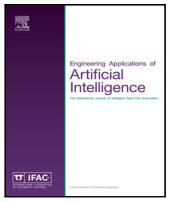
Engineering Applications of Artificial Intelligence 131 (2024) 107867

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Learning Bayesian networks with heterogeneous agronomic data sets via mixed-effect models and hierarchical clustering

Lorenzo Valleggi [a], Marco Scutari [b,*], Federico Mattia Stefanini [c]

[a] Department of Statistics, Computer Science and Applications, University of Florence, Florence, Italy
[b] Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland
[c] Department of Environmental Science and Policy, University of Milan, Milan, Italy

ARTICLE INFO

Dataset link: https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/IASSTN

Keywords:
Hierarchical data sets
Bayesian networks
Causal networks
Structure learning
Prediction of maize yield

ABSTRACT

Maize, a crucial crop globally cultivated across vast regions, especially in sub-Saharan Africa, Asia, and Latin America, occupies 197 million hectares as of 2021. Various statistical and machine learning models, including mixed-effect models, random coefficients models, random forests, and deep learning architectures, have been devised to predict maize yield. These models consider factors such as genotype, environment, genotype-environment interaction, and field management. However, the existing models often fall short of fully exploiting the complex network of causal relationships among these factors and the hierarchical structure inherent in agronomic data. This study introduces an innovative approach integrating random effects into Bayesian networks (BNs), leveraging their capacity to model causal and probabilistic relationships through directed acyclic graphs. Rooted in the linear mixed-effects models framework and tailored for hierarchical data, this novel approach demonstrates enhanced BN learning. Application to a real-world agronomic trial produces a model with improved interpretability, unveiling new causal connections. Notably, the proposed method significantly reduces the error rate in maize yield prediction from 28% to 17%. These results advocate for the preference of BNs in constructing practical decision support tools for hierarchical agronomic data, facilitating causal inference.

1. Introduction

The global economy relies on agriculture as a vital source of income and employment as well as food, ensuring food quality and safety, environmental preservation, fostering comprehensive rural development, and upholding social cohesion in rural areas. Given the projected global population growth, which is expected to reach 9.7 billion by 2050 (Pew Research Center, 2019), it is estimated that global agricultural production must increase by 60% (Alexandratos and Bruinsma, 2012) to meet the increase in demand. With these premises, improving crop management systems is essential to match future needs. Maize is one of the most widely cultivated crops in sub-Saharan Africa, Asia and Latin America, with a total area of 197 M ha (FAO et al., 2021). It provides almost all the caloric intake in the Americas (285 kcal/capita/day) and in Africa (374 kcal/capita/day; FAOSTAT, 2019). Predicting the grain yield of this cultivar provides valuable information about the expected crop output before harvest, enabling more effective management practices. To achieve accurate predictions, it is essential to consider the interplay between genotype, environment, and field management. Widely adopted statistical models for this task include linear mixed-effect and random coefficient models that use genome-wide association study (GWAS) to study the causal effects of genotype, environmental variables and their interactions (Zorić et al., 2022; Ndlovu et al., 2022; Tolley et al., 2023; Rotili et al., 2020).

More recently, machine learning models such as random forests (Yang et al., 2022; Leroux et al., 2019), naive Bayes and SVM (Mupangwa et al., 2020) have been applied to maize crop yield prediction using multi-temporal UAV remote sensing data. Deep learning architectures such as Long Short-Term Memory (LTSM; Zhang et al., 2021; Krishna et al., 2023) and Convolutional Neural Network (CNN; Yang et al., 2021) have also been explored. Despite their predictive performance, which rests on their ability to encode complex non-linear relationships, these models are not causal. Outside of randomized experiments, they are particularly vulnerable to confounding (Pearl, 2009), that is, learning spurious associations as causal relationships due to unobserved variables acting as common causes of treatment and outcome and or due to selection bias. They also often disregard the hierarchical structure that is typical of the data collected in agronomic studies, which is highly informative.

* Corresponding author.
E-mail address: journals@bnlearn.com (M. Scutari).

https://doi.org/10.1016/j.engappai.2024.107867
Received 11 August 2023; Received in revised form 9 December 2023; Accepted 6 January 2024

In general, studies encompassing heterogeneous collections of related data sets (RDs) in which the relationships between the covariates and the outcome of interest may differ (say, in slope or variance; Gelman and Hill, 2007) are widespread in many fields, from clinical trials to environmental science (Spiegelhalter et al., 2004; Qian et al., 2010). Hierarchical (multilevel) models are commonly adopted to pool information across different subsets of the data while accounting for their specific features (Gelman et al., 2014). However, heterogeneity is not the only challenge in fitting a model on such data: the variables involved are typically related by a complex network of causal relationships, making their joint distribution challenging to learn (especially) from small data sets.

In this work, we chose to learn a Bayesian Network (BN) from RDs focused on the agronomic performance of maize. BNs can be learned and used as *causal network models* whose arcs represent cause–effect relationships and which can be used for causal inference following the work of Judea Pearl (Pearl, 2009). In the case of RDs, learning BNs is also related to *transfer learning* (Pan and Yang, 2010), which is not widely documented in the literature. Transfer learning has mainly focused on applications involving deep learning, with very few publications involving BNs. Notably, recent work by Yan et al. (2023) proposed a structure learning approach based on conditional independence tests for operational adjustments in a flotation process characterized by a small data set with a limited sample size. To induce transfer learning, they considered the results of the independence tests performed on variables X_i and X_j in both the source and target data sets, which differed in terms of sample size. Other authors have suggested using order-search algorithms to learn BN structures, introducing a structural bias term to facilitate the transfer of information between data sets and achieve more robust networks (Oyen and Lane, 2015). BNs and structural equation models have proven successful in the agronomic sector, optimizing various management practices such as phytosanitary treatments (Lu et al., 2020), irrigation management strategies (Ilić et al., 2022) and soil management (Hill et al., 2017), to minimize environmental impact and mitigate climate change. However, in the agronomic literature, transfer learning has predominantly focused on crop disease classification using deep learning techniques like convolutional neural networks (Coulibaly et al., 2019; Paymode and Malode, 2022), with little research involving BNs. A thorough exploration of the literature reveals various statistical methods for predicting maize grain yield, summarised in Table 1. We limited ourselves to statistical approaches in keeping with our focus on modelling frameworks that support causal reasoning.

We learned the structure and the parameters of a Conditional Gaussian Bayesian Network (CGBN) from a real-world agronomic data set with a hierarchical structure. To account for the high heterogeneity that characterizes such data, we developed a novel approach that integrates random effects into the local distributions in the BN, building on Scutari et al. (2022). Random effects are the salient feature of *linear mixed-effects models* (LME; Pinheiro and Bates, 2000). LME models are hierarchical models that extend the classical linear regression model by adding a second set of coefficients called "random effects", which are jointly distributed as a multivariate normal. The other coefficients are called "fixed effects". The coefficients associated with the random effects have mean zero, and they naturally represent the deviations of the effects of the parents in individual data sets from their average effects across data sets, represented by the fixed effects.

The hierarchical estimation in BNs learned from RDs was initially introduced by Azzimonti et al. (2019), who proposed a novel approach to tackle this challenge for discrete BNs using a hierarchical multinomial-Dirichlet model. That approach outperforms a traditional multinomial-Dirichlet model and is competitive with random forests, but as the number of domains increases, the estimation becomes more complex, necessitating the use of approximations such as variational or Markov chain Monte Carlo inference.

The remainder of the paper is structured as follows. In Section 2, we briefly describe the data set (Section 2.1), we introduce the background of BN (Section 2.2), we introduce the local distributions and the structure learning approach used to learn the BN (Section 2.3) and we how we evaluated its performance (Section 2.4). In Section 3, we present and evaluate the BN, and in Section 4, we discuss its performance before suggesting possible future research directions.

2. Materials and methods

2.1. Background of Bayesian network

Bayesian networks (BNs; Koller and Friedman, 2009) provide a powerful tool to learn and model highly structured relationships between variables. A BN is a graphical model defined on a set of random variables $\mathbf{X} = \{X_1, \ldots, X_K\}$ and a directed acyclic graph (DAG) $\mathcal{G}$ that describes their relationships: nodes correspond to random variables, and the absence of arcs between them implies the conditional independence or the lack of direct causal effects (Pearl, 2009). In particular, a variable X_i is independent of all other non-parent variables in $\mathcal{G}$ given the set of variables associated with its parents $pa(X_i)$. A DAG $\mathcal{G}$ then induces the following factorization:

$$P(\mathbf{X} \mid \mathcal{G}, \Theta) = \prod_{i=1}^{K} P(X_i \mid pa(X_i), \Theta_{X_i}), \tag{1}$$

where Θ_{X_i} are the parameters of the conditional distribution of $X_i \mid pa(X_i)$. In Eq. (1), the *joint multivariate distribution* of $\mathbf{X}$ is reduced to a collection of univariate conditional probability distributions, the *local distributions* of the individual nodes X_i. If all sets $pa(X_i)$ are small, (1) is very effective in replacing the high-dimensional estimation of Θ with a collection of low-dimensional estimation problems for the individual Θ_{X_i}. Another consequence of (1) is the existence of the *Markov blanket* of each node X_i, the set of nodes that makes X_i conditionally independent from the rest of the BN. It comprises the parents, the children and the spouses of X_i, and includes all the knowledge needed to do inference on X_i, from estimation to hypothesis testing to prediction.

The process of learning a BN from data can be divided into two steps:

$$\underbrace{P(\mathcal{G}, \Theta \mid D)}_{\text{BN learning}} = \underbrace{P(\mathcal{G} \mid D)}_{\text{structure learning}} \cdot \underbrace{P(\Theta \mid \mathcal{G}, D)}_{\text{parameter learning}}.$$

Structure learning aims to find the dependence structure represented by the DAG given the data D. Several algorithms are described in the literature for this task. Constraint-based algorithms such as the PC algorithm (Spirtes et al., 2000) use a sequence of independence tests with increasingly large conditioning sets to find which pairs of variables should be connected by an arc (or not), and then they identify arc directions based on the difference in conditional independence patterns between v-structures (of the form $X_j \rightarrow X_i \leftarrow X_k$, with no arc between X_j and X_k) and other patterns of arcs. Score-based algorithms instead use heuristics (like hill climbing; Russell and Norvig, 2009) or exact methods (as in Cussens, 2012) that optimize a network score reflecting the goodness of fit of candidate DAGs to select an optimal one. *Parameter learning* provides an estimate of Θ through the parameters in the Θ_{X_i} conditional to the learned DAG.

Structure learning algorithms are distribution-agnostic, but the choice of the conditional independence tests and the network scores depend on the types of distributions we assume for the X_i. The three most common choices are *discrete BNs*, in which the X_i are multinomial random variables; *Gaussian BNs* (GBNs), in which the X_i are univariate normal random variables linked by linear dependence relationships; and *conditional Gaussian BNs* (CGBNs), in the X_i are either multinomial random variables (if discrete) or mixtures of normal random variables (if continuous). Common scores for all these choices are the Bayesian

Table 1
Statistical methods used for maize grain yield prediction from 2018 to 2023 in the literature. We reported the data structure, the variables and the method used, and whether the method is causal. Abbreviations used: greenness index (GI), modified simple ratio (MSR), normalized difference vegetation index (NDVI), spectral polygon vegetation index (SPVI), ratio vegetation index (RVI), chlorophyll index (CInir), soil-adjusted vegetation index (SAVI), triangular vegetation index (TVI), enhanced vegetation index (EVI), wide dynamic range vegetation index (WDRVI), temperature, hours of sunshine (RAD), Rainfall, Standardized Precipitation Evapotranspiration index (SPEI), Peters and Clark Momentary Conditional Independence (PCMCI), Linear mixed model (LMM), Linear regression (LM), Hierarchical linear model (HLM), Multiple linear regression (MLR), Principal component analysis (PCA), Polynomial regression model (PRM), Generalized linear model (GLM), and Bayesian network (BN).

Data structure	Variables	Method	Causal	Multi-response	Reference
Period: 2001–2019 N° locations: 5 N° trials: 130	FAO maturity groups Precipitation, Air temperature Solar radiation , Stress degree days	LMM	No	No	Zorić et al. (2022)
Period: 2011–2015 N° locations: 3 N° trials: 13	Soil condition, Genotype data	LMM, PCA	No	No	Ndlovu et al. (2022)
Period: 2014–2017 N° locations: 108 N° plots: 59,416	PAR, Temperature, Humidity, Surface pressure Wind speed, Precipitation Soil characteristics, Genotypic data	LMM,	No	No	Tolley et al. (2023)
Period: 2014–2016 N° locations: 9	Soil type, Sowing data, Hybrids Plant density, Row configuration	LMM	No	No	Rotili et al. (2020)
Period: 2016–2017 N° locations: 3 N° fields: 18	NDVI, NDVIG, NDVIre	LM, GLM	No	No	Schwalbert et al. (2018)
Period: 2000–2016	Temperature, Precipitation Vapour pressure, Shortwave radiant flux Soil water content, NDVI	MLR	No	No	Kern et al. (2018)
Period: 1981–2016 N° locations: 12 N° counties: 1051	Vapour pressure, Temperature Precipitation, EVI	PRM	No	No	Li et al. (2019)
Period: 2014–2018 N° locations: 23 N° fields: 94,000	Hybrids, Plant height, Tassel height	LMM	No	No	Anderson II et al. (2019)
Period: 1986/87, 2015/16	Temperature, Precipitation Heat Magnitude Day, SPEI	PCMCI	Yes	Yes	Simanjuntak et al. (2023)
Period: 2016–2019 N° locations: 10	GI, MSR, NDVI, SPVI, RVI, CInir, SAVI, TVI, EVI, WDRVI, Temperature, RAD, Rainfall	HLM	No	No	Zhu et al. (2021)
Period: 2000–2018 N° locations: 9 N° fields: 11	Extreme degree days, Growing degree days, Precipitation	Causal forest	Yes	Yes	Kluger et al. (2022)
Period: 3-year	Temperature, Precipitation	BN	Yes	Yes	Chamorro et al. (2023)

information criterion (BIC; Schwarz, 1978) or the marginal likelihood of $\mathcal{G}$ given $\mathcal{D}$ (Heckerman and Geiger, 1995). As for the conditional independence tests, we refer the reader to Edwards (2000), which covers various options for all types of BNs.

Parameter learning uses maximum-likelihood estimates or Bayesian posterior estimates with non-informative priors for all types of BNs (Koller and Friedman, 2009). All the conditional independence tests, the network scores and the parameter estimators in the literature referenced above can be computed efficiently thanks to (1) because they factorise following the local distributions.

2.2. The data set: Agronomic performance of maize

This study uses the data from Millet et al. (2019b), whose authors are well-known in plant science research. They conducted a *randomised* genome-wide association study to assess the genetic variability of plant performance under different year-to-year and site-to-site climatic conditions. The original analysis of these data in Millet et al. (2016) confirms the quality of this experimental design in terms of controlling both confounding and various sources of noise. Overall, 29 field experiments were arranged in Europe, nine sites, and in Chile, one site: each of

them was defined by a combination of year, site and water regime (watered or rain-fed), and 244 *varieties* of maize (*Zea mays L.)* were studied. Each experiment was designed as alpha-lattice design (Patterson and Williams, 1976), with two replicates of the watered regime and three for the rain-fed regime. The data were collected at experimental sites in France, Germany, Italy, Hungary, Romania, and Chile between 2011 and 2013. After filtering out incomplete observations, the study analysed eight *sites*, each with a different sample size: Gaillac (France, $n = 2437$), Nerac (France, $n = 1716$), Karlsruhe (Germany, $n = 2626$), Campagnola (Italy, $n = 1260$), Debrecen (Hungary, $n = 2181$), Martonvasar (Hungary, $n = 1260$), Craiova (Romania, $n = 1055$), and Graneris (Chile, $n = 760$). Many weather variables were measured for each site, such as air temperature, relative humidity (RH), wind speed and light; they were measured every hour at 2 m height. Soil water potential was measured daily at 30, 60, and 90 cm depths. For this analysis, we decided to use only air temperature and RH because they are the more basic variables that can describe plant growth. Since weather variables were measured for each site instead of for each plot, we decided to aggregate the weather data in order to have the *average temperature* (°C), the *diurnal temperature range*, the *average relative humidity* (%) and the *diurnal relative humidity range* (%) for each site and year for three different periods, which correspond to the main phenological stages of maize: seeding, germination, and emergence of the seeds, the vegetative phase, where leaves emerge (May to June), the flower development, pollen shedding, grain development (July to August), maturation of the grain, and harvest (September to October). Furthermore, random noise with a mean of 0 and a standard deviation of 0.1 was added to each weather observation to simulate the sensor's measurement error and avoid blocks of identical measurements in each individual site.

At the end of the experiment, the phenological variables listed below were measured for each plot at each site:

- The *grain yield* adjusted at 15% grain moisture, in ton per hectare (t/ha).
- The *grain weight* of individual grains (mg).
- The *anthesis*, male flowering (pollen shed), in thermal time cumulated since emergence (d20 °C).
- The *sinking*, female flowering (silking emergence), in thermal time cumulated since emergence (d20 °C).
- The *plant height* from ground level to the base of the flag leaf (highest) leaf (cm).
- The *tassel height*, plant height including tassel, from ground level to the highest point of the tassel (cm).
- The *ear height*, ear insertion height, from ground level to the ligule of the highest ear leaf (cm).

2.3. Learning algorithm

We learned the structure of the BN, denoted $\mathcal{B}_{LME}$, following the steps in Algorithm 1.

For the hill-climbing algorithm, we used the implementation in the *bnlearn* R package (Scutari, 2010) and the BIC score. We provided a list of arcs to be excluded (*blacklist*) or included (*whitelist*) by hill-climbing to avoid evaluating unrealistic relationships (such as the Average temperature of July–Aug → Average temperature of May–June).

Firstly, we regressed the grain yield against all the available variables for all combinations of site and variety. We used the residuals' mean and variance from the regression for each combination of site and variety to cluster them using the agglomerative Ward clustering algorithm (Murtagh and Legendre, 2014) from the *stats* R package. The resulting discrete variable was added to the data used to identify the RDs.

Algorithm 1: Structure learning $\mathcal{B}_{LME}$.

Data: data set $\mathcal{D}$, *blacklist*, and a *whitelist*
Result: The DAG $\mathcal{G}_{max}$ that maximises BIC($\mathcal{G}_{max}$,$\mathcal{D}$).

1. Run a linear regression on grain yield and extract the residuals ϵ_i.
2. For each Site × Variety combination, compute the mean and the standard deviation of ϵ_i.
3. Perform hierarchical clustering on the means and standard deviations of the residuals from each site-variety combination.
4. Add a new variable with the cluster labels to $\mathcal{D}$.
5. Compute the score of $\mathcal{G}$, $\mathcal{S}_{\mathcal{G}} = \text{BIC}(\mathcal{G}, \mathcal{D})$ and set $\mathcal{S}_{max} = \mathcal{S}_{\mathcal{G}}$ and $\mathcal{G}_{max} = \mathcal{G}$.
6. *Hill-climbing*: repeat as long as $\mathcal{S}_{max}$ increase:
   (a) Add, delete or reverse all possible arc in $\mathcal{G}_{max}$ resulting in a DAG.
      i. compute BIC of the modified DAG $\mathcal{G}^*$, $\mathcal{S}_{\mathcal{G}^*} = \text{BIC}(\mathcal{G}^*, \mathcal{D})$;
      ii. if $\mathcal{S}_{\mathcal{G}^*} > \mathcal{S}_{max}$ and $\mathcal{S}_{\mathcal{G}^*} > \mathcal{S}$ set $\mathcal{G} = \mathcal{G}_*$ and $\mathcal{S}_{\mathcal{G}} = \mathcal{S}_{\mathcal{G}_*}$.
   (b) Update $\mathcal{S}_{max}$ with the new value of $\mathcal{S}_{\mathcal{G}_*}$.
7. Return the DAG $\mathcal{G}$.

Following the approach and the notation described in Scutari et al. (2022), we assumed that each RD is generated by a GBN and that all GBNs share a common underlying network structure but different parameter values. To ensure the partial pooling of information between RDs, the clusters are made a common parent for all phenological variables and incorporated into the local distributions as a random effect. Therefore, we modelled the local distributions for those variables as a linear mixed-effect model using the *lme*4 R package (Bates et al., 2015):

$$
\begin{aligned}
&X_{i,j} = (\mu_{i,j} + b_{i,j,0}) + \boldsymbol{\Pi}_{X_i}(\beta_i + b_{i,j}) + \epsilon_{i,j}, \\
&\begin{pmatrix} b_{i,j,0} \\ b_{i,j} \end{pmatrix} \sim N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}_i), \\
&(\epsilon_{i,1}, \ldots, \epsilon_{i,j}, \ldots)^T \sim N(\mathbf{0}, \sigma_i^2 \boldsymbol{I}_{nj})
\end{aligned}
\tag{2}
$$

where bold letters indicate matrices. The only exception was grain yield because it also required a model for variances, which have been implemented using *nlme* R package (Heisterkamp et al., 2017) as follows:

$$
\begin{aligned}
&X_{i,j} = (\mu_{i,j} + b_{i,j,0}) + \boldsymbol{\Pi}_{X_i}\beta_i + \epsilon_{i,j}, \\
&b_{i,j,0} \sim N(0, \sigma^2_{b,i}), \\
&N(0, (\sigma^2_{i,1}, \sigma^2_{i,2}, \ldots, \sigma^2_{i,j}, \ldots)\boldsymbol{I}_{nj}), \\
&(\epsilon_{i,1}, \ldots, \epsilon_{i,j}, \ldots)^T \sim N(\mathbf{0}, (\sigma^2_{i,1}, \ldots, \sigma^2_{i,j}, \ldots)\boldsymbol{I}_{nj}), \\
&\sigma^2_{i,j}(v) = |v|^{2\theta_j}.
\end{aligned}
\tag{3}
$$

In both (2) and (3), the notation is as follows:

- $j = 1, \ldots, J$ are the clusters identifying the RDs;
- $\boldsymbol{\Pi}_{X_i}$ is the design matrix associated to the parents of X_i;
- $b_{i,j,0}$ is the random intercept;
- $b_{i,j}$ is the random slope parameter for the jth cluster;
- $\widetilde{\boldsymbol{\Sigma}}_i$ is the $n_j \times n_j$ block of $\boldsymbol{\Sigma}_i$ associated with the jth cluster;
- $\sigma^2_{i,j}\boldsymbol{I}_{nj}$ is the $n_j \times n_j$ matrix arising from the assumption that residuals are homoscedastic in (2);

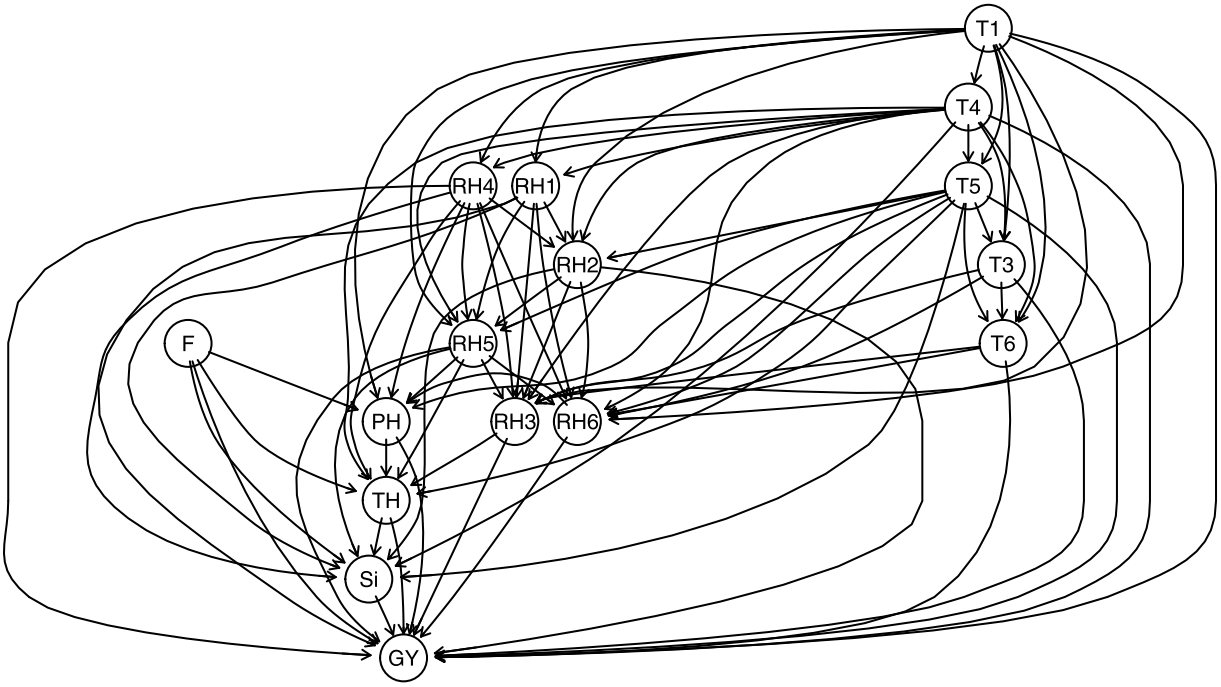


Fig. 1. Structure of the BN $\mathcal{B}_{LME}$ learned with Algorithm 1. The nodes represent: the average temperature May–June (T1), the average temperature Sept–Oct (T3), the diurnal temperature range May–June (T4), the diurnal temperature range July–Aug (T5), the diurnal temperature range Sept–Oct (T6), the average RH May–June (RH1), the average RH July–Aug (RH2), the average RH Sept–Oct (RH3), the diurnal RH range May–June (RH4), the diurnal RH range July–Aug (RH5), the diurnal RH range Sept–Oct (RH6), Silking (Si), TH (Tassel height), PH (Plant height), EH (Ear height) and F (Clusters).

- $\mu_{i,j}$ is the intercept;
- and β_i are the fixed effects.

In (3), we assumed the variance of residuals to be heteroscedastic and following a power function, where v is the variance covariate and θ_j is the variance function coefficient that changes for every level of the common discrete parent.

We modelled the weather variables using only fixed effects for simplicity:

$$X_i = \mu_i + \boldsymbol{\Pi}_{X_i}\beta_i + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_i^2\ \boldsymbol{I}_n). \tag{4}$$

We prevented the clusters from being their parent with the blacklist because the resulting arcs are not of interest from an agronomic perspective.

From these assumptions, the BN $\mathcal{B}_{LME}$ we learned from the data has a global distribution that is a mixture of multivariate normal distributions like a CGBN.

2.4. Predictive and imputation accuracy

The most important variable in this analysis was grain yield because it is one of the key quantities used to evaluate an agronomic season. To assess the predictive ability of $\mathcal{B}_{LME}$, we evaluated the Mean Absolute Percentage Error (MAPE) of:

- the *predictive accuracy* of grain yield predictions in the scenarios listed in Table A.2, which are meant to study the potential effect of measuring a reduced set of variables in future years;
- the *imputation accuracy* for the grain yield in each site–variety combination, which we removed in turn and imputed from the rest.

As a term of comparison, we used a CGBN learned from the data ($\mathcal{B}_{CGBN}$) and compared its performance with that of $\mathcal{B}_{LME}$. We implemented both prediction and imputation using likelihood weighting (Koller and Friedman, 2009; Darwiche, 2009).

To validate the learning strategy in Algorithm 1, we performed 50 replications of hold-out cross-validation where 20% of the site–variety combinations were sampled and set aside to be used as a test set. The remaining 80% was used as a training set to learn $\mathcal{B}_{LME}$ and $\mathcal{B}_{CGBN}$. We computed the predictions for each phenological node X_i (except for grain yield) from its Markov blanket and used these predictions to predict the grain yield in turn. We used the kernel densities of the predicted values and the resulting credible intervals with coverage 0.80 to assess the variability in prediction.

We are aware that predictive accuracy is not an adequate performance measure for a causal model, which is why we also validated it using expert knowledge from the literature in Section 4. However, it allows for comparisons with other machine learning models that cannot be assessed causally, and it can be used to evaluate specific loss functions for $\mathcal{B}_{LME}$ in different application settings. A visual summary of the above data analysis is reported as a mechanistic diagram as Supplementary Material.

3. Results

The complete BNs $\mathcal{B}_{LME}$ and $\mathcal{B}_{CGBN}$ learned from the data are shown in the Supplemental Material. Here, we show only the subgraph around the variable grain yield for each BN in Figs. 1 and 2. Following Section 2.3, we identified 60 site–variety clusters (with only 5 containing fewer than 100 observations) and used them as a discrete variable set to be the parent of the phenological nodes.

The structure of $\mathcal{B}_{LME}$ is more complex than that of $\mathcal{B}_{CGBN}$: $\mathcal{B}_{LME}$ has 118 arcs compared to the 92 of $\mathcal{B}_{CGBN}$, and the average Markov blanket size reflects that (17 for $\mathcal{B}_{LME}$, 12 for $\mathcal{B}_{CGBN}$). Notably, we discovered more relationships for the phenological nodes, particularly for the grain yield variable (Table A.3), which had eight more parents than in $\mathcal{B}_{CGBN}$.

The predictive accuracy for each of the scenarios reported in Table A.2 is shown in Fig. 3 for both $\mathcal{B}_{LME}$ and $\mathcal{B}_{CGBN}$. Overall, $\mathcal{B}_{LME}$ outperformed $\mathcal{B}_{CGBN}$ in terms of MAPE. The exception was in a few cases, specifically scenarios 7 to 11, 19, 20 and from 21 to 24, where

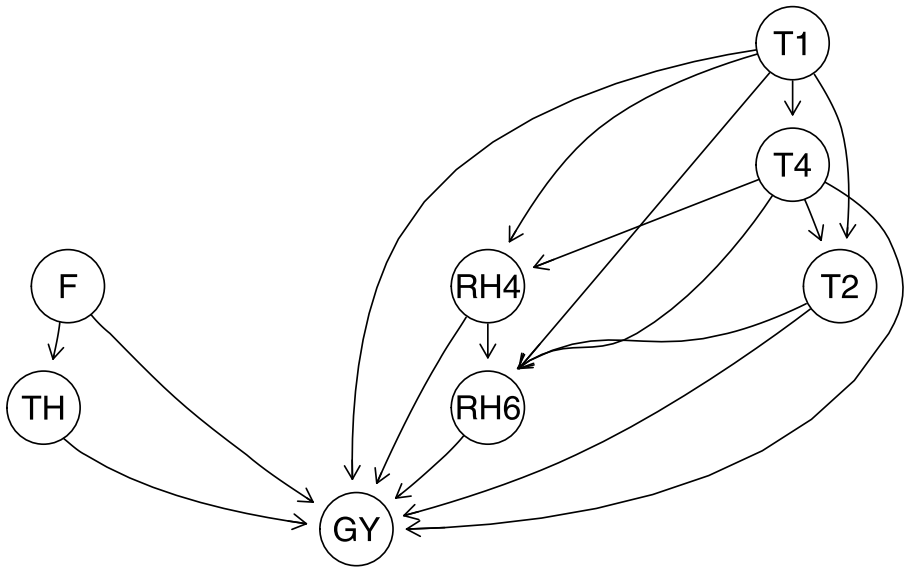


Fig. 2. Structure of the BN $\mathcal{B}_{CGBN}$. The nodes represent: the average temperature May–June (T1), the average temperature July–Aug (T2), the diurnal temperature range May–June (T4), the diurnal RH range May–June (RH4), the diurnal RH range Sept–Oct (RH6), TH (Tassel height) and F (Clusters).

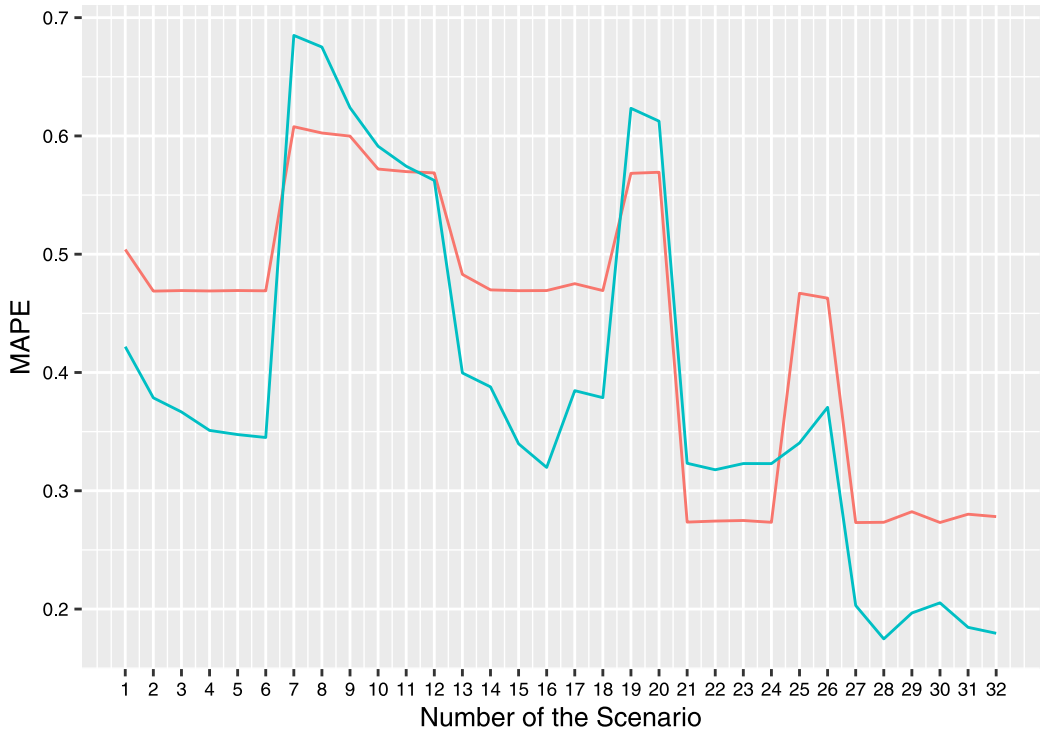


Fig. 3. Prediction accuracy of the learned BNs, $\mathcal{B}_{LME}$ (blue line) and $\mathcal{B}_{CGBN}$ (orange line), in terms of grain yield Mean Absolute Percentage Error (MAPE) of each scenario of evidence propagation (definitions of the scenarios are reported in Table A.2). Lower values are better.

$\mathcal{B}_{CGBN}$ demonstrated a lower MAPE than $\mathcal{B}_{LME}$, albeit with a difference in MAPE of only 0.06. In contrast, when $\mathcal{B}_{LME}$ outperformed $\mathcal{B}_{CGBN}$, the difference in MAPE was 0.14. This trend was particularly evident in scenarios 27 to 32, where an increasing usage of weather/phenological variables was provided. As expected, the scenarios with the lowest MAPE utilized the Markov Blanket (scenario 31) and the parents of grain yield (scenario 32).

The MAPE for the imputation of different site–variety combinations is shown in Fig. 4. We observe that $\mathcal{B}_{LME}$ and $\mathcal{B}_{CGBN}$ perform similarly for all combinations except those involving the sites of Craiova (numbered 250–500) and Campagnola (numbered 1250–1500), for which $\mathcal{B}_{CGBN}$ has a higher MAPE than $\mathcal{B}_{LME}$.

The kernel densities of grain yield for the first nine runs of cross-validation are shown in Fig. 5, full results are reported as Supplementary Materials. The densities for the training set exhibit somewhat heavier tails than those for the predicted values. Furthermore, the predictive densities have narrower credible intervals than those from the training set, particularly on the lower tail, and are more often positively skewed. The mean values are nearly identical for both, at approximately 7 t/ha, with a 0.8 credible interval [4 t/ha, 11 t/ha].

Finally, we employed a stepwise parent elimination algorithm to search for non-significant effects in each of the local distributions of the phenological variables. The BIC values consistently indicated that, within $\mathcal{B}_{LME}$, the best set of effects were those selected by our method.

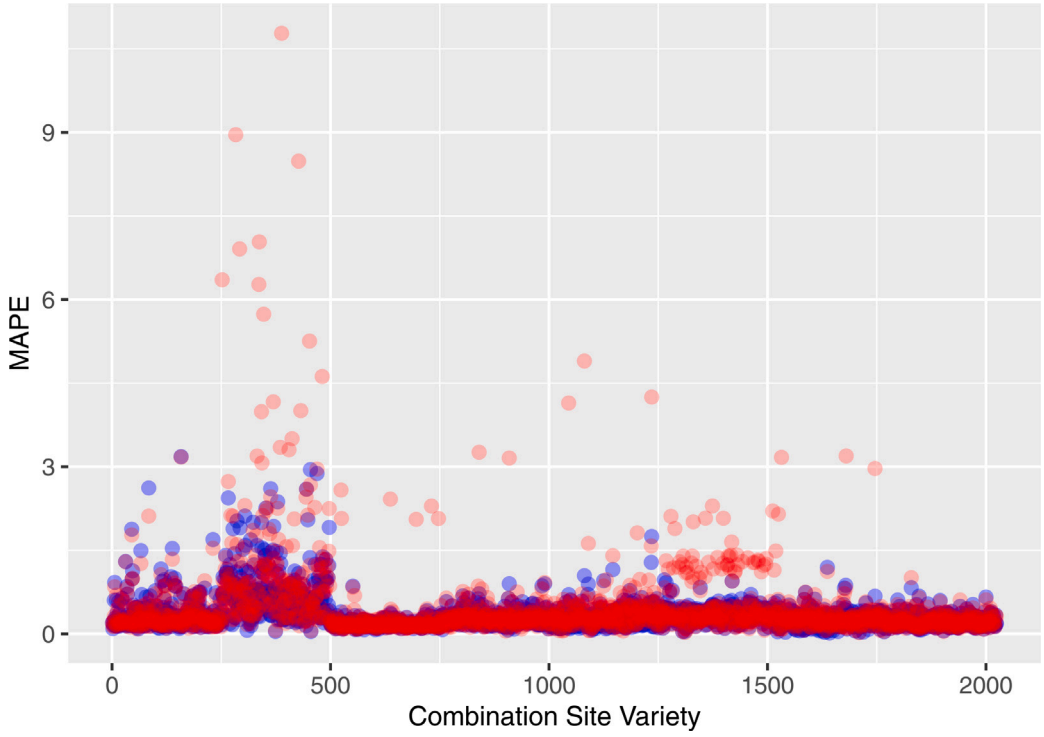


Fig. 4. Imputation accuracy of the learned BNs, $\mathcal{B}_{LME}$ (blue points) and $\mathcal{B}_{CGBN}$ (red points), in terms of grain yield Mean Absolute Percentage Error (MAPE) of each site–variety combination, shown sequentially for brevity. Lower values are better.

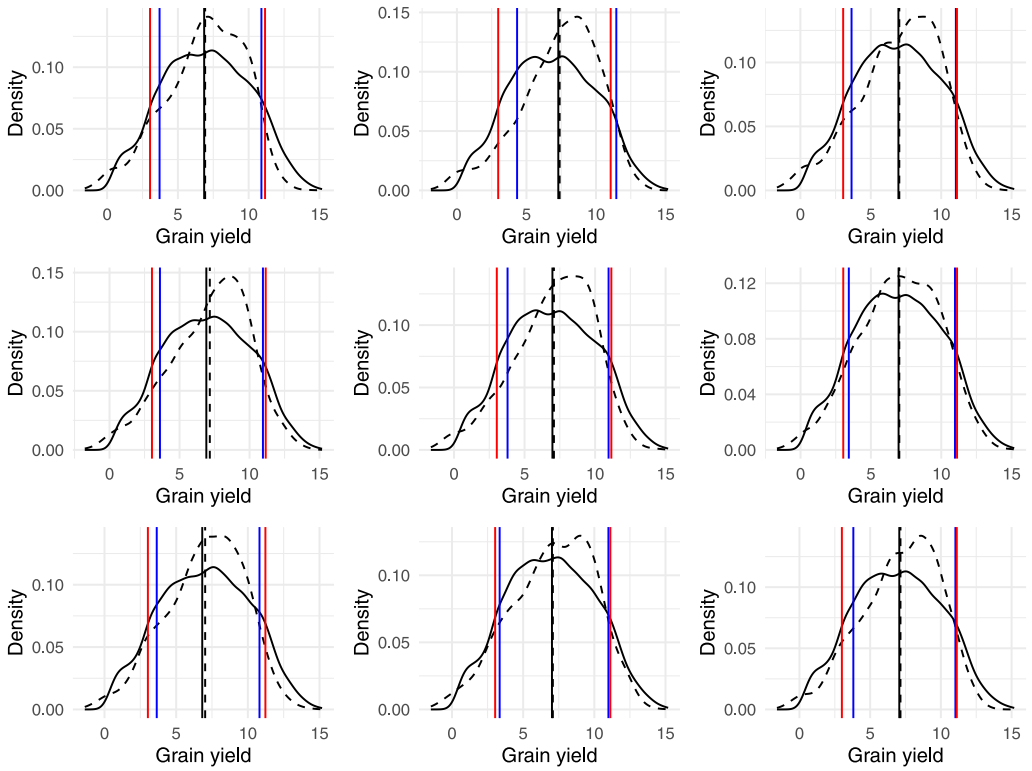


Fig. 5. Kernel densities of the grain yield in the training set are represented by the solid curve, while the dashed curve depicts the kernel densities of the predicted grain yield obtained through likelihood-weighted approximation during cross-validation. The kernel density-based credible interval at 80% for the grain yield in the training set is indicated by the red line and for the predicted grain yield by the blue line. The mean is reported with a solid line for the grain yield of the training set and a dashed line for the predicted grain yield.

The only exception was the variable "tassel height", for which the BIC was lower when the "diurnal RH range July–August" variable was omitted. The same procedure was applied with the removal of random effects from the local distributions. In this case, the set of effects selected by our method still yielded the best BIC values. Furthermore, we compared the BIC values for local distributions with and without the random effects. Generally, the presence of random effects improved goodness of fit, except for the variables "tassel height" and "silking", which exhibited better BIC values in the absence of random effects. All BIC results are reported in Tables A.4 and A.5. The BIC values of the first row correspond to the set of parents of the local distribution found with our method, and the other rows correspond to the BIC value found after each parent elimination.

4. Discussion

In this paper, we used a Bayesian network (BN) to analyse the results of a multi-site agronomic experiment comprising eight different sites in Europe and Chile. Our goal was to obtain a network model that can be used for causal inference, thus providing an ideal foundation to develop decision support systems to manage maize crops. To do that effectively, we modified the BN's structure learning to encode the data's hierarchical structure, thus addressing the violation of the exchangeability assumption that characterizes the RDs.

The data we used consisted of weather variables and phenological variables of maize measured from 2011 to 2013. In our study, we selected a subset of variables based on their agronomic relevance in addition to the weather variables for temperature and humidity. The latter was measured daily for each site, so we calculated their mean for specific time-slices corresponding to the key phenological phases of maize, namely seeding, germination, emergence (May–June); vegetation stage, tasselling, silking, ear formation (July–August); and grain filling, maturation and harvest (September–October). The reason for this choice was to capture the effect of each weather variable on the phenological variables. Based on strong prior knowledge, specific arcs were prohibited due to their lack of causal meaning. For example, it is not plausible for a weather variable from a later time slice to affect another in an earlier time slice. We applied the same logical reasoning to the connections between phenological variables recorded in different time slices. For instance, the arc from grain yield to silking was prohibited, as it is causally impossible for the grain yield to cause female flowering (silking). Moreover, all arcs that made the cluster variable a child of other variables were prohibited.

We compared the structures of the BN incorporating random effects ($\mathcal{B}_{LME}$) to the baseline including only fixed effects ($\mathcal{B}_{CGBN}$): the former contains 26 additional arcs compared to the latter, with a significant difference in the case of grain yield, which had eight more parents. We further assessed the predictive accuracy of phenological variables in both $\mathcal{B}_{LME}$ and $\mathcal{B}_{CGBN}$ using the Diebold–Mariano test (Diebold and Mariano, 2002). This statistical evaluation allowed us to determine that the predictive accuracy improvement observed in $\mathcal{B}_{LME}$ was statistically significant for all of the variables (p-value < 0.05, results not shown). We are aware that predictive accuracy is not an adequate measure of performance for a causal model, which is better assessed using expert knowledge as we do below, but it provides a term of comparison for our model in the wider context of machine learning models, which cannot be assessed causally.

Regarding grain yield, plant height emerged as a new parent: its role as a reliable predictor for maize grain yield is well-documented in the literature (Yin et al., 2011; Pugh et al., 2018). Additionally, its ease of measurement using remote sensing makes it a suitable candidate for predicting maize grain yield (Han et al., 2018; Chu et al., 2018). Supporting evidence comes from the work of Anderson et al. (2019), who studied 280 hybrids conducted in 1500 plots using unmanned aerial systems and found a positive correlation between plant height and maize grain yield. Another new parent identified in the analysis is silking. Existing evidence also supports this finding, as Malik et al. (2005) demonstrated a significant negative correlation between silking and grain yield. They posited that this negative relationship could be attributed to late female flowering, resulting in a less favourable photoperiod and low temperature induced by changing seasons. Considering variables related to temperature and relative humidity (RH), they are all the parents of grain yield in $\mathcal{B}_{LME}$ but not in $\mathcal{B}_{CGBN}$, where only diurnal RH range May–June (RH4), diurnal RH range Sept–Oct (RH6), average temperature May–June (T1), average temperature July–Aug (T2), diurnal temperature range May–June (T4) where present. This is plausible since environmental conditions are essential for maize growth: for instance, evidence shows that high humidity during flowering promotes the maize yield (Butts-Wilmsmeyer et al., 2019). Temperature plays a crucial role in influencing maize yield, particularly during the reproductive phase, where sub-optimal or supra-optimal values can have a significant impact. For instance, temperatures ranging from 33 °C to 36 °C during the pre-and post-flowering stages can result in a reduction of grain yield by 10% to 45% (Neiff et al., 2016). In a review by Waqas et al. (2021), the detrimental effects of thermal stress on maize growth were thoroughly examined from both an agronomic and a physiological perspective. They emphasized that high temperatures, especially during the flowering period, can have various adverse consequences on floret number, silk number, and grain development. Furthermore, the process of fertilization and grain-filling may also be compromised under such conditions. On the other hand, low temperatures below 10 °C can also be detrimental, negatively impacting the normal growth process of maize. Such cold temperatures can limit germination, adversely affect root morphology, and decrease the efficiency of photosystem II. These combined factors demonstrate the sensitivity of maize to temperature fluctuations, which can significantly influence its growth and overall productivity.

We applied hierarchical clustering to the mean and the variance of the residuals from a simple linear regression of the grain yield, which was selected due to its agronomic relevance against all other variables to avoid making any assumptions about the possible parent grain yield. After grouping the residuals by site–variety combination, hierarchical clustering produced 60 relatively-balanced clusters: they were included in the data as discrete variable that was set as a common parent of the phenological variables in a setup similar to a conditional Gaussian BN as described in Section 2.3. We decided to use the clusters, rather than just the site of origin or the maize variety as individual variables, for two reasons:

- When using either the site or the maize variety as a common discrete parent variable, we found the dispersion of residuals in the local distribution, particularly that of grain yield, to be non-homogeneous.
- Combining the site of origin and the maize variety without clustering their combinations produces a variable with approximately 2000 possible values, which would make BN structure learning computationally prohibitive.

As a result, we improved the model's computational feasibility and predictive accuracy. Using clustering as a pre-processing step has been proposed in the literature to find suitable scenarios in risk assessment analysis (Pettet et al., 2017) or to reduce the dimension of the estimation problem, learning the structure of one subgraph for each cluster (Gu and Zhou, 2020). Rodriguez-Sanchez et al. (2022) also proposed a multi-partition clustering that produces a set of categorical variables that encode clusters. These partitions represent a distinct clustering solution and were used as parents, leading to a more interpretable and flexible way to find clusters.

As discussed in Section 2.3, we assumed that the local distributions of phenological variables are linear mixed-effect regression models to

allow for the partial pooling of information across clusters: this model balances the individual cluster-specific estimates with the overall trend observed in the data, leading to more stable and reliable estimates (Scutari et al., 2022). We assumed different local distributions for grain yield and the weather variables. For grain yield, we introduced a power function to model the variance after observing a skewed residual distribution during the exploratory data analysis. Moreover, we modelled grain yield with a random intercept as the only random effect; in contrast, all other phenological variables have both random coefficients and intercepts. For the weather variables, we used a linear regression model containing only fixed effects as the local distribution. We made this decision based on visual inspection, which revealed that the weather variables appeared disconnected from the clusters. This observation implies that the values of these variables were independent of both the site of origin and the variety of maize.

These assumptions reduced the prediction error for grain yield from 28% to approximately 17% when its Markov blanket or its parents were used as predictors, as shown in Fig. 3. In two specific intervals, our model exhibits a higher Mean Absolute Percentage Error (MAPE) compared to the baseline model. These intervals are scenarios 7 to 11 and scenarios 19 to 24. In the first interval (7–11), our model's performance is affected as we gradually introduce variables related to relative humidity (RH). The second interval (19–24) corresponds to the gradual inclusion of phenotypic variables. The reason for our model's higher MAPE in these ranges could indicate that simply considering phenotypic and humidity variables is insufficient for using our model as a reliable Decision Support System (DSS). However, a notable trend emerges when we combine temperature and humidity variables (scenarios 13–18) and subsequently add phenotypic variables (24–32). During these scenarios, the MAPE of our model significantly improves compared to the baseline. Interestingly, this improvement does not occur with temperature alone, where our model's MAPE remains lower than the baseline. This suggests that introducing temperature as a new variable enhances the model's performance. This observation aligns with the causal and biological context since temperature plays a pivotal role in influencing the phenological stages of plants. Hence, incorporating temperature as a factor results in more accurate predictions. Furthermore, we assessed whether the incorporation of random effects in the structure learning procedure enhanced the local distributions, not just in terms of structure but also in terms of model specification using a backward algorithm that removes variables from the local distribution and tests the BIC, results are reported in Tables A.4 and A.5 in Appendix A. Millet et al. (2019a) used the same data set for grain yield prediction, employing grain weight and grain number as predictors. Their approach included modelling grain numbers through a factorial regression model with predictors such as specific intercepted radiation, soil water potential, night temperature, hybrids, and experimental location. Their cross-validation analysis involved testing new hybrids already evaluated in previous experiments and vice versa. The correlation between the observed and predicted grain yield ranged from 0.43 to 0.85 per experiment and 0.71 to 0.97 per hybrid. For new hybrids in tested experiments, correlation results ranged from 0.21 to 0.71 per experiment and 0.66 to 0.96 per hybrid. Our study employed a different cross-validation scheme than Millet et al. (2019a) to exclude simultaneous sites and varieties from the training set. This led to correlations between observed and predicted values ranging from 0.86 to 0.90, with an average of 0.88. We designed our sampling scheme based on clustering to group together sites and varieties with similar grain yield characteristics. Consequently, our model was tested by randomly removing site–variety combinations, simulating scenarios where an agronomist queries the model for grain yield predictions of sites and varieties akin to those used in model training for enhanced robustness assessment.

Our findings indicated that random effects have a favourable impact on both structure learning and model specification. They contribute to a more accurate explanation of the data without introducing undue complexity. However, exceptions were observed for "tassel height" and "silking", which were best modelled without random effects. This suggests that the maxima identified with our method might be local maxima rather than global ones. Additionally, it implies that the estimation of their local distributions did not benefit from the partial pooling information provided by random effects.

Our findings confirm that a CGBN incorporating mixed-effects models to exploit the hierarchical structure of the data provides better accuracy than a standard CGBN. Considering that it is a causal model as well, we argue that it can serve as an effective decision support system, particularly in domains with inherent hierarchical structures, such as the agronomic field (Burchfield et al., 2019; Li et al., 2020).

Even though our proposed method exhibits low prediction error, it has limitations. Firstly, the clustering pre-processing is based only on grain yield regression to simplify the model and to reduce the computational cost of learning it. To address this, future work will expand the clustering approach to specific clusters for each phenological variable, enabling a more detailed analysis. Another limitation is the time it took to learn it: approximately 13 h of linear time. To mitigate this, we may explore different approaches to model hierarchical data, such as the Integrated Nested Laplace Approximation (INLA; Rue et al., 2017). Moreover, in this study, we used the hill-climbing algorithm for structure learning because it compares favourably in terms of speed and structural accuracy (Scutari et al., 2019) and because our focus was on incorporating mixed effects. Hence, we wanted to avoid hyperparameter tuning issues common with more complex structure learning algorithms. Other algorithms, however, may very well be more suitable for this particular type of data, and we will explore them in future work. In particular, reformulating structure learning to share information on the covariance structure of the data between iterations and variables has the potential to vastly reduce computational complexity without impacting accuracy.

5. Conclusions and Future works

Maize stands as a crucial crop for both food and feed production. Predicting its yield can enhance farm practices and refine crop management systems. In agricultural datasets, a common hierarchical structure can be leveraged for predictive purposes; therefore, models that harness this structure tend to yield highly accurate predictions. In this study, we introduced a novel approach to learning Bayesian Networks integrating random effects into local distributions, obtaining such conclusions:

- The variables encoding the provenance of the field (*Site*) exhibit a dispersion of residuals in the local distribution.
- Introducing a group variable by performing hierarchical clustering based on the mean and variance of residuals from a complete linear regression of maize grain yield, conditioned on the Site–variety combination, can reduce the dimension of the Cartesian product of Site–Variety combinations.
- The application of this method led to a reduction in prediction errors compared to the baseline model.
- The reduction in errors was attributed to the partial pooling of information provided by the random effects.

We propose its applicability as a valuable decision support system, particularly in fields marked by inherent hierarchical structures like agronomy.From an agricultural engineering point of view, crop yield predictions provide good management of agricultural practices, optimally scheduling the irrigation system and phytosanitary treatments, leading farming systems resilient to climate change and economic losses. Future work, as we explained in the discussion section, will consider performing precise clustering for each phenotypical variable. Additionally, introducing alternative estimation methods, such as INLA, could reduce computational demands and empower experts to actively engage in the learning process within the Bayesian framework. Moreover, such methods could be used to take into account the spatial component of the datasets since INLA is typically used for such purposes.

Table A.2
Prediction scenarios for the grain yield of maize, identified by the set of variables used as predictors: the average temperature May–June (T1), the average temperature July–Aug (T2), the average temperature Sept–Oct (T3), the diurnal temperature range May–June (T4), the diurnal temperature range July–Aug (T5), the diurnal temperature range Sept–Oct (T6), the average RH May–June (RH1), the average RH July–Aug (RH2), the average RH Sept–Oct (RH3), the diurnal RH range May–June (RH4), the diurnal RH range July–Aug (RH5), the diurnal RH range Sept–Oct (RH6), Silking (Si), GW (Grain weight), An (Anthesis), TH (Tassel height), PH (Plant height) and EH (Ear height).

Scenario	T1	T2	T3	T4	T5	T6	RH1	RH2	RH3	RH4	RH5	RH6	Si	GW	TH	PH	An	EH
1	✓																	
2	✓	✓																
3	✓	✓		✓														
4	✓	✓		✓	✓													
5	✓	✓		✓	✓	✓												
6	✓	✓	✓	✓	✓	✓												
7							✓											
8							✓	✓										
9							✓	✓		✓								
10							✓	✓		✓	✓							
11							✓	✓		✓	✓	✓						
12							✓	✓	✓	✓	✓	✓						
13	✓						✓											
14	✓	✓					✓	✓										
15	✓	✓		✓			✓	✓		✓								
16	✓	✓		✓	✓		✓	✓		✓	✓							
17	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓						
18	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						
19													✓					
20													✓	✓				
21													✓	✓	✓			
22													✓	✓	✓	✓		
23													✓	✓	✓	✓	✓	
24													✓	✓	✓	✓	✓	✓
25	✓						✓						✓					
26	✓	✓					✓	✓					✓	✓				
27	✓	✓		✓			✓	✓		✓			✓	✓	✓			
28	✓	✓		✓	✓		✓	✓		✓	✓		✓	✓	✓	✓		
29	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	
30	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
31	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
32	✓	✓		✓	✓	✓	✓	✓	✓	✓			✓		✓			

CRediT authorship contribution statement

Lorenzo Valleggi: Methodology, Software, Validation, Writing – original draft. **Marco Scutari:** Conceptualization, Methodology, Software, Writing – original draft. **Federico Mattia Stefanini:** Conceptualization, Methodology, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset is available online at the following link: https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/IASSTN.

Appendix A. More details about scenarios, structure learning and validation tests

See Tables A.2–A.5.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.engappai.2024.107867.

Table A.3
New relationships found in $\mathcal{B}_{LME}$. Variables: the average temperature May–June (T1), the average temperature July–Aug (T2), the average temperature Sept-Oct (T3), the diurnal temperature range May–June (T4), the diurnal temperature range July–Aug (T5), the diurnal temperature range Sept–Oct (T6), the average RH May–June (RH1), the average RH July-Aug (RH2), the average RH Sept–Oct (RH3), the diurnal RH range May–June (RH4), the diurnal RH range July–Aug (RH5), the diurnal RH range Sept–Oct (RH6), Silking (Si), GW (Grain weight), An (Anthesis), TH (Tassel height), PH (Plant height) and EH (Ear height).

Parent		Child	Parent		Child
PH	→	GY	T5	→	Si
PH	→	EH	T5	→	TH
EH	→	Si	T5	→	PH
Si	→	GY	T6	→	GW
T1	→	EH	RH1	→	GY
T1	→	PH	RH2	→	GY
T2	→	An	RH3	→	GY
T2	→	TH	RH4	→	TH
T3	→	GY	RH4	→	PH
T4	→	GW	RH5	→	GY
T4	→	Si	RH5	→	EH
T4	→	TH	RH5	→	PH
T5	→	GY	RH6	→	GW

Table A.4
The BIC score values for the variables with random effects in $\mathcal{B}_{LME}$. The columns correspond to Silking (Si), GW (Grain weight), An (Anthesis), TH (Tassel height), PH (Plant height), EH (Ear height) and GY (Grain yield).

EH	PH	TH	An	Si	GW	GY
108 448.0	128 945.4	98 210.88	72 373.64	64 279.80	127 634.6	42 664.90
108 581.7	128 988.4	131 629.19	73 127.37	67 958.98	129 620.1	42 788.77
108 913.9	130 979.4	98 721.24	72 466.75	64 423.46	130 956.2	43 600.06
109 422.9	129 204.0	98 306.37	73 802.03	78 577.20	129 137.0	42 817.61

(*continued on next page*)

Table A.4 (*continued*).

EH	PH	TH	An	Si	GW	GY
109 076.0	128 987.3	98 975.48	73 569.12	64 617.86	128 784.5	47 312.45
108 793.3	128 996.4	99 125.48	74 726.54	64 740.27	129 580.9	42 683.18
–	130 634.3	98 957.10	73 009.51	65 812.12	127 641.1	44 288.90
–	–	98 134.49	74 775.35	65 272.43	129 603.5	43 055.93
–	–	–	75 937.24	64 331.62	130 433.6	42 733.74
–	–	–	–	64 426.97	128 022.4	43 319.44
–	–	–	–	–	129 434.5	43 919.43
–	–	–	–	–	–	43 890.76
–	–	–	–	–	–	44 146.02
–	–	–	–	–	–	42 899.58
–	–	–	–	–	–	42 926.77

Table A.5
The BIC score values for the variables without random effects in $\mathcal{B}_{LME}$. The columns correspond to Silking (Si), GW (Grain weight), An (Anthesis), TH (Tassel height), PH (Plant height), EH (Ear height) and GY (Grain yield).

EH	PH	TH	An	Si	GW	GY
109 382.7	129 069.7	97 997.56	72 921.13	63 953.28	128 026.4	50 158.96
109 517.0	129 130.9	132 132.71	73 968.06	67 636.88	129 661.0	50 252.83
110 087.1	131 885.7	99 046.89	73 012.02	64 104.01	131 334.4	50 586.06
109 464.6	129 540.4	98 415.05	74 703.04	78 859.63	129 263.1	50 196.68
109 928.2	129 201.7	99 377.01	74 227.08	64 271.16	130 586.0	53 345.75
109 577.8	129 234.8	99 329.96	75 441.82	64 386.79	131 320.9	50 161.03
–	131 600.8	99 481.83	73 857.10	65 460.04	128 286.2	51 321.11
–	–	98 304.39	75 487.06	65 002.78	130 356.2	50 378.48
–	–	–	76 760.46	64 166.55	132 314.7	50 178.26
–	–	–	–	64 281.80	128 579.4	50 610.04
–	–	–	–	–	130 592.4	51 098.38
–	–	–	–	–	–	51 108.30
–	–	–	–	–	–	50 974.64
–	–	–	–	–	–	50 306.71
–	–	–	–	–	–	50 364.45

References

Alexandratos, N., Bruinsma, J., 2012. World Agriculture: Towards 2030/2050. ESA Working Paper No. 12–03, FAO, Rome, Italy.

Anderson, S.L., Murray, S.C., Malambo, L., Ratcliff, C., Popescu, S., Cope, D., Chang, A., Jung, J., Thomasson, J.A., 2019. Prediction of maize grain yield before maturity using improved temporal height estimates of unmanned aerial systems. Plant Phenome J. 2 (1), 1–15. http://dx.doi.org/10.2135/tppj2019.02.0004.

Anderson II, S.L., Murray, S.C., Malambo, L., Ratcliff, C., Popescu, S., Cope, D., Chang, A., Jung, J., Thomasson, J.A., 2019. Prediction of maize grain yield before maturity using improved temporal height estimates of unmanned aerial systems. Plant Phenome J. 2, 190004. http://dx.doi.org/10.2135/tppj2019.02.0004.

Azzimonti, L., Corani, G., Zaffalon, M., 2019. Hierarchical estimation of parameters in Bayesian networks. Comput. Statist. Data Anal. 137, 67–91. http://dx.doi.org/10.1016/j.csda.2019.02.004.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Burchfield, E.K., Nelson, K.S., Spangler, K., 2019. The impact of agricultural landscape diversification on U.S. crop production. Agric. Ecosyst. Environ. 285, 106615. http://dx.doi.org/10.1016/j.agee.2019.106615.

Butts-Wilmsmeyer, C.J., Seebauer, J.R., Singleton, L., Below, F.E., 2019. Weather during key growth stages explains grain quality and yield of maize. Agronomy 9 (1), 16. http://dx.doi.org/10.3390/agronomy9010016.

Chamorro, O., Carlos, L., Barriga, C., Eduardo, J., 2023. Urban risks due to climate change in the Andean municipality of Pasto, Colombia: A Bayesian network approach. Risk Anal. 43, 2017–2032. http://dx.doi.org/10.1111/risa.14086.

Chu, T., Starek, M.J., Brewer, M.J., Murray, S.C., Pruter, L.S., 2018. Characterizing canopy height with UAS structure-from-motion photogrammetry—Results analysis of a maize field trial with respect to multiple factors. Remote Sens. Lett. 9 (8), 753–762. http://dx.doi.org/10.1080/2150704X.2018.1475771.

Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., Traore, D., 2019. Deep neural networks with transfer learning in millet crop images. Comput. Ind. 108, 115–120. http://dx.doi.org/10.1016/j.compind.2019.02.003.

Cussens, J., 2012. Bayesian network learning with cutting planes. In: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. pp. 153–160.

Darwiche, A., 2009. Modeling and Reasoning with Bayesian Networks. Cambridge University Press.

Diebold, F., Mariano, R., 2002. Comparing predictive accuracy. J. Bus. Econom. Statist. 20 (1), 134–144.

Edwards, D.I., 2000. Introduction to Graphical Modelling, second ed. Springer.

FAO, IFAD, UNICEF, WFP, WHO, 2021. The State of Food Security and Nutrition in the World. FAO: Food and Agriculture Organization of the United Nations, URL: https://policycommons.net/artifacts/1850109/the-state-of-food-security-and-nutrition-in-the-world-2021/2596732/.

FAOSTAT, 2019. Food balance sheets. URL: http://www.fao.org/faostat/en/data/FBS.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2014. Bayesian Data Analysis, third ed. CRC Press.

Gelman, A., Hill, J., 2007. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Gu, J., Zhou, Q., 2020. Learning big Gaussian Bayesian networks: Partition, estimation and fusion. J. Mach. Learn. Res. 21 (158), 1–31.

Han, X., Thomasson, J.A., Bagnall, G.C., Pugh, N.A., Horne, D.W., Rooney, W.L., Jung, J., Chang, A., Malambo, L., Popescu, S.C., Gates, I.T., Cope, D.A., 2018. Measurement and calibration of plant-height from fixed-wing UAV images. Sensors 18 (12), 4092. http://dx.doi.org/10.3390/s18124092.

Heckerman, D., Geiger, D., 1995. Learning Bayesian networks: a unification for discrete and Gaussian domains. In: UAI. pp. 274–284.

Heisterkamp, S.H., van Willigen, E., Diderichsen, P.-M., Maringwa, J., 2017. Update of the nlme package to allow a fixed standard deviation of the residual error. R J. 9 (1), 239–251. http://dx.doi.org/10.32614/RJ-2017-010.

Hill, E.C., Renner, K.A., Sprague, C.L., Fry, J.E., 2017. Structural equation modeling of cover crop effects on soil nitrogen and dry bean. Agron. J. 109 (6), 2781–2788. http://dx.doi.org/10.2134/agronj2016.12.0712.

Ilić, M., Mutavdžić, B., Srdević, Z., Srdević, B., 2022. Irrigation water fitness assessment based on Bayesian network and FAO guidelines. Irrig. Drain. 71 (3), 665–675. http://dx.doi.org/10.1002/ird.2676.

Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. Agricult. Forest Meteorol. 260–261, 300–320. http://dx.doi.org/10.1016/j.agrformet.2018.06.009.

Kluger, D.M., Owen, A.B., Lobell, D.B., 2022. Combining randomized field experiments with observational satellite data to assess the benefits of crop rotations on yields. Environ. Res. Lett. 17, 044066. http://dx.doi.org/10.1088/1748-9326/ac6083.

Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques. MIT Press.

Krishna, M.V., Swaroopa, K., SwarnaLatha, G., Yasaswani, V., 2023. Crop yield prediction in India based on mayfly optimization empowered attention-bi-directional long short-term memory (LSTM). Multimedia Tools Appl. Online first, 1–28. http://dx.doi.org/10.1007/s11042-023-16807-7.

Leroux, L., Castets, M., Baron, C., Escorihuela, M.-J., Bégué, A., Lo Seen, D., 2019. Maize yield estimation in west africa from crop process-induced combinations of multi-domain remote sensing indices. Eur. J. Agron. 108, 11–26. http://dx.doi.org/10.1016/j.eja.2019.04.007.

Li, Y., Guan, K., Yu, A., Peng, B., Zhao, L., Li, B., Peng, J., 2019. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S.. Field Crops Res. 234, 55–65. http://dx.doi.org/10.1016/j.fcr.2019.02.005.

Li, Z., Taylor, J., Yang, H., Casa, R., Jin, X., Li, Z., Song, X., Yang, G., 2020. A hierarchical interannual wheat yield and grain protein prediction model using spectral vegetative indices and meteorological data. Field Crops Res. 248, 107711. http://dx.doi.org/10.1016/j.fcr.2019.107711.

Lu, W., Newlands, N.K., Carisse, O., Atkinson, D.E., Cannon, A.J., 2020. Disease risk forecasting with Bayesian learning networks: Application to grape powdery mildew (Erysiphe necator) in vineyards. Agronomy 10 (5), 622. http://dx.doi.org/10.3390/agronomy10050622.

Malik, H.N., Malik, S.I., Hussain, M., Chughtai, S.U.R., Javed, H.I., 2005. Genetic correlation among various quantitative characters in maize (Zea mays L.) hybrids. J. Agric. Soc. Sci. 1 (3), 262–265.

Millet, E.J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F., Tardieu, F., 2019a. Genomic prediction of maize yield across European environmental conditions. Nature Genet. 51, 952–956. http://dx.doi.org/10.1038/s41588-019-0414-y.

Millet, E.J., Pommier, C., Buy, M., Nagel, A., Kruijer, W., Welz-Bolduan, T., Lopez, J., Richard, C., Racz, F., Tanzi, F., Spitkot, T., Canè, M.-A., Negro, S.S., Coupel-Ledru, A., Nicolas, S.D., Palaffre, C., Bauland, C., Praud, S., Ranc, N., Presterl, T., Bedo, Z., Tuberosa, R., Usadel, B., Charcosset, A., van Eeuwijk, F., Draye, X., Tardieu, F., Welcker, C., 2019b. A multi-site experiment in a network of European fields for assessing the maize yield response to environmental scenarios. http://dx.doi.org/10.15454/IASSTN.

Millet, E., Welcker, C., Kruijer, W., Negro, S., Nicolas, S., Praud, S., Ranc, N., Presterl, T., Tuberosa, R., Bedo, Z., Draye, X., Usadel, B., Charcosset, A., van Eeuwijk, F., Tardieu, F., Coupel-Ledru, A., Bauland, C., 2016. Genome-wide analysis of yield in Europe: Allelic effects as functions of drought and heat scenarios. Plant Physiol. 172 (2), 749–764. http://dx.doi.org/10.1104/pp.16.00621.

Mupangwa, W., Chipindu, L., Nyagumbo, I., Mkuhlani, S., Sisito, G., 2020. Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. SN Appl. Sci. 2 (5), 952. http://dx.doi.org/10.1007/s42452-020-2711-6.

Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? J. Classification 31 (3), 274–295. http://dx.doi.org/10.1007/s00357-014-9161-z.

Ndlovu, N., Spillane, C., McKeown, P.C., Cairns, J.E., Das, B., Gowda, M., 2022. Genome-wide association studies of grain yield and quality traits under optimum and low-nitrogen stress in tropical maize (Zea mays L.). Theor. Appl. Genet. 135 (12), 4351–4370. http://dx.doi.org/10.1007/s00122-022-04224-7.

Neiff, N., Trachsel, S., Valentinuz, O.R., Balbi, C.N., Andrade, F.H., 2016. High temperatures around flowering in maize: Effects on photosynthesis and grain yield in three genotypes. Crop Sci. 56 (5), 2702–2712. http://dx.doi.org/10.2135/cropsci2015.12.0755.

Oyen, D., Lane, T., 2015. Transfer learning for Bayesian discovery of multiple Bayesian networks. Knowl. Inf. Syst. 43 (1), 1–28. http://dx.doi.org/10.1007/s10115-014-0775-6.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. http://dx.doi.org/10.1109/TKDE.2009.191.

Patterson, H.D., Williams, E.R., 1976. A new class of resolvable incomplete block designs. Biometrika 63 (1), 83–92. http://dx.doi.org/10.1093/biomet/63.1.83.

Paymode, A.S., Malode, V.B., 2022. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. Artif. Intell. Agric. 6, 23–33. http://dx.doi.org/10.1016/j.aiia.2021.12.002.

Pearl, J., 2009. Causality: Models, Reasoning and Inference, second ed. Cambridge University Press.

Pettet, G., Nannapaneni, S., Stadnick, B., Dubey, A., Biswas, G., 2017. Incident analysis and prediction using clustering and Bayesian network. In: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI. pp. 1–8. http://dx.doi.org/10.1109/UIC-ATC.2017.8397587.

Pew Research Center, 2019. World's population is projected to nearly stop growing by the end of the century. URL: https://www.pewresearch.org/short-reads/2019/06/17/.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-Effects Models in S and S-PLUS. Springer.

Pugh, N.A., Horne, D.W., Murray, S.C., Carvalho Jr, G., Malambo, L., Jung, J., Chang, A., Maeda, M., Popescu, S., Chu, T., Starek, M.J., Brewer, M.J., Richardson, G., Rooney, W.L., 2018. Temporal estimates of crop growth in sorghum and maize breeding enabled by unmanned aerial systems. Plant Phenome J. 1 (1), 1–10. http://dx.doi.org/10.2135/tppj2017.08.0006.

Qian, S.S., Cuffney, T.F., Alameddine, I., McMahon, G., Reckhow, K.H., 2010. On the application of multilevel modeling in environmental and ecological studies. Ecology 91 (2), 355–361. http://dx.doi.org/10.1890/09-1043.1.

Rodriguez-Sanchez, F., Bielza, C., Larrañaga, P., 2022. Multipartition clustering of mixed data with Bayesian networks. Int. J. Intell. Syst. 37 (3), 2188–2218. http://dx.doi.org/10.1002/int.22770.

Rotili, D.H., de Voil, P., Eyre, J., Serafin, L., Aisthorpe, D., Maddonni, G.A., Rodríguez, D., 2020. Untangling genotype x management interactions in multi-environment on-farm experimentation. Field Crops Res. 255, 107900. http://dx.doi.org/10.1016/j.fcr.2020.107900.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: A review. Annu. Rev. Stat. Appl. 4 (1), 395–421. http://dx.doi.org/10.1146/annurev-statistics-060116-054045.

Russell, S.J., Norvig, P., 2009. Artificial Intelligence: A Modern Approach, third ed. Prentice Hall.

Schwalbert, R.A., Amado, T.J.C., Nieto, L., Varela, S., Corassa, G.M., Horbe, T.A.N., Rice, C.W., Peralta, N.R., Ciampitti, I.A., 2018. Forecasting maize yield at field scale based on high-resolution satellite imagery. Biosyst. Eng. 171, 179–192. http://dx.doi.org/10.1016/j.biosystemseng.2018.04.020.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.

Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. J. Stat. Softw. 35 (3), 1–22.

Scutari, M., Graafland, C.E., Gutiérrez, J.M., 2019. Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. Internat. J. Approx. Reason. 115, 235–253.

Scutari, M., Marquis, C., Azzimonti, L., 2022. Using mixed-effects models to learn Bayesian networks from related data sets. Proc. Mach. Learn. Res. 186, 73–84.

Simanjuntak, C., Gaiser, T., Ahrends, H.E., Ceglar, A., Singh, M., Ewert, F., Srivastava, A.K., 2023. Impact of climate extreme events and their causality on maize yield in South Africa. Sci. Rep. 13, 12462. http://dx.doi.org/10.1038/s41598-023-38921-0.

Spiegelhalter, D.J., Abrams, K.R., Myles, J.P., 2004. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley.

Spirtes, P., Glymour, C., Scheines, R., 2000. Causation, Prediction, and Search. MIT Press.

Tolley, S.A., Brito, L.F., Wang, D.R., Tuinstra, M.R., 2023. Genomic prediction and association mapping of maize grain yield in multi-environment trials based on reaction norm models. Front. Genet. 14, 1221751.

Waqas, M.A., Wang, X., Zafar, S.A., Noor, M.A., Hussain, H.A., Azher Nawaz, M., Farooq, M., 2021. Thermal stresses in maize: Effects and management strategies. Plants 10 (2), 293. http://dx.doi.org/10.3390/plants10020293.

Yan, H., Song, S., Wang, F., He, D., Zhao, J., 2023. Operational adjustment modeling approach based on Bayesian network transfer learning for new flotation process under scarce data. J. Process Control 128, 103000. http://dx.doi.org/10.1016/j.jprocont.2023.103000.

Yang, W., Nigon, T., Hao, Z., Dias Paiao, G., Fernández, F.G., Mulla, D., Yang, C., 2021. Estimation of corn yield based on hyperspectral imagery and convolutional neural network. Comput. Electron. Agric. 184, 106092. http://dx.doi.org/10.1016/j.compag.2021.106092.

Yang, B., Zhu, W., Rezaei, E.E., Li, J., Sun, Z., Zhang, J., 2022. The optimal phenological phase of maize for yield prediction with high-frequency UAV remote sensing. Remote Sens. 14 (7), 1559. http://dx.doi.org/10.3390/rs14071559.

Yin, X., McClure, M.A., Jaja, N., Tyler, D.D., Hayes, R.M., 2011. In-season prediction of corn yield using plant height under major production systems. Agron. J. 103 (3), 923–929. http://dx.doi.org/10.2134/agronj2010.0450.

Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., Li, S., 2021. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. Agricult. Forest Meteorol. 311, 108666. http://dx.doi.org/10.1016/j.agrformet.2021.108666.

Zhu, B., Chen, S., Cao, Y., Xu, Z., Yu, Y., Han, C., 2021. A regional maize yield hierarchical linear model combining landsat 8 vegetative indices and meteorological data: Case study in Jilin Province. Remote Sens. 13, 356. http://dx.doi.org/10.3390/rs13030356.

Zorić, M., Gunjača, J., Galić, V., Jukić, G., Varnica, I., Šimić, D., 2022. Best linear unbiased predictions of environmental effects on grain yield in maize variety trials of different maturity groups. Agronomy 12 (4), 922. http://dx.doi.org/10.3390/agronomy12040922.