

Modeling Climate Change Impacts on Some Selected Cereal Crops in Northern Ghana using Dynamic Bayesian Network Approach

William Kofi Nkegbe*, Abukari Alhassan, Alhassan Faisal

Department of Statistics, University for Development Studies, P. O. Box 1350, Tamale, Ghana

*Corresponding author: william.nkegbe@uds.edu.gh

Received May 16, 2025; Revised June 18, 2025; Accepted June 26, 2025

Abstract Statistical modeling is fundamental in understanding and predicting the variability of climate change and its effect on cereal crop yield in Northern Ghana. Most often, there is little or no data to study these relationships within the Northern Ghana enclave. The Dynamic Bayesian Network Modeling was employed to model the relationship using data on some selected cereal crops (Maize, Rice, and Sorghum) yield from Ghana's Ministry of Food and Agriculture, and some climate variables (rainfall, sunshine, temperature, humidity, and windspeed) from Ghana Meteorological Agency, for Northern Ghana for a thirty-one-year period. The results showed that climate variables are significantly related to yield at the intra-slide and inter-slide levels and that there exist positive and significant (link strength values > 1) relationships between cereal crop yield and climate change variables. The results imply that, with data on current climate variables under the study, the next cereal crop yield can be predicted with reduced uncertainty and better forecasting ability while holding other factors constant. On the basis of the analysis, the study affirmed the use of the Dynamic Bayesian Networks and highly recommended the extension of the study to include other variables such as vegetative cover, cultivated area, and chemical use, when data becomes available.

Keywords: *Dynamic Bayesian Networks, Climate Change, Link Strength, Cereal Crop Yield, Northern Ghana, Agriculture*

Cite This Article: William Kofi Nkegbe, Abukari Alhassan, and Alhassan Faisal, "Modeling Climate Change Impacts on Some Selected Cereal Crops in Northern Ghana using Dynamic Bayesian Network Approach." *American Journal of Applied Mathematics and Statistics*, vol. 13, no. 2 (2025): 38-46. doi: 10.12691/ajams-13-2-3.

1. Introduction

It is observed that the six ecological zones in Ghana that provide livelihood support (forestry and agricultural) and other critical natural resources for rural communities and the country's economy are not spared the clutches of climate change [1]. Agriculture, being a major source of employment in Ghana suffers the most from climate change; and predominantly affected by the distribution of rainfall. The increasing rate at which rainfall patterns vary result in increase in the risk linked to farming, since it makes prediction very difficult [2,3].

The Ghanaian economy to some extent, relies on sectors that are sensitive to climatic conditions, mainly agriculture, energy and forestry. The inadequate use of irrigation resources and high dependence on unfavorable climatic conditions for good harvest tend to introduce significant volatility in the standard of living of people. Climate change poses a serious socioeconomic risk to Ghana, especially in the Northern sector which has a long dry season of about seven months and a shorter rainy season of about five months with mostly intermittent

drought and floods [4]. The Northern sector is also observed to be the most agricultural part of Ghana with a high percentage of subsistent food crop farmers, and manifestations of climatic instability in the form of floods, late rains before planting season, and persistent droughts during the planting season [5].

Northern Ghana also contributes immensely to the production of rice and maize in the country and is a primary source of sorghum and millet in Ghana [6,7,8]. Cereals such as maize, rice, sorghum, and millet are sources of major staples in Northern Ghana in particular and the country as a whole [6], besides its other commercial uses in areas such as the poultry and brewery industries [9]. The decision to conduct the study in Northern Ghana is therefore worthwhile.

Further, the effects of climate change on food crop production is getting worse [4,10], and it is even affecting the achievement of the United Nation's Sustainable Development Goal Two; which seeks to end all forms of hunger, achieve food security, improve nutrition, and promote sustainable agriculture by 2030 [11,12]. This makes every contribution towards enhancing the efforts at ensuring the sustenance of food production to feed the rising population and the economy, especially in the

Northern sector of the country, worthwhile.

The socioeconomic importance of the agricultural sector and cereal production for that matter, in the country cannot be overemphasized, and for that matter, steps towards mitigating the impact of climate change in the Northern sector is crucial. To be able to effectively implement any mitigation scheme in this respect, climate change variability and its effects, especially on cereal crop yield, need to be understood and be able to be predicted. This makes it imperative to include statistical modeling in the process. The question on what the relationship between cereal crop yield and climate change in Northern Ghana arises. The attempt at finding an answer to this question results in the quest to investigate and model the impact of climate change on the yield of some selected cereal crops in Northern Ghana. This is to enable easy and accurate assessment of climate change impacts on cereal crop yield, to enhance understanding of the situation ahead of time, and to inform mitigation actions. It will further contribute to literature on climate change and its impact on cereal crop yield in the area, and will also show the appropriateness of the application of the Dynamic Bayesian Networks in the study.

Climate Change here is, a change of climate that is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable periods [13].

Some climate-related research works have been done in the study area. Dumenu and Obeng (2016), observed that the most experienced climate change impacts in the most vulnerable ecological zones (Sudan and Guinea Savanna zones) are prolonged drought, erratic rainfall, reduction in crop yield and frequent flooding [14]. Antwi-Agyei et al. (2012) found the Regions in Northern Ghana, as the most vulnerable to drought. These areas are also observed to have high temperatures and low levels of rainfall [15]. Prolonged drought and erratic rainfall are particularly seriously affecting the rural dwellers in these areas, who rely solely on rains for farming as their occupation [3]. Climate change is a danger to agriculture, food security, and national development, and as a result, there is a need to plan adaptation and mitigation actions that will help reduce its effects.

Studies have also been conducted to assess climate change impacts, vulnerability, and adaptation strategies in Ghana [16,17,18]; impacts on livelihoods and adaptation strategies of rural communities [19]. There are however, little or no sufficient studies on the relationship between climate change and cereal crop yield in Northern Ghana. Further, there is also the issue of inadequate data on cereal crop yield from the study area, hence making the application of a Bayesian approach most suitable.

Bayesian network (BN) models are a kind of probabilistic graphical models that have become very popular among practitioners mainly due to the powerful probability theory involved, which makes them able to deal with a wide range of problems [20]. In their work on groundwater quality assessment using data clustering, [21] adopted the mixture of truncated exponentials (MTEs) framework in the design and application of a probabilistic clustering, based on hybrid Bayesian networks [22]. [20] also applied the Bayesian Networks in their work on

climate change assessment using the evaluation of its impact on the olive system in Andalusia. Their results predicted that climate change will lead to changes in the territorial distribution of olive crops. They also observed from their results that Bayesian networks are a powerful tool that allows dealing with both discrete and continuous data [23].

Dynamic Bayesian Networks (DBNs) are extensions of Bayesian networks to model dynamic processes and consist of a series of time intervals that present the states of all variables at a given time and thus represent the evolution of a process over time [24,25]. They can be seen as a generalization of Markov Chains and Hidden Markov Models since they represent a space of states in a factorized way instead of as a single discrete random variable and can be classified as directed and probabilistic [26]. DBNs have been applied in some areas such as pattern detection of human brain behavior [27], speech recognition [28,29], medical diagnosis [30,31,32], visual activities and crime risk analysis [30,33,34], sensor validation, client analysis and video object tracking [30,34] to capture their dynamic behavior in the modeling process.

2. Materials and Methods

2.1. Study Area

The study covers climate change variables and cereal crop yield in Northern Ghana, which comprises five administrative regions, namely, the Northern, Savanna, North East, Upper East, and Upper West regions. It lies between latitudes 8°N and 11°N and has a land area of 97702 km², with rainfall characterized by a long dry period of about seven months from October/November to April/May with no appreciable rainfall level, and a rainy season from May to October with adequate levels of rainfall. Annual precipitation ranges between 400 and 1200mm. Agriculture is about 95% purely rain-fed and employs about 70% of the population. The region is one of the most agricultural parts of Ghana with a high percentage of subsistent food crop farmers; covering about 36% of the nation's cereal (maize, rice, and sorghum) yield over the last three decades [35]. The Region is sensitive and vulnerable to climate change and unpredictable climatic variability, especially erratic rainfall [36].

2.2. Data Sources

The study relied on secondary data from state institutions mandated to collect and handle such data. Data on climate variables (Rainfall, Temperature, Hours of Sunshine, Relative humidity, Wind speed) for Northern Ghana, was taken from seven weather stations in the North (Bole, Salaga, Tamale, Wa, Yendi, Walewale, and Bawku) from 1992 to 2022 was sourced from the Ghana Meteorological Agency (GMA) through its Client Service Office. Data on Annual Cereal Production figures was also sourced from the Statistics, Research and Information Directorate (SRID) of Ghana's Ministry of Food and Agriculture (MoFA) for the period 1992 to 2022 for the study area. Data was standardized to help eliminate the

effects of other influencing factors and also ensure standardized results for comparability and easy interpretability. Besides, since climate data are measured in different units of measurement, the scaled data is appropriate to synchronize for analysis.

2.3. Bayesian Network Models

Bayesian network (BN) is a statistical multivariate model for a set of random variables $X = \{X_1, \dots, X_n\}$ which is defined in terms of two components, namely, the qualitative component, comprising a directed acyclic graph (DAG) of random variables (vertices, X), linked by edges between them representing the existence of statistical dependence relationships between these variables in the model; and quantitative component, which is the conditional probability distribution for each variable $X_i, i = 1, \dots, n$, given its parents in the graph ($pa(x_i)$) expressed in conditional probability tables (CPTs) (for discrete variables) or probability functions (for continuous variables) [20,37]. The qualitative component allows Bayesian Network models to be easily understood by experts in other fields who are not familiar with the model's mathematical context; and also allows that, with no mathematical calculation involved, the variable(s) that are or are not relevant for a certain one or more variables can be known, and help to simplify the joint probability distribution of the variables necessary to specify the model [38,39]. Bayesian Networks therefore provide a compact representation of the joint probability distribution over all the variables, defined as the product of the conditional distributions attached to each node, so that

$$P(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | pa(x_i)) \quad (1)$$

where $pa(x_i)$ is a set of parents of variable x_i , according to the structure of the directed acyclic graph [37].

Various Bayesian Networks have been developed and used in modeling in different situations and on various data sets. Some of these include the Naïve Bayes (NB), Tree Augmented Naïve Bayes (TAN), Mixture of Truncated Exponentials (MTE) for regression [40,41,42,43,44,45] and Additive Bayesian Network (ABN) [46,47,48].

2.4. Dynamic Bayesian Networks

Dynamic Bayesian Networks (DBN) is a type of probabilistic graphical model that extends the static Bayesian Network structure to instances of time series, in which time series data are used for formulating causal relationships among random variables. A DBN is also comprised of a directed acyclic graph defining its structure and a set of parameters that define the probabilistic associations of the variables. In the case of DBNs, time is discretized into time slices that represent consecutive instances.

If $X^t = \{X_0^t, X_1^t, \dots, X_n^t\}$ a set of variables in a single time slice t , and for some horizon $t = T$, we can define the joint probability distribution of the network as

$$\begin{aligned} p(X^T, X^{T-1}, \dots, X^0) &= p(X^{T:0}) \\ &= p(X^T) \prod_{t=T-1}^0 p(X^t | X^{T:t+1}) \end{aligned} \quad (2)$$

where $p(X) = \prod_{i=1}^n p(x_i | Pa(x_i))$ represents the probability distribution of a set of nodes X , and $Pa(x_i)$ represents a set of parent nodes of X_i in the graph. Here, X^0 is the most recent instance.

In this setting, it is possible that nodes will have parents in previous time slices and for that matter variables in all instants $X^{T:0}$ have to be taken into consideration in computing the joint probability distribution as in Eq. (2). In this case, a common assumption is to suppose the DBN to be first-order Markovian [49,50]. This assumes that the future state of the system is independent of the past given the present, that is, $p(X^t | X^{T:t+1}) = p(X^t | X^{t+1})$.

2.5. Dynamic Gaussian Bayesian Networks

The Dynamic Gaussian Bayesian Networks (DGBN) is an extension of the Gaussian Bayesian Networks (GBN) model to a dynamic case, where the temporal component of a Dynamic Bayesian Network (DBN) is considered. In this context, time is discretized into time slices in a given period. There is a static GBN in each time slice which has the usual intra-slice parents, and at the same time has parents in the previous slices, and is also connected to the next slice. All the time slices are accounted for by the joint probability distribution until a certain horizon T :

$$\begin{aligned} p(X_0, X_1, \dots, X_T) &= p(X_{0:T}) \\ &= p(X_0) \prod_{t=0}^{T-1} p(X_{t+1} | X_{0:t}) \end{aligned} \quad (3)$$

where $X_t = \{X_{1t}, X_{2t}, \dots, X_{nt}\}$ is the vector of all the nodes in a time slice t and $t = 0, 1, \dots, T$. Eq. (3) requires that all the previous time slices should be considered in computing the product. The Markov assumption [50] is used to simplify this situation.

Definition 2.5.1. A Dynamic Bayesian Network fulfills the Markov assumption if, for all $t \geq 0$, the future time $t + 1$ is independent of the past $0:t - m$ given the last m time slices. This m is known as the Markovian order of the network [49,50].

The Markovian order determines the number of time slices required to assume with certainty that the present is independent of the past. The resulting joint probability distribution obtained when a Markovian order one is applied to the network will be

$$p(X_{0:T}) = p(X_0) \prod_{t=0}^{T-1} p(X_{t+1} | X_t) \quad (4)$$

When the temporal component is combined with the GBN structure, DGBNs are obtained, specified by linear Gaussian CPDs in each time slice. To calculate the joint density over time the nodes in the current time slice have to be taken into consideration together with the parent nodes in the previous time slice only.

It is a common practice to assume that the framework of the network in a DBN is homogeneous over time. This

implies that the structure remains constant and not independent on t . It means the inter-slice and intra-slice arcs and the parameters are reproduced whenever the network is unrolled. If a Markovian order one DGBN is created, it holds that given the present time slice, the future is independent from the past, and the whole Dynamic GBN can be represented by the first two time slices. This implies also that only inter-slice arcs are allowed from $t - 1$ to t . As the Markovian order increases, more arcs can emerge from earlier lags to the present. If it is detected that the state of the system depends on more than one temporal lag, the Markovian order of the network can be increased at the expense of greater complexity in learning its structure, its parameters, and for performing inference [49,51].

2.6. Structural Learning

The structure of the DGBN is learned to appreciate how the present is affected by the past values of the variables by adapting the dataset to the time discretization of DBNs and using the appropriate learning algorithm. Three structure learning algorithms are identified in learning GDBN. These are the dynamic max-min hill-climbing (DMMHC) algorithm [52], the binary particle swarm optimization (PSO) algorithm [53], and the natural number order invariant encoding PSO algorithm [49].

The structural learning in this study was done with our data set and the dynamic max-min hill climbing (DMMHC) algorithm [54]. The DMMHC is an extension of the max-min hill climbing (MMHC) algorithm [54], a hybrid method that searches the space of possible structures with a local search and then directs the arcs and scores the resulting networks with an evaluation criterion. The algorithm begins by performing a local search of the structure around the possible parents and children of each node, by determining the Markov blanket of the node. The group of nodes that make a node conditionally independent from the other nodes is the Markov blanket of that node. There are many estimates of conditional independence with continuous variables and given data. One of such is the exact t test for Pearson's correlation coefficient:

$$t(X, Y | Z) = \rho_{X,Y|Z} \sqrt{\frac{n - |Z| - 2}{1 - \rho_{X,Y|Z}^2}} \quad (5)$$

where $\rho_{X,Y|Z}$ is the partial correlation coefficient of X and Y given the set of variables Z and $|Z|$ represents the number of nodes in the set Z .

If G is the set of all nodes in the graph, and $Y \equiv G \setminus \{X, Z\}$, then what is essential is to find the set of nodes Z that makes a node X independent from all nodes $Y \in Y$. When the test is not significant, an undirected arc between X and all the nodes in Z is added to help find the local structure of the nodes. The skeleton of the network is obtained by combining the local structure of all the nodes found, which constrains the step of scoring and finding the best structure [55]. That is, to direct the arcs in the network, only edges that are present in the skeleton of the

net are allowed to be included, and the resulting structures are scored with the Bayesian information criterion (BIC) [56]. Once the local structure of all the nodes is found, we obtain the skeleton of the network by combining them, which constrains the step of scoring and finding the best structure [57]. Thus, to direct the arcs in the network, only edges that are present in the skeleton of the net are allowed to be included, and the resulting structures are scored with the Bayesian information criterion (BIC) [58]; which measures how well a model fits the data, and at the same time, penalizes complexity as measured by the number of parameters in the model. The model parameters are then fitted using maximum likelihood estimation, upon determining the structure of the network.

2.7. Inference on the GDBN

The exact inference method is implemented in this study by computing conditional probabilities in the equivalent multivariate Gaussian. It consists of converting the DGBN to its equivalent multivariate joint Gaussian $X = (X_1, X_2)$, where X_1 are the objective nodes and X_2 are the instantiated nodes. Once the evidence is instantiated, the objective node is marginalized. When the network is translated to a mean vector $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and a covariance matrix $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, the posterior mean and variance of the objective nodes can be calculated [59,60].

$$p(X_1 | X_2) = \mathcal{N}(X_1 | \mu_{1|2}, \Sigma_{1|2}) \quad (6)$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2) \quad (7)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad (8)$$

2.8. Bayesian Network Stability

It is recommendable to study a "stable" network structure; a set of edges and directions that is unlikely to vary; especially for cross-sectional studies. This is because the structure of Bayesian Networks could vary when learning algorithm is performed several times. Researchers need to study stable networks whose structure is likely to be influenced by factors such as sample size [61]. Bootstrapping methods can be used to evaluate the stability of network estimates in the case of partial correlation networks [62], and it is also possible to directly account for stability in structure learning in a similar way, where a large number of Bayesian Networks are learned from bootstrap samples and only edges that appear in a proportion of Bayesian Networks higher than some threshold are considered [61,63]. In a bootstrapping method, with at least 100 bootstrap samples, edges that appear in more than 85% of networks (edge strength), and whose direction appeared in more than 50% of networks (minimum direction) are considered accurately estimated and stable enough to be included in the network and reported [61,64,65].

2.9. Significant Arcs Determination

Arcs, also known as edges represent the causal dependencies among variables (nodes) in the probabilistic directed acyclic graph (DAG) of a Bayesian Network structure. They connect nodes that interact and indicate the direction of influence [66,67,68]. Significant arcs can be determined using a threshold estimated from the data and referred to as a significant threshold [69].

A systematic approach to identifying statistically significant features in a network was developed using bootstrap resampling and model averaging [70,71,72]; and summarized by [69] as:

1. For $b = 1, 2, 3, \dots, m$:
 - a. A new data set X_b^* is sampled from the data set X , using bootstrap;
 - b. The structure of the graphical model $G_b = (V, E_b)$ is learned from X_b^* .
2. The probability that each possible edge e_i , $i = 1, \dots, k$ is present in the true network structure $G_0 = (V, E_0)$ is estimated as

$$\hat{P}(e_i) = \frac{1}{m} \sum_{b=1}^m I_{\{e_i \in E_b\}} \quad (9)$$

Where $I_{\{e_i \in E_b\}}$ is the indicator function of the event $\{e_i \in E_b\}$; which implies that it is equal to 1 if $e_i \in E_b$, and 0, otherwise.

The observed probabilities $\hat{P}(e_i)$ are referred to as arc strengths, and explained as the degree of confidence that e_i is present in the structure of the network G_0 which describes the true dependence structure of X [73]. It is worth mentioning that the probabilities $\hat{P}(e_i)$ are an estimator of the expected value of the $\{0,1\}$ random vector which describes the presence of each possible edge in G_0 [74]. They therefore do not sum to one and are dependent on one another in a nontrivial way [73], [75].

The probability distribution of the structure of the graphical model G_b in the space of the network structures is unknown, hence it is difficult to evaluate these empirical probabilities. The value of the confidence threshold is also an unknown function of both the data and the structure learning algorithm. This threshold is the minimum degree of confidence for an edge to be considered significant and accepted as an edge of G_0 [75]. This leads to the use of ad hoc, pre-defined thresholds, which serve as a limitation in the identification of significant edges [76]; and a statistically motivated algorithm should be used for choosing a suitable threshold instead [73].

This study adopted a solution proposed by [69], which is implemented in bnlearn package to produce a default threshold value when a value is not specified by the analyst. They proposed that the threshold value could be computed through bootstrapping as;

If the value of the stored threshold is:

$$\hat{P} = \{\hat{P}_i, i = 1, 2, \dots, k\} \quad (10)$$

and

$$\hat{P}(\cdot) = \{0 \leq \hat{p}_1 \leq \hat{p}_2 \leq \dots \leq \hat{p}_k \leq 1\}$$

Then the arc strength that closely represent the unknown averaged network $G = (V, E_0)$ as

$$\hat{P}(i) = \begin{cases} 1, & \text{if } e_i \in E_b \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

that is to say, the set of strengths that designates any arc as being significant or not with certainty. An arc is considered significant if its arc strength value is greater than the computed threshold.

2.10. Link Strength

Link strength (LS), is a measure of the maximum amount that evidence at the parent node of the link can affect the belief at the child, given that all the other parents of the child have some evidence [77]. It is a single number measure for each link in a Bayesian Network.

The link strength of $A \rightarrow B$ is defined as:

$$LS(A \rightarrow B) = C \in \prod (C(B - \{A\})) CS(A, B | c) \quad (12)$$

where $C(B)$ is the set of parents of B and CS is the connection strength of $A \rightarrow B$.

From Figure 1, the strength of the $A \rightarrow B$ link is the highest value of connection strength that links A to B as all the parents $C_1, \dots, C_n$ take on varying evidence.

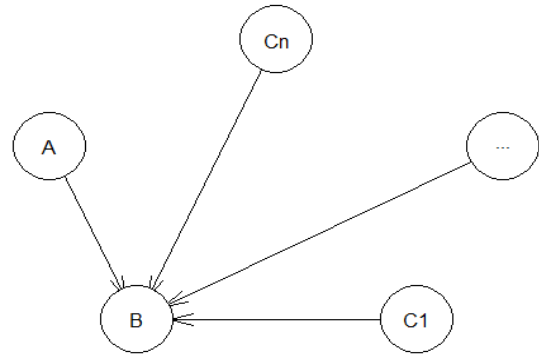


Figure 1. B and its parent nodes required to define the link strength of the link from A to B LS values vary from 0 for null links to infinity for the strongest links:

$$0 \leq LS \leq \infty \quad (13)$$

A link is significant enough to be included in a DAG if the link strength is greater than 0 [77,78]. This is used to determine the significance or otherwise of a link between cereal crop yield and climate change variables.

2.11. Forecasting with DBN Models

Forecasting with DBN models can be performed up to some horizon T , using a sliding window [79]. Given the initial state vector $s_0 = ((x_1^0, \dots, x_n^0), \dots, (x_1^t, \dots, x_n^t))$ with values of the variables in a system observed at instant t and at many previous instances as defined by the Markovian order of the network, inference is performed to

obtain the values, $\hat{x}^{t+1} = (\hat{x}_1^{t+1}, \dots, \hat{x}_n^{t+1})$, that the variables are predicted to take at the next instant. All the preceding evidence is then moved forward in time. The oldest evidence X^0 is forgotten from the system and $\hat{x}^{t+1}$ is introduced as the new evidence of the last instant to create the new state vector $s' = \left((x_1^1, \dots, x_n^1), \dots, (\hat{x}_1^{t+1}, \dots, \hat{x}_n^{t+1}) \right)$. The current inference step is completed at this stage, while s' is employed as the opening state vector of the next inference step, to predict the next state of the system. To reach the desired horizon T , it is required that one should perform as many inference steps as needed. When the forecasting is completed, the values of the target variables at each instant are returned, and the mean absolute error (MAE) is calculated with the original TS if some test data are provided [49,79].

2.12. Statistical Software Employed

The study will employ R version 4.3.3 software from the Comprehensive R Archive Network (CRAN) project, with its corresponding r packages.

3. Results and Discussions

3.1. DBN Network Structure

The DBN network structure as learned is shown below in Figure 2. It shows the network connection between climate variables and crop yield at the previous time slides. It also indicates a connection from the previous time slide to the current time slides, showing how previous and current time slides influence cereal crop production.

The structure indicates by the direction of the arrows that there is an association within the time slides, where climate variables in the previous slides show causal relationships with cereal crop yield at that time slide

(Cropc_t_1). The side with the blue balls represents the current time slide conditioned on previous time slides. There is a connection that links all the climate variables at previous time slides to the crop yield at the current time slide, including cereal crop yield in previous time slides. There is also an indication of climate variables at the current time slide (intra-slide) having links with crop yield. These links are all represented by directed arcs toward cereal crop yield, which indicate that the variables at the tail have some effect on those at the head. Climate variables at each time slide have an effect on cereal crop yield at the current time slides from our data, according to the model as fit. That is, cereal crop yield is associated with the previous and current climate variables considered.

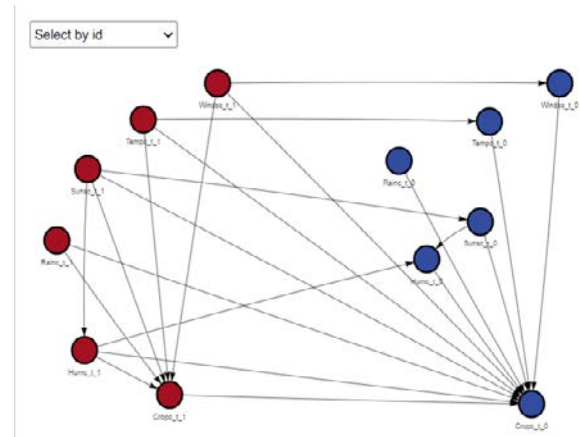


Figure 2. Network Structure of the Dynamic Bayesian Network

3.2. DBN Parameter Learning

Upon learning the structure of the DBN, the parameter learning is done and the estimates are presented in Table 1 below. The table indicates the variables at the previous and current time slides with their corresponding partial correlations and link strengths.

Table 1. Parameter Estimates of the DBN Model

Variable	Cropc_t_0			Cropc_t_1		
	Corr	Odds Ratio	Link Strength	Corr	Odds Ratio	Link Strength
(Intercept)	-0.0084	0.9916		-0.0049	0.9951	
Rainc_t_0	0.0466	1.0477	2.1340	-	-	-
Sunsc_t_0	0.1057	1.1115	1.7478	-	-	-
Tempc_t_0	-0.0175	0.9826	2.2243	-	-	-
Humc_t_0	0.0735	1.0763	2.1610	-	-	-
Windsc_t_0	0.1140	1.1207	1.6842	-	-	-
Cropc_t_1	0.4007	1.4928	-4.6544	-	-	-
Rainc_t_1	0.0542	1.0557	2.0700	0.0107	1.0108	2.1808
Sunsc_t_1	0.0632	1.0653	2.0680	0.1725	1.1882	1.0140
Tempc_t_1	0.0646	1.0667	2.0500	0.0662	1.0684	2.0086
Humc_t_1	0.1201	1.1277	2.0246	0.2329	1.2622	0.1685
Windsc_t_1	-0.2519	0.7773	-0.2162	0.0079	1.0079	2.1834
Res. SD	0.8868	2.4274		0.9746	2.6501	
Model Diagnostics						
logLik = -1184.3200		AIC = -1230.3200		BIC = -1283.0010		

From the Table 1, we observe the values of the various nodes at the different time slides under the partial correlation column which indicates how these nodes can be used as predictors of crop yield. The odds ratio column represents the classical odds ratio (OR), and it is a measure of the relationship between the random variables at the different time slides. It indicates that all the variables from the previous time slide are positively associated with the current crop yield with values greater than one, except windspeed which has a value less than one, and hence negatively associated with the current crop yield. The OR values for the intra-slide relationship at the current time slide indicate that all the climate values have a positive relationship with crop yield at the current time-slide, except temperature (temp_t_0) which has a negative relationship.

The link strength column shows values that represent the measure of the highest amount that evidence at the parent node of a link can effect the belief, given that every one of the other parents of the child has some evidence. It indicates the significance of the link if this value is either greater than or more than 0.5. From Table 1 therefore, the link strength indicates significant connections or influence of the climate variables at the different time slides with crop yield (Croc_t_0) at the current time slide except for crop yield (Croc_t_1) and windspeed (Windsc_t_1) at the previous time slides.

With the intra-slide interaction at the previous time slide, we observe that all the climate variables are positively and significantly associated with the cereal crop yield, except humidity (Humc_t_0) which is positively but not significantly associated with crop yield.

3.3. DBN Model Diagnostics

3.3.1. DBN Stability Checks

Figure 3 represents the DAG indicating the stability of the DBN model fitted. With a ten thousand resampling process, the DAG indicates that the arcs and arc directions remain stable. It indicates that in the face of increased data points and additional variables, the model will still fit well. Thick arcs in the diagram indicate that there is a negative and non-significant relationship.

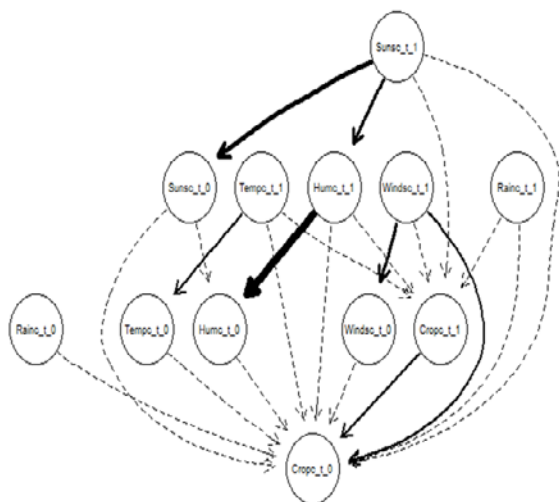


Figure 3. DBN model stability DAG

3.3.2. Forecast/ Prediction Accuracy

The purpose is to forecast the number of cereals in metric tons to be produced for a specified year, since obtaining accurate predictions will help enhance production efficiency decision-making ahead of time before and during cultivation. The model indicates a mean absolute error of 0.0429 which stabilizes towards zero as the number of predicted instances increases. Figure 4 shows predicted values (red) almost converging with true values with time.

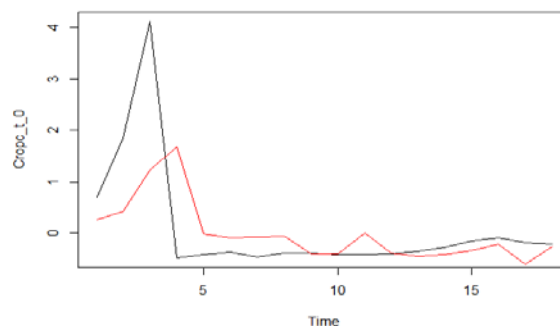


Figure 4. Plot of Prediction (red) from the DBN model

4. Conclusions

The objective of the study is to investigate and model the effect of climate change on cereal crop yield in Northern Ghana, using the Dynamic Bayesian Network model. Using the research data (cereal crop yield and climate variables) for a thirty-one-year period, the Dynamic Bayesian Networks model was fit and model diagnostics were assessed.

The results indicated the existence of intra and inter-slide connections in the model. It was determined that climate variables at each time slide had shown to have some relationship with cereal crop yield at the current time slide. That is, cereal crop yield is associated with the previous and current climate variables. It also revealed that all the climate variables were positively and significantly associated with the cereal crop yield, except humidity (Humc_t_0) which was positively but not significantly associated with crop yield. Model diagnostics processes indicated that the model fits well with the data. The model also has a Mean Absolute Error (MAE) value of 0.0429 which indicates better forecasting accuracy and has the tendency to perform well in understanding or assessing the relationship between cereal crop yield and climate change variables in Northern Ghana.

The model results showed that climate variables are significantly linked with yield at the intra and inter-slide levels. It was also proven to be sensitive enough to capture association even with small data and showed stability to out-of-sample variations. It was further revealed that there exist significantly positive relationships between cereal crop yield and climate change variables. The results imply that, with data on current climate variables, we can predict the next cereal crop yield with reduced uncertainty and better forecasting ability while holding other factors and inputs constant. Again, it was established that any misspecification in rainfall, temperature, and humidity has

a greater potential effect on yield since cereal crop yield has shown to be sensitive to these.

The study therefore recommends the use of this approach in understanding the relationship between cereal crop yield and climate change. Also, climate change impacts should always be taken into consideration and factored into decision-making processes before and during cultivation; since this study has shown that such relationships exist. Besides, since other factors could influence cereal crop yield, it is recommended that the study be extended to include other variables such as vegetative cover, area cultivated, chemical use, etc. when data becomes available.

References

- [1] EPA, "Climate change vulnerability and adaptation assessment of coastal Zone of Ghana. Final report. Accra," 2000.
- [2] UNEP and UNDP, "National Climate Change Adaptation Strategy," *Global Environmental Change*, vol. 5, no. 1, pp. 1–11, 2013, [Online].
- [3] C. Arndt, F. Asante, and J. Thurlow, "Implications of climate change for Ghana's economy," *Sustainability (Switzerland)*, vol. 7, no. 6, pp. 7214–7231, 2015.
- [4] J. Amikuzuno and S. A. Donkoh, "Climate Variability and Yields of Major Staple Food Crops in Northern Ghana," 2012.
- [5] R. Mendelsohn, P. Kurukulasuriya, A. Basist, F. Kogan, and C. Williams, "Climate analysis with satellite versus weather station data," *Clim Change*, vol. 81, no. 1, pp. 71–83, 2007.
- [6] B. Darfour and K. A. Rosentrater, "Maize in Ghana: An overview of cultivation to processing," 2016 American Society of Agricultural and Biological Engineers Annual International Meeting, ASABE 2016, pp. 1–16, 2016.
- [7] MoFA-IFPRI, "Ghana's Maize Market," International Food Policy Research Institute (IFPRI) in Ghana, no. 1, pp. 1–4, 2020.
- [8] USDA/IPAD, "USDA/IPAD Country Summary," 2024. [Online]. Available: <https://ipad.fas.usda.gov/countrysummary/default.aspx?id=GH>
- [9] D. Gage et al., "The Market for Maize, rice, Soy, and Warehouse in Northern Ghana," ASABE, no. January, 2012.
- [10] B. T. Anang and J. Amikuzuno, "Factors Influencing Pesticide Use in Smallholder Rice Production in Northern Ghana," *Agriculture, Forestry and Fisheries*, vol. 4, no. 2, p. 77, 2015.
- [11] S. Hwang and J. Kim, Sustainable development and SDGs. UN and SDGs: A handbook for youth. Yeonsu-gu, Incheon 22004, Republic of Korea: United Nations ESCAP East and North-East Asia Office, 2017. [Online]. Available: https://www.google.com/search?q=un+sdgs+pdf&sca_esv=12e82f26f75a1e30&sxsrf=ADLYWJDu8Q-L7wsF5zo9VpSFS0QISJT2Q%3A1728593700285&ei=JD8IZ_-PEe6rhbIPmLbmKQQ&oq=un+&gs_lp=Egxn3Mtd2l6LXNlcnAiA3VuICoCCAeyCxAAGIAEGJECGloFMgoQABiABBhDGloFMgoQABiABBhDGloFMgsQABiAB
- [12] UNOSD, "The 2030 Agenda for Sustainable Development's 17 Sustainable Development Goals (SDGs)," in 4th SDG Youth Summer Camp, – United Nations Office for Sustainable Development, 2020, p. 19. [Online]. Available: https://www.google.com/search?q=un+sdgs+pdf&sca_esv=12e82f26f75a1e30&sxsrf=ADLYWJDu8Q-L7wsF5zo9VpSFS0QISJT2Q%3A1728593700285&ei=JD8IZ_-PEe6rhbIPmLbmKQQ&oq=un+&gs_lp=Egxn3Mtd2l6LXNlcnAiA3VuICoCCAeyCxAAGIAEGJECGloFMgoQABiABBhDGloFMgoQABiABBhDGloFMgsQABiAB
- [13] UNFCCC, "Climate change science - the status of climate change science today," United Nations Framework Convention on Climate Change, no. February 2011, pp. 1–7, 2011.
- [14] W. K. Dumenu and E. A. Obeng, "Climate change and rural communities in Ghana: Social vulnerability, impacts, adaptations and policy implications," *Environ Sci Policy*, vol. 55, no. February 2018, pp. 208–217, 2016.
- [15] M. Derkyi, S. G. K. Adiku, V. Nelson, B. D. Dovie, S. Codjoe, and E. Awuah, "Smallholder farmers' perception of climatic and socio-economic factors influencing livelihoods in the transition zone of Ghana," *AAS Open Res*, vol. 1, no. 0, p. 7, 2018.
- [16] S. L. y John A. Stanturf, Melvin L. Warren, Jr., Susan Chamley, Sophia C. Polasky and Y. A. N. Goodrick, Frederick Armah, "Ghana Climate Change Vulnerability and Adaptation Assessment," 2011. [Online]. Available: [file:///C:/Users/user/Desktop/Climate/Climate Change Assessment_Ghana_FINAL.pdf](file:///C:/Users/user/Desktop/Climate/Climate%20Change%20Assessment_Ghana_FINAL.pdf)
- [17] J. A. Yaro, "The perception of and adaptation to climate variability/change in Ghana by small-scale and commercial farmers," *Reg Environ Change*, vol. 13, no. 6, pp. 1259–1272, 2013.
- [18] EPA, "Ghana ' s Third Biennial Update Report to United Nations Framework Convention on Climate Change," 2021.
- [19] B. Y. Fosu-Mensah, P. L. G. Vlek, and D. S. MacCarthy, "Farmers' perception and adaptation to climate change: A case study of Sekyedumase district in Ghana," *Environ Dev Sustain*, vol. 14, no. 4, pp. 495–505, 2012.
- [20] R. F. Roper, R. Rumi, and P. A. Aguilera, "Bayesian networks for evaluating climate change influence in olive crops in Andalusia, Spain," *Nat Resour Model*, vol. 32, no. 1, pp. 1–18, 2019.
- [21] P. A. Aguilera, A. Fernández, R. F. Roper, and L. Molina, "Groundwater quality assessment using data clustering based on hybrid Bayesian networks," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 2, pp. 435–447, 2013.
- [22] P. A. Aguilera, A. Fernández, R. F. Roper, and L. Molina, "Groundwater quality assessment using data clustering based on hybrid Bayesian networks," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 2, pp. 435–447, 2013.
- [23] R. F. Roper, R. Rumi, and P. A. Aguilera, "Bayesian networks for evaluating climate change influence in olive crops in Andalusia, Spain," *Nat Resour Model*, vol. 32, no. 1, pp. 1–18, 2019.
- [24] D. Koller and N. Friedman, *Probabilistic Graphical Models Principles and Techniques*. Massachusetts Institute of Technology, 2009.
- [25] P. Shiguihara, A. De Andrade Lopes, and D. Mauricio, "Dynamic Bayesian Network Modeling, Learning, and Inference: A Survey," *IEEE Access*, vol. 9, no. i, pp. 117639–117648, 2021.
- [26] K. P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning," University of California, Berkeley, 2002.
- [27] J. C. Rajapakse and J. Zhou, "Learning effective brain connectivity with dynamic Bayesian networks," *Neuroimage*, vol. 37, no. 3, pp. 749–760, 2007.
- [28] T. Charitos, P. R. de Waal, and L. C. van der Gaag, "Convergence in Markovian models with implications for efficiency of inference," *International Journal of Approximate Reasoning*, vol. 46, no. 2, pp. 300–319, 2007.
- [29] H. Lähdesmäki and I. Shmulevich, "Learning the structure of dynamic Bayesian networks from time series and steady state measurements," *Mach Learn*, vol. 71, no. 2–3, pp. 185–217, 2008.
- [30] P. Hernandez-Leal, J. A. Gonzalez, E. F. Morales, and L. Enrique Sucar, "Learning temporal nodes Bayesian networks," *International Journal of Approximate Reasoning*, vol. 54, no. 8, pp. 956–977, 2013.
- [31] T. Charitos, L. C. van der Gaag, S. Visscher, K. A. M. Schurink, and P. J. F. Lucas, "A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients," *Expert Syst Appl*, vol. 36, no. 2 PART 1, pp. 1249–1258, 2009.
- [32] T. Charitos, P. R. de Waal, and L. C. van der Gaag, "Convergence in Markovian models with implications for efficiency of inference," *International Journal of Approximate Reasoning*, vol. 46, no. 2, pp. 300–319, 2007.
- [33] Y. Zhang, Y. Zhang, E. Swears, N. Larios, Z. Wang, and Q. Ji, "Modeling temporal interactions with interval temporal bayesian networks for complex activity recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 10, pp. 2468–2483, 2013.
- [34] X. G. Gao, J. F. Mei, H. Y. Chen, and D. Q. Chen, "Approximate inference for dynamic Bayesian networks: Sliding window approach," *Applied Intelligence*, vol. 40, no. 4, pp. 575–591, 2014.
- [35] MoFA, "Cereal Production Figures from 1992," 2020.
- [36] R. M. Kranjac-Berisavljevic, G., Blench, A. S. Bayorbor, T. B., Turton, C. N., Abdulai, and E. Boyd, C., Obeng, F. and Drake, "Rethinking natural resource degradation in semi-arid subSaharan Africa: The case of semi-arid Ghana. London, Overseas

- Development Institute. http://www.odi.org.uk/rpeg/soil_degradation/ghlit.pdf,” 1999.
- [37] P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón, “Bayesian networks in environmental modelling,” *Environmental Modelling and Software*, vol. 26, no. 12, pp. 1376–1388, 2011.
- [38] J. Pearl, “Uncertainty in Ai Systems,” *Probabilistic Reasoning in Intelligent Systems*, pp. 1–28, 1988.
- [39] R. F. Ropero, R. Rumí, and P. A. Aguilera, “Bayesian networks for evaluating climate change influence in olive crops in Andalusia, Spain,” 2019.
- [40] A. Fernández and A. Salmerón, “Extension of Bayesian network classifiers to regression problems,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5290 LNAI, pp. 83–92, 2008.
- [41] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,” *Mach Learn*, 1997.
- [42] S. Moral, R. Rumí, and A. Salmerón, “Mixtures of truncated exponentials in hybrid Bayesian networks,” *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, vol. 2143, pp. 156–167, 2001.
- [43] R. V. Aguilera, K. Desender, M. K. Bednar, and J. H. Lee, “Connecting the Dots: Bringing External Corporate Governance into the Corporate Governance Puzzle,” *Academy of Management Annals*, vol. 9, no. 1, pp. 483–573, 2015.
- [44] H. Langseth, T. D. Nielsen, I. Pérez-Bernabé, and A. Salmerón, “Learning mixtures of truncated basis functions from data,” 2014.
- [45] S. L. Lauritzen, “Propagation of probabilities, means, and variances in mixed graphical association models,” *J Am Stat Assoc*, vol. 87, no. 420, pp. 1098–1108, 1992.
- [46] G. Kratzer, F. Lewis, A. Comin, M. Pittavino, and R. Furrer, “Additive Bayesian Network Modeling with the R Package abn,” *J Stat Softw*, vol. 105, no. 8, 2023.
- [47] F. I. Lewis and B. J. J. McCormick, “Revealing the complexity of health determinants in resource-poor settings,” *Am J Epidemiol*, vol. 176, no. 11, pp. 1051–1059, 2012.
- [48] M. Pittavino et al., “Comparison between generalized linear modelling and additive Bayesian network; identification of factors associated with the incidence of antibodies against *Leptospira interrogans* sv Pomona in meat workers in New Zealand,” *Acta Trop*, vol. 173, no. November 2016, pp. 191–199, 2017.
- [49] D. Quesada, C. Bielza, and P. Larranaga, “Structure Learning of High-Order Dynamic Bayesian Networks via Particle Swarm Optimization with Order Invariant Encoding Structure Learning of High-Order Dynamic Bayesian Networks via Particle Swarm Optimization with Order Invariant Encoding,” *Research Gate*, no. September, 2021.
- [50] D. Koller and N. Friedman, *Probabilistic Graphical Models Principles and Techniques*. Massachusetts Institute of Technology, 2009.
- [51] D. Quesada, C. Bielza, P. Fontán, and P. Larrañaga, “Piecewise forecasting of nonlinear time series with model tree dynamic Bayesian networks,” *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 9108–9137, 2022.
- [52] A. Tucker, F. Höppner, A. Siebes, and S. Swift, “Dynamic MMHC: A Local Search Algorithm for Dynamic Bayesian Network Structure Learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8207 LNCS, no. May 2023, 2013.
- [53] F. P. Santos and C. Dias Maciel, “A PSO approach for learning transition structures of Higher-Order Dynamic Bayesian Networks,” *ISSNIP Biosignals and Biorobotics Conference, BRC*, 2014.
- [54] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, “The max-min hill-climbing Bayesian network structure learning algorithm,” *Mach Learn*, vol. 65, no. 1, pp. 31–78, 2006.
- [55] D. Quesada, C. Bielza, and P. Larranaga, “Structure Learning of High-Order Dynamic Bayesian Networks via Particle Swarm Optimization with Order Invariant Encoding Structure Learning of High-Order Dynamic Bayesian Networks via Particle Swarm Optimization with Order Invariant Encoding,” *Research Gate*, no. September, 2021.
- [56] G. Schwarz, “Estimating the Dimension of a Model,” *Institute of Mathematical Statistics Stable*, vol. 6, no. 2, pp. 461–464, 1978.
- [57] D. Quesada and P. Larrañaga, “dbnR: Gaussian Dynamic Bayesian Network Learning and Inference in R,” *J Stat Softw*, vol. VV, no. Ii, 2023.
- [58] G. Schwarz, “Estimating the Dimension of a Model,” *Institute of Mathematical Statistics Stable*, vol. 6, no. 2, pp. 461–464, 1978.
- [59] K. P. Murphy, “Machine Learning - A Probabilistic Perspective,” *The MIT Press*, p. 1049, 2012.
- [60] D. Quesada, G. Valverde, P. Larrañaga, and C. Bielza, “Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks,” *Eng Appl Artif Intell*, vol. 103, no. May, p. 104301, 2021.
- [61] G. Briganti, M. Scutari, and R. J. McNally, “A Tutorial on Bayesian Networks for Psychopathology Researchers,” *Psychol Methods*, no. December, 2022.
- [62] S. Epskamp and E. I. Fried, “A tutorial on regularized partial correlation networks,” *Psychol Methods*, vol. 23, no. 4, pp. 617–634, 2018.
- [63] G. Briganti, M. Scutari, and P. Linkowski, “A machine learning approach to relationships among alexithymia components,” *Psychiatr Danub*, vol. 32, no. September, pp. S180–S187, 2020.
- [64] G. Briganti and P. Linkowski, “Exploring network structure and central items of the Narcissistic Personality Inventory,” *Int J Methods Psychiatr Res*, vol. 29, no. 1, pp. 1–7, 2020.
- [65] G. Briganti, M. Scutari, and P. Linkowski, “Network Structures of Symptoms From the Zung Depression Scale,” *Psychol Rep*, vol. 124, no. 4, pp. 1897–1911, 2021.
- [66] J. L. Puga, M. Krzywinski, and N. Altman, “Points of Significance: Bayesian networks,” *Nature Methods*, vol. 12, no. 9, pp. 799–800, 2015.
- [67] S. Baraty and D. A. Simovici, “Edge evaluation in Bayesian network structures,” *Conferences in Research and Practice in Information Technology Series*, vol. 101, pp. 193–202, 2009.
- [68] M. Scutari and J.-B. Denis, *Bayesian Networks with Examples in R*, vol. 12, no. 4. Taylor and Francis Group; Chapman and Hall/CRC, 2015.
- [69] M. Scutari and R. Nagarajan, “Identifying significant edges in graphical models of molecular networks,” *Artificial Intelligence in Medicine*, vol. 57, no. 3, pp. 207–217, 2013.
- [70] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging*. Cambridge University Press, 2008.
- [71] N. Friedman, M. Goldszmidt, and A. Wyner, “Data Analysis with Bayesian Networks: A Bootstrap Approach,” pp. 196–205, 2013.
- [72] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CHAPMAN & HALL/CRC, 1993.
- [73] M. Scutari and J.-B. Denis, *Bayesian Networks with Examples in R*, vol. 12, no. 4. Taylor and Francis Group; Chapman and Hall/CRC, 2015.
- [74] M. Scutari and R. Nagarajan, “Identifying significant edges in graphical models of molecular networks,” *Artif Intell Med*, vol. 57, no. 3, pp. 207–217, 2013.
- [75] M. Scutari and R. Nagarajan, “Identifying significant edges in graphical models of molecular networks,” *Artif Intell Med*, vol. 57, no. 3, pp. 207–217, 2013.
- [76] R. Nagarajan, M. Scutari, and S. Lèbre, *Bayesian Networks in R, with Applications in Systems Biology*. 2013.
- [77] B. Boerlage, “Link strength in Bayesian networks,” no. October, 1992.
- [78] J. N. Kok and SpringerLink (Online service), “Machine learning ECML 2007 : 18th European Conference on Machine Learning, Warsaw, Poland, September 17-21, 2007 : proceedings,” 2007.
- [79] D. Quesada and P. Larrañaga, “dbnR: Gaussian Dynamic Bayesian Network Learning and Inference in R,” *J Stat Softw*, vol. VV, no. Ii, 2023.

