collected from official Department of Information System and Statistical Analysis affiliated to the Ministry of Agriculture, Gedarif State. Among the 9 combination models (three vegetation indices X three sampling dates), NDVI at the third dekad of September showed the best fit with R^2-value of 0.77 ($P<0.05$) and assumed to be the most appropriate for yield prediction of sorghum cultivated in eastern Sudan (Fig. 5). Analysis showed that, estimated *versus* actual sorghum yields were highly correlated during 2013 and 2014 cropping seasons. R^2-values were 0.64 and 0.74, respectively. The average result of both seasons gave also higher R^2-value of 0.71 (Fig. 6). The slightly lower R^2-value during 2013 season could be attributed to other dominated factors that may happen after the prediction date.

For more verification of the predictive performance of the method, the established model was used to estimate sorghum yield during 2013 and 2014 cropping seasons based on actual yield to validate the model. The yield was gathered from large-scale mechanized schemes in the study area (Fig. 7). The validation was based on relative error (%) and RMSE. The results showed that the relative error (%) was 3.4% and that the whole RMSE was 63.26 kg/ha. The RMSE value of this study was much lower than the RMSE (214.16 kg/ha) of Ren *et al*. (2008) who estimated yield of winter wheat using the same set of data (MODIS-NDVI).

The MODIS-NDVI model derived was used to extract the spatial distribution map of sorghum yield during the two cropping seasons (2013 and 2014) as shown in Fig. 8. The spatial distribution map was able to detect several isolated areas of high yield (southern part of Gadarif State). However, the northern part of the State relatively showed lower yields. The MODIS-derived estimates of spatial distribution were in excellent agreement with the large-scale mechanized schemes statistics derived from Fig. 7.

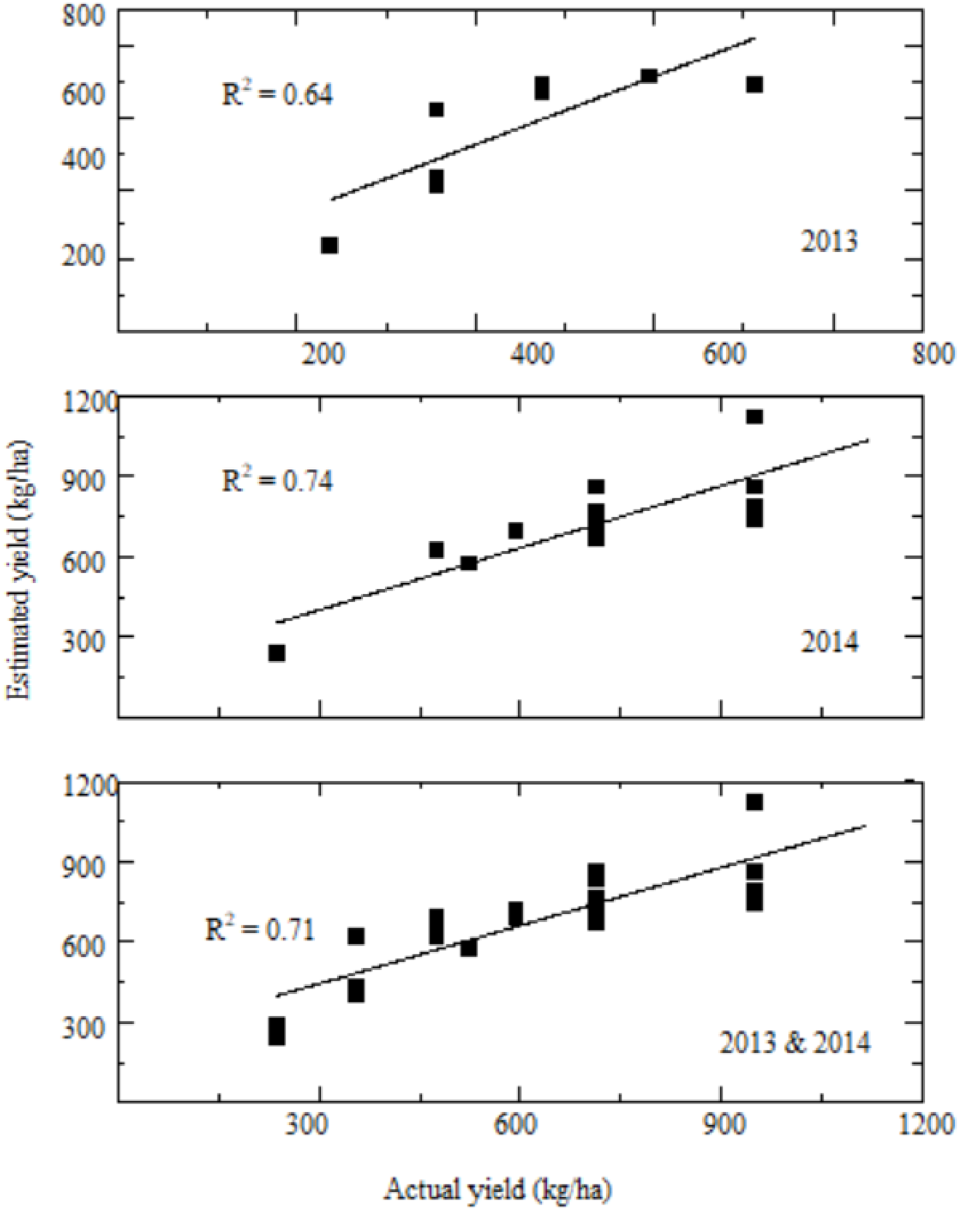

Fig. 6. Relationship between actual and estimated yield of sorghum in Gedarif area.

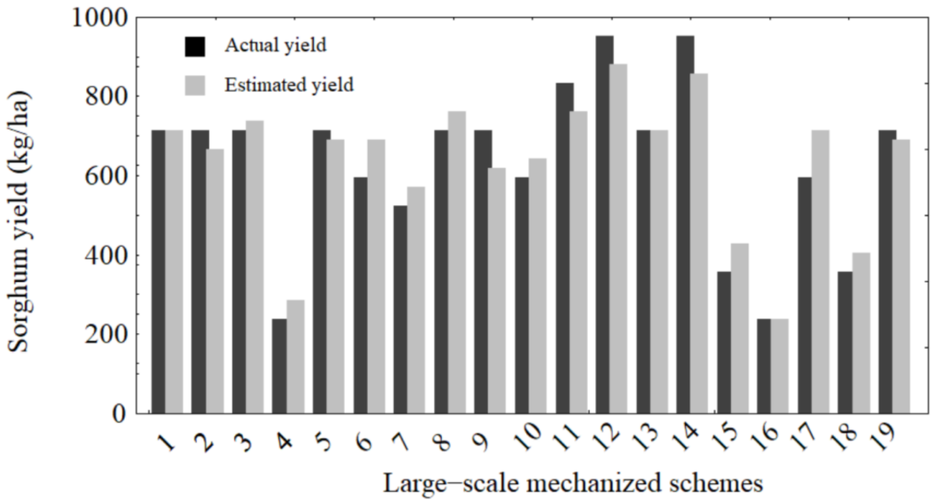

Fig. 7. Estimated and actual yield of sorghum for different large-scale mechanized schemes during the two cropping seasons (2013 and 2014).

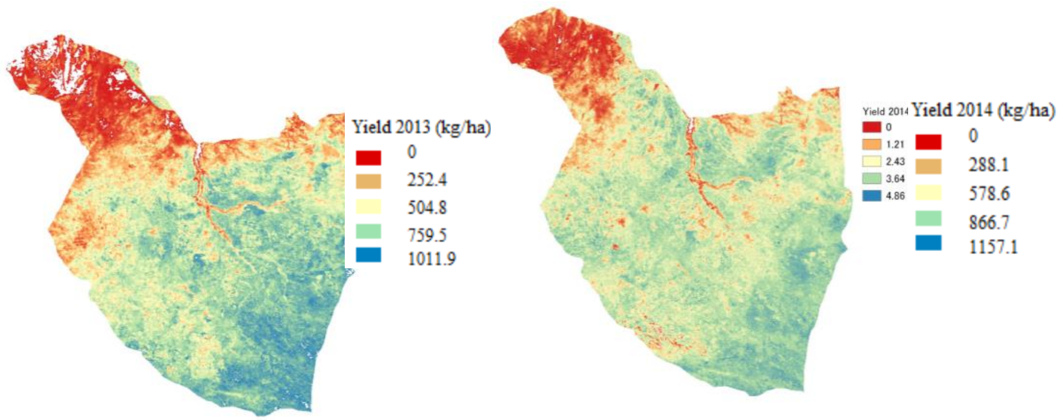

Fig. 8. Estimated yield of sorghum (kg/ha) map in Gedarif area.

DISCUSSION

In this study, satellite information coupled with limited ground data were used to improve sorghum yield prediction in Gadarif area, eastern Sudan. The NDVI-derived model gave higher R^2-value (0.64 and 0.74) in both seasons (2013 and 2014, respectively). This model can be used to predict crop yield more than 30 days before harvesting time (i.e. the prediction time could be made at the grain filling stage of sorghum.) with high accuracy and less cost compared to the current traditional methods. The study illustrated a distinct difference between the method used in this paper and the traditional way of predicting sorghum yield using several ground surveys. The conventional technique does not cover the whole cultivated area (≈4 million ha), however, it depends on a random crop cutting. While the proposed method covers the target area using remote sensing data. The traditional technique requires at least two surveys during the growing season, while the proposed method depends on a developed model established between sorghum yield and the relevant NDVI value to retrieve yield information at a pixel level. The traditional way of predicting yield depends on the experience of the survey staff; thus, the results always need to be adjusted using crop cutting methodology. The proposed method is swift, cost-effective and accurate as shown in many countries around the world (Bastiaanssen and Ali, 2003; Wang and Lin, 2005; Bashir *et al*. 2010; Mkhabela *et al*. 2011; Kowalik *et al*. 2014). Crop yields are sometimes affected by some factors which may dominate in certain seasons and/or locations causing yield reduction. Therefore, future research to improve crop productivity and sustainability are required and must be strengthened.

CONCLUSION

Based on the results achieved during the study period using MODIS-NDVI derived data and limited ground information, rainfed sorghum yield could be forecasted *via* a simple regression model at the third dekad of September (Sep.III). This means that acceptable sorghum yield forecast using the proposed model can be achieved more than 30 days prior to harvesting (i.e. at grain filling stage). These findings are not only fundamental to precision agriculture and food security issues, but can also be very useful to other end users and interested parties such as federal agricultural and state ministries, large-scale farmers, farmer's union, trading and insurance companies.

ACKNOWLEDGEMENT

This research was partially supported by Japan Student Services Organization (JASSO). The collection of field data was funded by Agricultural Research Corporation (ARC). In addition, authors express their appreciation to the staff of Gedarif Research Station, who supported the field data collection.

REFERENCES

Baez-Gonzalez, A.D., P. Chen, M. Tiscareno-Lopez and R. Srinivasan. 2002. Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. Crop Science 42: 1943-1949.

Baret F. and G. Guyot. 1991. Potential and limits of vegetation indices for LAI and APAR assessment. Remote Sensing and the Environment 35: 161-173.

Bashir, M.A., H. Tanakamaru and A. Tada. 2010. Application of remote sensing for estimating crop water requirements, yield and water productivity of wheat in the Gezira Scheme. International Journal of Remote Sensing 31: 4281-4294.

Bastiaanssen, W.G.M. and S. Ali. 2003. A new crop yield forecasting model based on satellite measurements applied across the Indus basin, Pakistan. Agriculture Ecosystem and Environment 94: 321-340.

Becker-Reshef, I., E. Vermote, M. Lindeman and C. Justice. 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukrine using MODIS data. Remote Sensing of Environment 45: 311-326.

Bolton, D.K. and M.A. Friedl. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology 173: 74-84.

Dadhwal, V.K. and S.S. Ray. 2000. Crop assessment using remote sensing. Part II. Crop condition and yield assessment. Indian Journal of Agricultural Economics 55: 54-67.

Doraiswamy, P.C., S. Moulin, P.W. Cook and A. Stern. 2003. Crop yield assessment from remote sensing. Photogrammetry Engineering and Remote Sensing 69(6): 665-674.

Elhag, M.M. and H.M. Sulieman. 2017. Rainfall erosivity, land-use and land-cover change analysis for Gadarif region, Sudan. International Journal of Water Resources and Arid Environments 6(2): 252-260.

Gnyp, M.L., Y. Miao, F. Yuan, S.L. Ustin, K. Yu, Y. Yao, S. Huang and G. Bareth. 2014. Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. Field Crops Research 155: 42-55.

Hassan, A.E. and A.E. Elasha. 2010. Crop Production under Mechanized Rain-fed Agriculture in Sudan. In: Proceedings of the National Symposium on Sustainable Rain-fed Agriculture in Sudan, 17- 18

November 2009, Al Sharga Hall, University of Khartoum, Sudan. M.A. Mustafa, (eds.). UNESCO Chair of Desertification Studies, University of Khartoum, Sudan. pp. 89- 112.

Hassan, I. M., M. Abdel Rahim, A. Seed and M.A. Omran. 2010. Utilization of water budget model for early season forecasting of sorghum yield and optimum sowing date in Gedarif mechanized rain fed areas, Sudan. Agriculture and Biology Journal of North America 1(4): 510-525.

Huete, A.R. 1988. A Soil-Adjusted Vegetation Index (SAVI). Remote Sensing of Environment 25: 295-309.

Justice, C.O., E. Vermote, J.R.G. Townshend, R. Defries, D.P. Roy, D.K. Hall, V.V. Salomon-son, J.L. Privette, G. Riggs, A. Strahler, W. Lucht, R.B. Myneni, Y. Knyazikhin, S.W. Running, R.R. Nemani, W. Zhengming, A.R. Huete, W. Van Leeuwen, R.E. Wolfe, L. Giglio, J.P.Muller, P. Lewis and M.J. Barbsley. 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. IEEE Transaction on Geoscience and Remote Sensing 36: 1228-1249.

Kidwell, K.B. 1998. Polar orbiter data users' guide (TIROS-n, NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11, NOAA-12, and NOAA-14). Washington, DC. {http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/podug/index.htm}.

Kowalik, W., K. Dabrowska-Zielinska, M. Meroni, T.U. Raczka and A. de Wit. 2014. Yield estimation using SPOT-VEGETATION products: A case study of wheat in European countries. International Journal of Applied Earth Observation and Geoinformation 32: 228-239.

Mkhabela, M.S., P. Bullock, S. Raj, S. Wang and Y. Yang. 2011. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. Agricultural and Forest Meteorology 151: 385-393.

Ren, J., Z. Chen, Q. Zhou and H. Tang. 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. International Journal of Applied Earth Observations 10: 403-413.

Rouse, J.W., R.H. Haas, J.A. Schell and D.W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. Earth Resources Technology Satellite-1 Symposium, Goddard Space Flight Center, Washington D.C. 309-317.

Rudorff, B.F.T. and G.T. Batista. 1991. Wheat yield estimation at the farm level using TM Landsat and agrometeorological data. International Journal of Remote Sensing 12: 2477-2484.

Soper, D.S. 2018. P-value calculator for correlation coefficient [Software]. Available from https://www.danielsoper.com/statcalc

Sulieman, H.M. 2008. Mapping and modeling of vegetation changes in the southern Gedarif region, Sudan, using remote sensing. PhD Thesis, TUD Press, Dresden, Germany.

Thenkabail, B.S., A.D. Ward and J.G. Lyon. 1994. Landsat-5 thematic mapper models of soybean and corn crop characteristics. International Journal of Remote Sensing 15: 49-61.

Wang, C. and W. Lin. 2005. Winter wheat yield estimation based on MODIS EVI. Transactions of the Chinese Society of Agricultural Engineering 21(10): 90-94.

Zhang, J.Y. and X.Q. Li. 1999. The characteristics of weather yield for global crop and its relationship with precipitation. Quarterly Journal of Applied Meteorology 10(3): 327-332.

Zwart, S.J. and W.G.M. Bastiaanssen. 2004. Review of measured crop water productivity and scope for improvement in eight irrigated wheat systems. Agricultural Water Management 89: 287-296.