Academic Journal of Applied Mathematical Sciences
ISSN(e): 2415-2188, ISSN(p): 2415-5225
Vol. 7, Issue. 3, pp: 158-166, 2021
URL: https://arpgweb.com/journal/journal/17
DOI: https://doi.org/10.32861/ajams.73.158.166

Academic Research Publishing Group

Original Research **Open Access**

Application of Bayesian Networks of Genotype by Environment Interaction Evaluation Under Plant Disease, Soil Types and Climate Condition-using BayesiaLab

Siraj Osman Omer
Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan
Email: sirajstat@yahoo.com

Article History
Received: 27 March, 2021
Revised: 30 April, 2021
Accepted: 27 May, 2021
Published: 2 June, 2021

Abstract

Genotype by environment interaction (GEI) linked to plant disease, soil properties and climate conditions add potential value for a breeding program to underpin decision making. In understanding genotype x environment interaction, the most challenging factors are the identification of genetic variation for a range of traits and their responsiveness to the climate change factors. In order to study the complex relationships with genetic and non-genetic factors, the application of Bayesian network tools will help understand and accelerate plant breeding progress and improve the efficiency of crop production. In this study, we proposed the application of Bayesian networks (BNs) to evaluate genotype by environment interaction under plant diseases, soil type, and climate variables. An adapted to simulate multiple environmental trial (MET) data of maize (corn) was used to examine the performance of the BN predictive modeling using BayesiaLab for deriving knowledge and graphical structure for exploring GEI diagnosis and analysis. The results highlighted that genotypes have the same probability and the frequentist of rainfall, temperature, soil type, and disease type occurred as <=88 (46%), 35 (37%), clay (27%), and MB (47%) respectively, which have to monitor reflects in each discretization. This study provided a roadmap to knowledge modeling of GEI using BayesiaLab software. On a broader scale, this study helps predict the yield of crop varieties by understanding agronomic and environmental factors under farm conditions rather than conducting long-term agricultural testing under well-controlled conditions of the on-station trials. Future improvements of BNs application of METs should consider working on a larger and more detailed soil and irrigation system linked to agro system.

Keywords: Bayesian networks; GEI; Climate condition; Diseases soils; Maize; BayesiaLab.

1. Introduction

Agricultural experiment station often involves a range of crops, livestock, and land variability on the natural soil variation. Accordingly, spatial variation of soil properties has implications for experimental station management in several ways: (1) accurate and reliable sampling for determining fertilizer application rates, (2) the degree of field uniformity influences the experimental design and plot size, and (3) possible carryover effects of residual nutrients where nutrients or soil moisture are not an experimental variable [1]. The classification of genotype and environmental interaction (GEI) occurs when differences between genotypes are not the same in all locations within and across years. Therefore, GEI is useful in the development and evaluation of plant varieties since it reduces the genotypic stability values under diverse environments [2]. Incorporation of different weather and soil physical variables into multi environmental trials can determine the potential causal factors of GEI modelling [3]. In order to better understand GEI from multi environmental trials (MET), it is important to understand the phenology of maize in relation to local environments [4]. Bayesian networks (BNs) are using a probabilistic propagation algorithm (Bayes Rule) to estimate the parameters of a network automatically from data. The degrees of uncertainty of the variables differed from various sources (such as expert opinion, empirical data or numerical data) for capturing the conditional dependence of a variable upon others [5]. BNs have been described by many authors (e.g. Laurila-Pant, *et al.* [6], Zhang, *et al.* [7], Do Amaral, *et al.* [8], Kocian, *et al.* [9]). BNs allow to study the causal relationships between variables and to calculate the probability of a variable when the other variables are known in the model, in BN, Monte Carlo analysis (MCA) can be considered a random sampling of probability distribution functions represent as inputs of Bayesian model to produce hundreds or thousands of possible outcomes [10]. BNs has been addressing in real applications in agricultural sciences, such as farming, water resources, reforestation and ecological modeling for examples; [11] has applied Bayesian networks prediction of coffee rust disease using Bayesian networks; [12] have investigated Bayesian network in wheat and growing barley without pesticides; [13] have discussed maize disease diagnosis system based on BN has (i.e. A BN software package), which provide an effective tool for maize disease diagnosis; [14] have used BN for predicting energy crop yield; [15] have recently applied to analyze port variables in order to make sustainable planning and management decision [16]. Applied Bayesian network to visualize the complex interrelationships between interdisciplinary variables resulting from the impacts of climate change scenarios in agriculture [17]. Explored Bayesian networks as a convenient and interpretable framework for the simultaneous modeling of multiple quantitative traits in the plant breeding program from a Multiparent Advanced Generation Intercross (MAGIC) winter wheat population. BNs finally became accessible to a wide range of scientific research, particularly, in agricultural perspective. However, BNs can be built from agricultural knowledge, i.e. from multiple environments trials using machine-learned from qualitative and quantities data [18].

Recently [19] addressed a machine learning approach for crop improvement by leveraging phenotypic and genotypic big data to predict agricultural production based on phenotypic traits used in genomic selection in breeding. Furthermore, [20] reviewed the latest studies on machine learning in the field of plant breeding and biotechnology. There were more than fifty examples of recently applied machine learning models in classical and modern plant breeding studies, including neural networks, partial least square regression, and random forest in plant breeding and biotechnological studies [21-24]. Machine learning techniques are capable of handling large amounts of data in various areas of plant breeding for improved interoperability with more precise outcomes [25]. Stöckle and Kemanian [26], mentioned that the crop model development to assess genotypes performance across diverse target genetic × environment × management interaction practices will help understanding genotype suitability, best agronomic management in plant breeding programs ut results. Also Lu, *et al.* [27] used Bayesian learning network model to explore the complex causal interactions between environments (i.e., climate, weather, drives and its impact severity), plant diseases for three different grape cultivars in Quebec, Canada. The problem is that uncertainty in agriculture is more extensive and complex; therefore, to create an effective intelligence system, uncertain knowledge must be dealt with. For the uncertainty in agriculture, there are some good model and application, but mainly rule-based adapted to knowledge represented by Bayes rule [28]. In many situations GEI has been constituted due to the association between the environment and the phenotypic traits of a genotype, because GEI factors present in the environment as temperature, rainfall as well as the genetic constitution of an individual (genotype), influence the phenotypic expression of a trait [29]. Resende, *et al.* [30]. justified that with successful applications of traditional statistical methods in genotype, environment, and their interaction (G × E) studies have low efficiency in analyzing data obtained from multiple environmental trials due to nature of plant characteristics. Therefore, an adaptive and applied machine learning techniques is more effective than traditional statistical models in dealing with large amounts of complex and deterministic information [31]. The primary objective of this paper was to apply the Bayesian network for genotype by environment interaction under plant disease, land type and climate condition. This was addressed through the analysis and dynasties the factors affected genotype and environment interaction model in a simulation study and an application to maize dataset. The application of BNs in agricultural research, especially in METs will enable us to build comprehending Bayesian network probabilities for adapting strategies. A Bayesian network is useful for better estimation and prediction of genotype by environment interaction outcome through visualization parametric distribution for describing interactions. The paper is attempting to apply BN at GEI model to provide an efficient support breeding in the decision-making process.

2. Materials and Methods

2.1. Experimental Data

In this paper, a balanced dataset was used to meet the settings of the GEI models. GEI model obtains from simulating datasets based on genotypes and environments (three successive years and four locations/situations). Field experiment-based on location, soil type, year, total seasonal rainfall, total temperatures, and grain production were used to set the Bayesian networks' values. All the experiments were conducted at research stations at different locations to represent maize a production region. The phonotypic parameter consists of seed yield values in kg/ha on five genotypes of maize evaluated for three years (2010, 2011 and 2012) in different four locations. A simulated multiple environmental trials data of maize (corn) were used to examine the performance of the BN model approaches under climate conditions, maize disease, soil types. In this paper we consider discrete cases as categorical variables and continue cases as a numerical variable. Climate conditions are commonly used in describing data consisting of variables, where the variable includes (temperatures, rainfall). Causal learning will be used to obtain a Bayesian network that predicts reasonably well (including outperforming NB. All statistical analyses were conducted with the BayesiaLab software. BayesiaLab can be used for data analytics, diagnosis and knowledge modeling. BayesiaLab has made Bayesian networks, convenient for obtaining in-depth information on problem areas. BayesiaLab uses the graphic structure inherent in Bayesian networks to explore and explain complex issues [32].

In this study, the BNs components includes are

I. Nodes represent variables of interest (the degree temperature, average rainfall, soil type, soil mixture, trail site, the seasons, the occurrence of an event (year).
II. Direct and indirect causal dependencies among the variables. The directions are used to determine the relationship, i.e. genotype, genotype by environment interaction with diseases and climate condition.

Understanding the behavior of complex traits involves modeling a Bayesian network of interactions among the effects of genes, environmental conditions, and other covariates

2.2. Joint Probability Distribution Model

The GEI probabilistic model represents the joint probability distribution (JPD) by the combination of the values of all the variables. The conditional distributions for each variable in Bayesian networks always factoring by JPD. If x_i denotes some value of the variable Xi and pai denotes some set of values for the parents ofX_i, then $P(X_i|pai)$ denotes this conditional probability distribution [33]. From the frequentist point of view, we modelled plot grain yield data collected from multiple environments in terms of environmental effects, block effects within environments, genotype effects, and genotype-by-environment interactions, under these assumptions, the local distribution $P(X_i|x_i)$ of each factor is a linear model. By using these results, the joint probability distribution P of a Bayesian network $B = (G,P)$ can be factorized as follows:

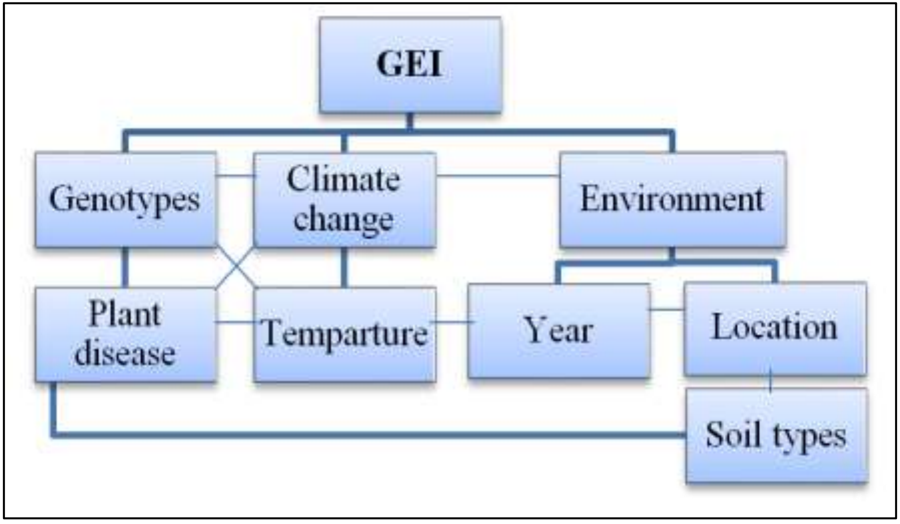
$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i|\, pa(X_i)),$$

Where $\mathrm{pa}(\mathrm{X_i})$ stands for the set of random variables associated with the parents of the node corresponding to the variable$\mathrm{X_i}$. Hence, the structure of the acyclic directed graph (ADG) that used to decompose the joint probability distribution into factors.

2.3. Bayesian Belief Networks

In Bayesian belief networks (BBN) are well known and established as a way of representing problems involving uncertain relations among a collection of random variables. The nodes in a Bayesian network are graphical representations of events that exist in real life and they are termed variables or states [34]. Relations between such nodes are represented with and an arc drawn between the nodes. If there is a causal relationship between two variables, and so the bow will be directional or directed from the cause variable to the result variable. Figure 1 shows a Bayesian network where the genotype, environment, climate variables, disease types and soil types are conditionally independent. In this BBNs structure, at that place are nodes and sub symptom nodes, which correspond to variable. For example, the diseases type's nodes are maize dwarf mosaic, maize sheath blight, maize northern blight and bipolarismaydis.

Figure-1. The BNs framework model



The BNs framework model is a proposed of GEI probabilistic modeling of the yield prediction system based on soil type, climate variables, effect of diseases applied to maize data. All the nodes in the network are associated with a probability distribution function which dimensions and definitions are dependent on the arcs that lead into the node. BNs can be considered as a special case of the more general class called graphical models where nodes represent random variables and the absence of arcs represents conditional independence assumptions between variables. In BNs, one node is used for each scalar variable, which may be discrete, continuous, or propositional (true/false). Moreover, results representation on GEI is a highly innovative point of this study, which allowed the assessment of a large amount of data gathering from multiple environmental databases to evaluate the model coupled with soil types, disease types and climate variables. The causal direction can be encoded by orienting the arcs from cause to effect.

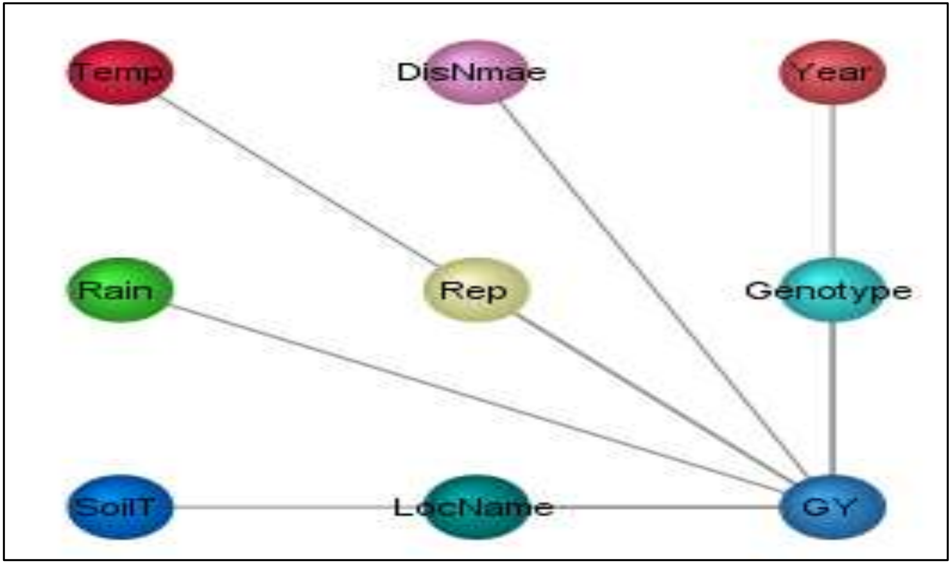
3. Results and Discussion

The Bayesian network using supervised machine learning considered for discovering probabilistic relationships between many variables, without having to specify input or output nodes. The results of BNs outcomes has been selected the grain yield with a selection of genotype by environments interaction in locations or years, with others environmental variability such rainfall and temperature. Likewise, the choice of genotypes based on performance indices of grain yield rate related to temperature, plant diseases and rainfall.

3.1. BNs Node Relationships

The relationships between the states of genotypes and environmental node in the GEI model were determined from the Bayesian estimation and learning. Figure 2 shows the conceptualization of BNs structure diagram developed around a selected priority issue by GEI model participants.

Figure-2. Mapping of GEI Bayesian network



Where: DisNmae= disease type (MD= Maize dwarf mosaic, SB= Maize sheath blight, NB=Maize northern blight and BM=Bipolarismaydis), Rep=replication/block, Rain=rainfall, Temp= Temperature SoilT=Soil type (Clay, Heavy clay, heavy and Heavy craning), Location name= Sinnar (SIN), Wad Medani (Wad), Rahad (RAH) and Gaderf (GEF) and GY=Maize grain yield.

Figure-3. Network consistency report of density function associated graph

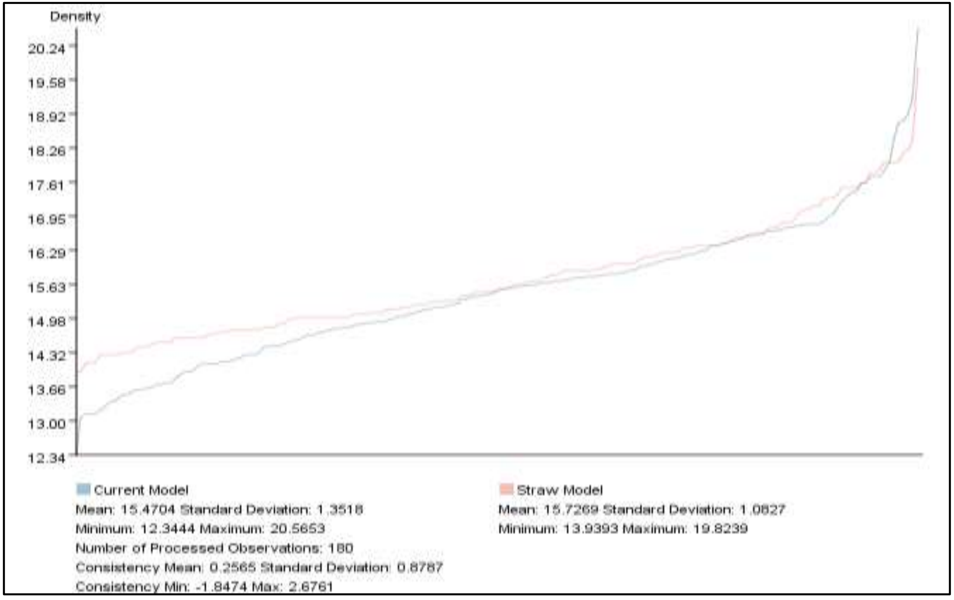


Figure 3 summarizes probability and density for overall current and straw model. Both are a good model fit with consistency mean, standard deviation, minimum, maximum, and consistency gain that indicating an adequate fit.

3.2. Probability Distributions of GEI Model

On Bayesian networks, the probability distributions are the foundation of statistical models' investigation and for building the probabilistic structure's relationship between the variable/notes. In figure 4, we get two graphs, histogram joint probability and mean values and histogram with interactive inference among the notes/variables that assigning a probability to a subset of the possible outcomes' relationship. The distribution and density function of the yield, temperature and rainfall are shown in figure 5, which presented a bunch of continuous density functions (probability distributions).

Figure-4. Histogram of genotypes, locations, years, rainfall, temperatures, soil type diseases and grain yield by bringing up its monitor

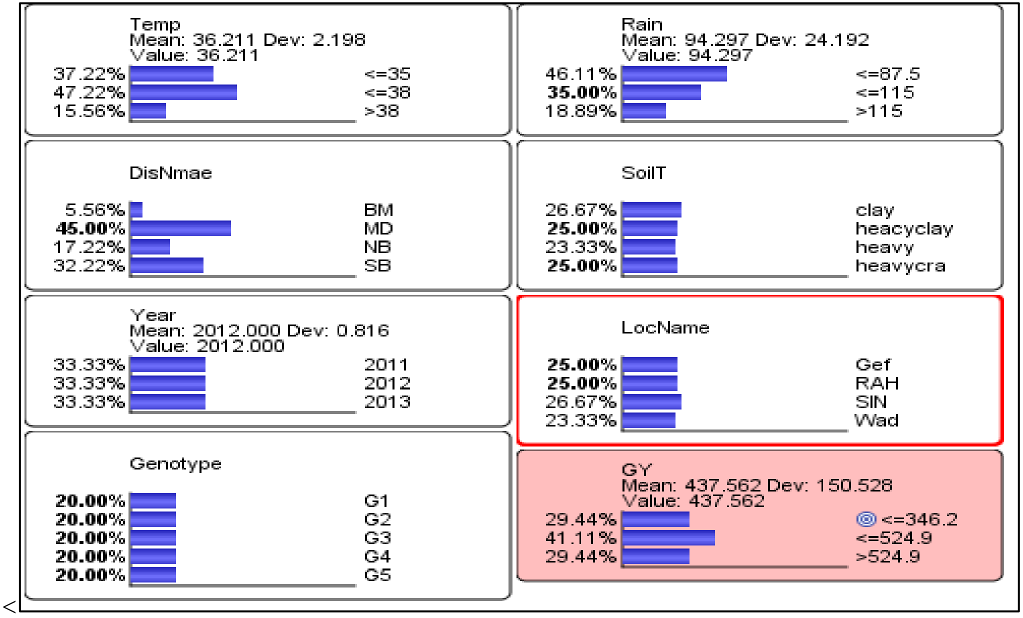


Figure 4 shows the histogram of genotypes, localizations, years, precipitation, temperature, and disease through its monitor. The results showed that genotypes had the same probability and that the frequency and probability of rainfall with <=88(46%), temperature with 35(37%), soil type of clay (27%), and disease were Manitoba (47%).

Figure-5. Distribution and Density function of the yield, temperature, and rainfall

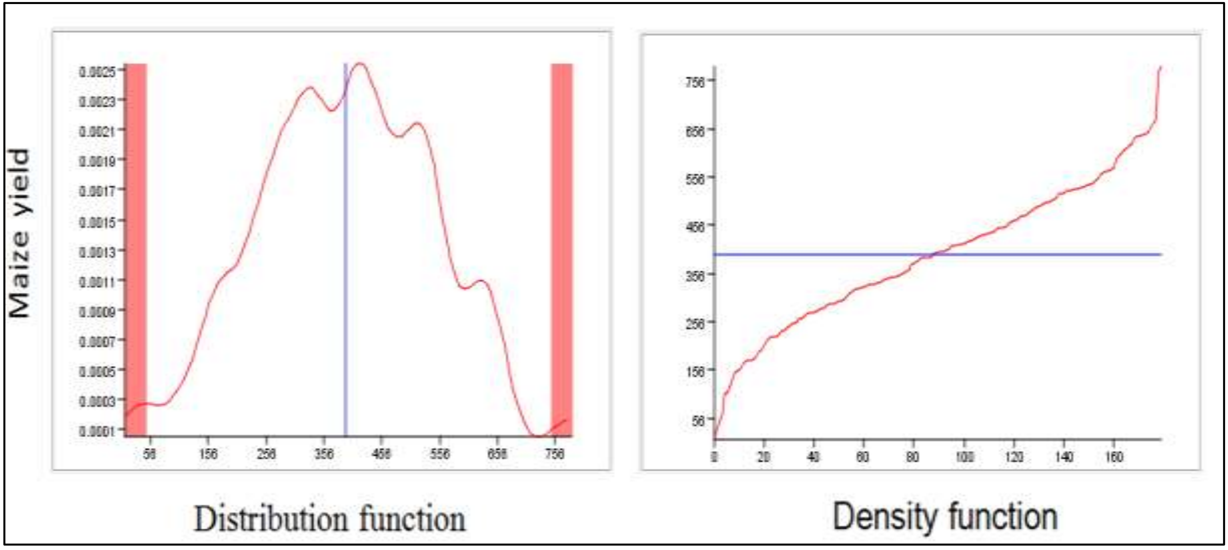


Figure 5 shows the distribution and density function of rainfall is mostly close to lognormal distribution, while maize yield is close to the normal distribution and temperature is close to the power lognormal distribution. However, a Bayesian network is considered to specific the real-world phenomenon of the parameters, which have actual numbers as values, which we can estimate from the data.

3.3. Bayesian Network Target Analysis of Maize Grain Yield

Simulated data information such as mean value, modal value, a priori modal value, Bayes factor variation, maximal positive variation and maximal negative variation were presented in different three level states of grain yield in Table 1, 2 and 3 respectively. We also describe the NBS results of the comparison of GEI performance of corn yields using three classes. The significance of the node in relation to the information that the node provides to the knowledge of grain yield.

Table-1. Performance indices of maize at grain yield (GY) = <=524.9 (2/3) (41.11%)

Node	Mean Value	Modal Value		A Priori Modal Value		Bayes Factor	Maximal Positive Variation		Maximal Negative Variation	
Genotype	1.93	G3	0.24	*G3*	*0.20*	0.28	G3	0.04	G5	0.05
Location	1.50	SIN	0.31	*SIN*	*0.27*	0.22	SIN	0.04	GEF	0.03
Soil type	1.43	clay	0.31	*Clay*	*0.27*	0.22	clay	0.04	Heavy clay	0.03
Temperature	36.18	<=38 (2/3)	0.42	*<=38 (2/3)*	*0.47*	-0.17	<=35 (1/3)	0.03	<=38 (2/3)	0.05
Replication	1.97	2.00	0.38	*2.00*	*0.33*	0.18	2.00	0.05	3.00	0.04
Year	2011.93	2011.00	0.38	*2011.00*	*0.33*	0.18	2011.00	0.05	2012.00	0.02
Rainfall	95.52	<=87.5 (1/3)	0.42	*<=87.5 (1/3)*	*0.46*	-0.14	<=115 (2/3)	0.04	<=87.5 (1/3)	0.04
Disease type	1.72	MD	0.49	*MD*	*0.45*	0.11	MD	0.04	NB	0.02

Table 1 presents performance indexes for all nodes (location, replication, years, genotypes, diseases, soil type, temperature and rainfall) identified by the node for grain yield (GY) <=524.9. The modal value of the cereal yield was localized at Sinnar in relication2/block2 with soil-type clay during the 2011 season at a temperature of 38c with MD-type diseases of genotypes G3 to 88 mm of rain. A performance index for the maximum positive variables was similar to the original modal value performance indexes, with temperatures and precipitation differing from 35c and 115mm. While the performance indexes of the maximum negative variability were entirely different from the original modal value. The maximum negative modal value of the grain yield was located at Gedarif with a heavy clay soil type in the 2012 season at temperature 38 with NB-type diseases of genotypes G5 to 88 mm of rain.

Table-2. Performance Indices of maize at grain yield (GY) = >524.9 (3/3) (29.44%)

Node	Mean Value	Modal Value		A Priori Modal Value		Bayes Factor	Maximal Positive Variation		Maximal Negative Variation	
Genotype	2.43	G5	0.36	G3	0.20	0.84	G5	0.16	G1	0.07
Year	2012	2011	0.45	2011.00	0.33	0.44	2011	0.12	2013	0.16
Location	1.32	GEF	0.36	SIN	0.27	0.52	GEF	0.11	SIN	0.12
Soil type	1.57	Heavy clay	0.36	clay	0.27	0.52	Heavy clay	0.11	clay	0.12
Rep	2.04	3.00	0.42	2.00	0.33	0.32	3.00	0.08	2.00	0.13
Disease type	1.74	MD	0.49	MD	0.45	0.12	MD	0.04	SB	0.02
Rainfall	95.60	<=87.5 (1/3)	0.43	<=87.5 (1/3)	0.46	-0.09	>115 (3/3)	0.02	<=87.5 (1/3)	0.03
Temperature	36.18	<=38 (2/3)	0.47	<=38 (2/3)	0.47	0.00	<=35 (1/3)	0.01	>38 (3/3)	0.00

Table 2 presents the performance indexes for all nodes (location, replication, years, temperature, disease, genotype and precipitation) reported by the node to grain yield (GY) <=524.9. The modal value of the cereal yield was located at Gedarif in relication3/block 3 with heavy clay soil type during the 2011 season at a temperature of 38ºC with DG disease type G5 to 88 mm of rain. The maximum positive change performance indexes were similar to the original forecast modal value performance indexes, with temperatures and precipitation differing from 35c and 115mm. The maximum negative modal value of cereal yield was located at Sinner with soil-type clay during the 2013 season at temperature 38 with SB-type diseases from genotypes G1 to 88 mm of rain.

Table-3. Performance Indices of maize at grain yield (GY) = <=346.2 (1/3) (29.44%)

Node	Mean Value	Modal Value		A Priori Modal Value		Bayes Factor	Maximal Positive Variation		Maximal Negative Variation	
Year	1.00	2012	2013	0.53	*2011*	*0.33*	0.66	2013	0.19	2011
Genotype	0.34	1.66	G1	0.28	*G3*	*0.20*	0.50	G1	0.08	G5
Disease type	0.17	1.85	MD	0.36	*MD*	*0.45*	-0.33	SB	0.04	MD
Rainfall	0.14	91.28	<=87.5 (1/3)	0.55	*<=87.5 (1/3)*	*0.46*	0.25	<=87.5 (1/3)	0.09	<=115 (2/3)
Location	0.12	1.62	SIN	0.32	*SIN*	*0.27*	0.27	SIN	0.05	Gef
Soil types	0.12	1.42	clay	0.32	*clay*	*0.27*	0.27	clay	0.05	heacyclay
Temperature	0.10	36.29	<=38 (2/3)	0.55	*<=38 (2/3)*	*0.47*	0.21	<=38 (2/3)	0.07	<=35 (1/3)
Replication	0.08	2.00	2.00	0.40	*2.00*	*0.33*	0.25	2.00	0.06	3.00

Table 3 presents the performance indexes for all nodes (location, replication, years, temperature, disease, genotype and precipitation) reported by the node to cereal yield (YG) <=346.2. The modal value of grain yield was sited at Sinnar locations with soil type clay during season 2013 at a temperature 38 degrees with diseases type MD of genotypes G1 88 mm rainfalls. Performance indexes of the maximal positive variation were like performance indices of original modal value expected the disease type was SB. The maximal negative modal value of grain yield was sited at Gedarif location in relication3/block3 with soil type Heavy clay during season 2011 at temperature 35 with diseases type MD of genotypes G5 at 115 mm rainfall. Generally, in terms of Bayes factor, the performance indices of maize at grain yield (GY) at GY =524.9 was performed better compared to other two values.

3.4. Multi Target Performance

Table 4 shows the precision and mean, coefficient of determination and mean square error of grain yield of the different percentages of values randomly suppressed (mean, minimum and maximum) for corn grain yield. The highest overall accuracy was achieved at 41% and the average was 36% with a 4.5% determination factor. This indicates that the Bayesian efficiency of grain yield is reasonable using suitable simulated data.

Table-4. Overall and means precision, coefficient of determination, and mean square error of GEI factors performance

Node	Overall Precision	Mean Precision	R2	RMSE	NRMSE
Disease type	45.00%	26.10%	0.0035	0.9665	32.22%
Grain yield	53.33%	53.24%	0.2176	133.1776	17.24%
Genotype	28.89%	28.89%	0.0456	1.3816	34.54%
Location Name	32.78%	31.39%	0.0112	1.0967	36.56%
Rainfall	46.11%	33.33%	0.0178	24.0101	18.61%
Replication	39.44%	39.44%	0.0011	0.8161	40.80%
Soil type	32.78%	31.39%	0.0033	1.1305	37.68%
Temperature	47.22%	33.33%	0.0003	2.1982	21.98%
Year	44.44%	44.44%	0.1011	0.7741	38.71%

R2= coefficient of determination, RMSE= root mean square error, NRMSE= the normalized root mean square error.

Table 4 provides the overall and average precision, coefficient of determination and mean square error of the GEI variables used in this study. Years and block site (replication) gave a high overall accuracy, while disease, grain yield rainfall gave a global accuracy and a temperature of about 60%, and locations (sites) have given an overall accuracy of about 50 percent. Genotypes had a general accuracy of 31%. On the basis of the coefficient of determination, grain yields and the season proved to be strongly linked. Different levels of precision were found for the different traits. Overall accuracy forecasts were 53.3% with a global coefficient of determination of 47%. However, grain yield and rainfall were subject to a high forecast error of 133.17 and 24, respectively. For quantitative evaluations, discrete output states may be replaced by NRMSE values of 17.24% and 18.6% for greater efficiency. The optimization tree has constructed like (grain yield with location, rainfall with (genotypes and temperatures), soil types and plant disease. The target interpretation tree prescribed in this study to obtain maximum information to a GEI model, also taking into account the fact that the data were adapted or simulated. Conceptual model of the links between genotype and environmental values (precipitation and temperature) and disease. The target grain yield (GY) influence analysis represents the management actions and attributes of a priority in response to changes in grain performance metrics. The study provided a road map for modeling knowledge on plant diseases and climatic conditions using factual reasoning with Bayesian networks. This study is being proposed as a Bayesian network framework for GEI model inference at managerial levels and to highlight the need of this interference as an integral part of the classification, optimization and predication in BNs. While the use of BNs is already becoming a standard practice, the coherence of the BayesiaLab software will be useful and appropriate in the application of Bayesian research and also in improving the capacity of decision analysis plans. The results demonstrated that applies the Bayesian network for genotype efficacy by the interaction of the environment under plant diseases, the model of land types and climate variables as a synthesis framework enables the agronomist to establish the uncertainty that characterizes the definition and implementation of the GEI analysis. As observed in other studies, Bayesian networks can investigate appropriate and interpretable framework for the simultaneous modeling of multiple quantitative traits in better predictive power result in the context of additive genetic models [35] Among the other studies that had good performance for example [36] used VW-4DEnSRF algorithm to study the area and winter wheat yield estimation based on the WOFOST crop model and a crop yield assimilation system. Wei, *et al.* [25], study highlighted that despite of the impact machine learning approaches in understanding and exploiting GEI for prediction, there is still some room for expanding and improving their use in applications not yet explored. [37] used the deep learning model such as convolutional neural network (CNN) and artificial neural network (ANN) to Self-Predictable Crop Yield Platform (SCYP) based on crop diseases using deep learning that collects weather information (temperature, humidity, sunshine, precipitation, etc.) and farm status information (harvest date, disease information, crop status, ground temperature, etc.), diagnoses crop diseases by using and predicts crop yield.

4. Conclusions

The study is well addressed, the diagnosis of maize grain yield under climate, soil types and disease types. Indeed, the role of Bayesian network has identified the elements that determine a management scheme that ensures sustainability of maize yield resources. Bayesian network is a potent tool for interpreting and dealing with uncertain knowledge. Thus, it is instinctive to use Bayesian network for yield prediction using GEI model. However, the results obtained by this method are dependent on with the structure and parameters of the GEI Bayesian network. The results of this study were the ideal example to be concerning of future research application of Bayesian network in plant breeding programs of the big data related to GEI problem. Future development of Bayesian network of multiple environmental trials based on data, experimentation in the agricultural project will highly affect the agricultural classification.

Acknowledgments

The author would like to thank Mr. Murari Singh from the Department of Mathematics and Statistics, Concordia University, Montreal, Quebec, Canada for his comments and review for his attentive reading of the manuscript.

References

[1] Ryan, J., Singh, M., Masri, S., and Ibrikci, H., 2012. "Spatial variation in soil organic matter, available phosphorus, and potassium under semi-arid conditions: Research Station Management Implications." *Communications in Soil Science and Plant Analysis*, vol. 43, pp. 2820-2833.

[2] Mamo, L. A., Ararsa, L., Zeleke, H., and Nigusse, M., 2016. "Genotype by environment interaction and yield stability of maize (Zea mays L.) hybrids in Ethiopia." *Journal of Natural Sciences Research*, vol. 6, p. 13.

[3] Elias, A. A., Robbins, K. R., Doerge, R. W., and Tuinstra, M. R., 2016. "Half a century of studying genotype × environment interactions in plant breeding experiments." *Crop Science*, vol. 56, pp. 2090-2105.

[4] Mafouasson, H. N., Gracen, V., Yeboah, M. A., Ntsefong, G. N., Tandzi, N. L., and Mutengwa, C. S., 2018. "Genotype-by-environment interaction and yield stability of maize single cross hybrids developed from tropical inbred lines." *Agronomy*, vol. 8, p. 62.

[5] Kaikkonen, L., Parviainen, T., Rahikainen, M., Uusitalo, L., and Lehikoinen, A., 2020. "Bayesian networks in environmental risk assessment." *A Review Integrated Environmental Assessment and Management*, vol. 17, pp. 62-78.

[6] Laurila-Pant, M., Mäntyniemi, S., Venesjärvi, R., and Lehikoinen, A., 2019. "Incorporating stakeholders' values into environmental decision support: A Bayesian Belief Network approach." *Science of the total Environment*, vol. 697, pp. 1-11.

[7] Zhang, L., Pan, Q., Wang, Y., Wu, X., and Shi, X., 2017. "Bayesian network construction and genotype-phenotype inference using gwas statistics." *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, pp. 475-489.

[8] Do Amaral, C. B., de Oliveira, G. H. F., Eghrari, K., Buzinaro, R., and Vitti Môro, G., 2019. "Bayesian network: a simplified approach for environmental similarity studies on maize." *Crop Breeding and Applied Biotechnology*, vol. 19, pp. 70-76.

[9] Kocian, A., Massa, D., Cannazzaro, S., Incrocci, L., Lonardo, S. D., Milazzo, P., and Chessa, S., 2020. "Dynamic Bayesian network for crop growth prediction in greenhouses." *Computers and Electronics in Agriculture*, vol. 169, p. 10516.

[10] Sperottoab, A., Molinac, J. L., Torresanab, S., Crittoab, A., Velazquezd, M. P., and Marcominiab, A., 2019. "A Bayesian networks approach for the assessment of climate change impacts on nutrients loading." *Environmental Science and Policy*, vol. 100, pp. 21-36.

[11] Pérez-Ariza, C. B., Nicholson, A. E., and Flores, M. J., 2012. "Prediction of coffee rust disease using Bayesian networks." In *Sixth European Workshop on Probabilistic Graphical Models, Granada, Spain.*

[12] Balducci, F., Impedovo, D., and Pirlo, G., 2018. "Machine learning applications on agricultural datasets for smart farm enhancement." *Machines*, vol. 6, p. 38.

[13] HernÃ¡ndez, S. and LÃ³pez, J. L., 2020. "Uncertainty quantification for plant disease detection using Bayesian deep learning." *Applied Soft Computing*, vol. 96, p. 10659.

[14] Newlands, N. K. and Smith, L. T., 2010. "Predicting energy crop yield using Bayesian networks." In *Proceedings of the Fifth IASTED International Conference, Computer Intelligence, Maui, Hawaii, USA.*

[15] Serrano, M. B., Cancelas, G. N., Flores, F. S., Nuñez, S. A., and Orive, A. C., 2018. "Use of Bayesian networks to analyze port variables in order to make sustainable planning and management decision." *Logistics*, vol. 2, p. 5.

[16] Musango, J. K. and Peter, C., 2007. "A Bayesian approach towards facilitating climate change adaptation research on the South African agricultural sector." *Agrekon*, vol. 46, p. 46.

[17] Ershadi, M. M. and Seifi, A., 2020. "An efficient Bayesian network for differential diagnosis using experts' knowledge." *International Journal of Intelligent Computing and Cybernetics*, vol. 13, pp. 103-126. Available: https://doi.org/10.1108/IJICC-10-2019-0112

[18] Boudhar, A., Belaid, B., and Bourk, A., 2016. "Cephalopod fisheries management and sustainable development in morocco: a Bayesian networks approach." *Asian Journal of Science and Technology*, vol. 7, pp. 2565-2573.

[19] Tong, H. and Nikoloski, Z., 2021. "Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data." *Journal of Plant Physiology*, vol. 257, p. 153354. Available: https://doi.org/10.1016/j.jplph.2020.153354

[20] Niazian, M. and Niedbała, G., 2020. "Machine learning for plant breeding and biotechnology." *Agriculture*, vol. 10, p. 436.

[21] Niazian, M., Sadat-Noori, S. A., and Abdipour, M., 2018. "Modeling the seed yield of Ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models." *Ind. Crops Prod.*, vol. 117, pp. 224-234. Available: https://doi.org/10.1016/j.indcrop.2018.03.013

[22] Niedbała, G., Nowakowski, K., Rudowicz-Nawrocka, J., Piekutowska, M., Weres, J., Tomczak, R. J., Tyksi´nski, T., and Pinto, A. Á., 2019. "Multicriteria prediction and simulation of winter wheat yield using extended qualitative and quantitative data based on artificial neural networks." *Applied Sciences*, vol. 9, p. 2773.

[23] Verma, S., Chug, A., and Singh, A. P., 2020. "Application of convolutional neural networks for evaluation of disease severity in tomato plant." *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 3, pp. 273-282. Available: https://doi.org/10.1080/09720529.2020.1721890

[24] Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., and Singh, A. K., 2019. "Machine learning approach for prescriptive plant breeding." *Scientific Reports*, vol. 9, p. 17132.

[25] Wei, M. C. F., Maldaner, L. F., Ottoni, P. M. N., and Molin, J. P., 2020. "Carrot yield mapping: A precision agriculture approach based on machine learning." *Artificial Intelligence*, vol. 1, pp. 229-241.

[26] Stöckle, C. O. and Kemanian, A. R., 2020. "Can crop models identify critical gaps in genetics, environment, and management interactions?" *Frontiers in Plant Sciences*, vol. 11, p. 737.

[27] Lu, W., Newlands, N. K., Carisse, O., Atkinson, D. E., and Cannon, A. J., 2020. "Disease risk forecasting with bayesian learning networks: Application to grape powdery mildew (erysiphe necator) in vineyards." *Agronomy*, vol. 10, p. 622.

[28] Guttorp, P., 2014. "The annual review of statistics and its application is online at statistics." *Annual Reviews, Statistics and Climate*, vol. 1, pp. 87-10.

[29] Omer, S. O., Abdalla, A. W. H., Singh, N. P., Kumar, H., and Singh, M., 2020. "Bayesian estimation of variance components, heritability and genetic advance from multi-year and location chickpea trials in Indian environments." *International Journal of Statistics and Applications*, vol. 10, pp. 150-159.

[30] Resende, R. T., Piepho, H. P., Rosa, G. J. M., Silva-Junior, O. B. E., Silva, F. F., De Resende, M. D. V., and Grattapaglia, D., 2020. "Enviromics in breeding: applications and perspectives on envirotypic-assisted selection." *Theoretical and Applied Genetics*, vol. 134, pp. 95-112.

[31] Avilés, A., Célleri, R., Solera, A., and Paredes, J., 2016. "Probabilistic forecasting of drought events using markov chain- and bayesian network-based models: A case study of an andean regulated river basin." *Water*, vol. 8, pp. 222-224.

[32] Conrady, S. and Jouffe, L., 2015. *Bayesian networks and bayesialab: A practical introduction for researchers*. 1st ed. Bayesia USA.

[33] Scutari, M. and Denis, J. B., 2015. *Bayesian networks with examples in R*. CRC Press Taylor and Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL pp. 33487-2742.

[34] Aguilera, A., Fernandez, A., Fernandez, R., Rumi, R., and Salmeron, A., 2011. "Bayesian networks in environmental modeling." *Environmental Modeling and Software*, vol. 26, pp. 11376-1388.

[35] Scutari, M., Howell, P., Balding, D. J., and Mackay, I., 2014. "Multiple quantitative trait analysis using Bayesian networks." *Genetics*, vol. 198, pp. 129-137.

[36] Wu, S., Yang, P., Ren, J., Chen, Z., and Li, H., 2021. "Regional winter wheat yield estimation based on the WOFOST model and a novel VW-4DEnSRF assimilation algorithm." *Remote Sensing of Environment*, vol. 255, p. 112276. Available: https://doi.org/10.1016/j.rse.2020.112276

[37] Lee, S., Jeong, Y., Son, S., and Lee, B. A., 2019. "Self-predictable crop yield platform (scyp) based on crop diseases using deep learning." *Sustainability*, vol. 1, p. 3637. Available: https://doi.org/10.3390/su11133637