IGAD Centre for Pastoral Areas and Livestock Development (ICPALD) Regional Pastoral Livelihoods Resilience Project (RPLRP)

December, 2020

Prototype Rangeland Prediction Model

INTERGOVERNMENTAL AUTHORITY ON DEVELOPMENT

IGAD

AUTORITÉ INTERGOUVERNEMENTALE POUR LE DÉVELOPPEMENT

PEACE, PROSPERITY AND REGIONAL INTEGRATION

Submitted by:

Jully O. Ouma

Table of Contents

List of Figures

TABLE

INTRODUCTION

1.1 MODEL PILOT AREA OF INTEREST

The geographic interest area of this pilot assignment is the cross-border areas of Ethiopia, Kenya, Somalia, South Sudan, and Uganda. These areas are known as Cluster 1 (Karamoja Cluster- cross-border shared by Uganda, South-Sudan, Kenya and Ethiopia) and IGAD Cluster 2 (cross-border area shared by Kenya, Ethiopia, and Somalia). Specifically, it covers the districts in Karamoja region of Uganda; West Pokot, Turkana, Marsabit, Wajir and Mandera counties in Kenya; South Omo, Borana and Liben Zones in Ethiopia; Gedo region in Somalia (Figure 1). Much of the area of interest is covered by shrubs and grassland which defines the area as predominantly a pastoral zone (Figure 1) and is the extent of the pilot rangeland prototype model.

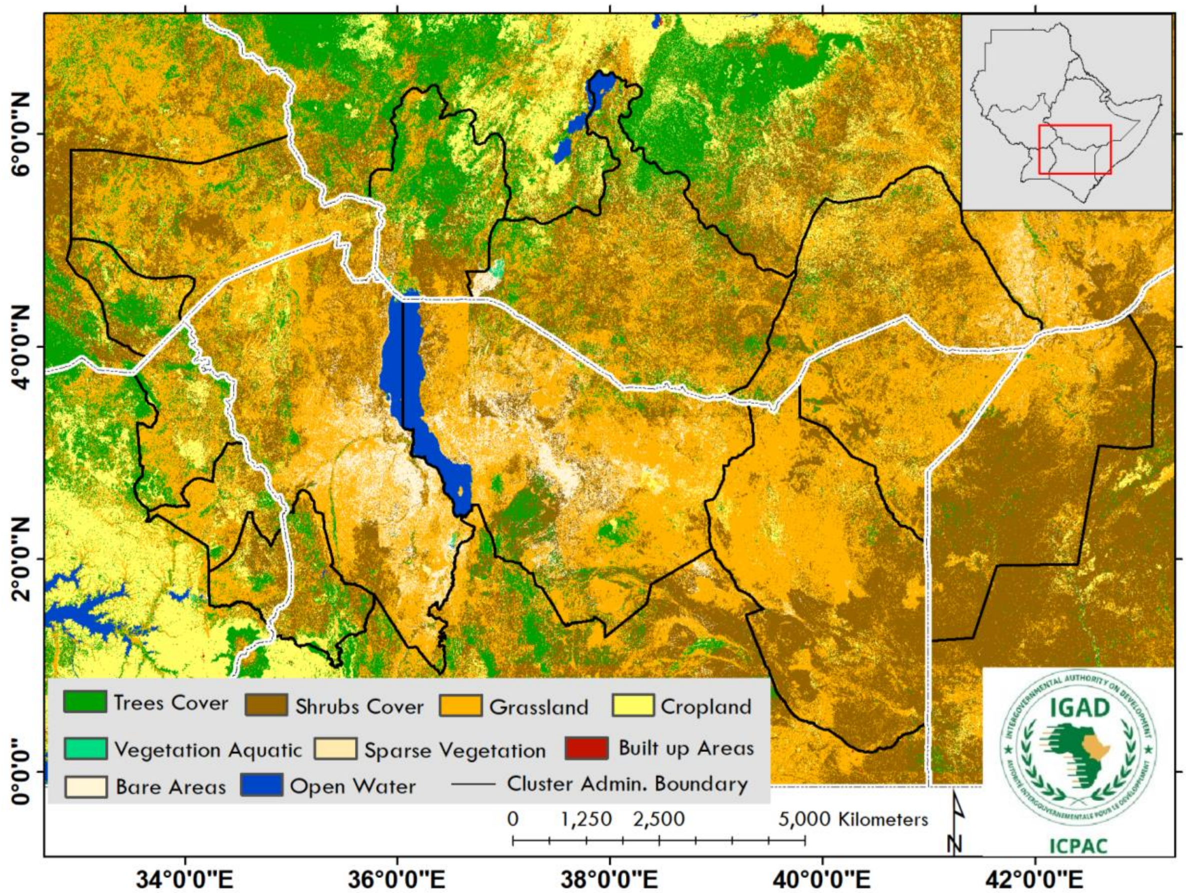


Figure 1: Map showing land cover over project area of interest which is the three IGAD cross-border areas (Data from ESA).

1.2 DATA SET USED

Datasets used in this model include Normalized Difference Vegetation Index (NDVI), Rainfall/Precipitation (PPT), Temperature (Temp), and Soil Moisture (Soilw). As a prerequisite for modeling, all the dataset will be grouped in two periods i.e. 1999-2013 for model development (training period) and 2014-2017 for model validation at seasonal time scale. The datasets were processed in seasonal timescales i.e. Mar-April-May (MAM), June-July-August (JJA), and October-November-December (OND) seasons.

The two statistical measure of mean and maximum are used to generate pixel data values at seasonal timescale, however, for rainfall the total rainfall of the respective seasons is used instead of maximum amount. These statistical parameters were then used in developing the model. In order to achieve this, inferential statistics were used to correlate variables and for testing significance. Result from correlation test indicates that, rainfall mean and the mean of soil moisture has high correlation with maximum NDVI at seasonal scale (Figure 2).

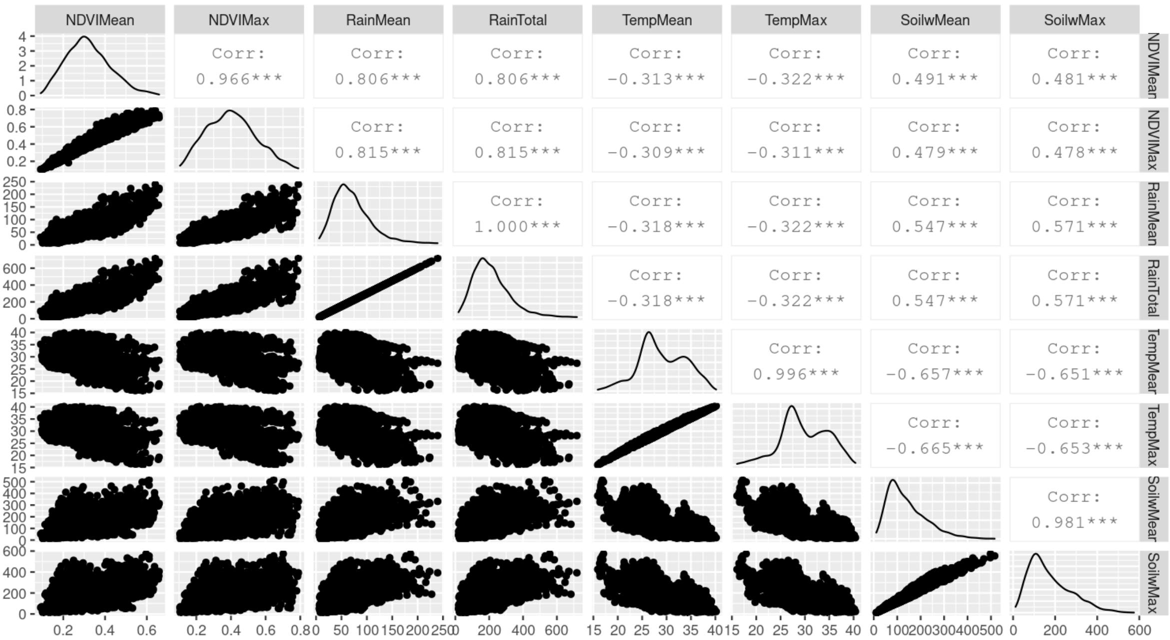


Figure 2: Correlation matrix of maximum, mean, and total for seasonal long term mean of NDVI, Temperature, Soil moisture, and Rainfall for March-April-May season.

The correlation matrix presented in the figure above was reduced to important variable in Table 1 which gave high correlation values. From this, a combination for developing the prediction model was selected, that is maximum NDVI, total Rainfall, maximum of Soil moisture (NDVImax, RainTotal, & SoilwMax).

Table 1: Independence variables with high positive correlation values.

Correlation	Rainfall.Mean	Rainfall.Total	Soilw.Mean	Soilw.Max
NDVI.Max	81.5%	81.5%	47.9%	47.8%

From the analysis, seasonal rainfall and soil moisture are well correlated spatially with maximum NDVI over the region, Figure 3 and **Error! Reference source not found.** for OND, and Figure 4 and Figure 6 for MAM respectively. This was observed as well JJA seasons. However, temperature on the other hand gave poor correlation with maximum NDVI, thus it was dropped from the prediction model development. Hence, seasonal rainfall and soil moisture were considered in this model development as predictands. Few areas with negative correlation especially with rainfall indicates that vegetation grew better accordingly with a continuous increase of rainfall in rainless areas (Leilei *et al.*, 2014).

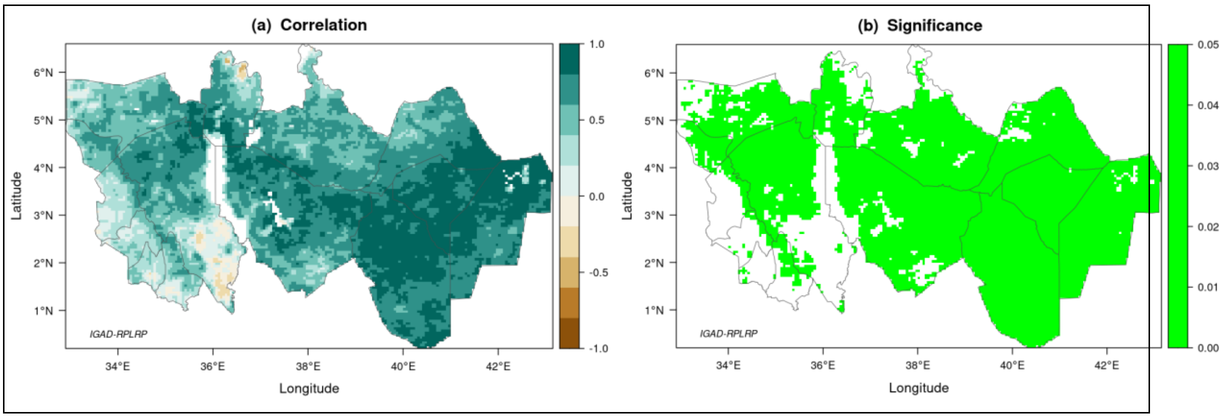


Figure 3: Spatial correlation for OND season between maximum NDVI and rainfall total over the area for the period 1999 to 2017.

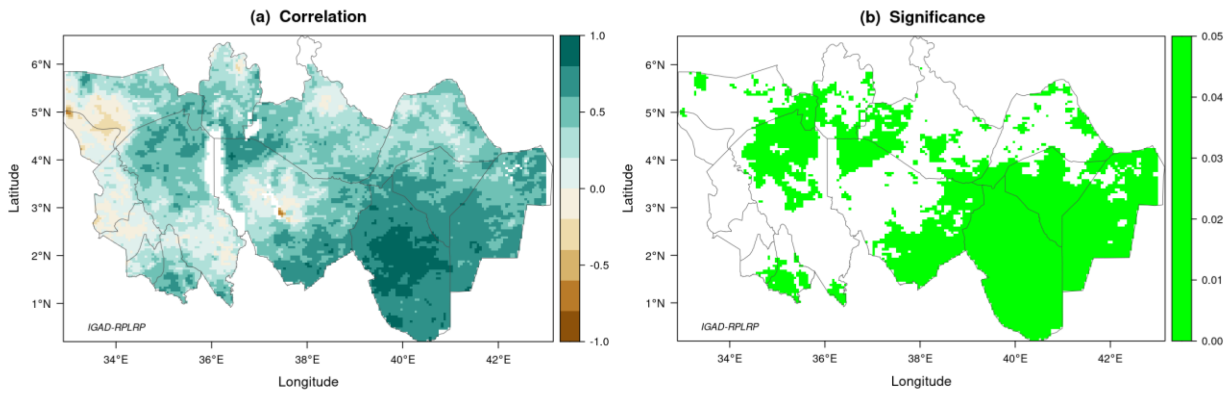


Figure 4: Spatial correlation for OND season between maximum NDVI and maximum soil moisture over the area for the period 1999 to 2017.

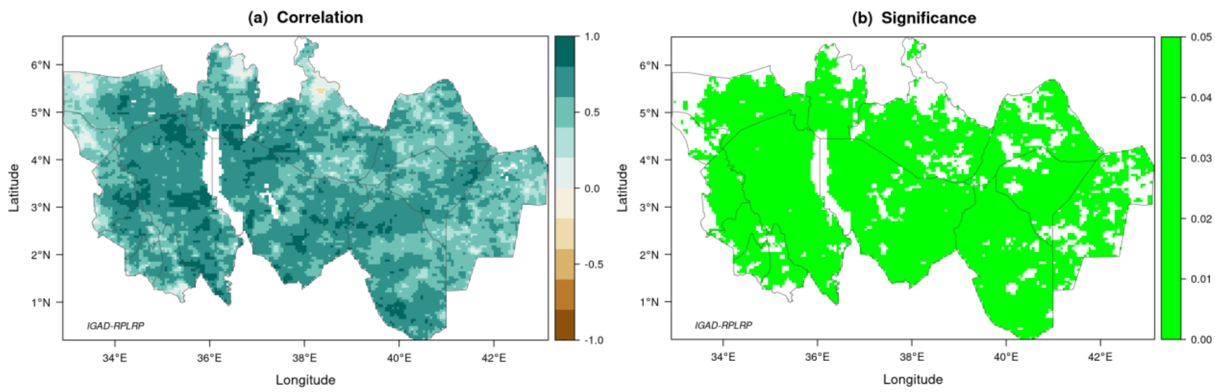


Figure 5: Spatial correlation for MAM season between maximum NDVI and rainfall total over the area for the period 1999 to 2017.

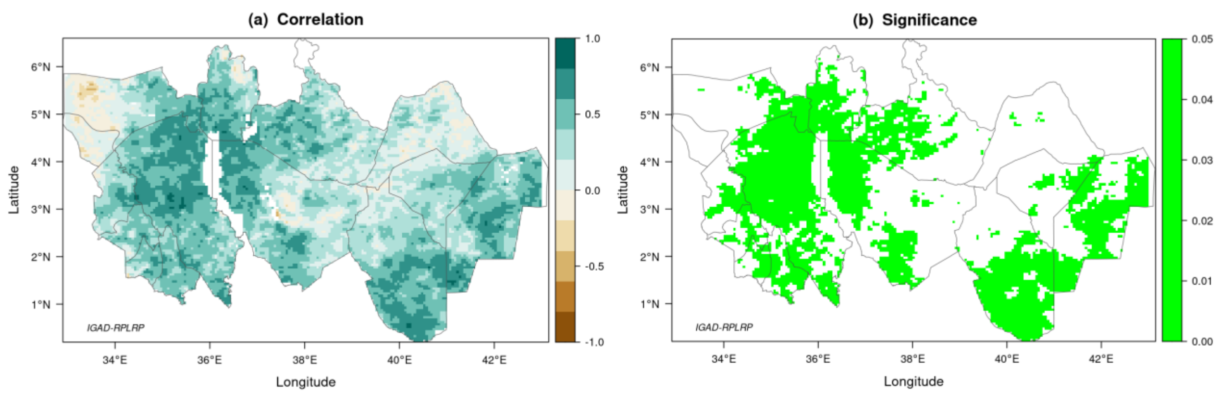


Figure 6: Spatial correlation for MAM season between maximum NDVI and maximum soil moisture over the area for the period 1999 to 2017.

1.3 MULTICOLLINEARITY TEST

Detecting multicollinearity before developing a model is critical and removing the same improves model performance. The easiest way for the detection of multicollinearity is to examine the correlation between each pair of explanatory variables (Imdadullah et al., 2016). If two of the variables are highly correlated, then this may be the possible source of multicollinearity. However, pair-wise correlation between the explanatory variables may be considered as the sufficient, but not the necessary condition for the multicollinearity (Wichers, 1975; Imdadullah et al., 2016).

Another way for detecting the multicollinearity among variables is to determine their coefficient in a regression model. As, the coefficient of determination in the regression model increases toward unity, that is, as the collinearity of a variable with the other regressors increases, variance inflation factor (VIF) also increases and in the limit it can be infinite (Adeboye et al., 2014; Imdadullah et al., 2016). Therefore, we can use the VIF as an indicator of multicollinearity. The larger the value of VIF_j, the more “troublesome” or collinear the variable X_j. As a rule of thumb, if the VIF of a variable exceeds 2.5 (Adeboye et al., 2014) then there is multicollinearity, which will happen if multiple correlation coefficient for j-th variable R^2j exceeds 0.90, that variable is said to be highly collinear.

The following tests were also used for multicollinearity testing as one method is not sufficient; Determinant |X'X| (Wichers, 1975; Imdadullah et al., 2016) help in normalizing the correlation matrix and if the value is zero then collinearity exists, Farrar Chi-Square test (Farrar and Glauber, 1967; Adeboye et al., 2014; Imdadullah et al., 2016), Red Indicator (Kovács et. Al., 2005; Imdadullah et al., 2016), Sum of Lambda Inverse (Imdadullah et al., 2016), Theil's Method (Theil, 1971; Imdadullah et al., 2016), and Condition Numbe (Adeboye et al., 2014; Imdadullah et al., 2016) which determines colliniearity if the value is greater than 1000. Klein rule (Imdadullah et al., 2016) which give the location of multicollinearity. One and zero were used to determine the test results as shown below.

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test

The multicollinearity test was done for the three seasons (MAM, JJA, and OND) as shown in Table 2 below. Out of the six diagnostic tests, MAM season had a multicollinearity detected by two test while the other seasons had one detection. This number is not enough to qualify multicollinearity in the data-set. The VIF values for all the seasons and individual data-sets were less than 2.5 which is the cut point to determine collinearity, Klein rule also had zero result indicating no multicollinearity in the data (Table 3).

Table 2: The overall multicollinearity diagnostics for the data-sets used for the three seasons.

	MAM		JJA		OND	
	Result	Detection	Result	Detection	Result	Detection
Determinant |X'X|	0.3958	0	0.4431	0	0.5986	0
Farrar Chi-Square	2639.1631	1	2317.386	1	1461.0423	1
Red Indicator	0.5267	1	0.4024	0	0.3890	0
Sum of Lambda Inverse	5.4774	0	5.6741	0	4.2535	0
Theil's Method	0.1580	0	0.1898	0	- 0.1382	0
Condition Number	24.5810	0	21.4841	0	21.1255	0

Table 3: All individual multicollinearity diagnostics result to determine the location of collinearity.

	MAM		JJA		OND	
	VIF	Klein	VIF	Klein	VIF	Klein
Rainfall	1.4355	0	1.6303	0	1.1556	0
Temperature	1.7686	0	1.8433	0	1.4565	0
Soil Moisture	2.2734	0	2.2005	0	1.6415	0

1.4 PROTOTYPE RANGELAND PREDICTION MODEL DEVELOPMENT

An exploratory technique known as Geographically Weighted Regression (GWR) was used in developing a prototype rangeland feed prediction model. The GWR technique is preferred over the Ordinary Least Square Regression (OLR) due to its capability of examining the existence of spatial non-stationarity in the relationship between a dependent variable and as set of independent variables (Fotheringham et al., 2003; Matthews and Yang, 2012; Wang et al., 2013; Georganos et al., 2017). This technique is fully described in a book by Fotheringham et al. (2003). This method allows estimation of local parameter by taking into account the location of observation as shown by the equation (Eqn. 1 and 2) below;

$$Y_i = \alpha(u_i,v_i) + \beta(u_i,v_i)X_i + \lambda(u_i,v_i)Z_i \qquad i = 1 : n \qquad \textbf{(Eqn. 1)}$$

$$Y_i = \alpha(u_i,v_i) + \beta(u_i,v_i)X_i + \lambda(u_i,v_i)Z_i - \varepsilon_i \qquad i = 1 : n \qquad \textbf{(Eqn. 2)}$$

Where coordinates of location i are represented by u_i, v_i while α, β, and λ are local parameters to be estimated particularly at location i (Georganos et al., 2017), X and Z are independent variables and ε is the mean error (Bias). Prediction using this model will be based on grid cells. Eqn. 1 is a model with Bias while Eqn. 2 has the bias removed.

In brief, the basis of this technique is the concern that the fitted coefficient values of a global model, fitted to all the data, may not represent detailed local variations in the data adequately. In this case, it follows other local regression implementations. It differs, however, in not looking for local variation in 'data' space, but by moving a weighted window over the data, estimating one set of coefficient values at every chosen 'fit' point. The kernel used here is Gaussian but one can also apply the bisquare kernel as well (Charlton et al., 2009). The fit points are very often the points at which observations were made, but do not have to be. If the local coefficients vary in space, it can be taken as an indication of non-stationarity (Bivand et al., 2020).

Like other mathematical models, this model has limitations as it doesn't account for both the auto correlations and cross correlations as the predictor variables are assumed to be independent of each other (Wang et al., 2013; Ivajnšič et al., 2014), it also has issues associated with multicollinearity, kernel bandwidth selection, and multiple hypothesis testing (Matthews and Yang, 2012).

Output from the prediction model was converted to pasture biomass using the equation below (Eqn. 3). This technique is well described in Hobbs (1995) with output in Kg/ha, and was divided by 1000 to convert it to tonnes per hectare (t/ha) i.e. 1 t/ha = 1000 kg/ha.

$$\textit{Pasture biomass} = (6480{\cdot}2 * \textit{Pred.Biomass}) - 958{\cdot}6 \qquad \textit{Eqn. 3}$$

This conversion outputs are in kg/ha at seasonal time scale but later converted to t/ha. This method was also used to convert observed maximum NDVI to pasture biomass which was then used to assess the skill of the prediction models.

The model development was grid based thus the constants used has a value per grid cell. A sample is shown below for MAM and OND seasons (Figure 7 and Figure 8 respectively). The intercept here represent ***α,*** rainfall coefficient represent ***β,*** and the soil moisture coefficient represents ***λ*** in Eqn. 1 and 2. The local correlation of these is significant and has a high value in much of the region i.e. over 60% (Figure 7). Each of the three seasons had its unique constant which was used in modelling the season.

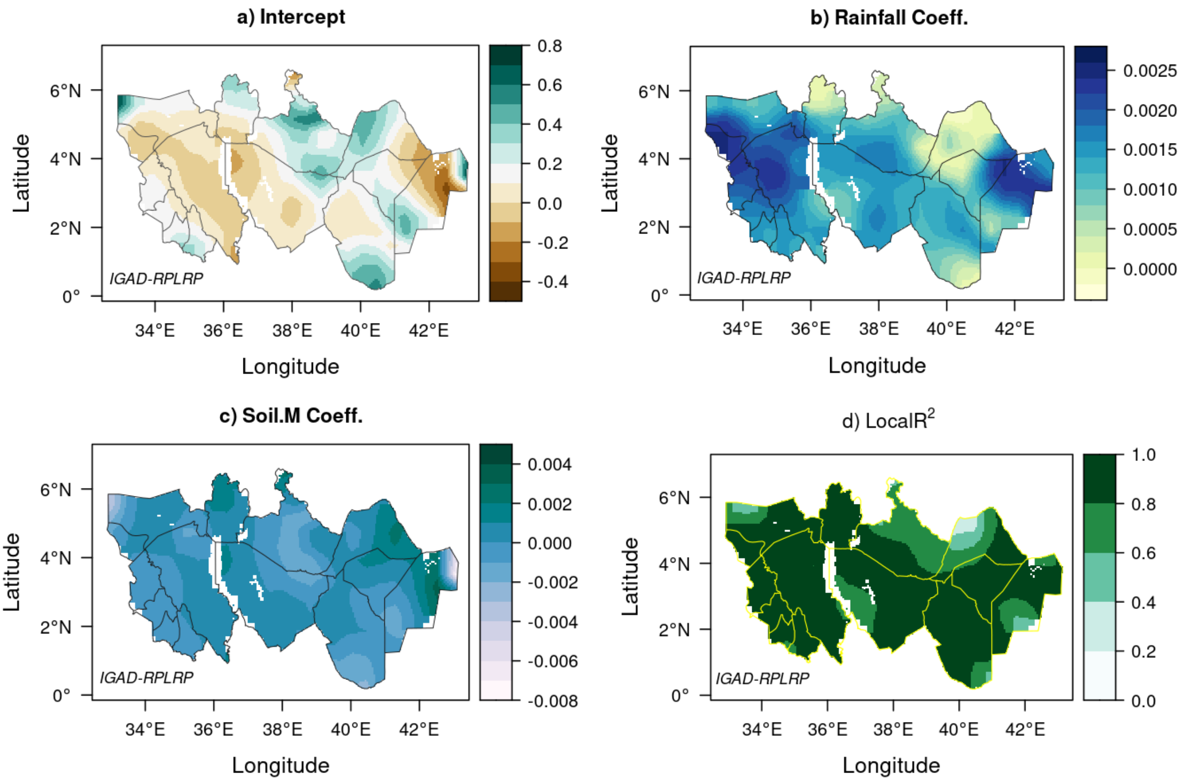


Figure 7: MAM season constants used in modelling MAM seasonal Biomass, (a) is the intercept, (b) is the rainfall coefficient, (c) is the soil moisture coefficient, and (d) is the local R square.

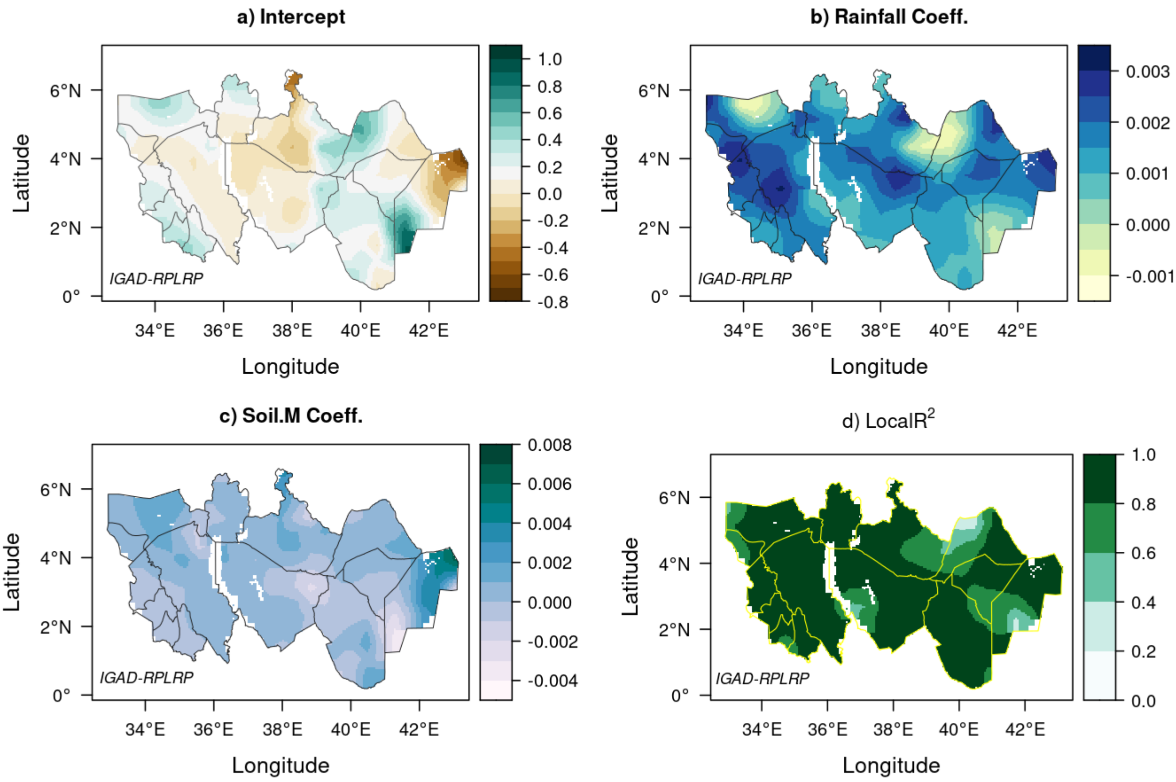


Figure 8: OND season constants used in modelling OND seasonal Biomass, (a) is the intercept, (b) is the rainfall coefficient, (c) is the soil moisture coefficient, and (d) is the local R square.

A first run of the model without considering the error term (Eqn. 1) gave the following result for MAM the main season shown in Figure 9. More pasture biomass (greater than 2 t/ha) was observed over the eastern, western, and northern part of the study area in 2018 and much in the northern part in 2019. The JJA season (not the main season) also had the same but on western and norther part of the study area. This is the expected pattern in the JJA season as NDVI performs best over the western and northern part of the study area during this season. When the error term is included, (Eqn. 2) we get a more realistic distribution of the the pasture biomass over the study area as shown in Figure 10. The performance of these two model is assessed in the next section.

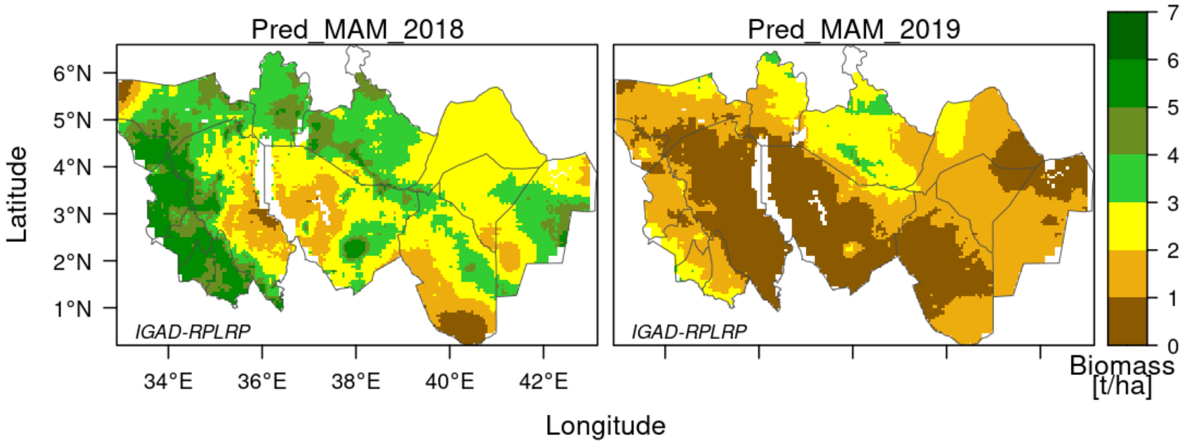


Figure 9: Predicted Biomass for MAM season over the study area from 2018 to 2019 without considering the error term.

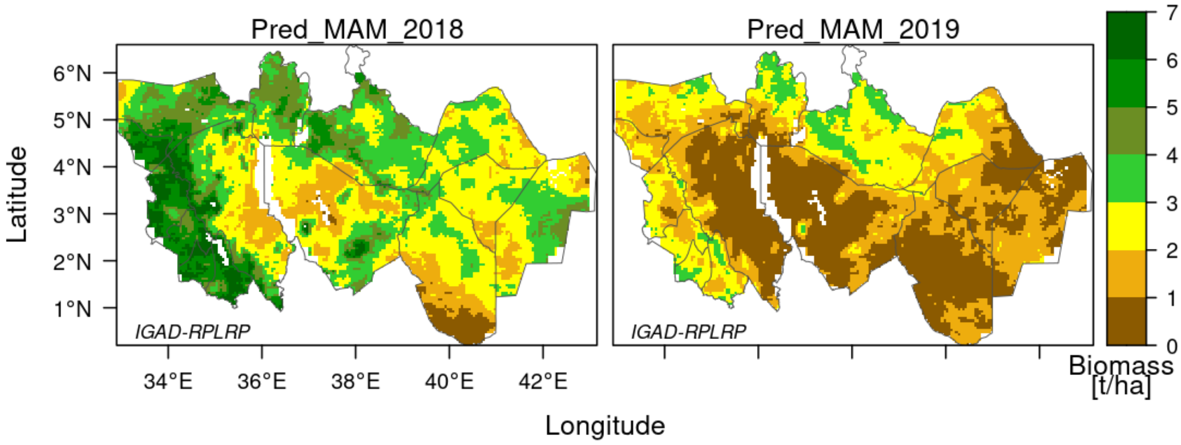


Figure 10: Predicted pasture biomass for MAM season over the study area from 2018 to 2019 with the error term considered.

METHODS USED IN MODEL SKILL ASSESSMENT

2.1 VISUAL VERIFICATION METHOD

The "eyeball" method is still the oldest and best verification method which look at the forecast and observations side by side and use human judgment to discern the forecast errors, however, this method is not quantitative (Stanski et al., 1989). The method is great if you only have a few forecasts, or you're not interested in quantitative verification statistics, but still useful when combined with statistics.

2.2 RELIABILITY DIAGRAM AND BOX PLOT

Reliability diagram (Wilks 1995) also known as attribute diagram was used to determine how well the predicted events correspond to their observed frequency. Reliability is indicated by the proximity of the plotted curve to the diagonal line (Wilks 1995; Hamill 1997; Bröcker and Smith 2007), and the deviation of this curve to the diagonal line gives a conditional bias. If the curve lies below the diagonal line, this indicates overforecasting; but above the line indicates underforecasting. In addition to the attribute diagram, an exploratory known as the box plot (Williamson et al., 1989) used to identify hidden patters in a group of data-sets. It has the power to summarize and compare groups of data-sets. The box plot show the range of data falling between the 25th and 75th percentiles (the difference between UQ (upper quartile) and LQ (lower quartile); Thirumalai et al., 2017). The horizontal line inside the box (**Error! Reference source not found.**) representing the median value, and the whiskers represents the complete range of the data (Jolliffe and Stephenson, 2012). This helps in assessing how well the distribution of predicted values correspond to the distribution of observed values. Box plot shows similarity between location, spread, and skewness of prediction and observed distributions (Jolliffe and Stephenson, 2012).

2.3 MEAN ERROR AND RELATIVE MEAN ABSOLUTE ERROR

The Mean Error (ME) and Relative Mean Absolute Error (RMAE) both estimate the average prediction error (Wang et al., 2013; Jolliffe and Stephenson, 2012), and give a perfect score when the value is zero. ME also known as the BIAS, gives the direction of the error while RMAE doesn't indicate the direction of the deviations. The closer the value is to zero the better the prediction model is (Table 4). The ME indicates how well the model estimate the corresponding observed value, it has a tendency to under-forecast when ME < 0 or over-forecast when ME > 0 events (Jolliffe and Stephenson, 2012).

Table 4: Model skill assessment using mean error and relative mean absolute error

Name	Values Range	Perfect Score

Mean Error	-∞ to ∞	0
Relative Mean Absolute Error	0 to ∞	0

2.3.1 Skill Assessment results

Using visual verification method as the first mode of model skill assessment, it is visually clear that Eqn. 2 results performs better than Eqn. 1 output when compared with the observed data. Figure 11 represents output of Eqn. 1 compared with observation while Figure 12 shows comparison of Eqn. 2 outputs with observation. Both figures show some pockets of overpredicting and underpredicting over the region in MAM (Figure 11), OND, and JJA (Annex 1) seasons. This was significantly corrected/improved with application of Bias in Eqn. 2 (MAM, Figure 12). Hence, under visual verification method, the model represented by Eqn. 2 performs well and should be used in pasture biomass prediction after validation in the field.

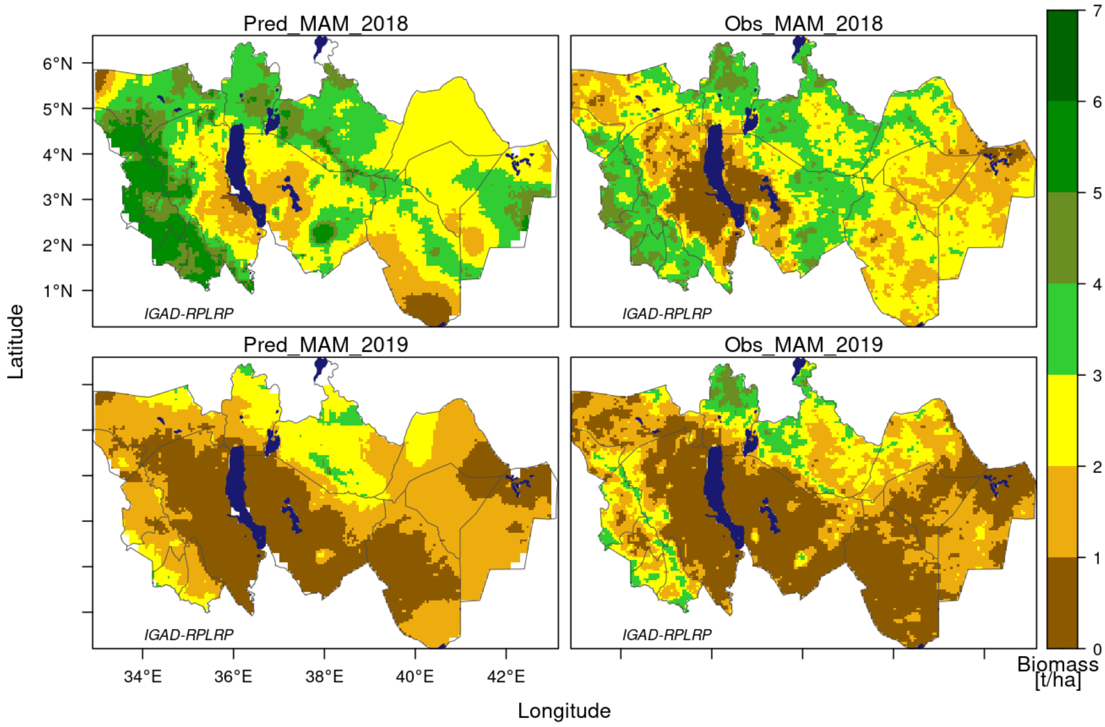


Figure 11: Predicted pasture biomass with bias compared with observed pasture biomass for the MAM seasons from 2018 to 2019.

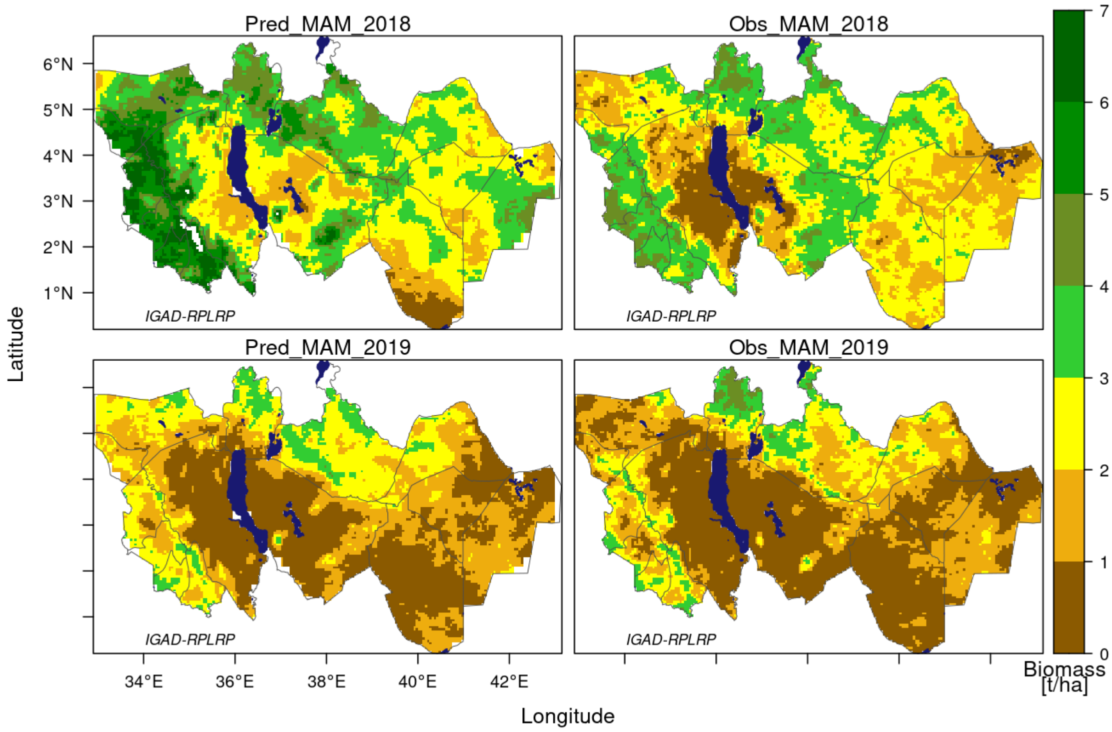
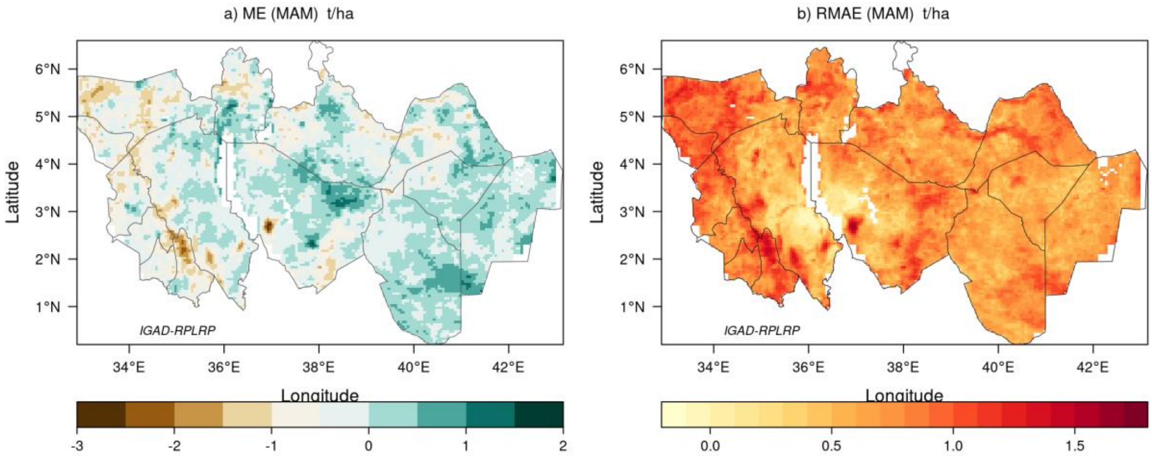


Figure 12: Bias corrected predicted pasture biomass compared with observed pasture biomass for the MAM seasons from 2018 to 2019.

The mean error (Figure 13 part (a)) clearly indicates areas that the model under estimated which is represented by the negative values and areas that were overestimated represented by the positive values in t/ha. The root mean absolute error (RMAE, Figure 13 part (b)) here is a measure of how accurate the model predicts the events. The lower values of RMAE indicates better fit, thus, this model performed generally well in predicting pasture biomass at seasonal time scale. The mean error shown here is what was used to improve the model performance in Eqn. 2 for each season i.e. MAM, JJA, and OND.



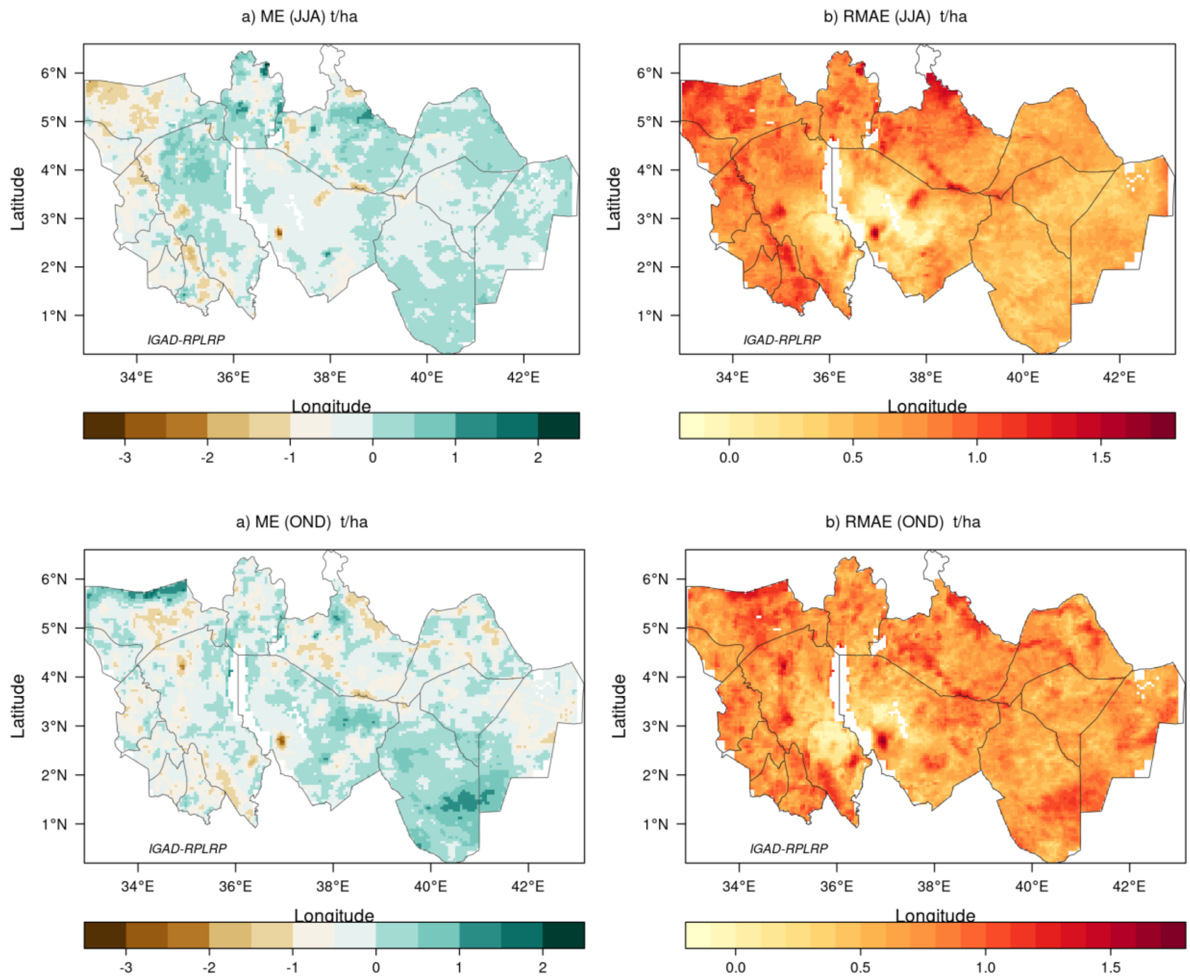


Figure 13: Mean Error (a) and Root Mean Absolute Error (b) for MAM season (first row), JJA season (second row), and OND season (third row).

Box plot was adopted to show the overall pattern of response for predicted values to observed values in all the three seasons. The median and the range which represents 50% of the data had a small shift when considering observed values and predicted ones. In addition to this, the reliability diagram was considered. The diagram was used to determine how well the predicted events correspond to their observed frequency. The model is said to be reliable if the curve falls on the diagonal line and skillful (in percentage) if it falls on the gray area of the plot. This model performed well after bias correction (Figure 14) which also increases the model skill score. The model skill score obtained using the reliability diagram is presented in Table 5 below for the three seasons in both scenarios of bias corrected and with bias. However, the data used to assess the skill score was only two years which as not enough to significantly conclude the performance. Thus its recommended that the model improvement to be a continuous as the model is being run in operational mode. The observed negative skill score in MAM 2018 and OND 2019 indicates that the score of the model is less than 33.3%. However, this can be significantly improved with at least a 30 year data used on model development. The model uncertainty in all the seasons and all scenarios does not exceed 25% (Table 5) thus this model performs well in predicting pasture biomass. The inter-annual variability are high in both cases due to the different seasonal rainfall diverse (Nicholson, 2017; Vellinga and Milton, 2018; Gudoshava *et al.*, 2020) as well as extreme events in the region which are attributed to El Nino and La Nina events.

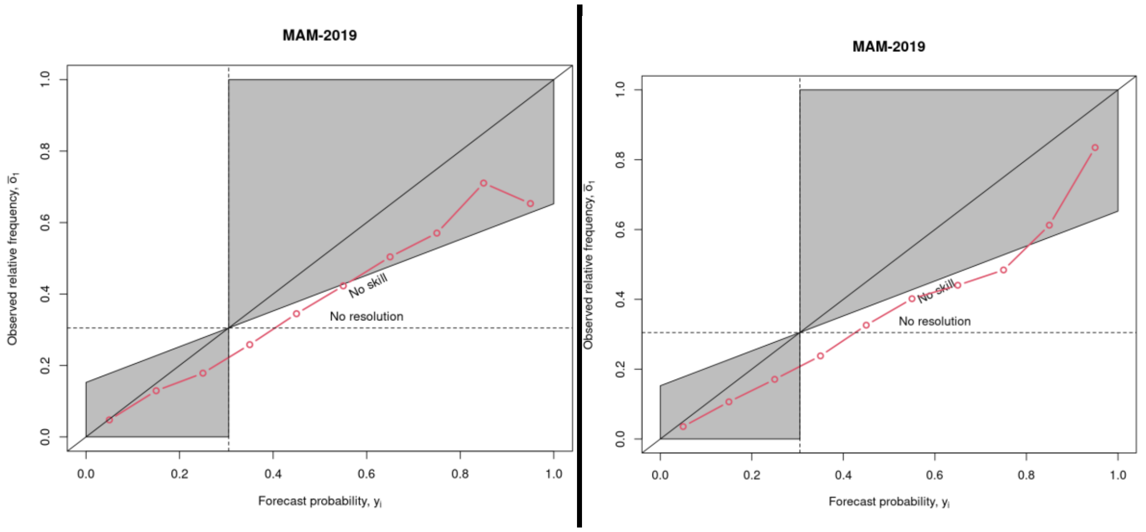


Figure 14: Reliability diagram for MAM 2019 with bias (left) and with bias corrected (right).

Table 5: A summary of reliability results (skill score and uncertainty) of the prediction model per year for the three seasons.

		MAM		**JJA**		**OND**	
		2018	*2019*	*2018*	*2019*	*2018*	*2019*
With Bias	Skill Score (%)	-9.8	56.7	46.7	93.6	65.5	-13.9
	Uncertainty (%)	24.7	23.6	22..2	24.7	24.4	23.9
Bias corrected	Skill Score (%)	-71.1	54.6	54.7	90.8	48.8	-44.1
	Uncertainty (%)	24.4	23.2	22.4	24.6	23.2	23.6

2.4 AVAILABLE FORAGE

Available pasture biomass can be computed from the total biomass pasture using a factor presented by Toxopeus, (2000; Eqn.4). This was arrived at after a review of different research in Ethiopia and Kenya rangelands (Van Wijngaarden, 1985; Cossins and Upton, 1987; FAO, 1988). If we are to consider the conversion to available forage/pasture then we can use the equation below by Toxopeus, (2000), the multiplication factor is 45%.

Available pasture = Pred. Pasture * 45% ***Eqn. 4***

Figure 15 shows result after conversion presented as the first column (a), and result before conversion as the second column (b). The available forage values are between 0 and 4 tonnes per hectare (t/ha) over the study area when considering the three seasons (MAM, JJA, and OND), while that of total forage/pasture biomass is between 0 and 7 t/ha. It's worth noting that the pattern for pasture availability is the same thus any can be used for early warning purposes as well as management planning.

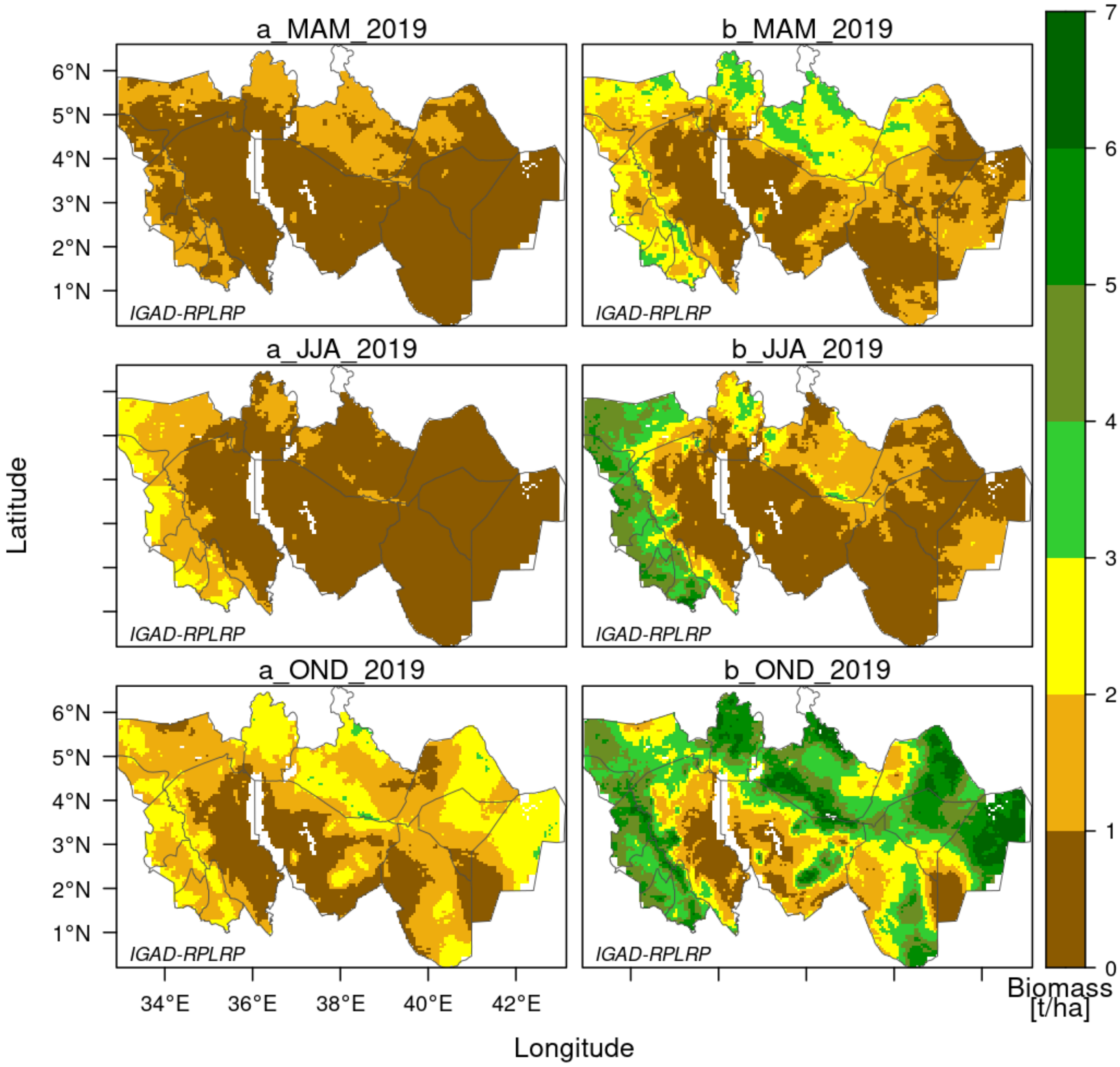


Figure 15: Available forage/pasture represented by the first column (a) while the second column (b) represent total pasture before conversion.

2.5 FIELD VALIDATION

The next and final stage of skill assessment was field validation (ground truthing). This was to further give confidence in the model which can then be rolled out for the whole region. Validation of the model though ground truthing (field activity) was done in December 2020 from the 8th to the 17th representing the last month of OND season. The aim of the field mission was to undertake local forage measurements, ground-truthing and estimation for the OND season. This was undertaken within each homogenous land cover unit in the project areas of interest.

2.5.1 Land cover of the study area

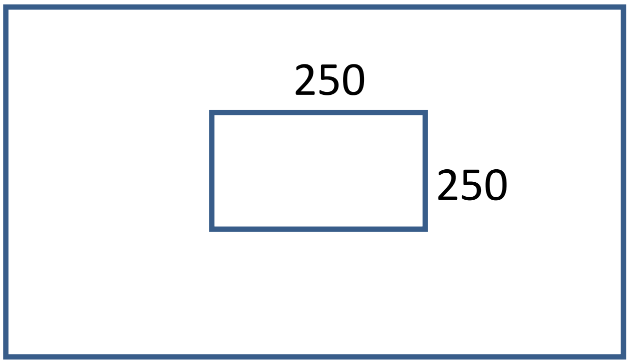
Savannah vegetation dominates the study area landscape and vegetation composition is closely related to rainfall. Vegetation cover increases with altitude thus scarce and sparse vegetation is found in low altitude areas compared to highlands that have dense vegetation. Ground cover throughout the study area varies seasonally depending on various grazing intensities and, overall, canopy cover ranges from less than 1% on heavily settled areas to about 30% on steep hills. Table 8 below gives a brief description of some of the different land cover types.

Table 6 Description of Land cover types

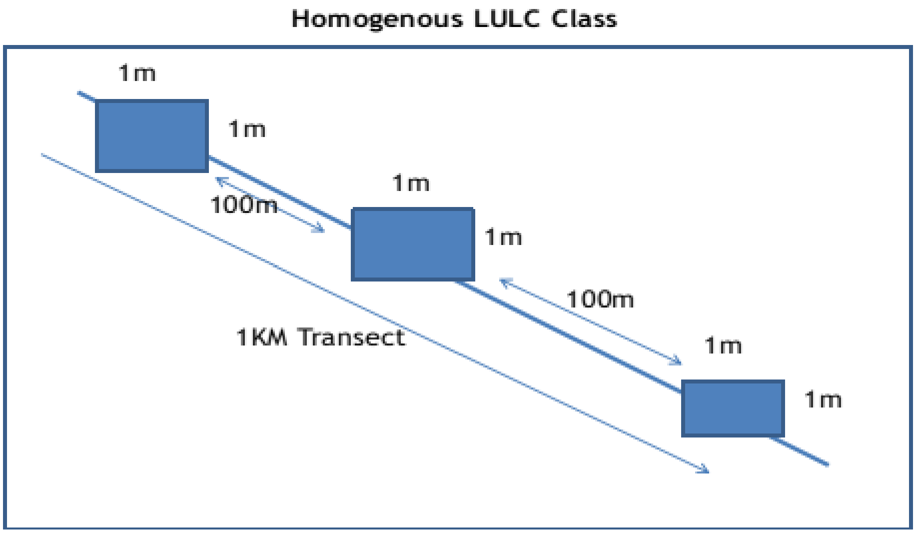
Land Cover Classes	Description
Barren/desert areas	Consists of bare desert lands, degraded Badlands and exposed rangelands/rocky areas
Mosaic Croplands/vegetation	Predominantly covered by vegetation but agriculture and other human activities are expanding
Sparse Vegetation	Consists of thorny shrubs mixed with grasslands. Is common in semi-arid parts of rangelands
Close to open shrub land	Consists of shrub land with thick canopy to areas where alteration to open grassland is taking place
Mixed Forest-shrub-grassland	Under this category pockets of remaining forests and mixed shrub/grasslands are included.

2.5.2 Sampling Frame

A fieldwork protocol has been developed to estimate green biomass (kg DM/ha) and available biomass (kg DM/ha) within different homogeneous land-use land cover (LULC) at local are field sites here in referred to as *stratas*. Within homogeneous LULC here in referred to as strata, randomly locate one (1) km line transect with sampling points of 250m^2 located around the pixel centroid to allow for plot-pixel comparisons and extrapolation (NDVI metrics).



Within grasslands land-cover types, a 1km line transect was used to randomly locate sampling points of $1m^2$ quadrants after every 100m intervals (see sampling layout below).



Yield measurements at each point was done by clipping the grass and herb layer within each subplot sampled using grass cutter and weighed using the digital kilograms automatic weighing machine to allow for plot-pixel comparisons (NDVI metrics). The resultant outcomes were extrapolated into total biomass (tons/ha) for every corresponding pixels within the different LULC and locations within the cluster areas.

2.5.3 Forage estimation and ground truthing findings

The forage estimation and ground truthing exercise as well as RPLRP water point ground truthing were undertaken in three areas within the Kenyan side of the cluster, i.e. Baringo, Marsabit, and Moyale during the period 8th – 17th of December 2020. The map below (Figure 16) shows the local areas visited, fieldwork major rout and sampling water points in blue.

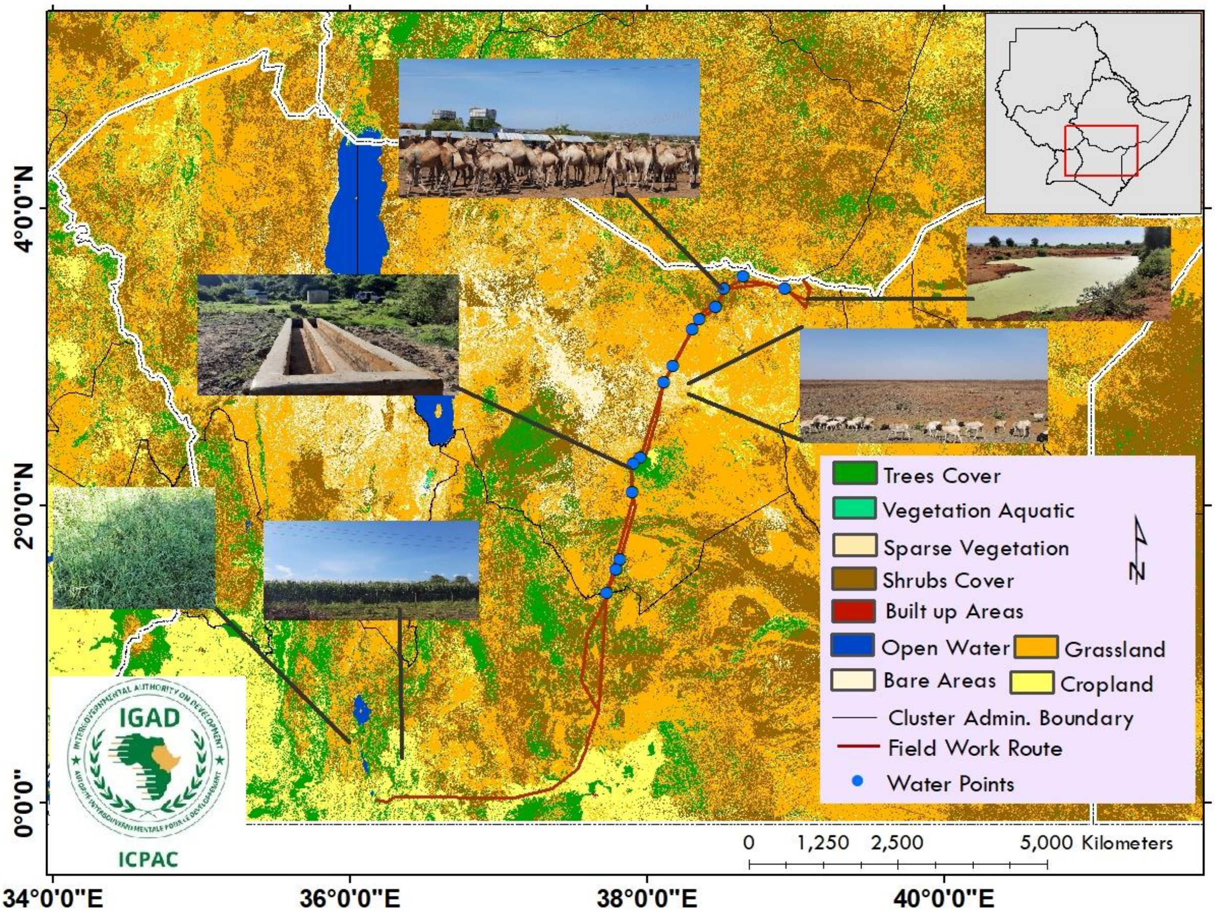


Figure 16: Fieldwork path along different land-cover types with sample photos.

The forage estimation and ground truthing startd after Lororo small scale irrgation scheme (Eldume Kailer) picture on the right, then Ngo'swe area, and finished with Kipcherere region. This covered different land-cover types along the field routes and transects.

Some of the images taken during this activity are shown below (Figure 17 to Figure 20). In all the places we visited, Baringo County was the one affected the most by *prosopis* (mathenge) plant species. Different land cover types were also recorded and sampled between Moyale, Sololo, and Marsabit. The results from the field estimations are discussed in the next section.

Figure 17: Training of the field assistance on how to take measurements in the field (Grassland).

Figure 18: Grass cutting and explanation to the community representative on the nature of work we are conducting in the area.

Figure 19: Taking field measurements at a shrub land cover type in Baringo. Noted that majorly goats feed on this.

Figure 20: Vegetation growing under prosopis affected areas

2.6 FIELD DATA

Field data estimates are presented in Table 7 Table 1below. The measurements were mainly done for shrubs and forage (grass) land cover types during ground truthing for purposes of model calibration. The measurements were done using a weighing machine with units in grams for different areas with a sample taken for square meter (g/m^2). It was then converted to t/ha by dividing the recorded values with 90.71 (using the formula *1 t/ha is equal to 90.71* g/m^2). High forage values were recorded for Barigo County, this was mainly in areas close to cropping areas. Otherwise, areas far from cropping zones are covered with shrubs and *prosopis* species in much of the county. The results obtained from field estimation were mainly towards the end of the season but we will still consider them for further analysis and validation of the prediction model output which is mainly seasonal.

Table 7: Forage estimation at different latitude and longitude in ton/ha.

Latitude	Longitude	Mixed Shrub (g/m^2)	Forage (grass) (g/m^2)	Forage (grass) (t/ha)
0^0 25’ 29.65”	36^0 0’ 33.77”	2,238	1,234	13.604
0^0 25’ 29.70”	36^0 0’ 34.38”	1,135	500	5.512
0^0 25’ 29.86”	36^0 0’ 34.19”	2,038	1,015	11.190
0^0 25’ 31.08”	36^0 0’ 33.78”	2,370	1,430	15.765
0^0 25’ 31.22”	36^0 0’ 34.06”	2,180	1,340	14.772
0^0 25’ 31.89”	36^0 0’ 33.92”	2,125	1,215	13.394
0^0 25’ 33.56”	36^0 0’ 39.64”	2,060	1,160	12.788
0^0 25’ 31.95”	36^0 0’ 42.77”	2,160	1,250	13.780
0^0 31’ 55.64”	35^0 59’ 11.75”	1,090 low lvl	-	-
0^0 31’ 55.37”	35^0 59’ 10.73”	450 hill top	-	-
0^0 31’ 56.52”	35^0 59’ 10.38”	302 hill top	-	-
0^0 32’ 2.98”	35^0 59’ 1.94”	356 hill top	-	-
0^0 34’ 29.76”	35^0 59’ 44.60”	1,209	640	7.055
0^0 34’ 29.09”	35^0 59’ 41.59”	1,570	1,015	11.190
0^0 34’ 29.77”	35^0 59’ 40.20”	1,250	690	7.607
0^0 34’ 29.27”	35^0 59’ 41.04”	1,770	1,285	14.166
1^0 50’ 32.63”	37^0 51’ 47.91”	-	321	3.534
1^0 50’ 35.09”	37^0 51’ 49.06”	-	345	3.803
2^0 57’ 59.11”	38^0 11’ 15.42”	-	31	0.342
2^0 57’ 59.24”	38^0 11’ 16.90”	-	24	0.265
2^0 58’ 4.34”	38^0 11’ 19.14”	-	26	0.287
3^0 17’ 21.10”	38^0 22’ 17.00”	-	19	0.209
3^0 27’ 18.64”	38^0 31’ 0.08”	740	433	4.773
3^0 29’ 9.62”	38^0 48’ 58.08”	451	-	-
3^0 28’ 9.24”	39^0 5’ 40.34”	1, 289	779	8.588
3^0 24’ 35.24”	39^0 4’ 48.06”	1,211	644	7.100
3^0 20’ 16.45”	39^0 4’ 0.98”	1,010	524	5.777
3^0 32’ 44.99”	38^0 38’ 46.03”	1,310	776	8.555
3^0 32’ 35.75”	38^0 38’ 47.40”	1,561	1,040	11.465

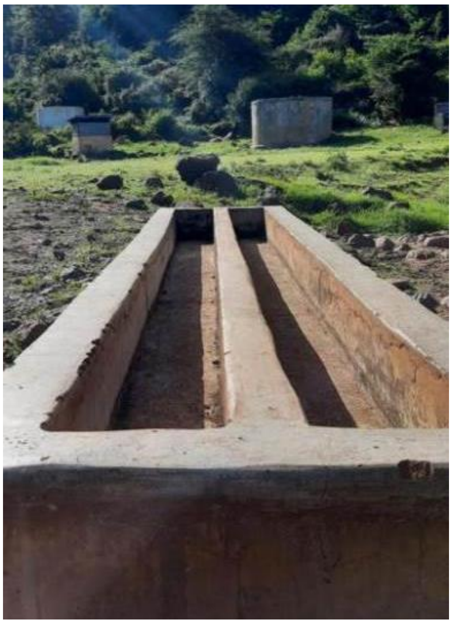
The data in Table 7 were collected in mid-December 2020, this was compared with the total pasture biomass predicted of OND season to have an idea on the comparability (Table 8). However, this was not a good representation for the season. Thus more data should be taken during the start, middle, and end of the rainy season for the three seasons after which an average is done per season.

This will give a good representation that can significantly compared with the predicted values.

Table 8: Field weight measurement for ground truthing for ten different plot with three sample areas in each.

Latitude (Deg.)	Longitude (Deg.)	Predicted OND [t/ha]	Observed December [t/ha]
1.8424	37.8633	1.08	3.53
1.8431	37.8636	1.08	3.80
2.9664	38.1876	2.08	0.34
2.9665	38.1880	2.08	0.27
2.9679	38.1887	2.08	0.29
3.2892	38.3714	3.43	0.21
3.4552	38.5167	5.81	4.77
3.4860	38.8161	5.90	-
3.4692	39.0945	6.23	8.59
3.4098	39.0800	5.85	7.10
3.3379	39.0669	4.56	5.78
3.5458	38.6461	6.41	8.56
3.5433	38.6465	6.41	11.47

2.6.1 Water Points Sampling



The status of water points that were within one kilometer along Isiolo, Marsabit, and Moyale route were sampled and recorded. The water points were majorly of two types i.e. Drilled water points and the pan types. Most of the drilled water points are powered by solar system (Solar Panels) which is a good initiative for sustainability compared to those powered by generator. One of the drilled water point next to Marsabit National park was not working when we visited it (Picture on the right). We asked the community members that were around on the status and they informed us that the "*generator that used to pump the water was stolen from the site and that was why it had been off for a while.*" Thus drilled water points in ASALs should only be powered by solar system or wind as they can serve the community much better than the generator powered water points.

Location and status of the visited water points are shown in Table 9 below which is then followed by pictured taken for some water points.

Table 9: **Water point location, type, and status along the fieldwork route.**

GPS No.	x	y	Elevation	Water Point Type	Status
276	383213	256327	1343	Drilled	Not working Generator Stolen
269	446296	381913	645	Drilled	Working
270	440199	368828	637	Drilled	Working
272	427654	359808	634	Drilled	Working
273	422781	352122	600	Drilled	Working
274	407676	324307	552	Drilled	Working
275	400785	312749	503	Drilled	Working
280	377119	230490	768	Drilled	Working
283	368335	180391	546	Drilled	Working
286	352212	70256	834	Ewasonyiro River	River Flowing
285	357944	155630	600	Merile River	Dry
258	491852	382321	672	Pan	Dry
266	460705	391643	697	Pan	Working
271	440061	368673	623	Pan	Working
277	377940	251814	1244	Pan	Working
284	365208	173246	609	Pan	Working

Figure 21: Protected water point (Pan) for domestic and animal use during dry periods.

Figure 22: Camels taking water in shifts at a drilled water point powered by solar panels

CONCLUSION

The prototype prediction model performs well in pasture biomass prediction and only needs more field validation for continuous improvement and calibration. The field activities including local forage measurements for prediction model calibration should be done three times in a season i.e. from the onset of the rains in a particular season, during the month of the peak rains, and the cessation week of the season. This will improve field data report on biomass production of the different land cover units for different seasons and, an accurate calibrated forage estimation model for estimating of the forage condition and performance of the model since the output is season based. Current water point locations will be added to the early warning system to give more information in the rangeland that is critical for pastoralist, sub-national key actors (government, NGOs), and other relevant policy makers.

RECOMMENDATION

Transhumance data need to be assessed for MAM and OND season as the current available transhumance data is for the JJA season. The available one will only be used when giving the early warning for JJA season only. The is also need for more investment on the water point as it is still a major resource for the pastoralist.

DISCLAIMER

The administrative boundary data used in this work are no warranted to be error free nor do it imply official endorsement and acceptance by IGAD.

REFERENCE

Adeboye, N. O., Fagoyinbo, I. S., & Olatayo, T. O. (2014). Estimation of the effect of multicollinearity on the standard error for regression coefficients. *Journal of Mathematics*, 10(4), 16-20.

Bivand, R., Yu, D., Nakaya, T., Garcia-Lopez, M. A., & Bivand, M. R. (2020). Package 'spgwr'. *R software package.*

Bröcker, J., & Smith, L. A. (2007). Increasing the reliability of reliability diagrams. *Weather and forecasting, 22*(3), 651-661.

Charlton, M., Fotheringham, S., & Brunsdon, C. (2009). Geographically weighted regression. *White paper. National Centre for Geocomputation. National University of Ireland Maynooth*.

Cossins, N.J. & M.Upton, (1987). The Borana Pastoral System of Southern Ethiopia. *Agricultural Systems 25*, pp: 199-218.

FAO, (1988). Guidelines: Land evaluation for extensive grazing. FAO Soils Bulletin No. 58, Rome, 90 pp

Farrar and Glauber (1967), "Multicollinearity in regression analysis" review of economics and statistics, 49, pp. 92-107

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Georganos, S., Abdi, A. M., Tenenbaum, D. E., & Kalogirou, S. (2017). Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression. *Journal of Arid Environments, 146*, 64-74.

Goswami, S., Gamon, J., Vargas, S., & Tweedie, C. (2015). *Relationships of NDVI, Biomass, and Leaf Area Index (LAI) for six key plant species in Barrow*, Alaska (No. e1127). PeerJ PrePrints.

Gudoshava, M., Misiani, H. O., Segele, Z. T., Jain, S., Ouma, J. O., Otieno, G., ... & Atheru, Z. (2020). Projected effects of 1.5 C and 2 C global warming levels on the intra-seasonal rainfall characteristics over the Greater Horn of Africa. Environmental Research Letters, 15(3), 034037.

Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather and forecasting, 12*(4), 736-741.

Hobbs, T. J. (1995). The use of NOAA-AVHRR NDVI data to assess herbage production in the arid rangelands of Central Australia. *International Journal of Remote Sensing, 16*(7), 1289-1302.

Imdadullah, M., Aslam, M., & Altaf, S. (2016). mctest: An R Package for Detection of Collinearity among Regressors. *R J., 8*(2), 495.

Ivajnšič, D., Kaligarič, M., & Žiberna, I. (2014). Geographically weighted regression of the urban heat island of a small city. *Applied Geography, 53,* 341-353.

Jolliffe, I. T., & Stephenson, D. B. (Eds.). (2012). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

Kovács, P., Petres, T., & Tóth, L. (2005). A new measure of multicollinearity in linear regression models. *International Statistical Review, 73*(3), 405-412.

Leilei, L., Jianrong, F., & Yang, C. (2014, March). The relationship analysis of vegetation cover, rainfall and land surface temperature based on remote sensing in Tibet, China. In IOP conference series: earth and environmental science (Vol. 17, No. 1, p. 012034). IOP Publishing.

Lumbierres, M., Méndez, P. F., Bustamante, J., Soriguer, R., & Santamaría, L. (2017). Modeling biomass production in seasonal wetlands using MODIS NDVI land surface phenology. *Remote Sensing, 9*(4), 392.

Matthews, S. A., & Yang, T. C. (2012). Mapping the results of local statistics: Using geographically weighted regression. *Demographic research*, 26, 151.

Nicholson, S. E. (2017). Climate and climatic variability of rainfall over eastern Africa. Reviews of Geophysics, 55(3), 590-635.

Stanski, H. R., Wilson, L. J., & Burrows, W. R. (1989). Survey of common verification methods in meteorology.

Theil, H. (1971). *Principles of Econometrics*. John Wiley & Sons, New York. [p495, 497].

Thirumalai, C., Vignesh, M., & Balaji, R. (2017, April). Data analysis using Box and Whisker plot for Lung Cancer. *In 2017 Innovations in Power and Advanced Computing Technologies* (i-PACT) (pp. 1-6). IEEE.

Toxopeus, A. G. (2000). Spatial and temporal modelling for sustainable management of semi-arid rangelands. In *Proceedings of the 28th*

International Symposium on Remote Sensing of the Environement: Information for Sustainable Development. March (pp. 27-31).

Vellinga, M., & Milton, S. F. (2018). Drivers of interannual variability of the E ast A frican “L ong R ains”. Quarterly Journal of the Royal Meteorological Society, 144(712), 861-876.

Wang, K., Zhang, C., & Li, W. (2013). Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. *Applied Geography, 42,* 73-85.

Wichers, C. R. (1975). The detection of multicollinearity: A comment. *The Review of Economics and Statistics*, 366-368.

Wijngaarden, W. van, (1985). Elephants, trees, grass, grazers; relationship between climate, soil, vegetation and large herbivores in a semi-arid savanna ecosystem (Tsavo, Kenya). *PhD thesis, Wageningen, 159 pp*.

Wilks, D. S. (1995). Statistical Methods in the Athmospheric Sciences, volume 59 of International Geophysics Series. *Academic Press*, (0.2), 0-4.

Williamson, D. F., Parker, R. A., & Kendrick, J. S. (1989). The box plot: a simple visual method to interpret data. *Annals of internal medicine, 110* (11), 916-921.

ANNEX 1

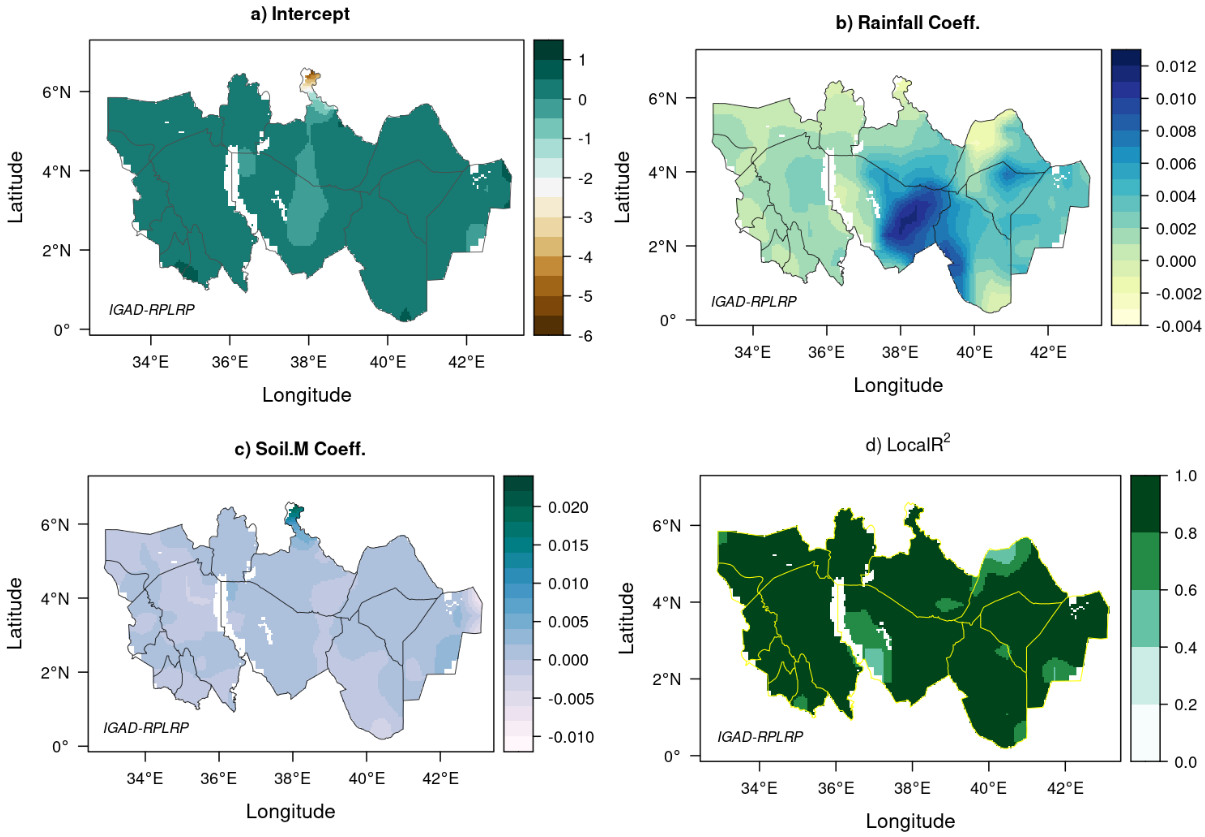


Figure 23: JJA season constants used in modelling JJA seasonal Biomass, (a) is the intercept, (b) is the rainfall coefficient, (c) is the soil moisture coefficient, and (d) is the local R square.

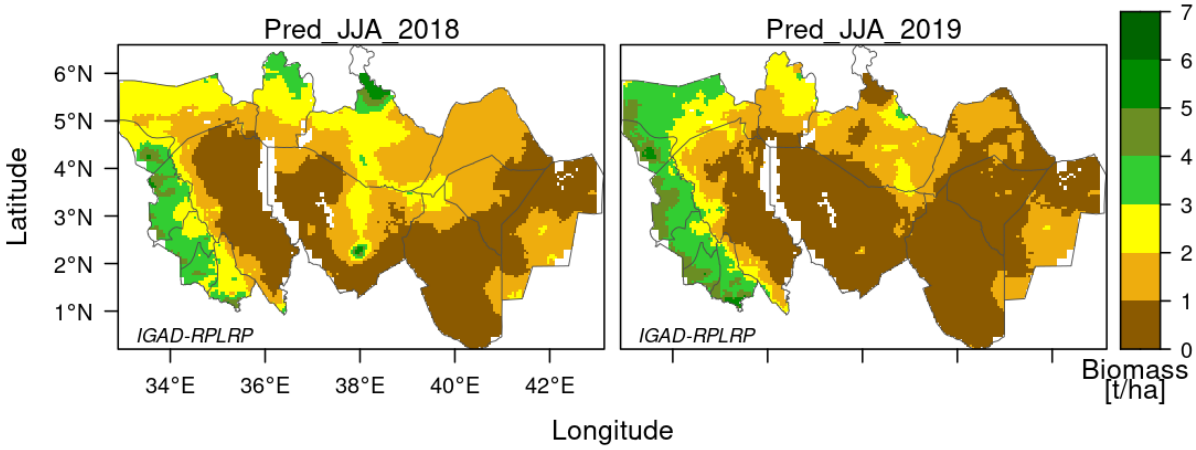


Figure 24: Predicted Biomass for JJA season over the study area from 2018 to 2019 without considering the error term.

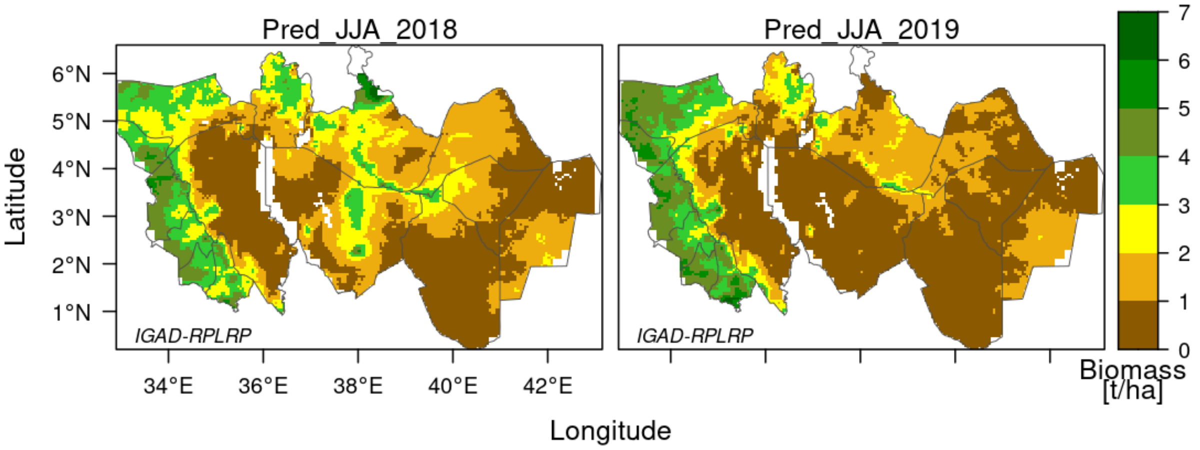


Figure 25: Predicted pasture biomass for JJA season over the study area from 2018 to 2019 with the error term considered.

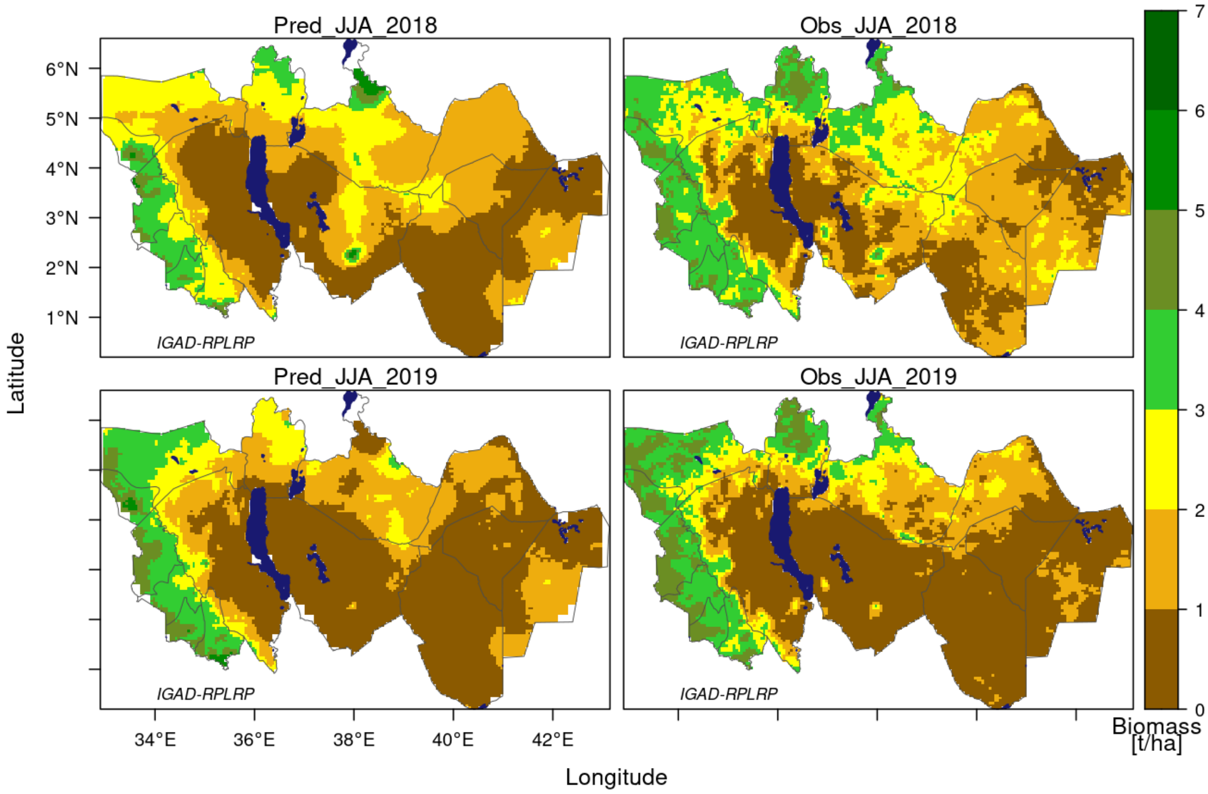


Figure 26: Predicted pasture biomass with bias compared with observed pasture biomass for the JJA seasons from 2018 to 2019.

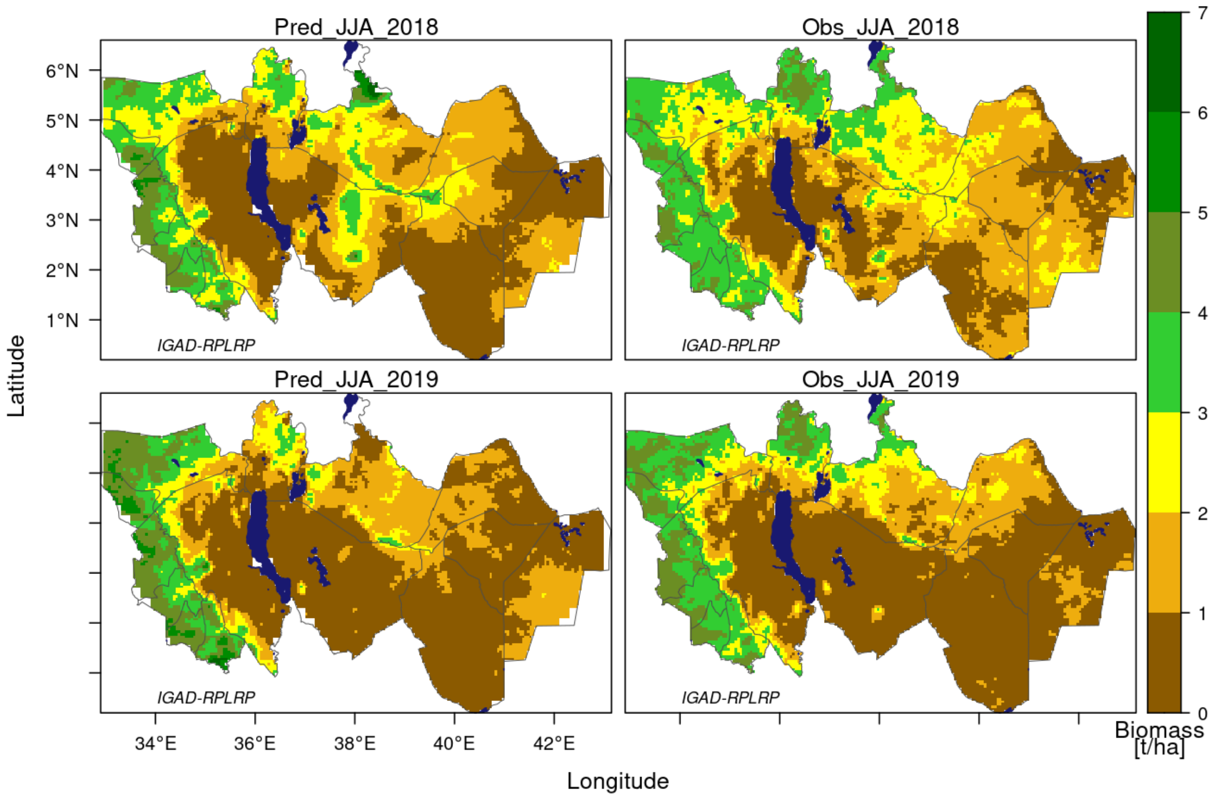


Figure 27: Bias corrected predicted pasture biomass compared with observed pasture biomass for the JJA seasons from 2018 to 2019.

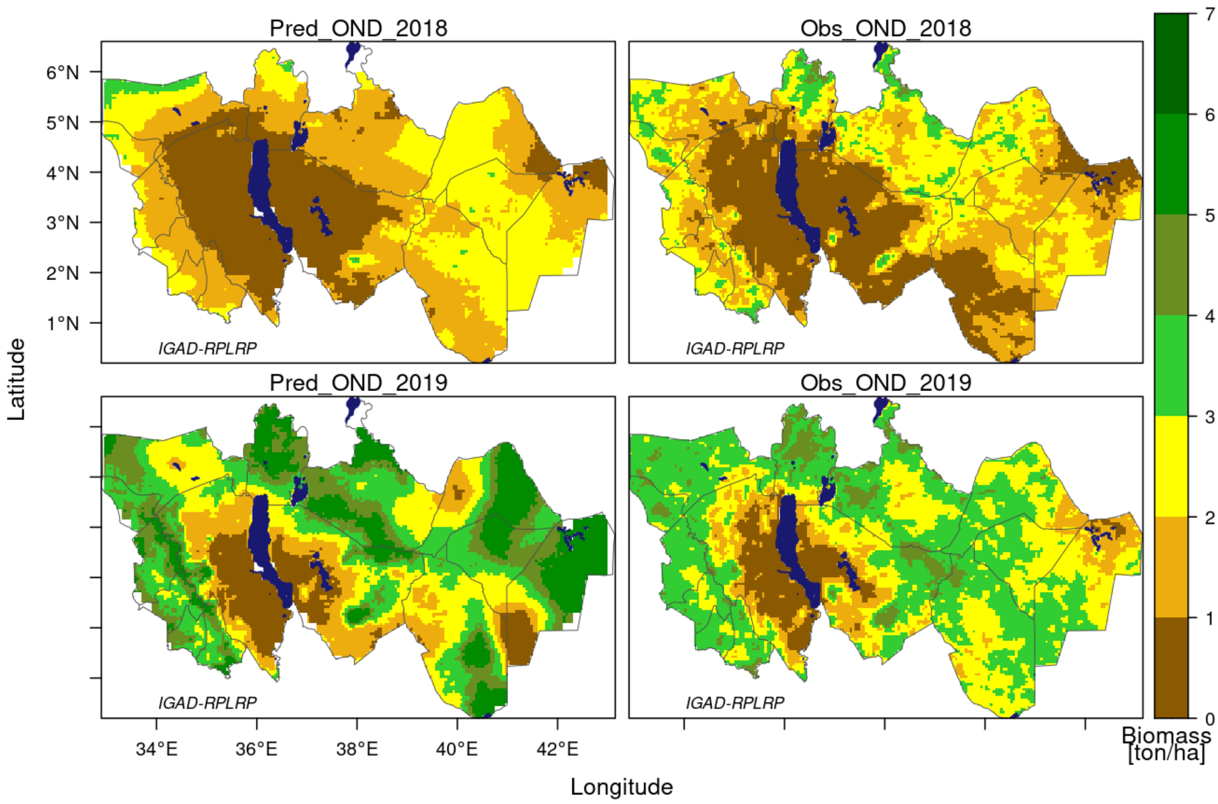


Figure 28: Predicted pasture biomass with bias compared with observed pasture biomass for the OND seasons from 2018 to 201

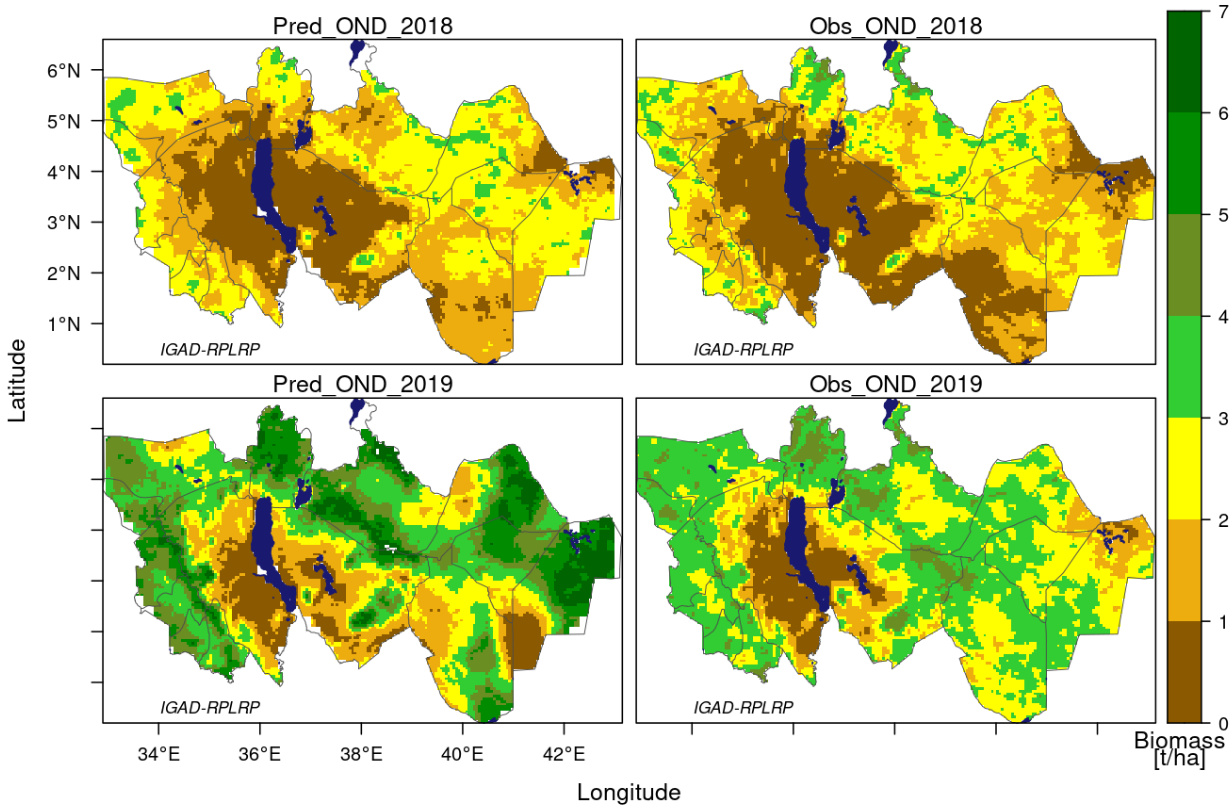


Figure 29: Bias corrected predicted pasture biomass compared with observed pasture biomass for the OND seasons from 2018 to 2019.

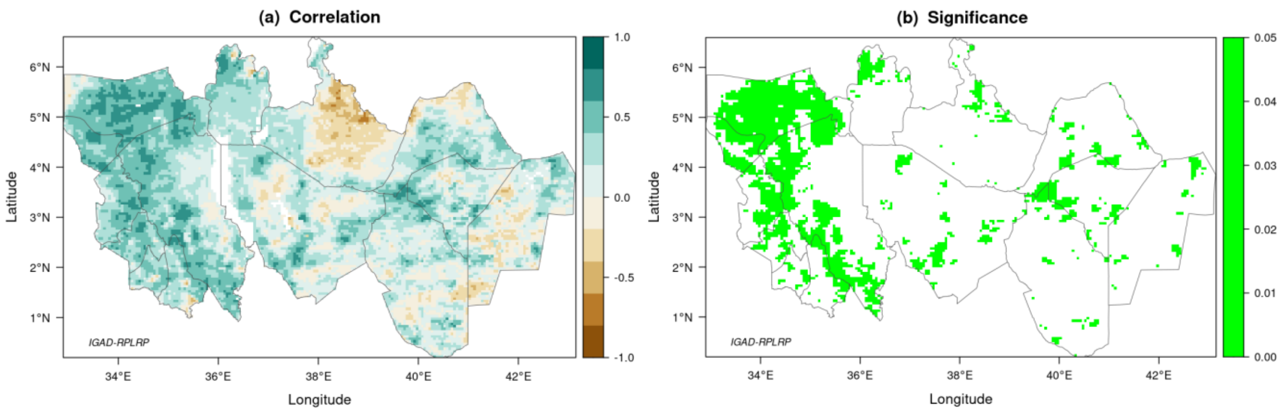


Figure 30: Spatial correlation for JJA season between maximum NDVI and rainfall total over the area for the period 1999 to 2017.

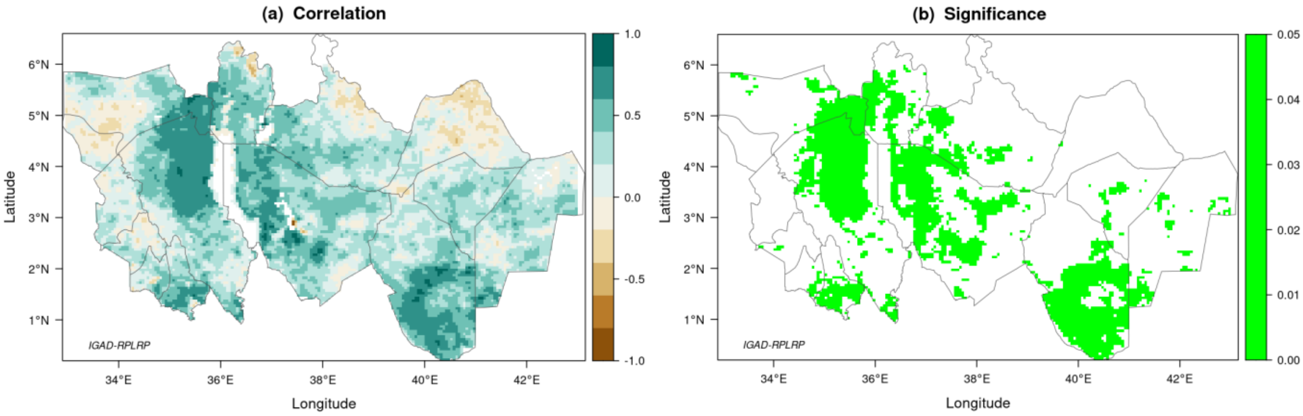


Figure 31: Spatial correlation for JJA season between maximum NDVI and maximum soil moisture over the area for the period 1999 to 2017.