Habitat suitability study for green gram production under present and future climatic scenarios in Kibwezi East Kenya

Maluvu Zipporah*, Oludhe Christopher†, Kisangau Daniel‡, Maweu Jacinta Mwende§

†Department of Earth and Climate Sciences, University of Nairobi, Nairobi, Kenya, ‡Department of Life Sciences, South Eastern Kenya University (SEKU), Kitui, Kenya, §Department of Journalism & Communication, University of Nairobi, Nairobi, Kenya

ABSTRACT

The species distribution model was used to predict the suitability of green gram production under the present, RCP 4.5 and 8.5 climate scenarios. An ensemble of a species distribution model comprising six models was developed. Validation of these models revealed that all models were robust with the best model being random forest (RF) with Area Under the Curve (AUC) = 0.98 and Deviance = 0.29 while the least was the generalized linear model (GLM) with AUC = 0.87 and Deviance = 0.71. The green gram habitat suitability greatly decreased under RCP 8.5 climate scenario prediction whereby about half of the agricultural land in the Kibwezi East Sub County was highly unsuitable for green gram production. The Habitat suitability predictions showed that Thange ward out of the four wards in the location was the most suitable for green gram production. However, as per the predictions its suitability for green gram production may be affected by climate change under all climate scenarios. Results from this study give decision-makers a foundational understanding of the likely effects of climate change in the 2050s compared to the present scenario on habitat suitability for green gram production and a basis for creating strategies and policies to enhance adaptation and create resilience to its effects.

KEYWORDS: Adaptation, Climate Change, Green gram, Habitat suitability, Kibwezi East Sub-County, Species Distribution Modelling

INTRODUCTION

Green gram is one of the pulse crops that are important for food and income generation in the arid and semi-arid lands (ASALs) of Kenya. The crop has been cultivated in these areas for decades. Predicted climate change scenarios in the future may cause adverse effects on the habitat suitability for green gram production in some areas.

Climate change is likely to affect the high potential and low potential agricultural areas making them unsuitable for crop production. Habitat/climate suitability analysis helps to determine areas that are suitable for species survival and find out if they will remain suitable in the future (Halder, 2013). Research on climate suitability is important in adapting to and mitigating the effects of climate change. Green gram is cultivated in more than six million hectares in the warmer locations of the world.

It is also adapted to a wide range of agroecological climates and requires low inputs during production (Nair et al., 2012). A study by Mugo et al. (2020), found Kenya to be suitable for green gram production whereas Kitui, Makuene and West Pokot Counties were highly suitable for green gram production.

Species distribution models (SDMs) are currently the most commonly used tools for predictions of species habitat suitability (Jarvie & Svenning, 2018). SDMs can have some degree of uncertainties related to inherent variability of natural systems (Gould et al., 2014; Noce et al., 2019). However, ensemble forecasting has been revealed as an effective method of modelling to reduce variability of SDMs. The ensemble forecasting employs more than one model and therefore it combines several sources of uncertainty to improve the accuracy of climate change forecasts (Buisson et al., 2010; Taleshi et al., 2019). SDMs can be combined with Global circulation models

Copyright: © The authors. This article is open access and licensed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted, use, distribution and reproduction in any medium, or format for any purpose, even commercially provided the work is properly cited. Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
and Greenhouse Gas (GHG) emission concentration pathways (RCPs) to quantify projections in climate change studies (Vieilledent et al., 2013; Zhang et al., 2015; Noce et al., 2017). RCPs are dependent trajectories of concentrations of GHGs and other environmental pollutants resulting from human activities (AR5) (IPCC, 2014). The RCP 4.5 stabilizes radiative forcing at 4.5 w/m² in the year 2100 when strategies and technologies to reduce GHG emissions are employed. RCP 8.5 is characterized by increasing GHG emissions and high concentration levels in the atmosphere. The increasing emissions present a rising radiative forcing pathway leading to 8.5 w/m² in 2100 approximately 1370 ppm equivalent (Noce et al., 2017; del Río et al., 2021).

Studies on green gram production and climate change have been carried out in Kenya (Mugo et al., 2016, 2020). The effects of climate change on habitat suitability have been analyzed from different forms of view (Kufa et al., 2022). To our knowledge, no study has been conducted in Kibwezi East Sub County using ensemble-forecasting models combining standard bioclimatic variables and biogeographic predictors to model habitat suitability of green grams production under present and future climate scenarios.

Therefore, this study analyzed the habitat suitability situation for green gram in Kibwezi East subcounty, Kenya. It was carried under the present (1970-2000) and the future climate scenarios in the 2050s (2041-2060) under RCP 4.5 and RCP 8.5 climate scenarios using the Species Distribution Modelling (del Río et al., 2021). The study combined an ensemble of species distribution models applied to Global circulation models driven by two representative concentration pathways.

MATERIALS AND METHODS

Study Location

This study was carried out in Kibwezi East Sub County in Makueni County, Kenya (Figure 1). Kibwezi East Sub County is approximately 200km south east of Nairobi and lies between longitude 37°58’4.25” E and latitude 2°24’37.89” S. Makueni County is approximately 8,034.7 Km² (GoK, 2013). It is divided into six sub-counties namely: Kaiti, Kilome and Mbooni, which are situated in the upper parts of the county while Makueni, Kibwezi East and Kibwezi West are situated in the middle and lower parts of the county. Kibwezi East Sub County has four wards namely; Thange, Mtito Andei, Ivingoni/Nzambani and Masongaleni.

Data Collection

World species presence data for green grams was obtained from the Global Biodiversity Information Facility (GBIF) using dismo package in R statistical software (R) (Fick & Hijmans, 2017). Using raster package in R, historical/present data from 1970-2000 that was released in 2020 was accessed from the Worldclim version 2 database. A total of 19 standard Worldclim bioclimatic variables were downloaded at a resolution of 30 seconds which is about an area of 1km² per pixel. More information about the standard bioclimatic variables is found on https://www.worldclim.org/data/bioclim.html. Bioclimatic variables with multicollinearity were identified and removed using usdm package in R (Chan et al., 2022). The remaining nine standard bioclimatic variables were used in this study (Table 1).

Future bioclimatic data was downloaded from Worldclim version 2 at a resolution of 30 seconds for the year 2050s, which are the years between 2041 and 2060 using raster package in R. This was downloaded from Coupled Model Intercomparison Project Phase 5 (CMIP5) downscaled data for future climate projections for the IPCC5 climate projections and global climate models (GCMs) for two representative concentration pathways (RCPs) i.e., RCP 4.5 and RCP 8.5. The Greenhouse Gas emission pathways provide a time and space dependent trajectory of concentrations of GHGs.

Other data included elevation, soil type and soil pH. Soil type and pH were obtained from (International Soil Reference and Information Centre (ISRIC) at a resolution of 250 m per pixel while elevation was obtained from Worldclim (Fick & Hijmans, 2017) at a resolution of 30 seconds. These were kept constant for the three simulations; current, RCP 4.5 and RCP 8.5 climate scenarios.

Development of green gram habitat suitability model

Using sdm package in R, six models namely; generalized linear model (glm), support vector machine (Svm), random forest (rf), boost regression tree (btr), multivariate adaptive regression splines (mars) and maximum entropy (maxent) were trained. The choice of these models was based on previous studies which stated their importance in providing robust predictions (Hastie et al., 2009; Latif et al., 2013). The algorithms were combined into an ensemble. The procedure was carried out using ten bootstrap replicate run types.
Table 1: Standard bioclimatic variables used for model development

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>B102</td>
<td>Mean Diurnal Range (Mean of monthly (max temp - min temp))</td>
</tr>
<tr>
<td>B103</td>
<td>Isothermality (B102/B107) (×100)</td>
</tr>
<tr>
<td>B108</td>
<td>Mean Temperature of Wettest Quarter</td>
</tr>
<tr>
<td>B109</td>
<td>Mean Temperature of Driest Quarter</td>
</tr>
<tr>
<td>B113</td>
<td>Precipitation of Wettest Month</td>
</tr>
<tr>
<td>B114</td>
<td>Precipitation of Driest Month</td>
</tr>
<tr>
<td>B115</td>
<td>Precipitation Seasonality (Coefficient of Variation)</td>
</tr>
<tr>
<td>B118</td>
<td>Precipitation of Warmest Quarter</td>
</tr>
<tr>
<td>B119</td>
<td>Precipitation of Coldest Quarter</td>
</tr>
</tbody>
</table>

B10 = Bioclimatic variable

Model Evaluation

The prediction quality of the six models were evaluated using area under the curve of receiver operating characteristic plot (AUC_{ROC}) (Hosmer & Lemeshow, 2013), Pearson’s correlation between predicted and observed species presence-absence data (COR), true skill statistics (TSS) (Allouche et al., 2006) and deviance (Lobo et al., 2008). This was important in evaluating the model specificity and sensitivity. Specificity is the ability of a model to predict absence in a location while sensitivity is the ability of a model to predict presence in a location (Zurell, 2020).

Green Gram Habitat Suitability Predictions

The ensemble of the developed models was used for predicting green gram habitat suitability under present scenario, RCP 4.5 and RCP 8.5 climate scenarios. This was carried out with the assumption that the species are at equilibrium with the environment and that they should also respond dynamically to global change show transient dynamics (Urban et al., 2016; Zurell, 2017). To make predictions for Kibwezi East, bioclimatic datasets from the predicted future data were cropped to scale down to Kibwezi East boundary coordinates accessed from Global Administrative Areas (GADM) version 4.1 (www.gadm.org). Green gram habitat suitability predictions were carried out using ensemble function of sdm package in R. These predictions were visualized using plots plotted from the predicted results using ggplot2 package in R. Three final suitability maps were presented under the present scenario, RCP 4.5 and RCP 8.5 climate scenarios as proposed by the IPCC AR5 (Mastrandrea et al., 2011; Noce et al., 2017). Percentage suitability change was computed in R.

RESULTS

Green gram distribution and presence data across the world indicated that the crop is majorly grown in the Asian countries and Australia (Figure 2). The green gram presence data used for modelling covered all global geographic areas suitable for its growth. The use of global green gram presence data increases environmental variability which enhances the reliability of the developed model. Species can only survive in geographic areas where both the abiotic and biotic conditions allow positive plant population growth of that particular species (Zurell, 2020).

Relative variable importance averaged over the six models was used to make an ensemble is shown in Figure 3. Out of nine bioclimatic variables, four had the highest contribution to the ensemble (Figure 3). Among these variables, bio9 (Mean Temperature of Driest Quarter) had the highest mean relative variable importance in the development of green gram habitat suitability ensemble model. This bioclimatic variable provides the mean temperature during the driest three months of the year. The second highest mean relative important variable in this study was bio3 (Isothermality) while bio14 (Precipitation of Driest Month) had the lowest mean relative variable importance.

Validation statistics showed that all models that were developed were robust (Table 2). Based on AUC_{ROC} values, the least was 0.87 for glm and highest was 0.98 for rf model. The highest COR values were (0.89) for the rf model while the least value was (0.65) for glm model. Trues Skill Statistic (TSS) values ranged from 0.59 for glm to 0.88 for rf. The Deviance values ranged from 0.29 for rf model to 0.94 for maxent model.

The AUC_{ROC} plots show AUC means for training and testing sets (Figure 4). From these plots, the training and testing AUC values for all models fitted well. The highest values were for the rf model where training set AUC value was 1 and testing AUC value was 0.98. The least values were in glm model with training set AUC value of 0.87 and testing AUC value of 0.87.

Figure 5 shows green gram habitat suitability plot for the present climatic scenario. This figure was plotted from results predicted using the developed ensemble model. The results showed that green gram habitat suitability was high in Ivingoni/Nzambani, Thange and Masongaleni wards. However, the western parts of Ivingoni/Nzambani and Thange wards had medium habitat suitability for green gram. A greater part of Mtito Andei ward had medium green gram habitat suitability with a small section of its southern part having very low habitat suitability.

In the year 2050s under RCP 4.5 climate scenario, the results (Figure 6) revealed loss in habitat suitability compared to present climatic scenario as shown in Figure 5. The habitat suitability for growing green grams in the western parts of Ivingoni/Nzambani and Thange wards greatly reduced from relatively...
Figure 3: Relative variable importance averaged over six models

Table 2: Model evaluation Statistics

<table>
<thead>
<tr>
<th>Models</th>
<th>AUC ROC</th>
<th>COR</th>
<th>TSS</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>0.87</td>
<td>0.65</td>
<td>0.59</td>
<td>0.71</td>
</tr>
<tr>
<td>SVM</td>
<td>0.95</td>
<td>0.81</td>
<td>0.8</td>
<td>0.45</td>
</tr>
<tr>
<td>RF</td>
<td>0.98</td>
<td>0.89</td>
<td>0.88</td>
<td>0.29</td>
</tr>
<tr>
<td>BRT</td>
<td>0.91</td>
<td>0.71</td>
<td>0.67</td>
<td>0.75</td>
</tr>
<tr>
<td>MARS</td>
<td>0.9</td>
<td>0.71</td>
<td>0.66</td>
<td>0.61</td>
</tr>
<tr>
<td>MAXENT</td>
<td>0.91</td>
<td>0.7</td>
<td>0.67</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Predicted climate data for the 2050s showed that green gram habitat suitability reduced greatly under RCP 8.5 compared to RCP 4.5 climate scenario (Figure 7). This was evident in Mtito Andei ward, where there was a loss of suitability making it unsuitable for green gram production. Habitat suitability for green gram production increased from medium to high in the central parts of the study area.
Andei and the eastern parts of Masongaleni wards. Green gram habitat suitability in the western parts of Ivingoni/Nzambani and Thange wards was predicted to reduce from Medium to low. In Figure 7, suitability in the central part of the sub county appears to decrease from high to medium. Overall, the predictions in the 2050s under RCP 8.5 climate scenario show that about 50% of the study location will have low habitat suitability for green gram production, about 25% will have medium suitability and the remaining 25% will have high habitat suitability for green gram production.

Table 3 shows the predicted habitat suitability change in percentage. The results show that the habitat suitability will change by 10.24% from the present climatic scenario to RCP 4.5 scenario in the 2050s. However, under the RCP 8.5 climatic scenario the suitability will change by 18.8% from the present scenario.

DISCUSSION

An ensemble of six species distribution models was developed and used to predict future green gram habitat suitability under RCP 4.5 and RCP 8.5 climate scenarios. In model development, variable importance is used to indicate the degree at which that variable affects the response value (Inglis et al., 2022). Among the standard bioclimatic (bio) variables used, bio9 had the highest variable importance averaged across the 6 models while bio14 had the lowest. Therefore, bio9 which contains the mean temperature of driest quarter significantly varied across geographical regions and consequently, it had the highest contribution to the model (Xie & Zhang, 2023). The developed model had 6 out of 9 predictors with high variable importance. This implied that the model was robust. This result on relative variable importance revealed the relative importance of each environmental factor in affecting species distributions (O’Donnell & Ignizio, 2012). Based on this result, it was concluded that the mean temperature during the driest three months of the year (bio9) contributes most to green gram habitat suitability.

Model validation results revealed that all models in the ensemble were robust in predicting green gram habitat suitability. The model validation statistics showed that the least robust model had AUC_{ROC} = 0.85. Validity of a model can be effectively measured using its sensitivity and specificity components (Kumar & Indrayan, 2011). This method relies on receiver operating characteristic curve (ROC) and depicts the trade-off between sensitivity and (1-specificity) across a series of points (Gajowniczek et al., 2014). A high AUC_{ROC} value was an indication that the developed model was robust to use in the prediction of habitat suitability (Kumar & Indrayan, 2011). Based on AUC_{ROC} statistics, all models were reliable. The COR statistics is the Pearson’s correlation between the predicted and observed species presence-absence data. A model is more robust when the COR values approaches one and less robust when this value approaches zero (Han & Liu, 2017). High COR values indicates that the models fitted the predicted and observed values well (Smith & Santos, 2020). In the developed models, the COR values ranged between 0.65 and 0.89 indicating that the models were robust.

Another metric used for model evaluation is True Skill Statistic (TSS). This is the most effective metric that is employed in validating species distribution models (Yoon & Lee, 2023). Calculation of TSS involves a confusion matrix composed of number of correct and incorrect predictions of species presence-
absence data (Allouche et al., 2006). This method accounts for sensitivity, specificity and accuracy and therefore, it is the most practical metric. In this study, the results show that the TSS values ranged between 0.59 and 0.88 and therefore, the models were reliable for habitat suitability prediction. The last metric used to evaluate the models was Deviance. Deviance is a measure of error and therefore, lower values means that the model fits well (Di Mari et al., 2023). In the results, out of the six models, four had Deviance values of more than 0.60. Based on this metric, the individual models did not perform well. It is on this premise that ensemble of models becomes helpful. An ensemble combines all sources of error from the various models and works on it to improve the accuracy of prediction (Buisson et al., 2010; Taleshi et al., 2019). Therefore, based on high accuracy of developed models and the use of the ensemble of the six models to fix uncertainties, the final ensemble model was highly reliable to predict green gram habitat suitability in Kibwezi East and other regions with similar agro climatic conditions. This accuracy is attributed to the use of predictors from wide geographical regions across the world and the application of an ensemble of models.

This research shows that climate change will have a great impact on green gram habitat suitability in Kibwezi East sub-county. Effects of climate change will significantly contribute to green gram habitat suitability loss. The result shows that habitat suitability will reduce by 18.8% in RCP 8.5 and 10.24% in RCP4.5. High habitat suitability loss in RCP 8.5 compared to RCP 8.5 scenario is due to high emissions of greenhouse gas from anthropogenic sources. This leads to increased temperatures in RCP 8.5 climate scenario compared to RCP 4.5 scenario. Green gram habitat suitability will greatly reduce in Mtito Andei and the western part of Ivingoni/Nzambani and Thange.

Habitat suitability models assess the environmental conditions and factors that influence the growth and development of crops. These models can identify areas with the most favorable conditions for crop cultivation. This information is invaluable for farmers in selecting suitable sites for planting, maximizing crop yield potential and minimizing risks associated with environmental constraints and climate change. They also aid in land use planning and management, helping to allocate resources efficiently and sustainably (Dastres et al., 2023).

The models facilitate risk assessment by identifying potential yield-limiting factors and guiding adaptation strategies to adapt to their impacts (James et al., 2017).

Applying habitat suitability models in green gram farming is essential for optimizing production, ensuring sustainability, and building resilience against risks associated with climate change. These models empower farmers with valuable insights into site suitability, crop performance and adaptation strategies, ultimately contributing to improved livelihoods and food security in green gram-producing regions.

### CONCLUSION

The results indicated that the habitat suitability for green gram production in Kibwezi East was well described by the selected predictors. Based on the evaluation statistics, all single models were included in the final ensemble forecasting model. The model that was developed will provide information on the current situation in green gram production and prediction on future performance of the crop under the various climate change scenarios and emission trajectories. It is a very important tool for decision making on green gram cropping management systems in the arid and semi-arid lands as well as in development of climate change mitigation and adaptation strategies. Climate change effects are experienced differently in different places. The analysis for Kibwezi east Sub County predicts a decline in habitat suitability. This calls for adaptation strategies as well as climate mitigation strategies. Joint efforts by all stakeholders are needed to devise the strategies because green gram is a major crop in this location. The study presents the importance of conducting the analysis at the biogeographic level since the effects of climate change are experienced differently in every area and require engagement and conservation policies at the local level.

### ACKNOWLEDGEMENTS

The authors wish to thank the county agriculture officers in Kibwezi East Sub-County who helped in institutional support and administrative assistance throughout the study as well as the respondents in the Sub County who participated in data provision through interviews and focus group discussions. Special thanks to Wilfred Abincha and Dr. Murenga Mwimali for the insights offered in putting this study into perspective.

### REFERENCES


Di Mari, R., Ingrassia, S., & Punzo, A. (2023). Local and Overall Deviance