Applying a Bivariate Frequency Ratio Technique for Potato High Yield Susceptibility Mapping

Khalid A. Al-Gaadi[1,2], Abdalhaleem A. Hassaballa[3], ElKamil Tola[2], Rangaswamy Madugundu[2], Ahmed G. Kayad[1]

[1]Department of Agricultural Engineering, College of Food and Agriculture Sciences, King Saud University, Riyadh, Saudi Arabia; [2]Precision Agriculture Research Chair, King Saud University, Riyadh, Saudi Arabia; [3]Department of Agricultural Engineering, Faculty of Engineering, University of Khartoum, Khartoum, Sudan

A paper from the Proceedings of the

14th International Conference on Precision Agriculture

June 24 – June 27, 2018

Montreal, Quebec, Canada

Abstract. *Spatial variation of soil characteristics and vegetation conditions are viewed as the most important indicators of crop yield status. Therefore, this study was designed to develop a crop yield prediction model through spatial autocorrelation between the actual yield of potato (Solanum tuberosum L.) crop and selected yield status indicators (soil N, EC, pH, texture and vegetation condition), where the vegetation condition was represented by the cumulative normalized difference vegetation index (CNDVI). The study was conducted, during the period from December 2016 to March 2017, on a 30 ha center-pivot irrigated field located in Wadi-Ad-Dawasir area, Riyadh Province, Saudi Arabia. GPS assisted soil samples and yield data were collected from 120 sampling locations and analyzed for soil EC, N, pH and texture. The CNDVI; however, was derived from Sentinel-2A satellite images. Bivariate frequency ratio statistical approach was employed to assess the spatial correlation between crop yield and both of soil parameters and CNDVI. The spatial autocorrelation results were interpreted and further utilized for the generation of potato high yield susceptibility map. The accuracy of the obtained susceptibility (prediction) maps was investigated using the area under the curve (AUC) spatial validation technique. Soil pH was the most predictive factor in high yield zones, followed by texture, N, CNDVI and EC with prediction rates (PR) of 5.0, 4.2, 2.5, 1.9 and 1.0, respectively. The generated high susceptibility map represented the actual yield map with an AUC of 90%.*

Keywords. *Frequency ratio, yield prediction, potato, mapping, soil parameters, NDVI.*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 14th International Conference on Precision Agriculture. Al-Gaadi, K.A., Hassaballa, A.A., Tola, E., Madugundu, R. & Kayad, A. G. (2018). Title of paper. Applying a Bivariate Frequency Ratio Technique for Potato High Yield Susceptibility Mapping. In Proceedings of the 14th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

The spatial variability of crop productivity within a field is usually the result of spatial variation of different soil characteristics (Jin and Jiang, 2002). Identifying the most influential causes of crop yield variability and exploring their magnitude of variation within agricultural fields can greatly help in achieving efficient site-specific management. Precision agriculture, or site-specific management, aims to tailor the agricultural inputs to fit different field locations based on the spatial variability of soil and crop parameters (Fraisse et al., 1999). These management techniques require a good knowledge of the spatial variability of soil and crop across the field. The spatial tendency of high yield is mostly attributed to factors that are related to field soil parameters (fertility, texture, etc.), agricultural practices (land preparation, irrigation, drainage, etc.), or to other factors (land topography, weeds, pesticide application, etc.).

It is common to use the bivariate statistical analysis to express the idea of overlaying a landslide inventory map with maps of the parameters that influence this landslide. This is to rank the corresponding classes based on their participation in landslide formation (Lee and Pradhan, 2007). The area under curve (AUC) is a process that is employed to qualitatively assess the prediction accuracy by obtaining the relative ranks for each prediction pattern (Lee et al., 2007).

Therefore, this study was designed to develop a crop yield prediction model, through establishing a spatial autocorrelation between the actual yield of potato (*Solanum tuberosum* L.) crop and soil parameters. These parameters included nitrogen (N), electrical conductivity (EC), pH, texture and the cumulative normalized difference vegetation index (CNDVI).

Materials and Methods

The study was conducted on a 30-ha center-pivot irrigated field in the Saudi Agricultural Development Company farm located in Wadi-Ad-Dawasir, Riyadh Province, Saudi Arabia, between the latitudes of 19.90° and 20.33° N and the longitudes of 44.81° and 44.95° E. The study area was under arid climates with daily temperature ranging from 6 °C (winter) to 43 °C (summer). The relative humidity was normally stable at 24%. The average wind speed was 13 km h^{-1}. The major cultivated crops in the farm included potatoes, corn, wheat and watermelon.

Sampling on 50 m by 50 m grids was adopted and a total of 120 sample locations were identified. Sampling points were georeferenced using a Hand-held GPS receiver (Trimble GeoXH). Georeferenced potato yield (Y_A, t ha^{-1}) data points were also collected two to three days prior to harvesting from a 3 m^2 area that represented each data point. The collected soil samples were analyzed for N, pH, EC and texture. The soil N was determined in the laboratory adopting the Kjeldahl method (Lynch and Barbano, 1999). However, the soil pH and EC values were determined using a pH/conductivity meter (Mi805). The soil texture, on the other hand, was determined using the hydrometer method (Bouyoucos, 1951).

Six cloud-free Sentinel-2A images, corresponding to the study period, were downloaded from the USGS portal (https://earthexplorer.usgs.gov/). The Q-GIS (Ver. 2.18) software program was used for image analysis. Initially, Sentinal-2A images were pre-processed for atmospheric corrections, radiometric calibration and reflectance value calculation. Subsequently, the area of interest was extracted using the “semi-automatic Classification Plugin (SCP)” of the Q-GIS. The NDVI was calculated from the corrected Sentinel-2A bands of 4 and 8. Multi-date NDVI layers were aggregated and the cumulative NDVI (CNDVI) was computed accordingly. Surface maps of the analyzed soil properties and yield were generated using ArcGIS software program.

The generated maps of soil N, EC, pH and CNDVI were considered as the independent variables and the yield (Y_A) as the dependent variable. High-yield zone of Y_A map was extracted and intersected with the independent parameters. The frequency ratio (FR), the relative frequency (RF) and the prediction rate (PR) coefficients were calculated for each parameter (soil N, EC, pH and CNDVI) with respect to Y_A. The accuracy of the obtained tendency map was performed using the AUC method.

Results and Discussion

Upon assessing the PR coefficient, it was observed that the soil pH represented the factor with most correlation to the potato yield, followed by soil texture. Within the soil texture, a sandy loam class produced the highest RF of high yield association, which was in agreement with the previous studies that assured the validity of the sandy loam in producing relatively high potato yields (Ramirez, 1992). In fact, the class imbalance in soil pH nominated the soil texture to become the most effective independent parameter in yield variability. This is further supported by the study of Redulla et al. (2002), which concluded that the soil texture components (sand, silt and clay) had a stronger influence on yield compared to that imposed by soil chemical properties.

For accuracy assessment, the testing points (high yield portion) were used and correlated spatially against the susceptibility map. The results proved a strong yield prediction provided by the resultant tendency map, producing an AUC of 90%. However, when the independent variables were separately correlated, low AUC values were obtained. This can be due to the admixed association caused by the non-representative classes at each factor map. However, the calculated PR of the independent factors together indicated that the soil pH was the most predictive factor in high yield zoning, followed by soil texture, N, CNDVI, and EC, with PR values of 5, 4.2, 2.45, 1.9 and 1.0, respectively. This can be attributed to the homogenous nature of the pH over the yield; where, a high prediction rate was dominant with almost 94%.

Conclusions

Bivariate frequency ratio statistical approach was employed in order to assess the factors influencing the distribution of potato yield. A yield tendency map was generated based on the weighted factors. The AUC spatial validation was used to assess the accuracy of the generated prediction map of potato yield. The independent factors highlighted that soil pH exhibited the highest association with high-yield zoning (with a prediction factor of 5.0).

Acknowledgement

The authors are grateful to the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia for funding this study through the Vice Deanship of Scientific Research Chairs. The extensive cooperation and support extended by Saudi Agricultural Development Company farm, Wadi-Ad-Dawasir, Saudi Arabia in carrying out the field research work are gratefully acknowledged.

References

Bouyoucos, G. J., (1951). A recalibration of the hydrometer method for making mechanical analysis of soils. *Agronomy Journal* 43(9), 434-438. doi:10.2134/ agronj1951.00021962004300090005x

Fraisse, C. W., Sudduth, K. A., Kitchen, N. R., & Fridgen, J. J. (1999). Use of unsupervised clustering algorithms for delineating within field management zones. ASAE 1999: Paper No. 993043. International Meeting, Toronto, Ontario, Canada. July 18-21.

Jin, J., & Jiang, C. (2002). Spatial variability of soil nutrients and site specific nutrient management in the P.R. China. *Computers and Electronics in Agriculture*, 36(2-3), 165-172.

Lee, S., & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides,* 4(1), 33-41.

Lee S. (2007). Application and verification of fuzzy algebraic operators to landslide susceptibility mapping*. Environmental Geology*, 52(4), 615-623.

Lynch, J. M., Barbano, D. M. (1999). Kjeldahl nitrogen analysis as a reference method for protein determination in dairy products. *Journal of AOAC International,* 82, 1389-98.

Pradhan, B. (2010). Remote sensing and GIS-based landslide hazard analysis and cross validation using multivariate logistic regression model on three test areas in Malaysia. *Advances in Space Research*, 45(10), 1244-1256.

Ramirez, G. P. (1992). Cultivation harvesting and storage of sweet potato products. *Seed,* 6, 6.

Redulla, C. A., Davenport, J. R., Evans, R. G., et al. (2002). Relating potato yield and quality to field scale variability in soil characteristics. *American Journal of Potato Research*, 79(5), 317-323.