

Spatial Modelling of Weather Variables for Plant Disease Applications in Mwea Region

Paul Onyango Ouma^{1*}, Patroba Achola Odera¹, John Bosco Mukundi²

¹Department of Geomatic Engineering and Geospatial Information Systems, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

²Department of Horticulture and Food Sciences, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

Email: *oumapaul94@gmail.com

Received 27 March 2016; accepted 17 May 2016; published 20 May 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Climate change is expected to affect the agricultural systems, such as crop yield and plant disease occurrence and spread. To be able to mitigate against the negative impacts of climate change, there is a need to use early warning systems that account for expected changes in weather variables such as temperature and rainfall. Moreover, providing such information at high spatial and temporal resolutions can be useful in improving the accuracy of an early warning system. This paper describes a methodology that can be used to produce high spatial and temporal resolutions of minimum temperature, maximum temperature and rainfall in an agricultural area. We utilize MarkSim GCM, a weather file generator that incorporates IPCC based climate change models to downscale the weather variables at monthly intervals. An ensemble of 17 GCM models is used within the RCP 8.0 emission scenario within the latest model based CMIP5. We first assess the usability of the model, by comparing results produced to what has been recorded at weather station level over a vast region. Then, we estimate the correction factors for model results by implementing a linear regression that is used to assess the relationship between the variables and the deviation of model outputs to the weather station data. Finally, we use kriging geostatistical technique to interpolate the weather data, for the year 2010. Results indicated that the model overestimated the results of maximum temperature, while underestimating the result of minimum temperature. Variability in the recorded weather variables was also evident, indicating that the response variables such as plant disease severity dependent on such weather information could vary in the area. These datasets can be useful especially in predicting the occurrence of plant diseases, which are affected by either rainfall or temperature.

Keywords

Climate Change, Rice Blast, GIS, Geostatistics, MarkSim GCM

*Corresponding author.

1. Introduction

The changes in weather conditions in Mwea region have been shown to have an impact on rice crop yield [1] [2]. In a period when the effects of climate change are becoming more evident, there is a need to provide early warning systems that take into account changes in weather for proper mitigation of unwarranted occurrences [3]. Already, the recent increased occurrence of rice blast disease has been attributed to the changes in weather patterns over the recent years [4]. More specifically, the changes in weather attributes such as temperature, rainfall and humidity are expected to have an impact, either negative or positive on the crop yields in the future [5]. The consequences are expected to be far more serious in areas that rely on agriculture as the major source of livelihood.

African countries are expected to experience some of the most severe negative impacts associated with climate change [6]. The most common variables are expected to rise, fall or have seasonal changes thus affecting various dynamics of food production, more so disease occurrence intensity and lifecycle of disease agent. These weather variables are well established correlates of severity and abundance of common plant diseases [7]. Modelling the spatial variation, especially in the future would therefore be useful in mitigating against the occurrence of plant diseases in the process providing empirical evidence that can be used for optimally targeted control measures [8]. More so, providing such information at high resolutions can help in improving the accuracy of disease occurrence prediction models. Indeed, many farmers in the area attribute the frequent occurrence of the disease in the recent past to the changing weather patterns caused by climate change [4].

This study aims to make a prediction of weather using MarkSim GCM [9], a stochastic downscaling web based tool that can be used to generate future distribution of weather variables across a rice growing region. First, we assess the suitability of using the model for spatial prediction by testing it against weather station data covering a large region. Then the correctly tested weather variables are predicted in space across the area of study. Using well established climate change models, the prediction is done for the future to provide a set of continuous weather variables that can be used for of plant disease prediction applications [10].

2. Materials and Methods

2.1. Study Area

The greater Mwea region is approximately within longitudes 37°13'E and 37°30'E and latitudes 0°32'S and 0°46'S [11]. The area is traversed by three agro-climatic zones, with varying weather characteristics. The soils are imperfectly drained, very deep, dark grey to black vertisols [11]. The predominant crop being planted in the area is rice, mainly through irrigation with a gazetted area of about 30,000 acres. Currently, about 16,000 acres has been developed for rice paddy production in the area. In the irrigation scheme, water is abstracted from two main rivers that traverse the area, which are the Thiba and Nyamindi rivers.

2.2. Analysis

MarkSim GCM is a web based weather file generator tool that is used to downscale climate information using a variety or all of the models available for downscaling. Details for the construction of the tool and its operation are found in [9]. In brief, the tool provides a generalised downscaling methodology that uses inputs from General Circulation Models to generate futuristic weather variables that to some extent account for changes in climatologies. MarkSim GCM uses a third order Markov model, in addition to special stochastic resampling of the model parameters to generate rainfall and temperature variances for any location. These variances are then used in conjunction with a set of interpolated climate surfaces to downscale and predict the weather variables.

The tool was chosen for a variety of reasons; First, it was developed specifically for applications of agricultural modelling [7] [10], with already wide applications in weather variable prediction [12]. Secondly, the tool is freely available (<http://gismap.ciat.cgiar.org/MarkSimGCM>) and can be used as an online application without any restrictions. Finally, its worldwide availability means the tool can be easily used to generate weather information for any location on earth. In areas characterised by a paucity in weather station data for accurate modelling, such Markov processes are useful for generating weather variables.

The results of the tool are downscaled, georeferenced weather variables, simulated at different time periods in two different formats. One is the annual charts of daily rainfall, temperature range and solar radiation. The other is a data format compatible with DSSAT (Decision Support System for Agro-technology Transfer) crop modelling suite. The file can also be opened in a normal text file and converted to any easy to use tabular format such as

comma separated variable format. In the modelling process, prediction of the monthly climatic data was based on the Intergovernmental Panel on Climate Change (IPCC), with different emission scenarios are supported by the model based, bias-corrected CMIP5 and CMIP3 simulations [11]. For this study the latest CMIP5 was used. The list of climate models used for this study is shown in **Table 1**.

The four existing Representative concentration pathways supported are RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5. Details of the construction, composition and differences between the emission scenarios are found elsewhere [13]. They provide simulations of future changes in greenhouse gas emissions, and also account for the effects of Land use land cover changes, and air pollutants in the environment [14].

Thus, the web based tool was used to predict minimum temperature, maximum temperature and Rainfall, at any location. In addition, the tool integrates a Google earth interface that makes it easier for selection of locations whose variables are to be extracted. These four variables have been shown to account for 96.14% of the severity of rice blast disease [7]. Thus when modelling the spatial occurrence of rice blast disease using weather variables, the four variables can as sufficiently approximate distribution of the disease.

2.3. Assessing the Model Accuracy

This stage was carried out in order to provide certainty and confidence to the data being used. For the purposes of comparison and calibration of the outputs, the model was run at several locations where weather station data had already been collected. Monthly data on the variables under investigation were then obtained from the Kenya Meteorological department weather data. Since there were deviations between the model results and the weather station data, the differences between the corresponding points were then derived. To model the relationship between the weather station data and the model data, the two datasets were subjected to a regression analysis. A linear regression analysis implemented in the statistical package R version 3.0, was conducted at the 95% confidence interval. Finally, regression fit plots using the datasets were also derived.

Table 1. List of climate models used for this study.

Model	Meaning
BCC-CSM1.1	Beijing Climate Center Climate System Model version 1
BCC-CSM1.1 (M)	Beijing Climate Center Climate System Model version 2 with a moderate resolution
CSIRO-Mk3.6	Commonwealth Scientific and Industrial Research Organisation
FIO-ESM	First Institute of Oceanography-Earth System Model
GFDL-CM3	Geophysical Fluid Dynamics Laboratory Climate Model version 3
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory Climate Model
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Climate Model
GISS-E2-R	Goddard Institute of Space Studies
GISS-E2-H	Goddard Institute of Space Studies
HadGEM2-ES	Hadley Global Environment Model 2-Earth System
IPSL-CM5A-LR	Institut Pierre Simon Laplace-Low Resolution
IPSL-CM5A-MR	Institut Pierre Simon Laplace-Mid Resolution
MIROC5	Model for Interdisciplinary Research on Climate
MIROC-ESM	Model for Interdisciplinary Research on Climate Earth System Model
MIROC-ESM-CHEM	An atmospheric chemistry coupled version of MIROC-ESM
MRI-CGCM3	Meteorological Research Institute
Nor ESM1-M	Norwegian Earth System Model

2.4. Sampling Locations

Consequently, random points covering the entire Mwea area region was chosen for sampling the downscaled weather data. An interval of 2km was chosen to give a representative coverage of the area during the sampling. A gridded structure for the sampling was chosen to ensure complete coverage of the area under study.

Modelling was first done for the year 2010, for the simple reason that it was the year that mapping for the rice blast disease distribution was carried out [4]. To allow for variability and total coverage of the whole area, a sampling strategy was used that involved using a selected grid interval of 2 km. The model was then run for each of the locations at the desired times. **Figure 1** shows the locations sampled for the weather variable generation

2.5. Geostatistical Modelling

Geostatistics is the process of using sampled georeferenced data of a phenomenon to predict into areas that were not observed [15]. Geostatistical procedures rely on Toblers’ law, which states that near things are more related than things further apart. In the geostatistical process, the problem of predicting the linear function of a Gaussian process $S(\mathbf{x})$ based on the observations

$$Y = S(x_i) + Z_i$$

where Z_i defines the zero-mean Gaussian random variables [15].

The term kriging is widely used to refer to the process of performing spatial interpolation at unsampled locations. For this process, the ordinary kriging process [16], which has been shown to be best at spatial prediction of weather variables such as temperature and rainfall was used to interpolate the variables under study. First, the empirical semivariogram, which is a means of exploring the spatial relationships between points is constructed. In

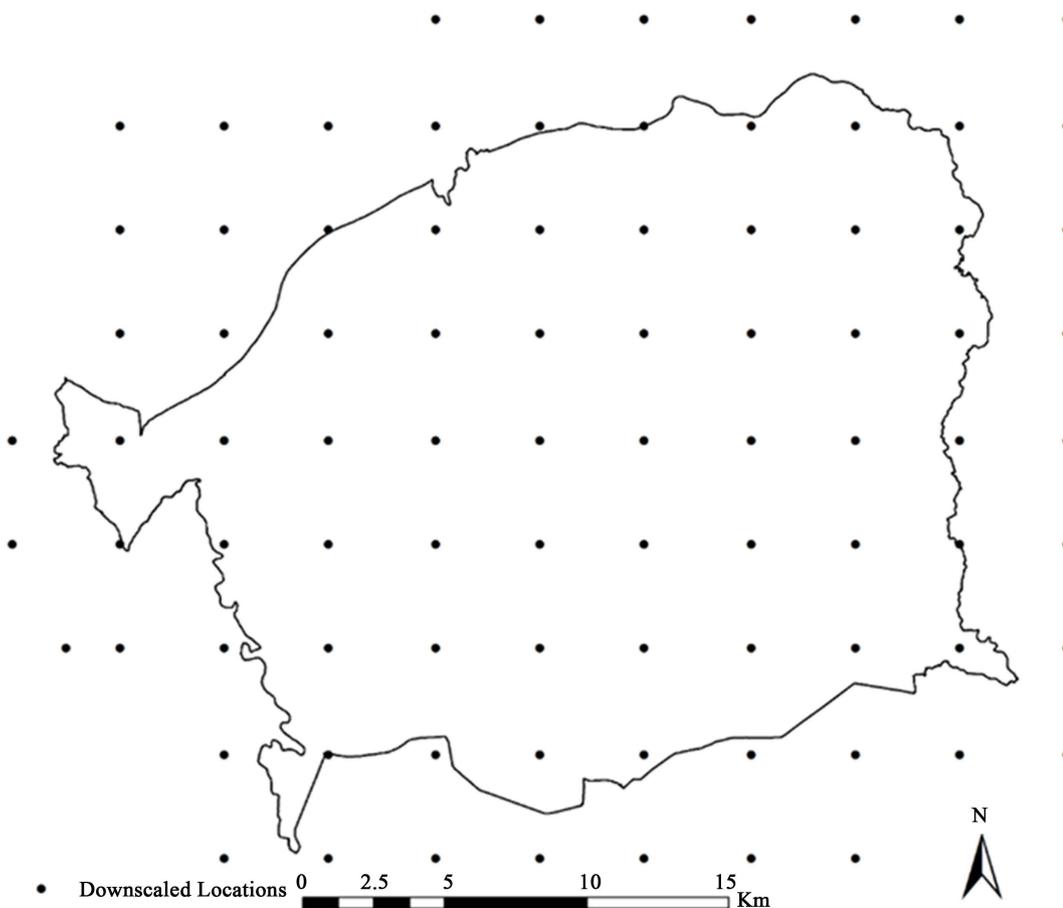


Figure 1. Shows the locations sampled for the weather variables generation.

brief, the semivariogram model examines the extent to which Tobler's law is true. That is near things being more related than things further apart, thus quantifying statistical correlation as a function of distance. In modelling the semivariogram, the short range variability in each of the modelled dataset, which warrants the application of a nugget effect, was examined. A nugget effect is the non-zero y intercept of the semivariogram plot, which if large enough can indicate the absence of spatial autocorrelation. The presence of the nugget effect is normally attributed to measurement errors, and in this case, an incorrectly calibrated model would cause the occurrence of such situations. For this case, the nugget effect was quite small and hence ignored in the kriging process.

Secondly, a line that best fits through the points of the semivariogram was modelled. This line defines spatial autocorrelation in the data and is provides the best fit through the points. The autocorrelation values are then defined from the semivariogram model to derive the kriging weights assigned to each of the measured value. The kriging weights are then used for the prediction process.

Cross validation was also done using a methodology that holds out some datasets during the prediction and uses the held out data for checking the accuracy.

3. Results

There were approximately 78 locations that were used to sample the weather variables. In addition, weather station data was obtained from the closest 19 meteorological stations to the area of study. For the climate change analysis, the ensemble of the 17 GCMs was used because of the speed in processing the model outputs, compared to performing the runs in any single model.

In the year 2010, the area is characterised by varying temperature and rainfall estimates. Minimum temperature changes from about 10 to 15 degrees Celsius as shown in [Figure 2](#).

Maximum temperature also varies with a similar range, changing from 29 to 33 degrees Celsius, as shown in [Figure 3](#).

The greatest variability was experienced in rainfall estimates where values changed from 85mm to more than 152 mm as shown in [Figure 4](#).

Maximum temperature

The results indicate a good fit for the regression analysis. The R^2 square is a good fit of 0.83. The ANOVA test was also carried out at 95% confidence interval and the results are also shown in [Figure 5](#). The significance value F is zero, which shows that the results are statistically significant.

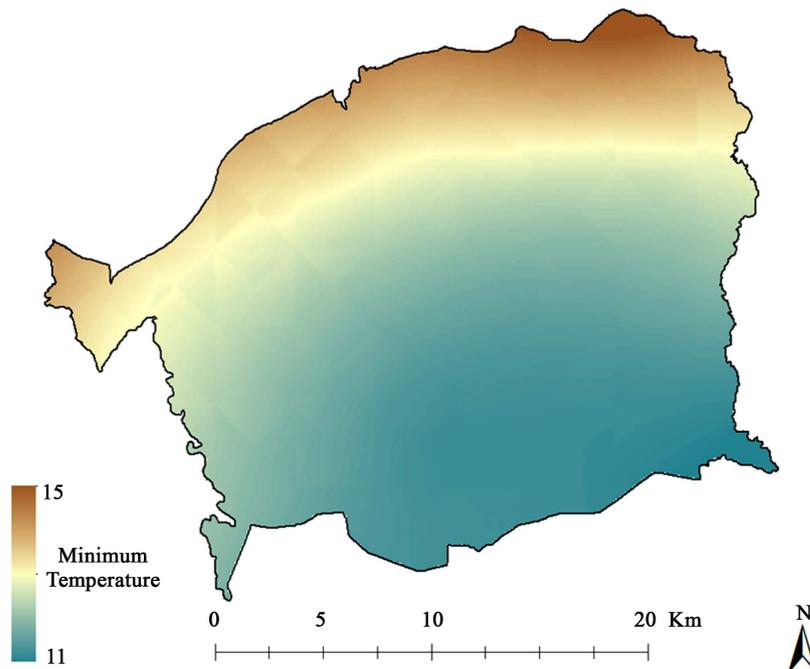


Figure 2. Spatial variation of Minimum temperature for the year 2010.

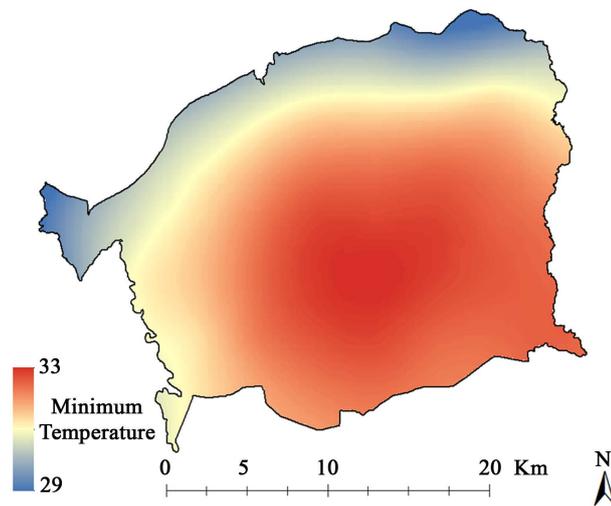


Figure 3. Spatial variation of maximum temperature for the year 2010.

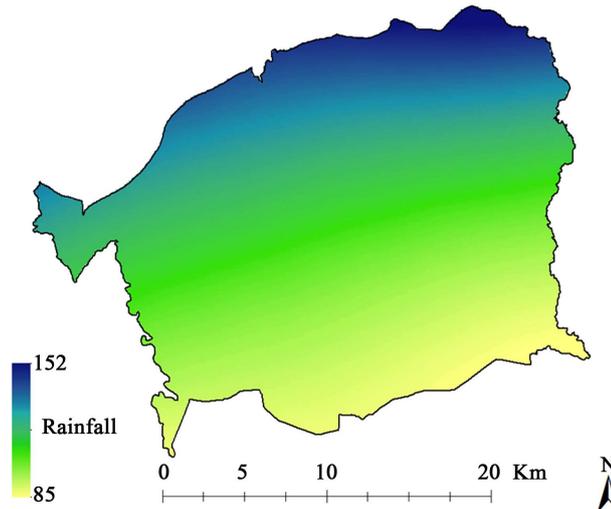


Figure 4. Spatial variation of rainfall distribution for the year 2010.

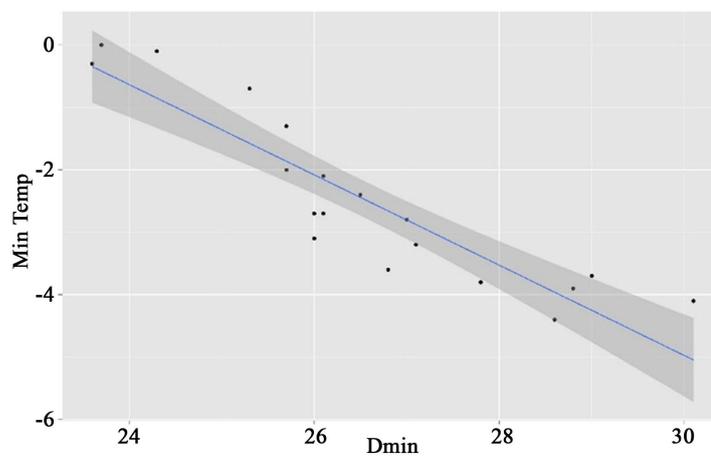


Figure 5. Regression between Maximum temperature values of the model and difference in temperature to that of the weather station. Because it was done at the 95% confidence interval, the shaded region shows the 95% confidence region.

Overall, from the regression results, the intercept for the regression equation was obtained as 16.732 CI [12.233 to 21.232, $P < 0.001$] and the coefficient of the X value maximum temperature obtained as -0.724 CI $[-0.893$ to 0.554 , $P < 0.001$], indicating a tendency to overestimate the maximum temperature. Consequently, there was strong evidence to support the regression equation for correcting the maximum temperature results as;

$$D \text{ MaxTemp} = (-0.724 \times \text{Max Temp}) + 16.732$$

Minimum Temperature

Similar analyses were carried out for the minimum temperature results. As it has been shown from the table, the residuals increase as the minimum temperature reduces, thereby under estimating the minimum temperature as shown in [Figure 6](#).

For the minimum temperature, the results were much less statistically significant with less correlation as shown in the R square value of 0.26. However, there was strong statistical evidence to accept the regression outputs. These were, the intercept as 8.053 CI [2.814 to 13.291, $P = 0.006$] and the X coefficient as -0.461 CI $[-0.927$ to 0.004 , $P = 0.052$].

The regression equation was therefore obtained as;

$$D \text{ Min Temp} = (-0.461 \times \text{Min Temp}) + 8.053$$

Rainfall

For rainfall data, the large recorded masked much of the differences between the model results and the weather station data. Therefore, the values were not changed as they would have been minimal.

Interpolation

Evaluation of the data distribution for all the datasets, minimum temperature, maximum temperature and rainfall, there was a trend towards normal distribution, but with little skewness. [Figure 7](#) highlights these results.

Cross validation results are shown in [Figure 8](#).

4. Discussion

This research explored the possibility of generating high resolution weather variables for agricultural application in the Mwea region. The region is largely an agricultural area while it also has the largest rice irrigation scheme in the country. The provision of present and future weather information would be useful in applications that require the analysis of such variables to generate other useful information. In particular, agricultural systems such as

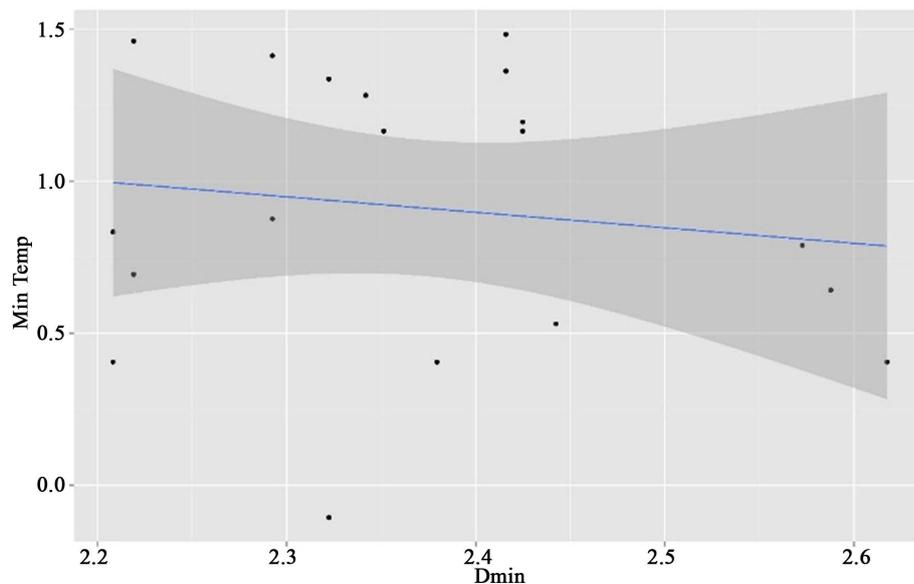


Figure 6. Regression between model minimum temperature and the difference between the minimum temperature and the weather station data. The shaded region highlights the 95% confidence interval region.

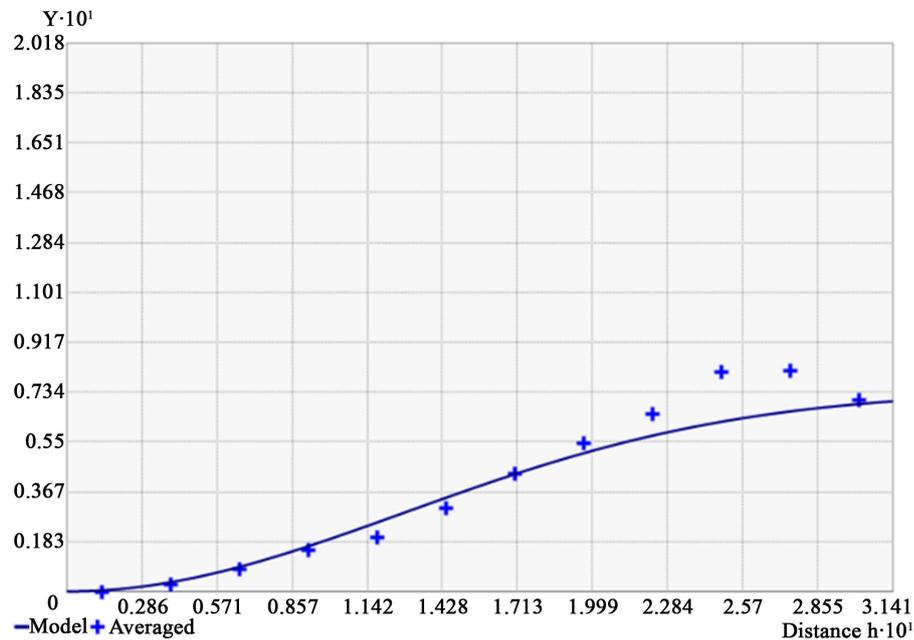


Figure 7. Modelled semivariogram used for the spatial interpolation. This is an example for the month of October rainfall data. The less spatial correlation means that less of the spatial structure was used for the modelling process.

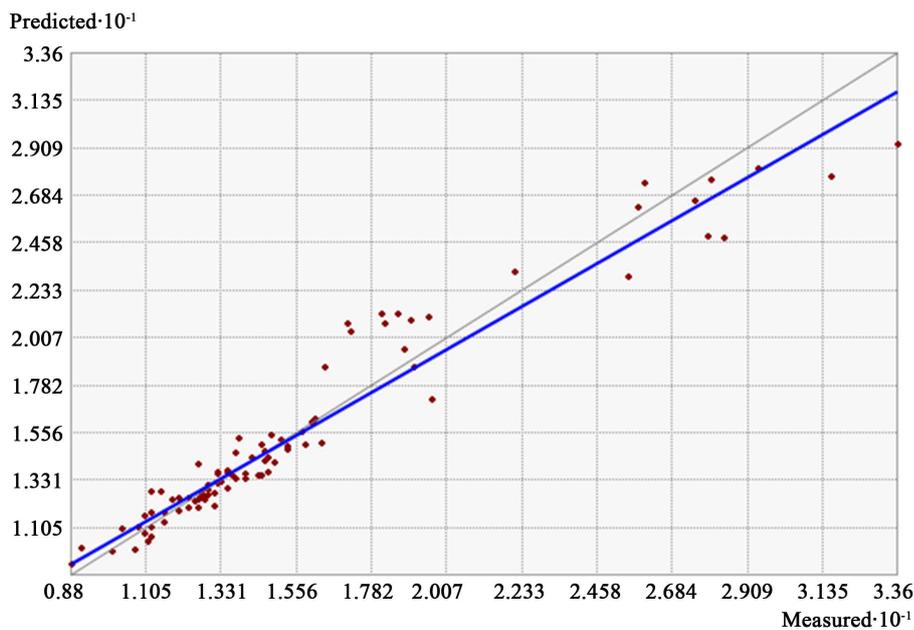


Figure 8. Cross validation results, showing the relationship between the input data in the X axis and the prediction in the Y axis.

disease spread and variation in the intensity is dependent on the spatial distribution of weather, already, it has been shown that changes in climate are expected to affect the spread of diseases such as rice blast [8] [17]. Therefore, to be able to appropriately implement disease forecasting models for effective controlling of the disease [18], there is need to provide such weather variables at high spatial and temporal resolutions [10]. First, a weather generating tool, that accounts for the climate change was used to systematically point based information on monthly rainfall, maximum temperature, minimum temperature and Solar radiation. Then kriging, a best linear unbiased estimator was used to interpolate the weather variables in space. To our knowledge, this is the first attempt to defining

spatially continuous variables of weather using data generated from MarkSim GCM. Consequently, weather data was generated for present and future.

Weather variables were successfully generated for the present and the future. Using an ensemble of 17 established General Circulation models (GCMs) and accounting for the greenhouse gas (GHGs) emission scenarios, data was extracted at 2 km intervals covering the whole area. The tool has already been used to generate point based weather variables for any position in the world [9] [12]. To assess the usability of the tool over the region, weather information was generated at the exact same locations where weather stations are located. The comparison between the model results and recorded weather variables were then compared using a regression analysis to derive a trend. This was very important in determining whether the tool underestimates or overestimates the weather information. The derived trend was then used to refine the results of the model, which were then used to produce contemporary weather information. The results highlight a consistent observation of the differences between the model results and weather station data, highlighting the ability to improve model outputs. In addition, the linear relationship between the model outputs and the weather station data all produced statistically significant relationships. This means that to some extent, the model is able to provide realistic guesses, which when once adjusted, can provide useful proxies for the weather variables under investigation.

The results highlight a new approach that can be used in providing synoptic monthly weather variables such as rainfall, maximum temperature and minimum temperature. These datasets can be useful when there is need to identify the intrinsic effects on agricultural systems that dynamic weather variables can reveal compared to static variables normally used. In addition, areas having less information from the “gold standard” weather station data can use such methods and data to fill such existing gaps.

The crux here being the ability to produce spatially defined weather variables for each month, of every year from the current period up to the year 2100. Most importantly, the methodology and results can be used in areas having a dearth in weather information necessary for agricultural applications. For example, such data can be used in mitigation of the future spread of plant diseases such as rice blast. This would be useful in planning for control measures based on the spatial distribution of the predicted weather information. It has already been shown that the spread of rice blast disease is expected to be affected by climate change in the long run [19], thus a change in how measures of disease control are implemented. In addition, the methodology allows for accounting for different emission scenarios in the future and therefore accounts for much of the uncertainty that may exist in terms of future changes.

The produced continuous surfaces are useful when weather information is required for any agricultural application. Other previous studies have already focused on using climate change to predict future occurrences of diseases, for example malaria [20]. Such applications can also be extended to agricultural areas while focusing on the ecological conditions required for the spread of a disease. A good example would be estimating the impact of climate change on the spatial distribution of any plant disease whose occurrence is affected by weather variables. For instance, the spread of rice blast disease in the area can be predicted in future to aid in controlling and mitigating its occurrence. Moreover, the datasets produced have already been shown to be the major causes of variability in the occurrence. Therefore, prediction can be done using these kind of datasets to produce contemporary distributions of the disease.

Some of the limitations in the study were lack of enough weather station datasets, particularly rainfall in the area to test the model results. The model results are also dependent on the inputs, which can certainly be improved. For instance, the model uses a total of about 10,000 weather stations while it has already been shown that even 50,000 stations would not be enough for such global applications. Also, determining the humidity level was not easy, especially for the future because the area is also depends on irrigation, and therefore, evapotranspiration cannot be ignored. Future studies should focus on applications of such datasets in predicting spatial distribution of plant diseases and producing such variables on a daily level to improve the disease prediction results.

5. Conclusion

This study was able to demonstrate a methodology that could be used to predict the spatial distribution of future weather variables. Most importantly, the weather variables are important in characterising the variability in plant diseases. The results are variables at high resolution monthly datasets for rainfall, maximum temperature and minimum temperature for the present and the future. These can be used in agricultural applications that require such spatial datasets particularly in predicting the future distribution of plant diseases.

References

- [1] FAO (2008) Climate Change and Food Security. *Food and Agriculture Organization*, **1**, 291-306. <http://dx.doi.org/10.1017/S002081830000607X>
- [2] Nyang'au, W.O., Mati, B.M., Kalamwa, K., Wanjogu, R.K. and Kiplagat, L.K. (2004) Estimating Rice Yield under Changing Weather Conditions in Kenya Using CERES Rice Model. *International Journal of Agronomy*, **2014**, 1-12. <http://dx.doi.org/10.1155/2014/849496>
- [3] Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., *et al.* (2014) Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3268-3273. <http://dx.doi.org/10.1073/pnas.1222463110>
- [4] Kihoro, J., Bosco, N.J., Murage, H., Ateka, E. and Makihara, D. (2013) Investigating the Impact of Rice Blast Disease on the Livelihood of the Local Farmers in Greater Mwea Region of Kenya. *SpringerPlus*, **2**, 308. <http://dx.doi.org/10.1186/2193-1801-2-308>
- [5] Luck, J., Spackman, M., Freeman, A., Tre, P., Griffiths, W. and Finlay, K. (2011) Climate Change and Diseases of Food Crops. *Plant Pathology*, **60**, 113-121. <http://dx.doi.org/10.1111/j.1365-3059.2010.02414.x>
- [6] Dokken, Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R., LLW. IPCC (2014) Summary for Policymakers. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, New York, 1-32.
- [7] Malicdem, A.R. (2015) Rice Blast Disease Forecasting for Northern Philippines College of Information Technology. *Transactions on Information Science and Applications*, **12**, 120-129.
- [8] Raquel Ghini, R.B. (2014) Rice Blast Disease in Climate Change Times. *Rice Research Journals*, **3**, e111. <http://dx.doi.org/10.4172/2375-4338.1000e111>
- [9] Jones, P.G. and Thornton, P.K. (2013) Generating Downscaled Weather Data from a Suite of Climate Models for Agricultural Modelling Applications. *Agricultural Systems*, **114**, 1-5. <http://dx.doi.org/10.1016/j.agsy.2012.08.002>
- [10] Johansson, M.A. and Glass, G.E. (2008) High-Resolution Spatiotemporal Weather Models for Climate Studies. *International Journal of Health Geographics*, **7**, 52. <http://dx.doi.org/10.1186/1476-072X-7-52>
- [11] Mati, B.M., Wanjogu, R., Odongo, B. and Home, P.G. (2011) Introduction of the System of Rice Intensification in Kenya: Experiences from Mwea Irrigation Scheme. *Paddy and Water Environment*, **9**, 145-154. <http://dx.doi.org/10.1007/s10333-010-0241-3>
- [12] Srinivasa Rao, M., Swathi, P., Rama Rao, C.A., Rao, K.V., Raju, B.M.K., Srinivas, K., *et al.* (2015) Model and Scenario Variations in Predicted Number of Generations of *Spodoptera litura* Fab. on Peanut during Future Climate Change Scenario. *PLoS ONE*, **10**, e0116762. <http://dx.doi.org/10.1371/journal.pone.0116762>
- [13] Rhee, J. and Cho, J. (2015) Future Changes in Drought Characteristics: Regional Analysis for South Korea under CMIP5 Projections. *Journal of Hydrometeorology*. <http://dx.doi.org/10.1175/JHM-D-15-0027.1>
- [14] van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., *et al.* (2011) The Representative Concentration Pathways: An Overview. *Climatic Change*, **109**, 5-31. <http://dx.doi.org/10.1007/s10584-011-0148-z>
- [15] Diggle, P.J. and Ribeiro, P.J. (2007) *Model-Based Geostatistics*. Springer, Berlin.
- [16] Frazier, A.G., Giambelluca, T.W., Diaz, H.F. and Needham, H.L. (2015) Comparison of Geostatistical Approaches to Spatially Interpolate Month-Year Rainfall for the Hawaiian Islands. *International Journal of Climatology*. <http://dx.doi.org/10.1002/joc.4437>
- [17] Pautasso, M., Döring, T.F., Garbelotto, M., Pellis, L. and Jeger, M.J. (2012) Impacts of Climate Change on Plant Diseases-Opinions and Trends. *European Journal of Plant Pathology*, **133**, 295-313. <http://dx.doi.org/10.1007/s10658-012-9936-1>
- [18] Iwata, M. (2001) Rice Blast Control. *The Royal Society of Chemistry*, **12**, 28-31.
- [19] Ghini, R., Hamada, E. and Bettiol, W. (2008) Climate Change and plant Diseases. *Scientia Agricola*, **65**, 98-107. <http://dx.doi.org/10.1590/S0103-90162008000700015>
- [20] Caminade, C., Kovats, S., Rocklov, J., Tompkins, A.M., Morse, A.P., Colón-González, F.J., *et al.* (2014) Impact of Climate Change on Global Malaria Distribution. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3286-3291. <http://dx.doi.org/10.1073/pnas.1302089111>