

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,300

Open access books available

130,000

International authors and editors

155M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Prediction of Crop Yields under a Changing Climate

Godfrey Shem Juma and Festus Kelonye Beru

Abstract

The impact of increasing climate variability on crop yield is now evident. Predicting the potential effects of climate change on crops prompts the use of statistical models to measure how the crop responds to climate variables. This chapter examines the usage of regression analysis in predicting crop yield under a changing climate. Data quality control is explained and application of descriptive statistics, correlation analysis and contingency tables discussed. Methodological aspects of crop yield modeling and prediction using climate variables are described. Estimation of yield via a multilinear regression approach is outlined and an overview of statistical model verification introduced. The study recommends the usage of regression models in estimating crop yield in consideration of many other externalities that can contribute to yield change.

Keywords: crop, yield, prediction, climate change, regression model

1. Introduction

In this chapter, we describe an experimental approach that can be employed in predicting crop yield in a changing climate. An introductory applied approach to linear statistical modeling and correlation analysis is examined.

Climate change is now evident with well documented socio-economic impacts that will affect food production [1, 2]. The decline in food production corresponding to reduction in crop yields can be investigated using statistical models [3, 4]. While climate related factors can affect yield of crop, there are other externalities that can impact on yield production that include the quality of soil, usage of commercial fertilizers or organic manures and residual effects of chemical substances in soils [5, 6].

Figure 1 below shows the projected changes in crop yield due to the impacts of increasing climate variability.

Increasing climate variability and associated uncertainties, its impacts on food production and general livelihoods prompts the usage of prediction models to estimate future food production for early warning and planning. Projection of crop yield in a changing climate has been identified with uncertainties that are continuously being reduced by improvement in climate parameter response functions, including temperature [7]. Developing countries have also been identified with weaker monitoring and reporting of crop health which can lead to absence of early warning systems and slow responses to droughts and potential food shortages [8]. The prediction models employed are broadly classified into statistical or dynamic (mechanistic). However, the modeling has been in some instances enhanced by artificial neural network technology that has been applied to generate regional

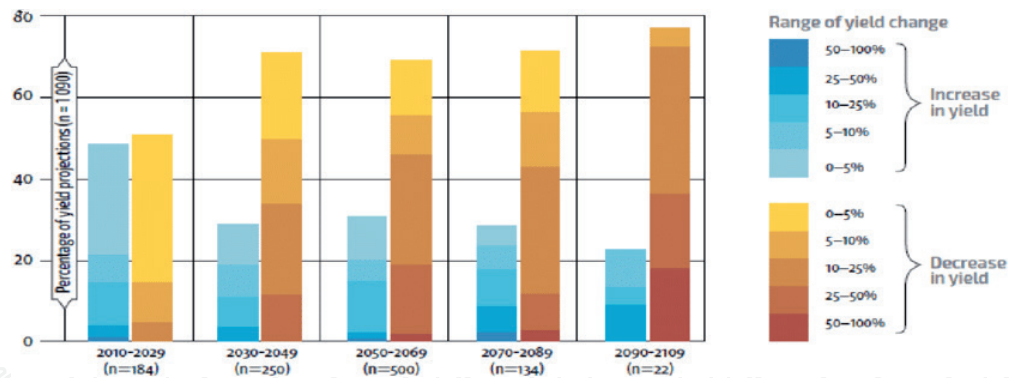


Figure 1.

Projected changes in crop yield associated with increasing climate variability (source: [7]).

time series of crop yield utilizing highly resolved output of the global and regional models. The Global Circulation Models (GCMs) give coarse output of important climate parameters preferably applicable on larger spatial scales while the Regional Circulation Models (GCMs) give fine scale resolutions of GCMs on shorter spatial scales. Where the variables of interests cannot be described by standard linear models, nested-error regression model having both fixed and random effects are usually applied to include Monte Carlo simulation methodology in order to enhance representative precision. Nested-error modeling is better at regional spatial scales and performs poorly at large scale spatial coverage [9].

Application of computing technology has seen historical climate data sets of at least 30 years of a given crop used with artificial neural network technology to investigate, simulate and predict historical time series of crop yields in climate zones over regions. Resultant neural networks are trained with data sets of selected climate zones and tested against an independent zone in order to enhance the power of crop yield predictability [10]. A combination of neural networks and fuzzy set theory has also been applied to construct Fuzzy Neural Network (FNN) and Granular Neural Network (GNN) that have been used for predicting crop yields over different spatial locations with inputs from spot vegetation cover data [11].

Assessment of vegetative cover using standardized vegetative indices has also been adopted to give estimates of crop yield over regions either separately or in combination of the above approaches. In this method, easily measurable proxies are applied that include Normalized-Difference Vegetation Indices (NDVI), Green vegetation indices (GVI), Soil Adjusted Vegetation Indices (SAVI), Back-propagation Neural Network (BPNN) that are positively correlated with crop yield [12]. Statistical models that combine the vegetative and thermal indices from satellite data have performed better in predicting crop yield compared to those that are based on vegetative cover indices alone [13]. While mechanistic models have been applied alongside statistical models, the later have been able to reproduce key features of crop responses to warming and precipitation changes using a process-based model approach [3, 14]. The Crop Environment Resource Synthesis (CERES) model has been applied as a decision making tool in crop yield estimation [15] while PRECIS (Predicting REgional Climate for Impact Studies) has been used to assess the impacts of climate change on crop yield [14].

Therefore, the need to continuously enhance understanding on yield estimation and prediction in a changing climate is continuous and cannot be over emphasized. It is in this back drop that this study makes attempts to adopt a simplified linear statistical approach applicable in crop yield estimation. Basic statistical description is defined and methodology discussed. Hybrid models that are both statistical and mechanistic, integrated by neural network technology based on

multiple variables of climate and crop physiological importance are found to have higher crop yield predictive power.

2. Statistical determination of crop yield

2.1 Descriptive statistical analysis

This entails describing data in statistical summaries meaningfully without making conclusions beyond the data. Measures of central tendency and measures of spread are widely adopted in describing data. The former involves the determination of the mean, median, mode, skewness and kurtosis while the later includes variance and standard deviation. These parameters can be used to describe climate and crop yield data as a preliminary approach alongside data quality control.

Data quality control involves approaches that are used to detect defections and inconsistencies in data sets. Various methods are used to determine the quality of climate data including linear regression approaches. In one such method, a single mass curve technique is applied where cumulative values of climate variable are plotted against a linear scale. The tendency of the resulting curve to shift towards linearity is identified with better quality data. This method is also called the data homogeneity test. Data that fails this test or data with more than 10 percent of missing values is judged to be of poor quality and not fit for inferential statistics.

2.2 Crop yield modeling and prediction

Statistical models have been applied in predicting crop yield and their ability to accurately predict yield responses to changes in mean temperature and precipitation has been determined by process-based crop models. Prototype models include Crop Environment Resource Synthesis (CERES) that can be applied to a crop to simulate corresponding yield and can be used for projecting future yield responses, with their usefulness higher at broader spatial scales [16]. *Mechanistic* models are also used alongside statistical models to predict crop yields [17]. Crop Yield Simulation and Land Assessment Models (CYSLAM) are applied to model the interaction of environmental variables, physiological responses, inputs, yields and land management mechanistic simulations of crop yield [17].

Yields constrained by radiation and temperature within 10 day periods (dekads) are initially estimated in order to account for effective rainfall, evapotranspiration, percolation, and soil moisture. The procedure is followed by a simulation of crop/soil water balance through the cycle of crop growth accounting for periods of moisture stress and consequently, estimation of crop yield [17]. The moisture-dependent yield is adjusted for nutrient supply, toxicities and drainage conditions of the soil [17]. However, validation of modules for moisture limited yield, nutrient yield and radiation and temperature limited yield is carried out separately in comparison with historical crop data.

2.2.1 Multilinear regression yield estimation

Single mass curve technique is used for data quality control where Cumulative values of data are plotted against a temporal scale. The nature and variability of climate elements is determined including the mean, skewness, standard deviation, students' t-test and correlation analysis. Trend is determined by dividing the data into two sets of equal length, and the difference in the means of the two sets is tested using the t-test [4].

The Relationship between crop Yield and Variations in Climatic Elements is carried out by Correlation Analysis, the degree of relationship between at least two variables. The Pearson's correlation coefficient (r) is used to determine the correlation between the climate elements and the crop yield according to the following expression (1) below:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2 \right]^{1/2}} \quad (1)$$

where

N => Total number of observations

\bar{x} => Mean of the variable 'x'

\bar{y} => Mean of the variable 'y'

Hypothetical approach is employed to test for statistical significance of the degree of association in (1) above. The null hypothesis that the correlation is zero and the alternative hypothesis that the correlation is nonzero is assumed. In this case, if the null hypothesis is valid, the relevant test variable (t) from Eq. (2) is a realization of student (t) random variable with mean (zero) and (n) degrees of freedom. P values are computed where $p < 0.05$ prompts the probability of rejecting the null hypothesis and vice versa. The student t-statistic can be used as given by the Eq. (2) below:

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (2)$$

Multiple Linear Regression Analysis gives models that involve more than one independent variable and one dependent variable. This gives an analytical model, which is used to develop a model for predicting crop yield from climatic elements at various time lags.

This relationship is given by the Eq. (3) below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon \quad (3)$$

Where β s are coefficients, X_i are the predictors, Y is the crop yield (predictand) and β_0 is a constant.

Climate variables are the independent variable while yield is the dependent variable. The data is imported into statistical analysis software (SYSTAT or R) where regression analysis is carried out in order to get coefficients β_0 , β_1 and β_n . These values are fitted in a multiple regression model of the form (3) above.

The regression model for predicting crop yield is arrived at via a series of enhancing steps where the initial step entails all climate variables specified in the data file. Climate variable with a p-value greater than 0.05 are judged as statistically significant in the model at 95% confidence level. The second step is a repeat of the

first one excluding the non-significant variable in the data file. A model that entails the statistically significant climate variables is specified and adopted.

The third step entails plotting the model residuals with keen interest on the normal Q-Q plot to detect the outliers. Where outliers are detected, the model has to be “re-built” without outliers. The error terms which are the differences between the observed value of the dependent variable and the predicted value are called residuals. The final outlier-less model is specified with the following key assumptions namely: Homoscedasticity of residual based on equal variance; Normality of residual; Leverage based on distance of plots to the center and the cook’s distance; Positive variance and non-perfect multicollinearity. Homoscedasticity is defined by a scatter plot and assumes equal distribution of the residuals. Normality of residual assumes that the regression follows a normal distribution. Cook’s distance provides an idea on influential data points that are worth checking for validity. Non-perfect multicollinearity occurs when one of the regressors is highly correlated with, but not equal to a linear combination of other regressors.

Contingency table can be used for verifying the model. Data is split into two data sets where one set is used in training the model in a statistical analysis tool (e.g. SYSTAT). Model verification statistics including percent correct (PC), Post Agreement (PA), False-Alarm Ratio (FAR), critical success index (CSI), probability of detection (POD), bias, and Heidke’s skill score (HSS) were determined.

The methodology was applied by [4] in their assessment of crop yield over Nandi East Sub-County in Kenya.

3. Conclusion and recommendation

This book chapter described a basic regression approach applied in predicting crop yield in a changing climate. Introductory concepts of descriptive statistics, data quality control, correlation analysis and multilinear modeling are discussed. A typical regression method to estimate yield is examined. The study concludes that crop yield prediction and estimation is married with uncertainties of both natural and anthropogenic nature and requires continuous improvement with more focus on externalities that affect crop yield. This study recommends hybrid models that are both statistical and mechanistic, integrated by neural network technology based on multiple variables of climate and crop physiological importance.

Acknowledgements

We acknowledge the contribution of the late professor Joseph Mwalichi Ininda of the Department of Meteorology, University of Nairobi for technical support.

Thanks

Much thanks to the office of the deputy vice chancellor, PPRI; Kibabii University for much support.

IntechOpen

Author details

Godfrey Shem Juma^{1*} and Festus Kelonye Beru²

1 Department of Mathematics, Kibabii University, Kenya

2 Department of Biological and Environmental Sciences, Kibabii University, Kenya

*Address all correspondence to: godfrey.juma@kibu.ac.ke

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, et al. Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. 2013;**1535**
- [2] Kumar M. Impact of climate change on crop yield and role of model for achieving food security. Environmental Monitoring and Assessment. 2016;**188**(8):465
- [3] Lobell DB, Burke MB. On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology. 2010;**150**(11):1443-1452
- [4] Sitienei BJ, Juma SG, Opere E. On the use of regression models to predict tea crop yield responses to climate change: A case of Nandi east, sub-county of Nandi county, Kenya. Climate. 2017;**5**(3):54
- [5] Uzen, N., Cetin, O., & Unlu, M. (2016). Effects of domestic wastewater treated by anaerobic stabilization on soil pollution, plant nutrition, and cotton crop yield.
- [6] Asare, E., & Scarisbrick, D. H. (1995). Rate of nitrogen and sulphur fertilizers on yield, yield components and seed quality of oilseed rape (*Brassica napus* L.). Field Crops Research, 44(1), 41-46. Environmental monitoring and assessment, 188(12), 664.
- [7] Dwivedi, A., Naresh, R. K., Kumar, R., Kumar, P., & Kumar, R. (2017). Climate smart agriculture. no. December.
- [8] Petersen LK. Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. Remote Sensing. 2018;**10**(11):1726
- [9] Herrador M, Esteban MD, Hobza T, Morales D. A modified nested-error regression model for small area estimation. Statistics. 2013;**47**(2):258-273
- [10] Heinzow, T., & Tol, R. S. (2003). Prediction of crop yields across four climate zones in Germany: an artificial neural network approach (No. FNU-34).
- [11] Savin IY, Stathakis D, Negre T, Isaev VA. Prediction of crop yields with the use of neural networks. Russian Agricultural Sciences. 2007;**33**(6):361-363
- [12] Panda SS, Ames DP, Panigrahi S. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. Remote Sensing. 2010;**2**(3):673-696
- [13] Holzman ME, Rivas R, Piccolo MC. Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. International Journal of Applied Earth Observation and Geoinformation. 2014;**28**:181-192
- [14] Erda L, Wei X, Hui J, Yinlong X, Yue L, Liping B, et al. Climate change impacts on crop yield and quality with CO₂ fertilization in China. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005;**360**(1463):2149-2154
- [15] Popova Z, Kercheva M. CERES model application for increasing preparedness to climate variability in agricultural planning—calibration and validation test. Physics and Chemistry of the Earth, Parts A/B/C. 2005;**30**(1-3):125-133
- [16] Estes LD, Bradley BA, Beukes H, Hole DG, Lau M, Oppenheimer MG, et al. Comparing mechanistic and empirical model projections of

crop suitability and productivity:
Implications for ecological forecasting.
Global Ecology and Biogeography.
2013;22(8):1007-1018

[17] De Wit PV, Tersteeg JL,
Radcliffe DJ. Crop yield simulation
and land assessment model for
Botswana (CYSLAMB). Part I theory
and validation. FAO/government of
Botswana. Land resource assessment for
agricultural land use planning project
TCP/BOT/0053. Field Document.
1993;2(72):1993