Technische Universität München
Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt
Lehrstuhl für Pflanzenzüchtung

Covariance structures in multivariate genome-enabled prediction models

Katrin Alexandra Töpner

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Aurélien Tellier
Prüfende der Dissertation:	1.	Prof. Dr. Chris-Carolin Schön
	2.	Prof. Daniel Gianola, Ph. D. University of Wisconsin – Madison, USA

Die Dissertation wurde am 03.09.2018 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 08.01.2019 angenommen.

The highest activity a human being can attain is learning for understanding, because to understand is to be free.

Baruch Spinoza

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

BGE	Bayesian Gaussian equivalent (score)
BIC	Bayesian information criterion
BN	Bayesian network
CDF	cumulative distribution function
CV	cross-validation
DAG	directed acyclic graph
DH	doubled haploid
DMC	dry matter content
DMY	dry matter yield
DtSILK	days to silking
DtTAS	days to tasseling
GS	Grow-Shrink (algorithm)
IC	inductive-causation (algorithm)
logL	logarithm of Bayesian marginal likelihood
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings (algorithm)
ML	marginal likelihood
MTM	multiple-trait model
PH	plant height
REML	restricted maximum likelihood
RKHS	reproducing kernel Hilbert space (regression model)
SD	standard deviation
SEM	structural equation model
SNP	single nucleotide polymorphism (marker)
TABU	Tabu-search (algorithm)
TVR	transition from vegetative to reproductive growth

List of Variables

α	significance level or acceptance probability
$\boldsymbol{\mu}$	intercept
ϑ	bandwidth parameter for spatial kernel adjustment
$\boldsymbol{\rho}$	residuals (independent and identically distributed)
$\boldsymbol{\Lambda}$	eigenvector matrix or structure matrix
$\boldsymbol{\Sigma}$	spatial trait covariance matrix (micro-environment)
$\boldsymbol{\Psi}$	diagonal matrix
$\boldsymbol{\Omega}$	general kernel
$\boldsymbol{e}$	error term
$\boldsymbol{g}$	genomic component (random or fixed)
$\boldsymbol{s}$	random spatial component (micro-environment)
$\boldsymbol{u}$	general random component
$\boldsymbol{y}$	phenotypic values (one per plot) or adjusted means (one per genotype)
$\mathbf{E}$	residual trait covariance matrix
$\mathbf{G}$	genomic trait covariance matrix
$\mathbf{K}$	genomic kernel matrix (kinship)
$\mathbf{S}$	spatial kernel matrix (block diagonal)
$\mathbf{S}_\vartheta$	spatial kernel matrix of one field
$\mathbf{V}$	trait covariance matrix (specified by index where appropriate)
$\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$	random variables, nodes in a DAG
$\boldsymbol{block}$	random component for blocks
$\boldsymbol{field}$	random component for fields
$\boldsymbol{loc}$	random component for locations (macro-environment)

1 Introduction

Traits of animals and plants, *e.g.*, height or age at maturity, are influenced both by genetic (heritable) and environmental factors. Selection exploits genetic properties of a population by passing favorable alleles onto subsequent generations. For breeding animals or plants, it is therefore crucial to distinguish between genetic contributions and environmental factors influencing traits. Thus, decomposition of phenotypic variances and covariances between traits into their genetic and environmental components has been under investigation for decades (*e.g.*, Fisher 1918; Hazel 1943; Falconer 1952; Robertson 1959; Searle 1961; Roff 1995; Lynch and Walsh 1998).

Identifying genetic relationships between traits is an important research focus when investigating genetic covariances; an example is a study of genetic correlations among 16 quantitative traits of maize (Malik et al. 2005). Insights into the connections between developmental or phenological traits with target traits such as stress resistance might assist breeding decisions. While such relationships may be antagonistic when two target traits affect each other adversely, they can be useful in genome-enabled prediction of traits when a complex trait is influenced by a second trait that is easier to measure, assess, or breed for. In addition, it has been found that selection for traits with low heritability can be enhanced by joint modeling with genetically correlated traits. These and other advantages of multiple-trait prediction models have been corroborated by experimental and simulation studies (*e.g.*, Jia and Jannink 2012; Pszczola et al. 2013; Guo et al. 2014; Jiang et al. 2015; Maier et al. 2015).

Discovering and understanding phenotypic relations among traits is therefore important in this context. Statistical methods have been developed that specify connections among traits as directed influences, in contrast to standard multiple-trait models where connections among traits are represented by unstructured covariance matrices. One effective approach, structural equation models (SEM), can describe systems of phenotypes connected via feedback or recursive relationships (Gianola and Sorensen 2004). SEM are regression models that allow for a structured dependence among variables, *e.g.*, by a structure matrix $\mathbf{\Lambda}$. Elements of $\mathbf{\Lambda}$ represent effects of dependence and can be either freely varying, or the structure can be constrained by setting some entries to zero *a priori*.

For a given structure $\mathbf{\Lambda}$, the effect of one variable on another can be estimated using likelihood-based or Bayesian approaches. SEM are well established and many studies in the animal sciences have investigated causal relationships among traits using various SEM approaches and data sets, *e.g.*, body composition and bone density in mouse intercross populations (Li et al. 2006), calving traits in cattle (de Maturana et al. 2009, 2010), or gene expression in mouse data by incorporating genetic markers (Schadt et al. 2005; Aten et al. 2008). Methodology, applications, and potential advantages of SEM in prediction of traits were reviewed by Rosa et al. (2011) and Valente et al. (2013). One of the limitations of SEM is that connections between traits and their directions are assumed to be known to be able to estimate their magnitude (*e.g.*, Valente et al. 2013).

To investigate connections among traits, Bayesian network (BN) learning methods can be used. BN are models representing the joint distribution of random variables (*e.g.*, traits) in terms of their conditional independencies. There are two main types of algorithms for learning a BN: constraint-based and score-based algorithms. The former use a sequence of conditional independence tests to learn the network among variables, while the latter compare the fit of many (ideally all) possible networks to the empirical data using a score (Nagarajan et al. 2013).

BN have been used for many purposes in quantitative genetics, for example, to predict individual total egg production of European quails using earlier expressed phenotypic traits (Felipe et al. 2015) and to study linkage disequilibrium using single nucleotide polymorphism (SNP) markers (Morota et al. 2012). BN have also been employed in genome-assisted prediction of traits, with a performance that was at least as good as that of other methods such as genomic best linear unbiased prediction or elastic nets (Scutari et al. 2014). In addition, many studies have investigated connections among several traits via a BN analysis incorporating quantitative trait loci and phenotypic data (Neto et al. 2008, 2010; Winrow et al. 2009; Hageman et al. 2011; Wang and van Eeuwijk 2014; Peñagaricano et al. 2015). BN can also search for connections among SNPs and traits for feature selection, or in genome-wide association studies by finding the Markov blanket for one or several traits (Porth et al. 2013; Scutari et al. 2013). However, interpretation of BN connections as causal influences is a delicate issue. Pearl (2000) offers a modern perspective on causal inference, methodology and issues in causal inference were reviewed by Rockman (2008). Within the scope of this thesis, learning BN is a hypothesis-generating tool with respect to the causal nature of the connections found.

Valente et al. (2010) investigated causal connections among phenotypic traits using an approach combining BN and SEM. They learned network structure from Markov chain Monte Carlo (MCMC) samples of the residual covariance matrix of a Bayesian multiple-trait model using the inductive-causation (IC) algorithm. This was a preliminary step before fitting SEM with the selected structure to estimate the magnitude of influences among traits. These authors illustrated the effectiveness of their approach with a simulation study.

While modeling of relationships with BN and SEM has been shown to be very effective, all applications discussed so far have considered both the genetic and residual trait relationships jointly. In this thesis, a novel approach combining BN and SEM analysis is presented to investigate several traits simultaneously, and for the first time with respect to their genomic and residual trait relationships separately. Arguably, causes for genomic connections among traits must be different from those for residual relationships in designed plant breeding experiments, the latter mainly due to environmental causes. The novel approach developed works as follows: First, different BN algorithms were used to learn genomic and residual networks. Second, the BN algorithms were compared and assessed by fitting SEM to the structures learned from them. The learned networks may provide deeper insight into relationships among traits than pairwise-association measures (such as correlations or covariances) on the genetic and residual levels.

Instead of using the IC algorithm to analyze MCMC samples of covariance matrices, it was decided to use predicted (fitted) genomic values and residuals from a multiple-trait model as input variables for structure-learning algorithms. This was enabled by finding a suitable transformation of the genomic values to meet sample-independence assumptions of the BN learning algorithms. Using effects for BN learning instead of MCMC samples of covariance matrices has the following advantages: it allows using available software for learning BN, is computationally feasible, and use of bootstrapping to assess uncertainty regarding the edges of the connections is straightforward (Scutari and Nagarajan 2011, 2013; Nagarajan et al. 2013). Using this strategy, it was possible to study the behavior of various BN algorithms on experimental data and compare them to each other.

In addition, effects are expected to be more accurately estimated than covariance matrices in mixed models because they are one hierarchy higher within the Bayesian approach. Results from bootstrapping can flag confounders affecting the networks (Edwards 2000; Nagarajan et al. 2013). This is an important opportunity when analyzing experimental data which cannot be assured to be free of confounders. Experimental data also poses a challenge to evaluate results since the true connections remain unknown. This issue is addressed in this thesis by proposing using SEM as a structure-evaluation tool, comparing the found structures with the standard, unstructured model.

The novel separation of genomic from residual structure allows understanding which trait connections might be manipulated by selective breeding because they have a genetic origin, whereas phenotypic structures or residual structures include many other causes for trait connections, especially environment and morphology. If two favorable traits are adversely genetically correlated and a third trait is found to be inducing this connection, this third trait might help with breeding decisions regarding the two antagonistic traits. In line with this argument, a genomic network analysis can also reveal spurious connections, *e.g.*, those that have a high genomic correlation, but no direct genomic connection with each other in a Bayesian network (Nagarajan et al. 2013). After illustrating trait connections in such a case, a clearer picture on which traits should be included in a multiple-trait prediction model or in an indirect selection process may be gained.

Traditionally, in quantitative genetics and plant breeding phenotypic variation is decomposed into random genetic and environmental contributions; genotype-by-environment interaction and genotype-by-environment correlation can confound both of these contributions. In experimental data from plant breeding, phenotypic data points are usually averaged across locations representing different environmental conditions, and every location contains replications of experiments. This approach reduces noise from environmental effects and also effectively decreases the influence of genotype-by-environment effects in the data (Falconer and Mackay 1996). With such data, genomic trait covariances fitted with a multiple-trait genome-enabled prediction model represent mainly genetic connections among traits, induced *inter alia* by pleiotropy and linkage disequilibrium. Residuals estimated with such a model include, *e.g.*, nonadditive genetic effects and nonaccounted-for (micro-)environmental effects, such as nongenetic physiological and morphological trait dependencies.

In all methodology presented so far, the residual structure contains trait covariances from several different causes and separating macro-environmental from micro-environmental trait connections has not been possible so far. Later in the thesis, it is therefore proposed to use a refined decomposition of phenotypic covariances into genomic, micro-, and macro-environmental components. Micro-environmental trait correlations might be helpful for understanding and improving the impact of the field design (including distribution of treatments) on trait correlations. So, this approach helps to identify alterable nongenetic trait connections (*e.g.*, induced by field design), and to distinguish them from inalterable factors concerning a certain site (*e.g.*, induced by climate, weather, presence of certain diseases, and soil quality).

In order to model these micro-environmental effects, the use of a single random component with a Gaussian kernel including one bandwidth parameter for modeling trait covariances due to spatial arrangement of plots in experimental trials is an appealing approach presented in this thesis (*cf.* Pérez-Elizalde et al. 2015 concerning the genomic component). Using a single random component instead of several is expected to represent a robust modeling approach for this purpose. Further, summarizing all micro-environmental effects into a single random component allows for BN structure analysis of this component, *e.g.*, in order to identify spurious correlations.

Spatial models are well known, both in terms of their ability of adjusting for random micro-environmental confounding in experiments (*e.g.*, Gilmour et al. 1997 or Piepho et al. 2008) and in their statistical formulation (*e.g.*, Pérez-Elizalde et al. 2015 concerning the genomic component). In experimental trials, random micro-environmental factors such as unevenly distributed fertilizer, light, soil moisture, or diseases may lead to covariance of phenotypic values among neighboring plants or plots in the trial. In addition, larger and nonrandom spatial trends, *e.g.*, water or light gradients caused by slopes in the field, can confound experimental data. Nearest neighborhood adjustments such as the Papadakis method, random walk, and the linear variance model (Piepho et al. 2008) have been widely applied in these circumstances. More complex models accounting for random confounding and gradients have been introduced recently (*e.g.*, Bernal-Vasquez et al. 2014, Rodríguez-Álvarez et al. 2016, De Faveri et al. 2017, Velazco et al. 2017). These methods usually include a random component per dimension in the experiment (also referred to as "rows" and "columns" in field design) and supplement them by accounting for various kinds of spatial effects, *e.g.*, derivatives of these dimensions (*e.g.*, diagonal) or components for gradients.

For exploring trait connections due to micro-environmental trait covariances, however, a single random component can be used to model all micro-environmental effects together. The approach of Pérez-Elizalde et al. (2015) was therefore amended in this thesis, using physical distances in the trials as features of joint random micro-environmental factors of neighboring plants. A Gaussian feature-based kernel with a bandwidth parameter that follows *a priori* a Gamma distribution was then applied to experimental data, thereby mapping physical distance onto spatial correlation between neighboring plots. Using a Bayesian approach, the bandwidth parameter is sampled via a Metropolis-Hastings (MH) algorithm implemented in this thesis, embedded in an already known

Gibbs sampler (de los Campos 2015; Lehermeier et al. 2015). This is a more efficient alternative than a grid search, and enables investigation of sensitivity to hyper-parameters.

Experimental trials may be laid out at several climatically diverse locations, *e.g.*, one site in a dry, hot place, one in a dry, cold place, one in a wet, hot place, and one in a wet, cold place. Locations can also vary in soil quality, presence of diseases, plot size and shape, altitude and many other factors. Each of these locations may contain more than one experimental unit, *i.e.*, crop fields. Fields in a given location are designed and treated as similarly as possible and placed at close distance because, in each of these fields, some or all of the genotypes planted in the other field are replicated (*e.g.*, Lehermeier et al. 2014). Hence, the bandwidth parameter is expected to be different in every field, and therefore the posterior mean of distribution of the bandwidth parameter is obtained for every field separately in a first step, using the aforementioned MH implementation. Afterward, all fields are integrated into a single model.

This single model contains *inter alia* a genomic component, a location component, and a spatial component, modeling genomic, macro-environmental, and micro-environmental effects, respectively, complemented with independent and identically distributed errors per plot. This model is referred to as "one-stage model" throughout the thesis (*cf.* Bernal-Vasquez et al. 2014). In contrast, the multiple-trait genome-enabled prediction model discussed earlier is fitted after a preliminary phenotypic analysis, where adjusted means for each genotype across locations and replications are estimated. These adjusted means are then decomposed into a genomic component and an error term, modeling genomic and residual effects and their covariance matrices per genotype, respectively. This approach is referred to as "two-stage model".

Regarding the genomic component, the one-stage and the two-stage models produce similar results; however, only the one-stage model allows for an analysis of micro- and macro-environmental effects and covariance matrices. The one-stage approach is seen as the gold standard (Bernal-Vasquez et al. 2014). However, practitioners often prefer the two-stage approach as it is computationally more robust and simpler when dealing with large experimental designs. If the error term's variances from the phenotypic analysis are used as weights in the genome-enabled prediction models, the difference between the one-stage and two-stage models is reduced (Möhring and Piepho 2009; Bernal-Vasquez et al. 2014).

In this thesis, the advantages of the novel approaches are illustrated using a maize data set. Biological objectives of this thesis focus on the investigation of connections among maize traits, on the comparison of networks with respect to two underlying sources of information (genomic and residual) and genetically different, heterotic maize groups (Dent and Flint). Later, the decomposition of the environmental covariance into micro- and macro-environmental covariance is investigated. Five well-characterized complex traits were studied: biomass yield, plant height, maturity, male and female flowering time.

This thesis starts with an overview of the plant material used for experimental trials. It continues with developing the methodology employed for investigating genomic and residual trait connections. Subsequently, the methodology for constructing the spatial component and the one-stage model is described. A third section summarizes methods for model evaluation and lists the software resources. Finally, results are presented and discussed.

2 Material

The data contain a sample of phenotypes and genotypes representing two important heterotic maize pools, Dent and Flint. European hybrid breeding exploits the genetic divergence between these pools, which arose from their geographical separation and different environmental adaptation. Dent, coming from subtropical climate, was introduced into Europe's southern regions already at the end of the 15^{th} century. Flint, adapted to temperate conditions and shorter vegetation periods, followed around 45 years later and enabled corn production in northern regions (Tenaillon and Charcosset 2011; Mir et al. 2013). Production of the material used in this thesis was similar to modern maize hybrid breeding programs including crossing within pools, derivation of doubled haploid (DH) lines, and assessing performance in testcrosses with a member of the opposite heterotic pool.

2.1 Genetic background of Dent and Flint

A panel of half-sib families was available for each pool, where each family contained several full-sib lines (Bauer et al. 2013). In the Dent (Flint) panel, 10 (11) founder lines were crossed to a common central Dent (Flint) line. DH lines were derived from these crosses in both panels (Table A1), where DH lines from one full-sib family are referred to as a DH population. All DH lines were genotyped with the Illumina MaizeSNP50 BeadChip containing 56,110 SNP markers (Ganal et al. 2011). After imputation and SNP quality control, 34,116 high-quality SNPs remained in the Dent and Flint panels. Only markers that were polymorphic within a panel were considered, which were 32,801 (30,122) SNPs in the Dent (Flint) panel. A total of 831 (805) Dent (Flint) DH lines were used for all analyses conducted in this thesis.

2.2 Phenotyping and field design

All DH lines (genotypes) were evaluated as testcrosses with the central line from the opposite pool as tester (Lehermeier et al. 2014). F353 (UH007) was the central Dent (Flint) line, so for example, performance of the Dent DH line CFD02-003 was assessed using testcrosses CFD02-003 × UH007 and the Flint DH line CFF03-001 was evaluated from testcrosses CFF03-001 × F353 (*cf.* Table A1).

Experiments were conducted according to an augmented p-rep design (Williams et al. 2011). In the Dent (Flint) panel, 480 (960) blocks were used all together for phenotyping the DH line testcrosses and other material like fillers, checks, or parental testcrosses. These 480 (960) blocks were distributed in four (six) European locations. Two fields were created at each location for partially replicating plant material, *i.e.*, one fourth (one third) of the Dent (Flint) material was replicated in each location. These fields were formed from 60 (80) incomplete blocks each, and every block consisted of ten (eight) plots. The term "replication" instead of field is avoided in this thesis since genotypes are only partially replicated in a p-rep design.

In both panels, five phenotypic traits were recorded: biomass dry matter yield (DMY) (decitons per hectare), biomass dry matter content (DMC) (percentage), plant height (PH) (centimeters), days to tasseling (DtTAS) (days), and days to silking (DtSILK) (days). Plots containing less than 70% of the median number of plants per location and those being identified as outliers (all traits) on the basis of maximum deviate residuals (Grubbs 1950) were classified as missing data and discarded (Lehermeier et al. 2014). For the multiple-trait spatial analyses, only plots with records available for all traits were used (Table A2).

2.3 Data availability

Phenotypic data are available from file S1 of Lehermeier et al. (2014) at http://www.genetics.org/content/suppl/2014/09/17/198.1.3.DC1. Genotypic data (Bauer et al. 2013) were deposited at the National Center for Biotechnology Information Gene Expression Omnibus as data set GSE50558, accessible via http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50558.

3 Methods for learning covariance structure

In the following subsections, the methods for obtaining and evaluating genomic and residual structures are described. First, adjusted phenotypic means are calculated employing a model in which all DH lines were assumed fixed. These were provided by a single-trait phenotypic analysis of a previous project (details are below). Second, these adjusted means are used in a linear mixed model to obtain estimates of genomic breeding values, both in a single-trait and a multiple-trait approach. Third, Bayesian networks learn genomic and residual structures from the genomic values and residuals estimated in the multiple-trait model. As a novel approach, structures are learned from the estimated effects, and for genomic and residual effects separately. Last, these structures are evaluated in a novel approach of structural equation models, separately and jointly. At the end of this section, the workflow of this methodology is summarized.

3.1 Phenotypic analysis before genomic value estimation

Adjusted means across locations and fields for each genotype were obtained from a previous study (Lehermeier et al. 2014), in which means were estimated using a single-trait model separately for each panel (Dent and Flint) and trait:

$$\boldsymbol{y}_{pbrlg} = \boldsymbol{\mu} + \boldsymbol{loc}_l + \boldsymbol{field}_{r(l)} + \boldsymbol{block}_{b(rl)} + \boldsymbol{g}_g + (\boldsymbol{loc} \cdot \boldsymbol{g})_{lg} + \boldsymbol{e}_{pbrlg} \tag{1}$$

where $\boldsymbol{y}$ is the vector of phenotypic values of all plots, $\boldsymbol{\mu}$ is a fixed intercept, $\boldsymbol{loc}$ and $\boldsymbol{field}$ are random components for the locations and the fields (nested within locations), $\boldsymbol{block}$ is a random component containing the blocks (nested within location and field), $\boldsymbol{g}$ is a fixed vector representing the genotypes, and $\boldsymbol{e}$ is the error term. $(\boldsymbol{loc} \cdot \boldsymbol{g})$ is a random component for the genotype-by-location interaction. The number of levels are summarized per component in Table 1. All random components and the error term were assumed to follow independent normal distributions with specific variances.

Table 1: Number of levels in phenotypic analysis

Index	Meaning	Levels		Range
		Dent	Flint	
t	trait	5	5	$1:d$
p	plot	4800	7680	$1:N$
b	block	480	960	—
r	field	8	12	—
l	location	4	6	—
g	genotype	841	811	$1:n$

Variance components were estimated by restricted maximum likelihood (REML) using the R package *ASReml* (Butler et al. 2009), and estimates and predictions were then calculated according to Henderson's mixed model equations (*e.g.*, Henderson 1977).

For predicting genomic breeding values, the adjusted means were mean-centered in each DH population, as focus of investigation was on allelic rather than on population effects. They were also standardized to unit variance based on their sample variance within each trait.

3.2 Obtaining genomic breeding values

A Bayesian univariate Gaussian model and a Bayesian multivariate Gaussian model were fitted to the $d = 5$ traits and a reduced set of $n = 831$ ($n = 805$) genotypes within the Dent (Flint) panel to separate genomic values from random residual contributions to phenotypic values (adjusted means). Not all genotypes from the phenotypic analysis could be used due to quality issues of the genomic data. The genomic and residual correlations between traits were estimated with the multiple-trait model.

An MCMC approach based on Gibbs sampling was used to explore posterior distributions. A burn-in of 30,000 MCMC samples was followed by an additional 300,000 MCMC samples. The MCMC samples were thinned with a factor of two, resulting in 150,000 MCMC samples for inference. Posterior means were then calculated for all effects and their covariance matrices, separately for the Dent and Flint panels. Genomic and residual trait correlations and their standard deviations (SDs) were estimated from samples of the posterior distributions of the respective covariance matrices.

3.2.1 Single-trait analysis

Genomic breeding values for a single trait can be obtained from

$$\boldsymbol{y} = \mu \cdot \mathbf{1}_n + \boldsymbol{g} + \boldsymbol{e} \tag{2}$$

where $\boldsymbol{y}$, $\boldsymbol{g}$, and $\boldsymbol{e}$ are $(n \times 1)$-dimensional column vectors of scaled adjusted means, genomic values, and residuals. $\mu \cdot \mathbf{1}_n$ is the product of an intercept scalar μ and an $(n \times 1)$-dimensional vector of ones ($\mathbf{1}_n$). Since adjusted means ($\boldsymbol{y}$) are mean-centered within their respective DH population, only a trait-specific intercept is included in the model.

The following mutually independent distributions were assumed for $\boldsymbol{g}$ and $\boldsymbol{e}$:

$$\begin{aligned} \boldsymbol{g} &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{K}\sigma_{\boldsymbol{g}}^2) \\ \boldsymbol{e} &\sim \mathcal{N}_n(\mathbf{0}, \mathbb{I}_{n\times n}\sigma_{\boldsymbol{e}}^2) \end{aligned} \tag{3}$$

where $\sigma_{\boldsymbol{g}}^2$ and $\sigma_{\boldsymbol{e}}^2$ are the genomic and residual variances, respectively, $\mathbf{K}$ is an $(n \times n)$-dimensional realized-kinship matrix estimated using simple matching coefficients (Sneath and Sokal 1973), and the $(n \times n)$- dimensional identity matrix $\mathbb{I}_{n\times n}$ represents independence of model residuals.

A flat prior was assigned to intercept μ. Prior assumptions for $\sigma_{\boldsymbol{g}}^2$ and $\sigma_{\boldsymbol{e}}^2$ were chosen analogously to the following multiple-trait analysis , *i.e.*, independent scaled-inverse χ^2 distributions with four degrees of freedom and a scale parameter equal to three:

$$\begin{aligned} \sigma_{\boldsymbol{g}}^2 &\sim \text{scale-inv-}\chi^2(4,3) \\ \sigma_{\boldsymbol{e}}^2 &\sim \text{scale-inv-}\chi^2(4,3) \end{aligned} \tag{4}$$

according to the parametrization of the *BLR* function (de los Campos and Pérez-Rodríguez 2012). In general, prior distributions for dispersion parameters were chosen to be relatively uninformative, and such that the variance of the scaled response variable was equally distributed among all random components, including error terms.

3.2.2 Multiple-trait analysis

Genomic breeding values were also estimated using the multiple-trait model (MTM)

$$\boldsymbol{y} = \boldsymbol{\mu}\otimes\mathbf{1}_n + \boldsymbol{g} + \boldsymbol{e} \tag{5}$$

where $\boldsymbol{y}$, $\boldsymbol{g}$, and $\boldsymbol{e}$ are $(nd \times 1)$-dimensional column vectors of scaled adjusted means, genomic values, and residuals sorted by trait and genotype within trait, respectively. $\boldsymbol{\mu}\otimes\mathbf{1}_n$ is a Kronecker product of a $(d \times 1)$-dimensional intercept vector $\boldsymbol{\mu}$ and an $(n \times 1)$-dimensional vector of ones. The trait-specific intercepts are included in the model as fixed effects represented by the d coefficients in $\boldsymbol{\mu}$.

For $\boldsymbol{g}$ and $\boldsymbol{e}$ the following mutually independent distributions were assumed:

$$\begin{aligned} \boldsymbol{g} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{G}\otimes\mathbf{K}) \\ \boldsymbol{e} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{E}\otimes\mathbb{I}_{n\times n}) \end{aligned} \tag{6}$$

where $\mathbf{G}$ and $\mathbf{E}$ are $(d \times d)$-dimensional covariance matrices containing the genomic and residual variances and covariances. $\mathbf{K}$ is again the $(n \times n)$-dimensional realized-kinship matrix estimated using simple matching coefficients (Sneath and Sokal 1973) and independence of residuals was again represented by the $(n \times n)$-dimensional identity matrix $\mathbb{I}_{n \times n}$ (independence among genotypes, not traits).

Flat priors were assigned to the elements of the intercept vector $\boldsymbol{\mu}$. The covariance matrices $\mathbf{G}$ and $\mathbf{E}$ were assumed to follow *a priori* independent inverse-Wishart distributions with specific degrees of freedom ν and scale matrix $\boldsymbol{\Phi}$, regarded as hyper-parameters, *i.e.*,

$$\begin{aligned} \mathbf{G} &\sim \mathscr{W}^{-1}(\boldsymbol{\Phi}_{\mathbf{G}}, \nu_{\mathbf{G}}) \\ \mathbf{E} &\sim \mathscr{W}^{-1}(\boldsymbol{\Phi}_{\mathbf{E}}, \nu_{\mathbf{E}}) \end{aligned} \tag{7}$$

The hyper-parameters ν and $\boldsymbol{\Phi}$ were chosen such that: (1) they implied relatively uninformative prior distributions of matrices $\mathbf{G}$ and $\mathbf{E}$; (2) the prior mode of genomic and residual trait variances was 0.5 for each trait, and (3) the choice of hyper-parameters needed to ensure equivalence with the choice of hyper-parameters for a scaled-inverse χ^2 prior defined later in the SEM models (*cf.* Equation 28) for specific trait variances. Considering (1), (2), and (3), $\nu_{\mathbf{G}} = \nu_{\mathbf{E}} = 8$ and $\boldsymbol{\Phi}_{\mathbf{G}} = \boldsymbol{\Phi}_{\mathbf{E}} = \frac{14}{2} \cdot \mathbb{I}_{n \times n}$, with ν and $\boldsymbol{\Phi}$ in accordance with the parametrization of the respective distributions in the R package *BGLR* and the *MTM* function of its multiple-trait extension (see software availability for details).

3.3 Learning genomic and residual Bayesian networks

A BN models connections among random variables. It consists of two components: a statement about the joint distribution of these random variables, and its graphical representation as directed acyclic graph (DAG) (Nagarajan et al. 2013). The statement about the joint distribution $f(\mathbf{X}) = f(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_d)$ of the d random variables $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_d$ is that $f(\mathbf{X})$ can be decomposed into a product of conditional distributions:

$$f(\mathbf{X}) = f(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_d) = \prod_{k=1}^{d} f(\mathbf{X}_k \mid \text{pa}(\mathbf{X}_k)). \tag{8}$$

Above, $\text{pa}(\mathbf{X}_k)$, the "parents" of $\mathbf{X}_k$, denotes the subset of random variables in $\mathbf{X}$ on which $\mathbf{X}_k$ depends. Accordingly, $\mathbf{X}_k$ is then called the "child" of all elements in $\text{pa}(\mathbf{X}_k)$. In the respective DAG, all random variables $\mathbf{X}_k$ are represented as nodes, and arcs point from parents to their children. The joint distribution is also referred to as a global distribution and the conditional distributions are called local distributions. If the global and local distributions are normal and the variables are linked by linear constraints, the BN is a Gaussian BN.

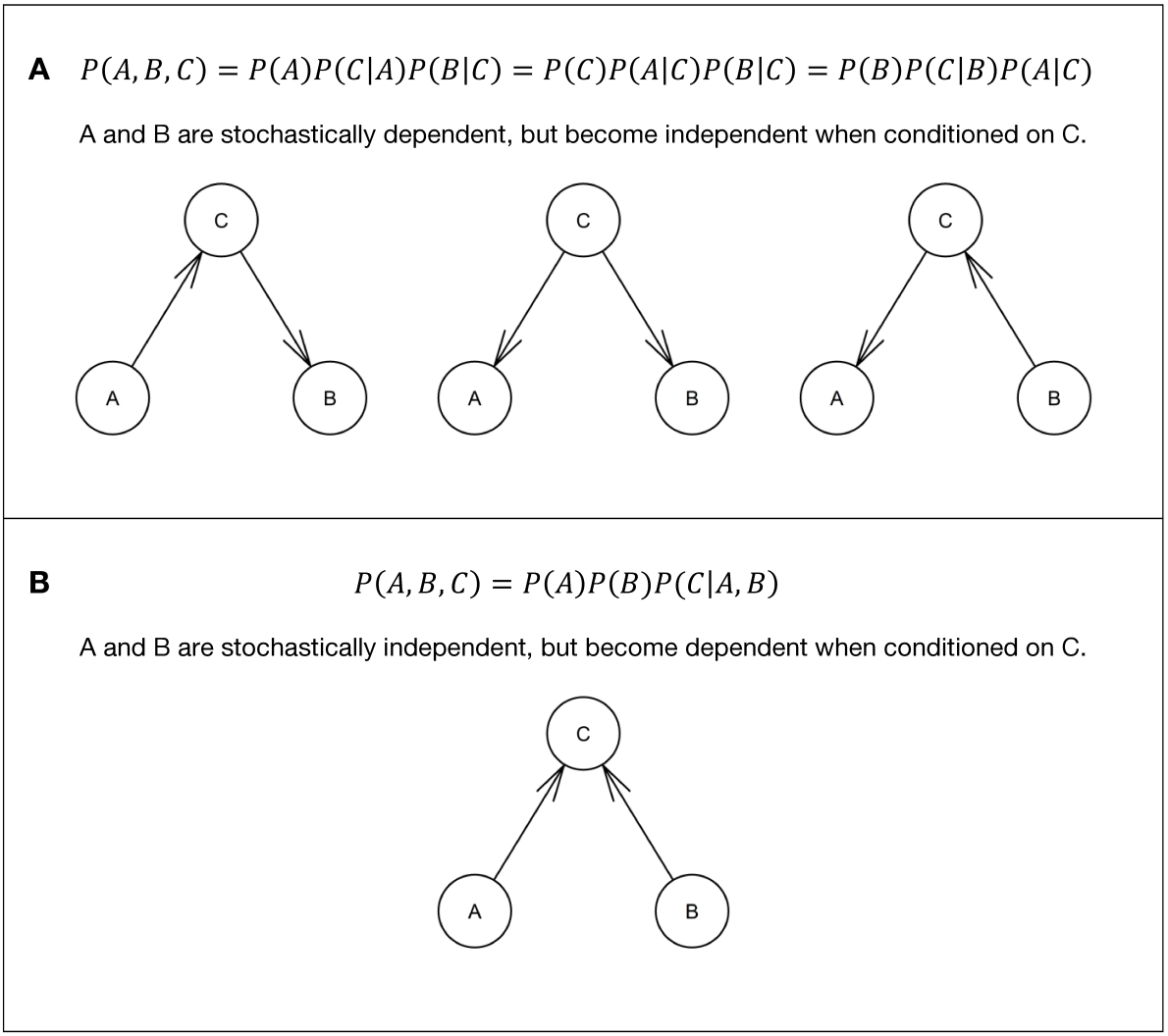


Figure 1: BN constraint-based structure learning: constraint-based algorithms search for trios of variables and test their relations with marginal and conditional independence tests. Thereby, they distinguish between the stochastic decomposition of the joint distribution of all variables as in (A) and the decomposition as in (B). These two decompositions can be represented in a DAG as shown. Whereas the directions are clear and unique in (B), the influences' directions in (A) can form three different graphs. Which of these equivalent structures is most likely, one can only deduce from the context of other already tested decompositions in their neighborhood and from the rule that no cycles must be formed in the graph.

There are two types of algorithms to learn structure of networks: constraint-based and score-based algorithms (Nagarajan et al. 2013). The constraint-based algorithms employ a series of conditional and marginal independence tests to infer potential connections and directions between each pair of variables. First, an undirected structure linking variables directly related to each other is constructed. Next, trios of variables where one is the outcome of the other two ("v-structures") are sought. Last, the remaining connections are oriented whenever possible, such that neither cycles nor new v-structures are created. For illustration, core structures with three nodes are shown in Figure 1. In contrast, score-based approaches search through the space of possible networks (including direction of edges) and compare them by a model-fit score, returning the network with the best score. Both approaches were employed in this thesis for comparison.

3.3.1 Assumptions and algorithms

This and the following subsection (3.3.1 and 3.3.2) summarize the assumptions and algorithms of BN learning that are relevant for this thesis, mainly following Nagarajan et al. (2013).

The correspondence between the DAG and conditional independence of the random variables is ensured by the directed separation (d-separation) criterion (Pearl 1988; Nagarajan et al. 2013). The Markov property of a BN can be directly derived from this criterion and enables the representation of the global distribution as product of local distributions by application of the chain rule (Korb and Nicholson 2010; Nagarajan et al. 2013). A definition of the d-separation is given in Nagarajan et al. (2013):

> **Definition (D-separation)**. If A, B, and C are three disjoint subsets of nodes in a DAG, then C is said to d-separate A from B, if along every sequence of arcs between a node in A and a node in B there is a node **X** satisfying one of the following two conditions:
>
> 1. **X** has converging arcs (*i.e.*, there are two arcs pointing to **X** from the adjacent nodes in the path) and none of **X** or its descendants (*i.e.*, the nodes that can be reached from **X**) are in C.
>
> 2. **X** is in C and does not have converging arcs.

By repeated application of Bayes' theorem, some factorizations of a global distribution can be transformed into others (*e.g.*, as in part A of Figure 1). As these transformations are symmetric, reflexive, and transitive, there are equivalent classes of structures, so called (Markov) equivalent structures (Nagarajan et al. 2013).

D-separation leads to the concept of the Markov blanket of a node, which is the set of nodes that completely d-separates this node from the remaining network. Markov blankets help to facilitate algorithmic approaches for learning BN and are also defined in Nagarajan et al. (2013):

Definition (Markov Blanket). The Markov blanket of a node **X** is the minimal subset S of all nodes such that **X** is probabilistically independent from all nodes not in S and itself, given S. In any BN, the Markov blanket of a node **X** is the set of the parents of **X**, the children of **X**, and all the other nodes sharing a child with **X** ("spouses").

Learning a BN naively includes exploring all probabilistic independencies of the given variables, where the number of possible conditional independence relationships grows exponentially with the number of variables (*cf.* IC algorithm by Verma and Pearl 1990). Most constraint-based algorithms therefore reduce search space by first learning the skeleton of the BN via Markov blankets and then orienting the edges (Nagarajan et al. 2013). A simple forward-selection Markov-blanket-detection approach was introduced by Margaritis (2003). It is called Grow Shrink (GS) for its growing and shrinking phases in the Markov-blanket detection (Algorithm A2).

Based on the Markov blankets of all nodes, the GS (Algorithm A3) reduces the Markov blanket of each node **X** ($B(\mathbf{X})$) to the set of direct neighbors of **X** ($N(\mathbf{X})$), *i.e.*, selecting parents and children and neglecting spouses (Margaritis 2003; Nagarajan et al. 2013). This is done by a series of independence tests between all pairs of nodes **X** and $\mathbf{Y} \in B(\mathbf{X})$ conditioned on all subsets of the smaller of their Markov blankets (excluding the nodes themselves), *i.e.*, basically testing if a connection via parents, children or spouses of **X** or **Y** can explain a probabilistic dependence between **X** and **Y** or not; if not so, they are assumed to be directly probabilistically dependent (Margaritis 2003). Thereby, the skeleton of the graph is derived. Afterward, the principle of Figure 1 helps to orient edges on this structure. Any edge that cannot be oriented by this principle will be oriented if its counter-direction would introduce a cycle (Margaritis 2003).

Alternatively, algorithms follow the idea of a heuristic optimization. The basic concept is to start from a certain network structure (*e.g.*, empty or full graph) and then adding, deleting, or reversing one arc at a time, trying to improve a certain score (Bouckaert 1995). Hence, they are called score-based. Again, not all possible networks can be explored as their number grows exponentially with the number of nodes, so only those networks that arise from the current network S by changing one arc without introducing a cycle are considered at a time ($N(S)$) (Bouckaert 1995; Nagarajan et al. 2013). In this process, algorithms become easily stuck in local optima (Bouckaert 1995). For leaving a local optimum, the algorithm must be able to relax the effectiveness of the current optimal score value and immediate re-visiting of the local optimum must be prohibited during this time. Therefore, local optima are stored in a tabu list (Bouckaert 1995), hence the name tabu search (TABU) for such an algorithmic approach. The length of this tabu list determines the minimal number of steps the algorithm needs to undertake before it can re-visit a local optimum (Algorithm A4).

3.3.2 Tests and scores

Generally, any global or local distribution can be part of a BN and an appropriate score or independence test might be – at least approximately or heuristically – derived. As data for the BN analysis is

assumed to be normally distributed in the MTM (*cf.* Equations 5 and 6), the choice for independence tests in GS and scores in TABU was limited to those consistent with this assumption. All these tests and scores assumed independent samples of the random variables as input, an issue that will be discussed and addressed later in this thesis.

Simple correlation coefficient with an exact Student's t-test. For testing conditional independence of **X** and **Y** given **Z**, one can use the partial correlation coefficient $\rho_{\mathbf{XY|Z}}$. It can be calculated as the standard Pearson correlation coefficient between residuals derived from **X** regressed against **Z** and residuals derived from **Y** regressed against **Z** (Legendre 2000). Nagarajan et al. (2013) defines the statistic for the exact t-test as

$$T(\mathbf{X},\mathbf{Y} \mid \mathbf{Z}) = \rho_{\mathbf{XY|Z}} \sqrt{\frac{\nu}{1-{\rho_{\mathbf{XY|Z}}}^2}} \tag{9}$$

T is distributed as a Student's t random variable with $\nu = n-3$ degrees of freedom, where n is the number of samples. This also holds if Z is a set of variables, then T is distributed as a Student's t with $\nu = n-q-2$ degrees of freedom, where q is the number of elements in Z (Nagarajan et al. 2013).

Mutual information investigated with an asymptotic χ^2 test. Mutual information is the general concept of how much information on one entity informs about the state of another entity, and vice versa. In statistics, mutual information relates to mutual dependence between two random variables (*e.g.*, Cover and Thomas 2006). Any measure of it should be 0 if and only if the two variables are independent, and positive otherwise. The term

$$\log\left(\frac{p(\mathbf{X},\mathbf{Y})}{p(\mathbf{X})p(\mathbf{Y})}\right)$$

is well established for this purpose (Cover and Thomas 2006). For measuring overall mutual information, this term is integrated over all possible events, using the probability density as weight, such that

$$\mathrm{MI}(\mathbf{X},\mathbf{Y}) = \iint_{\mathbf{X},\mathbf{Y}} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) d(x,y) \tag{10}$$

Conditional mutual information is defined analogously by

$$\mathrm{MI}(\mathbf{X},\mathbf{Y} \mid \mathbf{Z}) = \iiint_{\mathbf{X},\mathbf{Y},\mathbf{Z}} p(x,y,z) \log\left(\frac{p(x,y \mid z)}{p(x \mid z)p(y \mid z)}\right) d(x,y,z) \tag{11}$$

For **X**, **Y** and **Z** being normal random variables, MI becomes (Kullback 1959; Scutari 2010)

$$MI_g(\mathbf{X},\mathbf{Y} \mid \mathbf{Z}) = -\frac{1}{2}\log\left(1-{\rho_{\mathbf{XY|Z}}}^2\right) \tag{12}$$

MI_g is proportional to the log-likelihood ratio test and can be tested accordingly. This test can also be generalized to Z being a set of variables (Nagarajan et al. 2013).

MC permutation tests. Distributions of test statistics are often approximated when the real distribution of the random variables is unknown. For assessing an empirical distribution of a test statistic for the data given, one can permute the data, thereby reflecting random dependencies in the data instead of structural ones. Values of the test statistic for all permutations are then expected to represent samples of its empirical distribution under the null hypothesis, and tests are performed using empirical quantiles of the empirical distribution of the test statistic (Good 2005; Scutari and Brogini 2012). Without loss of generality, one can permute $\mathbf{X}$ when testing conditional independence between $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ and derive statistics (Legendre 2000; Nagarajan et al. 2013).

Bayesian information criterion (BIC) as score. BIC on a BN is defined using the local distributions (Rissanen 1978; Lam and Bacchus 1994; Nagarajan et al. 2013)

$$\mathrm{BIC} = \sum_{k=1}^{d} \log f\left(\mathbf{X}_k \mid \mathrm{pa}(\mathbf{X}_k)\right) - \frac{D}{2} \log n \tag{13}$$

where D is the number of parameters of the global distribution, or the sum of parameters of the local distributions, respectively.

Bayesian Gaussian equivalent (BGE) score. The BGE score is the Wishart posterior density of the network and the data when assigning a uniform prior to the space of the network structures and to the parameters of the local distributions (Geiger and Heckerman 1994, 1995; Nagarajan et al. 2013). *Inter alia*, BGE assigns the same scores to equivalent network structures (Chickering 1995; Nagarajan et al. 2013). Equivalent network structures refer to the same decomposition of the global distribution with different edge directions (*e.g.*, part A of Figure 1). Such network structures form a so called equivalent class of networks (Nagarajan et al. 2013).

3.3.3 Assumptions for causal interpretation of Bayesian networks

Nagarajan et al. (2013) explain which assumptions are necessary to learn causal effects instead of describing networks of conditional independence:

1. Each variable $\mathbf{X}_k \in \mathbf{X}$ used in BN analysis is conditionally independent of its non-effects, both direct and indirect, given its direct causes (causal Markov assumption). Non-effects are all variables having no direct or indirect causal effect on $\mathbf{X}_k$.

2. There must exist a network structure which is faithful to the dependence structure of $\mathbf{X}$, *i.e.*, the dependence structure of $\mathbf{X}$ must be decomposable into local distributions according to Equation 8 with a corresponding representation as DAG.

From these two assumptions it follows that there must not be any unobserved variables influencing the variables in the network (Nagarajan et al. 2013). Since confounders can hardly be excluded when dealing with experimental data, causal interpretation of BN is always a delicate issue.

Edwards (2000) illustrated that confounders can result in different networks in sub-groups of the data (Nagarajan et al. 2013). It therefore is important to measure robustness of network structures by sub-sampling from experimental data and measuring thereby uncertainty of edges and directions. This generally reduces the impact of locally optimal, but globally suboptimal networks, but might also indicate influences from unknown variables (Edwards 2000; Nagarajan et al. 2013). Therefore, a bootstrap approach was applied in this thesis that is explained in the next subsection.

3.3.4 Bootstrapping and averaging over networks

Bootstrapping is a broadly applied re-sampling strategy, with which one can empirically determine the distribution of a statistic, preserving structure of the data (*e.g.*, Good 2005). First, one samples a subset from observations of the random variable under investigation with replacement. Evaluations of the statistic from these samples then approximate the distribution of the respective statistic given the data and its structure (*e.g.*, Good 2005). With regard to learning BN, this means sampling from $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_d$ and learning BN from each sample (bootstrap networks). Appearance of edges and directions in the learned BN can be reported in percent and thereby uncertainty is measured (Scutari and Nagarajan 2011, 2013).

To infer one network from all bootstrap networks (*i.e.*, averaging over networks), arcs are oriented in the more preferred direction (in this thesis more than 250 out of 500 networks) and edges are included if their support in percent lies above a certain threshold. One can simply set such a threshold (*e.g.*, to 50% or 75%) or use the empirical distribution of support of all edges in the network for determining a statistically suggested threshold (Scutari and Nagarajan 2011, 2013). As sensitivity to the edge significance threshold was different among networks (results not shown), the latter approach was chosen in this thesis.

Scutari and Nagarajan (2011, 2013) proposed to consider all empirical probabilities (*i.e.*, frequencies) of the edges' appearance (percent) and order them in a vector

$$\hat{\boldsymbol{p}} = \left\{0 \leq \hat{p}_1 \leq \hat{p}_2 \ldots \leq \hat{p}_j \leq 1\right\} \tag{14}$$

with $j = 0.5 \cdot d(d-1)$ being the number of potential edges in the network (d being the number of variables). The significant edges have the larger values in $\hat{\boldsymbol{p}}$ (on the right-hand side of the vector). The true probability of an edge belonging to the BN is

$$\boldsymbol{p} = \{0, \ldots, 0, 1, \ldots, 1\} \tag{15}$$

In order to find a useful cut-off position in $\hat{\boldsymbol{p}}$ for deciding whether or not the respective edge belongs to the BN, the idea of Scutari and Nagarajan (2011, 2013) was to minimize the L_1 norm between

the cumulative distribution function (CDF) of the observed confidence levels ($\hat{\boldsymbol{p}}$) and the CDF of the confidence levels of the unknown BN ($\boldsymbol{p}$). He defined the empirical CDFs as

$$F_{\hat{\boldsymbol{p}}}(x) = \frac{1}{j}\sum_{i=1}^{j} \mathbb{1}_{\hat{\boldsymbol{p}}_i < x} \tag{16}$$

$$F_{\boldsymbol{p}}(x) = \begin{cases} 0 & \text{if } x \in (-\infty, 0) \\ t & \text{if } x \in [0, 1) \\ 1 & \text{if } x \in [1, \infty) \end{cases} \tag{17}$$

where $\mathbb{1}$ denotes the characteristic function and t corresponds to the significance threshold sought. Then they formulated the minimization problem as

$$\hat{t} = \underset{t \in [0,\,1]}{\operatorname{argmin}}\, L_1(t; \hat{\boldsymbol{p}}) \tag{18}$$

with

$$L_1(t; \hat{\boldsymbol{p}}) = \int \left| F_{\hat{\boldsymbol{p}}}(x) - F_{\boldsymbol{p}}(x; t) \right| dx = \sum_{x_i \,\in\, \{0\} \cup \hat{\boldsymbol{p}} \cup \{1\}} \left| F_{\hat{\boldsymbol{p}}}(x_i) - t \right| (x_{i+1} - x_i) \tag{19}$$

An edge is then significant if its value in $\hat{\boldsymbol{p}}$ is larger than $F^{-1}_{\hat{\boldsymbol{p}}}(\hat{t})$ (Scutari and Nagarajan 2011, 2013).

It was found in this thesis that one might obtain an inconsistent DAG (*e.g.*, containing cycles) from applying bootstrapping and then averaging the networks as described here. A manual correction is then necessary, advisably a minimal change such as re-directing an edge with uncertain direction (*i.e.*, both directions were appearing at the same or nearly same frequency in the bootstrap samples of the network).

3.3.5 Transformation of inter-dependent samples before learning a Gaussian Bayesian network

Scores and tests of BN learning algorithms assume independent observations, as mentioned above. In the MTM (*cf.* Equation 5), independence of residuals between genotypes was assumed. In contrast, the genomic components in $\boldsymbol{g}$ were correlated between genotypes due to kinship (represented by the matrix $\mathbf{K}$) in addition to being dependent within genotypes due to between-trait genomic covariances (represented by the matrix $\mathbf{G}$). Before learning BN from the genomic component, a transformation was applied to vector $\boldsymbol{g}$ such that the transformed vector $\boldsymbol{g}^*$ was distributed as $\mathcal{N}_{dn}(\mathbf{0}, \mathbf{G} \otimes \mathbb{I}_{n\times n})$, *i.e.*, elements of $\boldsymbol{g}^*$ were independent between genotypes while still preserving genomic relationships among traits as encoded by $\mathbf{G}$.

For this purpose, the kinship matrix $\mathbf{K}$ was decomposed into its Cholesky factors, $\mathbf{K} = \mathbf{L}\mathbf{L}'$, where $\mathbf{L}$ is an $(n \times n)$-dimensional lower triangular matrix. Define a $(dn \times dn)$-dimensional matrix $\mathbf{M} = \mathbb{I}_{d\times d} \otimes \mathbf{L}$, so that $\mathbf{M}^{-1} = \mathbb{I}_{d\times d} \otimes \mathbf{L}^{-1}$ and $(\mathbb{I}_{d\times d} \otimes \mathbf{L})\left(\mathbb{I}_{d\times d} \otimes \mathbf{L}^{-1}\right) = \mathbb{I}_{dn\times dn}$. Hence, for $\boldsymbol{g}^* = \mathbf{M}^{-1}\boldsymbol{g}$

$$\begin{aligned}\mathrm{Var}(\boldsymbol{g}^*) = \mathrm{Var}\left(\mathbf{M}^{-1}\boldsymbol{g}\right) &= \mathbf{M}^{-1}\mathrm{Var}(\boldsymbol{g})(\mathbf{M}^{-1})' = \mathbf{M}^{-1}(\mathbf{G}\otimes\mathbf{K})(\mathbf{M}^{-1})' \\ &= \left(\mathbb{I}_{d\times d} \otimes \mathbf{L}^{-1}\right)(\mathbf{G}\otimes\mathbf{K})\left(\mathbb{I}_{d\times d} \otimes \mathbf{L}^{-1}\right)' = \mathbf{G}\otimes\mathbb{I}_{n\times n} \end{aligned} \tag{20}$$

because $\mathbf{L}^{-1}\mathbf{K}(\mathbf{L}^{-1})' = \mathbb{I}_{n\times n}$, which follows directly from the definition of the Cholesky factor (*cf.* Equation A2). This transformation and the resulting covariance structure was used by Vázquez et al. (2010) for a single-trait model.

For a standard Cholesky decomposition of $\mathbf{K}$, a positive definite realized kinship matrix is needed. Here, realized kinship matrices were positive definite, but are not generally guaranteed to be positive definite. However, the proposed approach here can still be used as adjustments to assure positive definiteness are available (Nazarian and Gezan 2016).

3.3.6 Application of methodology to data

Here, the random variables investigated by BN were the estimates of the genomic and residual components of phenotypic traits in the MTM (*cf.* Equation 5), with the former transformed as described above. A Gaussian BN was chosen as a consequence of assuming that quantitative traits, and particularly genomic values and residuals, follow a normal distribution.

So the decomposition of the following global distributions was sought:

$$f(\boldsymbol{\Gamma}) = f(\boldsymbol{\Gamma}_{\cdot 1}, \boldsymbol{\Gamma}_{\cdot 2}, \ldots, \boldsymbol{\Gamma}_{\cdot 5}) = \prod_{k=1}^{5} f(\boldsymbol{\Gamma}_{\cdot k} \mid \mathrm{pa}(\boldsymbol{\Gamma}_{\cdot k})) \tag{21}$$

and

$$f(\mathbf{H}) = f(\mathbf{H}_{\cdot 1}, \mathbf{H}_{\cdot 2}, \ldots, \mathbf{H}_{\cdot 5}) = \prod_{k=1}^{5} f(\mathbf{H}_{\cdot k} \mid \mathrm{pa}(\mathbf{H}_{\cdot k})) \tag{22}$$

where $\boldsymbol{\Gamma}$ and $\mathbf{H}$ denote $(n \times d)$-dimensional matrices formed from the vectors $\boldsymbol{g}^* = \mathbf{M}^{-1}\boldsymbol{g}$ and $\boldsymbol{e}$ by arranging genotypes and traits in rows and columns, respectively. The index "$\cdot k$" denotes the k^{th} column of the respective matrix.

For learning BN, the GS algorithm (Margaritis 2003) was chosen as an example of a constraint-based approach, and TABU (Daly and Shen 2007) as an example of a score-based algorithm. For comparing the constraint-based with the score-based approach, these two algorithms are suited when dealing with small networks of up to 10–15 traits. For larger networks, runtime-optimized network learning algorithms might be preferred.

Table 2: BN learning settings: tests and scores used for GS and TABU

Abbreviation for Setting	Measure of Conditional Independence	Test / Score
Grow-Shrink (constraint-based)		
GS 1	partial correlation coefficient (Pearson)	exact Student's t test
GS 2	partial correlation coefficient (Pearson)	MC permutation test (Legendre 2000)
GS 3	mutual information	asymptotic χ^2 test (Kullback 1959)
GS 4	mutual information	MC permutation test
TABU (score-based)		
TABU 1		BIC (Lam and Bacchus 1994; Rissanen 1978)
TABU 2		BGE (Geiger and Heckerman 1994)

The GS algorithm was combined with four different tests for marginal and conditional independence and TABU was combined with two scores (see Table 2). Scores and tests were described above. Regarding the constraint-based approaches, the number of permutations used for each of the permutation tests was 450 (slightly more than half of the genotypes in each of the two panels).

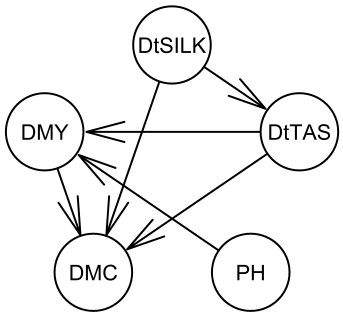
Multiple testing issues in the context of a constraint-based algorithm have been addressed by Aliferis et al. (2010a,b). According to these authors, for a sample size of about 1000, a significance level of $\alpha = 0.05$ resulted in a worst-case false positive rate of 6×10^{-4} for arc inclusion. Choosing $\alpha = 0.01$ in the more exhaustive GS, false positives should be avoided effectively with five variables (traits) and about 800 genotypes.

Each of the six settings was run with 500 bootstrap samples. After structure learning from the bootstrap samples, an averaged network was derived from the 500 bootstrap networks as described above using the R package *bnlearn* (Scutari 2010; Nagarajan et al. 2013).

3.3.7 Translation of learned structures from Bayesian networks to a structure matrix

The structure learned by the BN needed to be translated into a structure matrix for SEM analysis. Thus, a $(d \times d)$-dimensional matrix $\mathbf{\Lambda}$ was formed from each learned structure. Each arrow of the DAG pointing from a parent trait to a child trait becomes a freely varying coefficient in the structure matrix in the column of the parent and the row of the child. All other entries were set to zero *a priori*. The resulting structure matrices are lower triangular matrices, since loops are not allowed in a BN structure. The lower triangular form might not be produced by an arbitrary order of traits but can

always be formed by rearranging the order of the traits in columns and rows. Figure 2 shows an example of how a DAG is transformed into a structure matrix.



$$\mathbf{\Lambda} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \lambda_{\text{DtTAS, DtSILK}} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{\text{DMY, PH}} & \lambda_{\text{DMY, DtTAS}} & 0 & 0 \\ \lambda_{\text{DMC, DtSILK}} & 0 & \lambda_{\text{DMC, DtTAS}} & \lambda_{\text{DMC, DMY}} & 0 \end{bmatrix}$$

Figure 2: Transformation from a DAG to a structure matrix $\mathbf{\Lambda}$. An example of a directed, acyclic graph and its respective structure matrix is shown. The coefficients are ordered as DtSILK – PH – DtTAS – DMY – DMC.

If a learned structure was an element of an equivalent class of several structures, only one representative of this equivalent class was translated into a structure matrix and assessed in SEM analyses. SEM based on different $\mathbf{\Lambda}$ from the same equivalent class of networks may lead to different sets of coefficients, however, they lead to the same model fit of the respective SEM.

3.4 Fitting structural equation models

The following paragraph presents how the structures found in the BN analysis translate into SEM; doing so also clarifies how MTM (*cf.* Equation 5) and SEM are related. Genomic or residual trait dependence can be described by the recursive linear regression models

$$\begin{aligned} \boldsymbol{g}^* &= \left(\mathbf{\Lambda}_{\boldsymbol{g}^*} \otimes \mathbb{I}_{n\times n}\right)\boldsymbol{g}^* + \boldsymbol{\rho}_{\boldsymbol{g}^*} \\ \boldsymbol{e} &= \left(\mathbf{\Lambda}_{\boldsymbol{e}} \otimes \mathbb{I}_{n\times n}\right)\boldsymbol{e} + \boldsymbol{\rho}_{\boldsymbol{e}} \end{aligned} \tag{23}$$

where $\mathbf{\Lambda}_{\boldsymbol{g}^*}$ and $\mathbf{\Lambda}_{\boldsymbol{e}}$ denote the $(d \times d)$-dimensional genomic and residual structure matrices and $\boldsymbol{\rho}_{\boldsymbol{g}^*}$ ($\boldsymbol{\rho}_{\boldsymbol{e}}$) are vectors of dn independent and identically normal-distributed residuals. Independence among genotypes in $\boldsymbol{\rho}_{\boldsymbol{g}^*}$ follows from the transformation of $\boldsymbol{g}$ into $\boldsymbol{g}^*$; elements of $\boldsymbol{\rho}_{\boldsymbol{e}}$ are independent since $\boldsymbol{e}$ are independent among genotypes by the model assumptions of the MTM (*cf.* Equation 5). If the genomic and residual structures in Equations 23 explain all genomic and residual trait covariances fully, then $\boldsymbol{\rho}_{\boldsymbol{g}^*}$ and $\boldsymbol{\rho}_{\boldsymbol{e}}$ are also independent between traits.

For genotype i Equations 23 are

$$\begin{aligned} \boldsymbol{g}_i^* &= \mathbf{\Lambda}_{\boldsymbol{g}^*}\boldsymbol{g}_i^* + (\boldsymbol{\rho}_{\boldsymbol{g}^*})_i \\ \boldsymbol{e}_i &= \mathbf{\Lambda}_{\boldsymbol{e}}\boldsymbol{e}_i + (\boldsymbol{\rho}_{\boldsymbol{e}})_i \end{aligned} \tag{24}$$

Equations 24 can be rearranged into

$$\begin{aligned} \boldsymbol{g}_i^* &= \left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{g}^*} \right)^{-1} (\boldsymbol{\rho}_{\boldsymbol{g}^*})_i \\ \boldsymbol{e}_i &= \left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{e}} \right)^{-1} (\boldsymbol{\rho}_{\boldsymbol{e}})_i, \end{aligned} \tag{25}$$

which imply that

$$\begin{aligned} \mathrm{Var}(\boldsymbol{g}_i^*) &= \mathrm{Var}\left(\left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{g}^*} \right)^{-1} (\boldsymbol{\rho}_{\boldsymbol{g}^*})_i \right) = \left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{g}^*} \right)^{-1} \mathrm{Var}\left((\boldsymbol{\rho}_{\boldsymbol{g}^*})_i \right) \left(\left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{g}^*} \right)^{-1} \right)' \\ \mathrm{Var}(\boldsymbol{e}_i) &= \mathrm{Var}\left(\left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{e}} \right)^{-1} (\boldsymbol{\rho}_{\boldsymbol{e}})_i \right) = \left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{e}} \right)^{-1} \mathrm{Var}\left((\boldsymbol{\rho}_{\boldsymbol{e}})_i \right) \left(\left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{e}} \right)^{-1} \right)' \end{aligned} \tag{26}$$

where $\mathrm{Var}\left((\boldsymbol{\rho}_{\boldsymbol{g}^*})_i \right)$ and $\mathrm{Var}\left((\boldsymbol{\rho}_{\boldsymbol{e}})_i \right)$ are unknown diagonal matrices. Note that $\mathrm{Var}(\boldsymbol{g}^*{}_i) = \mathrm{Var}(\boldsymbol{g}_i)$ when considering individual genotypes and, therefore, $\boldsymbol{\Lambda}_{\boldsymbol{g}^*} = \boldsymbol{\Lambda}_{\boldsymbol{g}}$.

Let $\boldsymbol{\Psi}_{\boldsymbol{g}} = \mathrm{Var}\left((\boldsymbol{\rho}_{\boldsymbol{g}^*})_i \right)$ and $\boldsymbol{\Psi}_{\boldsymbol{e}} = \mathrm{Var}\left((\boldsymbol{\rho}_{\boldsymbol{e}})_i \right)$ for all i, *i.e.*, assume that both genomic and residual variances of traits are the same for all genotypes, so the genomic and residual covariance matrices $\mathbf{G}$ and $\mathbf{E}$ can be replaced by

$$\begin{aligned} \mathbf{G}^* &= \mathrm{Var}(\boldsymbol{g}^*{}_i) = \mathrm{Var}(\boldsymbol{g}_i) = \left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{g}} \right)^{-1} \boldsymbol{\Psi}_{\boldsymbol{g}} \left(\left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{g}} \right)^{-1} \right)' \quad \forall i \\ \mathbf{E}^* &= \mathrm{Var}(\boldsymbol{e}_i) = \left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{e}} \right)^{-1} \boldsymbol{\Psi}_{\boldsymbol{e}} \left(\left(\mathbb{I}_{d\times d} - \boldsymbol{\Lambda}_{\boldsymbol{e}} \right)^{-1} \right)' \quad \forall i \end{aligned} \tag{27}$$

Since the true structure matrices $\boldsymbol{\Lambda}_{\boldsymbol{g}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{e}}$ are unknown, the inferred structures from the BN are used instead.

Then, the MTM (*cf.* Equation 5) can be reformulated as the following SEM in its reduced form (Gianola and Sorensen 2004; Valente et al. 2010):

$$\boldsymbol{y} = \boldsymbol{\mu} \otimes \mathbf{1}_n + \boldsymbol{g} + \boldsymbol{e} \tag{28}$$

where genomic values and residuals are assumed to be independently distributed according to

$$\begin{aligned} \boldsymbol{g} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{G}^* \otimes \mathbf{K}) \\ \boldsymbol{e} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{E}^* \otimes \mathbb{I}_{n\times n}) \end{aligned} \tag{29}$$

The Bayesian Gibbs sampling implementation is as for the MTM (*cf.* Equation 5), but, additionally, $\mathbf{G}^*$ ($\mathbf{E}^*$) is sampled from its posterior distribution based on Equations 27. We fitted SEM (*cf.* Equation 28) with both or only one structured component, *i.e.*, with $\mathbf{G}^*$ and $\mathbf{E}^*$, $\mathbf{G}^*$ and $\mathbf{E}$, or $\mathbf{G}$ and $\mathbf{E}^*$ for assessing the structures and comparison with the MTM (*cf.* Equation 5).

The structures of $\mathbf{\Lambda}_g$ and $\mathbf{\Lambda}_e$ affect whether or not the use of $\mathbf{G}^*$ and $\mathbf{E}^*$ instead of $\mathbf{G}$ and $\mathbf{E}$ induces a reduction in the number of parameters. For example, $\mathbf{G}^*$ and $\mathbf{G}$ ($\mathbf{E}^*$ and $\mathbf{E}$) have the same number of nonnull parameters if the structure is fully recursive.

The diagonal elements of matrices $\mathbf{\Psi}_g$ and $\mathbf{\Psi}_e$ in SEM (*cf.* Equations 27) were assigned independent scaled-inverse χ^2 prior distributions with four degrees of freedom and a scale of three. As mentioned above, this choice of hyper-parameters yields the same prior distribution as for the MTM (*cf.* Equation 5) with respect to the diagonal elements of the covariance matrices. Also, the trait variances' prior mode is again 0.5 and the priors are relatively uninformative. The nonnull coefficients of $\mathbf{\Lambda}_g$ and $\mathbf{\Lambda}_e$ in SEM (*cf.* Equations 27) were assigned independent normal distributions with null mean and variance of 6.3 and 0.07, respectively.

The prior distribution's variance for the nonnull coefficients of $\mathbf{\Lambda}_g$ and $\mathbf{\Lambda}_e$ in SEM (*cf.* Equations 27) was chosen to align with the estimated genomic (residual) covariance components in the MTM (*cf.* Equation 7), which lay in the interval between -3.7 (-0.13) and 5.0 (0.52). When a normal prior distribution with null mean is assumed for the SEM structure matrices' coefficients, then 5.0 (0.52) lies within two times its standard deviation (SD) when its variance (a hyper-parameter) is chosen to be larger than $\left(\frac{5}{2}\right)^2 = 6.25 \left[\left(\frac{0.52}{2}\right)^2 = 0.0676\right]$. This assures that the prior distribution for the structure matrices' coefficients are flat enough to impose no additional prior restriction compared to the prior assumptions of the MTM (*cf.* Equation 5).

3.5 Summary and overview

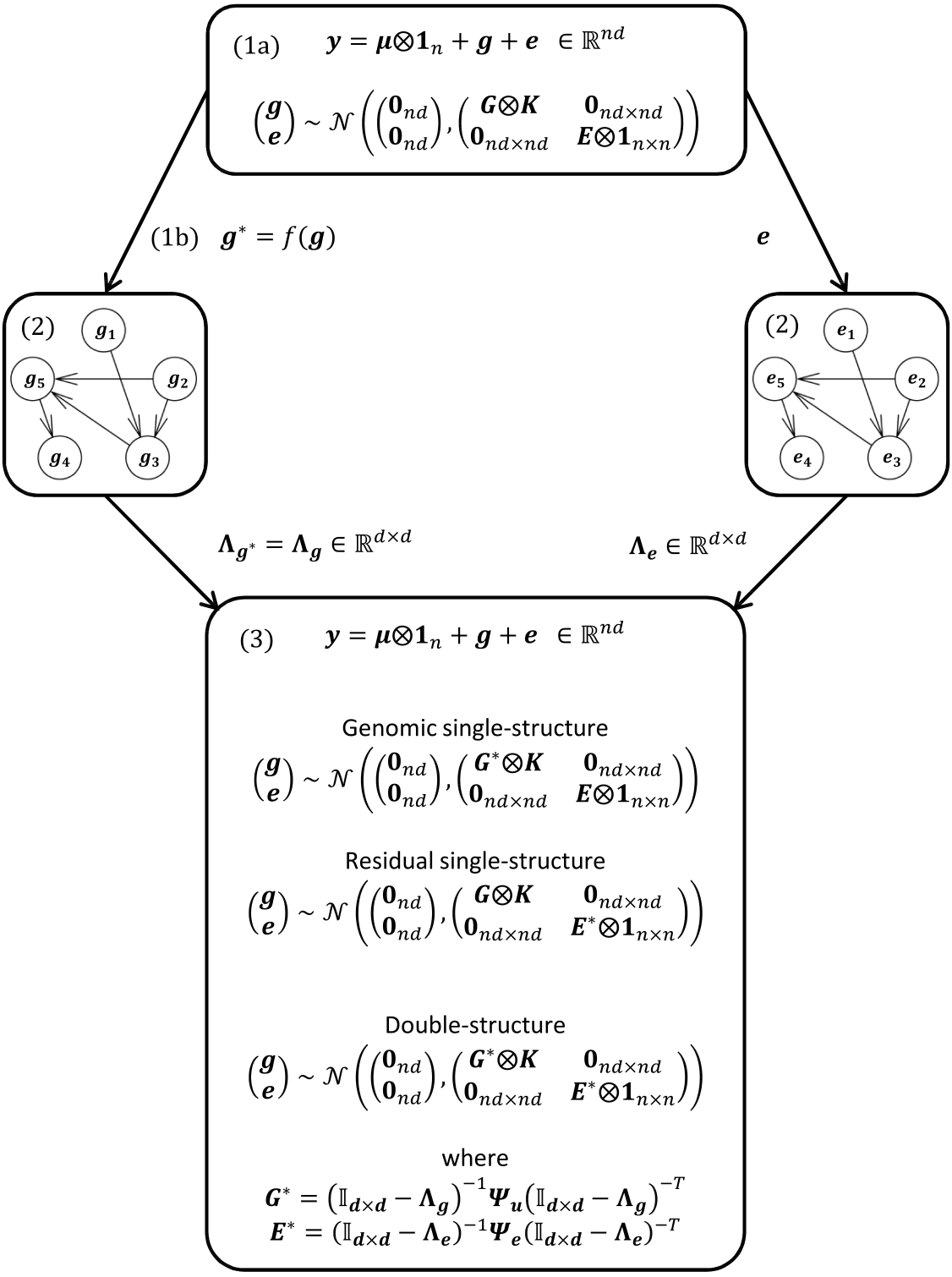


Figure 3: Investigating the structure of covariance components: (1a) fitting a Bayesian multiple-trait Gaussian model (*cf.* Equation 5) to all five traits to obtain posterior means of genomic values and of model residuals for each panel, and (1b) transforming the predicted genomic values to meet assumptions on sample independence required by a BN analysis; (2) applying the BN analysis to the residual and transformed genomic trait components (Equations 21 and 22); and (3) assessing the quality of the inferred structures by a structure-comparative SEM analysis (*cf.* Equation 28).

4 Methods for learning components of environmental covariances

In the following subsections, the methods for decomposing the environmental trait covariance into micro- and macro-environmental components are described. Micro-environmental factors are jointly and indirectly modeled using physical distance between plots in the experiments via a single spatial component. First, a bandwidth parameter for mapping physical distance into spatial correlation coefficients needs to be estimated for every field of the experiment separately. This is done by embedding a Metropolis-Hastings algorithm in an existing multivariate Gibbs sampler, where an R implementation was developed for the Metropolis-Hastings step in this thesis. Second, the posterior means of these bandwidth parameters and the respective spatial correlation coefficients are used to form a micro-environmental component in a one-stage approach across all fields of the experiment. In this one-stage model, macro-environmental factors are modeled by locations and, consequently, micro- and macro-environmental trait covariances can be estimated separately in this model. At the end of this section, the workflow of this methodology is summarized.

4.1 Estimating the spatial bandwidth parameter per field

Scaled phenotypic values of d traits and n genotypes in N plots were analyzed separately for each field by using the model

$$\boldsymbol{y} = \mathbf{1}_n \otimes \boldsymbol{\mu} + \boldsymbol{g} + \boldsymbol{e} \tag{30}$$

where $\boldsymbol{y}$, $\boldsymbol{g}$, and $\boldsymbol{e}$ are $(nd \times 1)$-dimensional column vectors (phenotypic values, genomic values, and residuals, respectively), all sorted by plot and trait within plot. $\mathbf{1}_n \otimes \boldsymbol{\mu}$ is a Kronecker product of an $(n \times 1)$-dimensional vector of ones and a $(d \times 1)$-dimensional intercept vector $\boldsymbol{\mu}$. No fixed effects except the d elements in $\boldsymbol{\mu}$ were included, and flat priors were assigned to these elements. Note that $n = N$ here as no genotype is replicated within a field, see Table A2 for genotype (plot) numbers in the respective fields. The model described in Equation 30 was applied to the data of each field separately, with the response being raw phenotypic values per plot (instead of adjusted means for each genotype as in the MTM, *cf.* Equation 5).

For $\boldsymbol{g}$ and $\boldsymbol{e}$ the following mutually independent distributions were assumed:

$$\begin{aligned} \boldsymbol{g} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{K} \otimes \mathbf{G}) \\ \boldsymbol{e} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbb{I}_{n \times n} \otimes \mathbf{E}) \end{aligned} \tag{31}$$

Here, $\mathbf{K}$ is an $(n \times n)$-dimensional realized kinship matrix estimated from the genomic data according to Astle and Balding (2009). Note that in this model ordering of elements within response and

covariable vectors is different than in the MTM (*cf.* Equation 5) and, therefore, the factors of the Kronecker product forming the respective covariance matrix of the prior distributions are switched.

The matrices $\mathbf{G}$ and $\mathbf{E}$ were assumed to follow *a priori* independent inverse-Wishart distributions:

$$\begin{aligned} \mathbf{G} &\sim \mathscr{W}^{-1}(\mathbf{\Phi}_{\mathbf{G}}, \nu_{\mathbf{G}}) \\ \mathbf{E} &\sim \mathscr{W}^{-1}(\mathbf{\Phi}_{\mathbf{E}}, \nu_{\mathbf{E}}) \end{aligned} \tag{32}$$

For obtaining prior modes for the genomic and residual trait variances of 0.5 for each trait and assuming a large variance of the prior distribution densities of $\mathbf{G}$ and $\mathbf{E}$, $\nu_{\mathbf{G}} = \nu_{\mathbf{E}} = 8$ and $\mathbf{\Phi}_{\mathbf{G}} = \mathbf{\Phi}_{\mathbf{E}} = \frac{14}{2} \cdot \mathbb{I}_{d \times d}$ were set with ν and $\mathbf{\Phi}$ in accordance with the parametrization of the respective distributions in the *BGLR* package and *MTM* function.

Adding a spatial component into Equation 30, one obtains

$$\boldsymbol{y} = \mathbf{1}_n \otimes \boldsymbol{\mu} + \boldsymbol{g} + \boldsymbol{s} + \boldsymbol{e} \tag{33}$$

where $\boldsymbol{s}$ is an $(nd \times 1)$-dimensional column vector of spatial effects, also sorted by plot and trait within plot. Independence of $\boldsymbol{s}$ from $\boldsymbol{g}$ and $\boldsymbol{e}$ and the following prior distribution were assumed for $\boldsymbol{s}$

$$\boldsymbol{s} \sim \mathscr{N}_{nd}(\mathbf{0}, \mathbf{S}_\vartheta \otimes \mathbf{\Sigma}) \tag{34}$$

where $\mathbf{\Sigma}$ is a $(d \times d)$-dimensional positive-definite symmetric matrix containing spatial variances of traits and covariances between traits. $\mathbf{S}_\vartheta$ is an $(n \times n)$-dimensional Gaussian kernel matrix, of which the entries depended on an unknown bandwidth parameter ϑ and on the physical distances between plots.

An element of $\mathbf{S}_\vartheta$ for plots i and j is

$$(\mathbf{S}_\vartheta)_{ij} = \exp(-\vartheta \cdot \mathrm{d}(i,j)^2) \tag{35}$$

where $\mathrm{d}(i,j)$ is the distance (meter) between the centers of plots i and j. ϑ was assigned a Gamma prior distribution (*cf.* Pérez-Elizalde et al. 2015) with shape parameter k and scale parameter ψ as hyper-parameters, *i.e.*,

$$\vartheta \sim \Gamma(k, \psi) \tag{36}$$

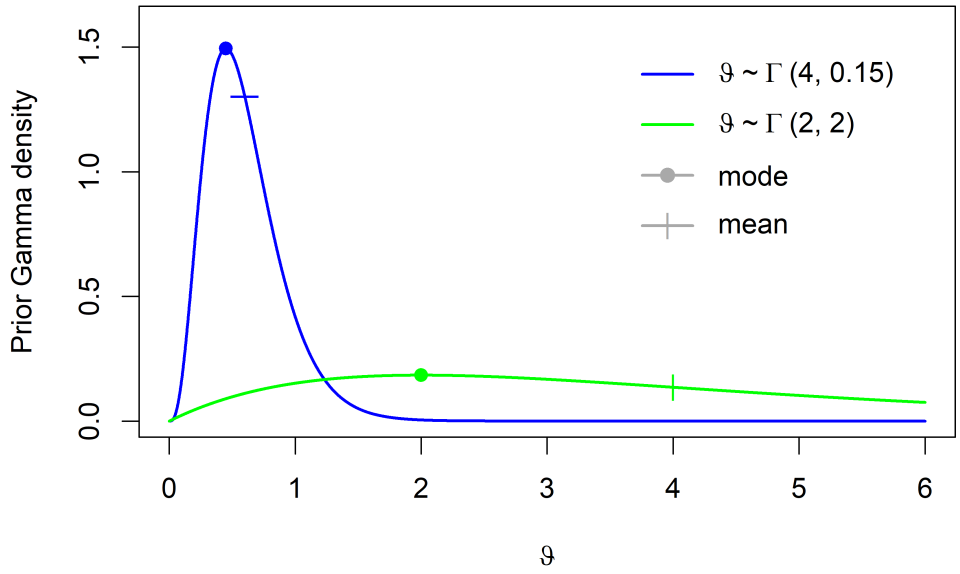


Figure 4: Density of two prior Gamma distributions for the bandwidth parameter ϑ

For assessing sensitivity with respect to the choice of hyper-parameters of the model described by Equation 33, two prior distributions of ϑ were compared. These were $\vartheta \sim \Gamma(k = 2, \psi = 2)$ (weak) and $\vartheta \sim \Gamma(k = 4, \psi = 0.15)$ (strong). The respective prior mode (mean) of ϑ was then 2 (4) and 0.45 (0.60), respectively (see Figure 4 for an illustration).

Prior independence of $\boldsymbol{\Sigma}$ from $\mathbf{G}$ and $\mathbf{E}$ was assumed and the following prior distribution was assigned to $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{\Sigma}}, \nu_{\boldsymbol{\Sigma}}) \tag{37}$$

where $\nu_{\mathbf{G}} = \nu_{\mathbf{E}} = \nu_{\boldsymbol{\Sigma}} = 8$ and $\boldsymbol{\Phi}_{\mathbf{G}} = \boldsymbol{\Phi}_{\mathbf{E}} = \boldsymbol{\Phi}_{\boldsymbol{\Sigma}} = \frac{14}{3} \cdot \mathbb{I}_{d \times d}$, resulting in modes of $\frac{1}{3}$ for the variance of each trait and random component (including the error term).

4.1.1 Multivariate Gibbs Sampler

A Gibbs sampling approach (with an MH algorithm embedded for sampling ϑ) was used to infer features of the posterior distributions of parameters in the models described by Equations 30 and 33 (Algorithms A1 and 1). After a burn-in of 200,000 iterations, the MCMC analysis was run for an additional 200,000 iterations. 100,000 MCMC samples were stored after thinning using a factor of two.

The Gibbs sampler from the R package *MTM* uses a quasi-orthogonal representation (de los Campos 2015) for each of the random effects, which is based on the eigenvalue decomposition of the kernel

matrices of each of the random effects. Let

$$\boldsymbol{y} = \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{u}_k + \boldsymbol{e} \tag{38}$$

where

$$\begin{aligned} \boldsymbol{u}_k &\sim \mathcal{N}_{nd}(\mathbf{0}, \boldsymbol{\Omega}_k \otimes \mathbf{V}_k) \\ \boldsymbol{e} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbb{I}_{n\times n} \otimes \mathbf{E}) \\ \mathbf{V}_k &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_k, \nu_k) \\ \mathbf{E} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}, \nu) \end{aligned} \tag{39}$$

be the model under investigation with intercept vector $\boldsymbol{\mu}$, p random effects $u_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_k \otimes \mathbf{V}_k)$ with kernel $\boldsymbol{\Omega}_k$ and covariance matrix $\mathbf{V}_k$, error term $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_{n\times n} \otimes \mathbf{E})$, and inverse-Wishart prior distributions assigned to each of the covariance matrices. $\boldsymbol{\Omega}_k$ could be a kinship matrix or a spatial kernel, $\mathbf{V}_k$ would then be the respective genomic or spatial trait covariance matrix.

A quasi-orthogonal representation of Equation 38 is given by using an eigenvalue decomposition of each of the kernel matrices $\boldsymbol{\Omega}_k$ (de los Campos 2015). Let m_k be the number of positive eigenvalues of the respective kernel matrix. The kernels are then decomposed as $\boldsymbol{\Omega}_k = \boldsymbol{\Lambda}_k \boldsymbol{\Psi}_k \boldsymbol{\Lambda}_k{}'$, where $\boldsymbol{\Psi}_k$ are $(m_k \times m_k)$-dimensional diagonal matrices containing m_k eigenvalues (> 0) of kernel $\boldsymbol{\Omega}_k$ on the diagonal, and $\boldsymbol{\Lambda}_k$ are $(n \times m_k)$-dimensional matrices of eigenvectors satisfying $\boldsymbol{\Lambda}_k{}'\boldsymbol{\Lambda}_k = \mathbb{I}_{m_k \times m_k}$ (de los Campos 2015). In the implementation of de los Campos (2015), all eigenvalues smaller than a small positive scalar are considered to be zero, hence the naming quasi-orthogonal.

In de los Campos (2015) it was shown that the distribution of $\boldsymbol{u}_k$ equals that of $\tilde{\boldsymbol{u}}_k = (\boldsymbol{\Lambda}_k \otimes \mathbb{I}_{d\times d})\boldsymbol{\nu}_k$ with $\boldsymbol{\nu}_k \sim \mathcal{N}_{m_k d}(\mathbf{0}, \boldsymbol{\Psi}_k \otimes \mathbf{V}_k)$ and $m_k \leq n$, and

$$\mathbb{E}(\tilde{\boldsymbol{u}}_k) = (\boldsymbol{\Lambda}_k \otimes \mathbb{I}_{d\times d})\mathbb{E}(\boldsymbol{\nu}_k) = \mathbf{0} \tag{40}$$

$$\mathrm{Var}(\tilde{\boldsymbol{u}}_k) = (\boldsymbol{\Lambda}_k \otimes \mathbb{I}_{d\times d})\mathrm{Var}(\boldsymbol{\nu}_k)(\boldsymbol{\Lambda}_k{}' \otimes \mathbb{I}_{d\times d}) = (\boldsymbol{\Lambda}_k \boldsymbol{\Psi}_k \boldsymbol{\Lambda}_k{}') \otimes \mathbf{V}_k = \boldsymbol{\Omega}_k \otimes \mathbf{V}_k \tag{41}$$

Then, the model represented by Equation 38 can be formulated as (de los Campos 2015)

$$\boldsymbol{y} = \boldsymbol{\mu} + \sum_{k=1}^{p} (\boldsymbol{\Lambda}_k \otimes \mathbb{I}_{d\times d})\boldsymbol{\nu}_k + \boldsymbol{e} \tag{42}$$

where

$$
\begin{aligned}
\boldsymbol{\nu}_k &\sim \mathcal{N}_{nd}(\mathbf{0}, \boldsymbol{\Psi}_k \otimes \mathbf{V}_k) \\
\boldsymbol{e} &\sim \mathcal{N}_{nd}(\mathbf{0},\ \mathbb{I}_{n\times n} \otimes \mathbf{E}) \\
\mathbf{V}_k &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_k,\ \nu_k) \\
\mathbf{E} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi},\ \nu)
\end{aligned}
\tag{43}
$$

As $\boldsymbol{\Psi}_k$ are diagonal matrices, sampling from the multivariate normal distributions of $\boldsymbol{\nu}_k$ is greatly facilitated (de los Campos 2015). The Gibbs sampler draws from the joint posterior distribution by sampling from a row of fully conditional distributions (Algorithm A1) (de los Campos 2015).

4.1.2 Embedding a Metropolis-Hastings algorithm

The fully conditional posterior distribution of ϑ is (Pérez-Elizalde et al. 2015)

$$
\mathrm{p}(\vartheta \mid \text{else}) \propto \mathrm{p}(\boldsymbol{s} \mid \vartheta) \cdot \mathrm{p}(\vartheta) = \mathcal{N}_{nd}(\mathbf{0}, \mathbf{S}_\vartheta \otimes \boldsymbol{\Sigma}) \cdot \Gamma(k, \psi) \tag{44}
$$

In this thesis, the fully conditional posterior distribution of ϑ was represented with a transformed spatial component for runtime acceleration, *i.e.*, the implementation of Equation 33 uses

$$
p(\vartheta \mid \text{else}) \propto \mathrm{p}(\boldsymbol{s} \mid \vartheta) \cdot \mathrm{p}(\vartheta) \propto \mathrm{p}(\boldsymbol{s}^* \mid \vartheta) \cdot \mathrm{p}(\vartheta) = \mathcal{N}_{nd}(\mathbf{0}, \mathbf{S}_\vartheta \otimes \mathbb{I}_{d\times d}) \cdot \Gamma(k, \psi) \tag{45}
$$

Derivation of a suitable $\boldsymbol{s}^*$ and proof of Equation 45 can be found in the Appendix. Sampling was done by an MH step (implemented for this thesis, see Algorithm 1 and code in Appendix) embedded in the *MTM* function. As $\vartheta > 0$, its proposal distribution cannot be symmetric and a Metropolis algorithm would draw biased samples, hence an MH approach was necessary.

Algorithm 1 MH algorithm for the bandwidth parameter ϑ

Let $\vartheta^{(t)}$ be the current state of ϑ and σ_ϑ^2 an arbitrary, but positive scalar.

Draw $\vartheta^* \sim \mathcal{N}_{(0,\infty)}(\vartheta^{(t)}, \sigma_\vartheta^2)$. ▷ proposal

Compute $\mathbf{S}_{\vartheta^*}$ from $(\mathbf{S}_{\vartheta^*})_{ij} = \exp(-\vartheta^* \cdot \mathrm{d}(i,j)^2)$. ▷ *cf.* Equation 35

Compute the densities of ϑ^* and $\vartheta^{(t)}$ by evaluating the conditional posteriors

$$\mathrm{p}(\vartheta^* \mid \text{else}) \propto \mathrm{p}(\boldsymbol{s}^* \mid \vartheta^*) \cdot \mathrm{p}(\vartheta^*)$$
$$\mathrm{p}(\vartheta^{(t)} \mid \text{else}) \propto \mathrm{p}(\boldsymbol{s}^* \mid \vartheta^{(t)}) \cdot \mathrm{p}(\vartheta^{(t)})$$

▷ fully conditional posterior

Compute the probability of moving to a new state

$$\alpha = \min\left(\frac{\mathrm{p}(\vartheta^* \mid \text{else}) \cdot \mathrm{p}(\vartheta^{(t)} \mid \vartheta^{(t)} \sim \mathcal{N}_{(0,\infty)}(\vartheta^*, \sigma_\vartheta^2))}{\mathrm{p}(\vartheta^{(t)} \mid \text{else}) \cdot \mathrm{p}(\vartheta^* \mid \vartheta^* \sim \mathcal{N}_{(0,\infty)}(\vartheta^{(t)}, \sigma_\vartheta^2))},\ 1\right)$$

Draw $x \sim \mathrm{U}(0,1)$.

if $x \leq \alpha$ **then**
  $\vartheta^{(t+1)} \leftarrow \vartheta^*$ and $\mathbf{S}_\vartheta{}^{(t+1)} \leftarrow \mathbf{S}_{\vartheta^*}$
else
  $\vartheta^{(t+1)} \leftarrow \vartheta^{(t)}$ and $\mathbf{S}_\vartheta{}^{(t+1)} \leftarrow \mathbf{S}_{\vartheta^{(t)}}$.
end if

4.2 Estimating micro- and macro-environmental trait covariances across all fields

A Bayesian multiple-trait Gaussian model was fitted to d traits and N plots of all fields of the Flint panel. This one-stage model can be expressed as

$$\boldsymbol{y}_{tprlg} = \boldsymbol{\mu}_t + \boldsymbol{loc}_{tl} + \boldsymbol{field}_{tr(l)} + \boldsymbol{s}_{tp(rl)g} + \boldsymbol{g}_{tg} + \boldsymbol{e}_{tprlg} \tag{46}$$

where $\boldsymbol{y}$ is the scaled vector of phenotypic values of all plots, $\boldsymbol{\mu}$ is a fixed intercept, $\boldsymbol{loc}$ and $\boldsymbol{field}$ are random components for the locations and fields, $\boldsymbol{s}$ is the spatial random component, $\boldsymbol{g}$ is the genomic random component, and $\boldsymbol{e}$ is the error term. Fields are nested within locations, and plots are nested within fields. Parameters and levels are summarized in Table 3. Scaling of the response variable was done by mean-centering and using the estimate of the sample variance.

A priori, the following mutually independent distributions were assumed:

$$\begin{aligned}
\boldsymbol{\mu} &\propto 1 \\
\boldsymbol{loc} &\sim \mathcal{N}_{6d}(\mathbf{0},\, \mathbb{I}_{6\times 6} \otimes \mathbf{V}_{\boldsymbol{loc}}) \\
\boldsymbol{field} &\sim \mathcal{N}_{12d}(\mathbf{0},\, \mathbb{I}_{12\times 12} \otimes \mathbf{V}_{\boldsymbol{field}}) \\
\boldsymbol{s} &\sim \mathcal{N}_{Nd}(\mathbf{0}, \mathbf{S} \otimes \mathbf{V}_{\boldsymbol{s}}) \\
\boldsymbol{g} &\sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{K} \otimes \mathbf{V}_{\boldsymbol{g}}) \\
\boldsymbol{e} &\sim \mathcal{N}_{Nd}(\mathbf{0},\, \mathbb{I}_{N\times N} \otimes \mathbf{V}_{\boldsymbol{e}})
\end{aligned} \tag{47}$$

$\mathbf{K}$ is again an $(n \times n)$-dimensional realized kinship matrix estimated from the genomic data (Astle and Balding 2009) as in the models described by Equations 30 and 33. $\mathbf{V}_{\boldsymbol{loc}}$, $\mathbf{V}_{\boldsymbol{field}}$, $\mathbf{V}_{\boldsymbol{s}}$, $\mathbf{V}_{\boldsymbol{g}}$, and $\mathbf{V}_{\boldsymbol{e}}$ are the location, field, spatial, genomic, and residual trait covariance matrices.

The bandwidth parameter ϑ for each field was assumed to be known in the model described by Equation 46. Values of ϑ for each field were determined as posterior means (denoted by $\bar{\vartheta}$) when

Table 3: Number of levels in the one-stage model

Index	Meaning	Levels Flint	Range
t	trait	5	$1:d$
p	plot	6323	$1:N$
b	block	960	—
r	field	12	—
l	location	6	—
g	genotype	805	$1:n$

estimating parameters in the model described by Equation 33. When $\boldsymbol{y}$ is sorted by field, and then, within each field, sorted by plot and trait within plot (*cf.* Equations 30 and 33), $\boldsymbol{S}$ in Equation 47 is a block diagonal matrix formed from the posterior means of $\mathbf{S}_{\bar{\vartheta}}$ for each of the twelve fields. $\boldsymbol{S}$ would then be

$$\boldsymbol{S} = \begin{bmatrix} \mathbf{S}_{\bar{\vartheta},1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\bar{\vartheta},2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{\bar{\vartheta},12} \end{bmatrix} \tag{48}$$

All trait covariance matrices $\mathbf{V}_{\boldsymbol{loc}}$, $\mathbf{V}_{\boldsymbol{field}}$, $\mathbf{V}_{\boldsymbol{s}}$, $\mathbf{V}_{\boldsymbol{g}}$, and $\mathbf{V}_{\boldsymbol{e}}$ were again assumed to follow *a priori* independent inverse-Wishart distributions:

$$\begin{aligned} \mathbf{V}_{\boldsymbol{loc}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{loc}}, \nu_{\boldsymbol{loc}}) \\ \mathbf{V}_{\boldsymbol{field}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{field}}, \nu_{\boldsymbol{field}}) \\ \mathbf{V}_{\boldsymbol{s}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{s}}, \nu_{\boldsymbol{s}}) \\ \mathbf{V}_{\boldsymbol{g}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{g}}, \nu_{\boldsymbol{g}}) \\ \mathbf{V}_{\boldsymbol{e}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{e}}, \nu_{\boldsymbol{e}}) \end{aligned} \tag{49}$$

For equivalent shrinkage on all components, $\nu_{\boldsymbol{loc}} = \nu_{\boldsymbol{field}} = \nu_{\boldsymbol{s}} = \nu_{\boldsymbol{g}} = \nu_{\boldsymbol{e}} = 8$ and $\boldsymbol{\Phi}_{\boldsymbol{loc}} = \boldsymbol{\Phi}_{\boldsymbol{field}} = \boldsymbol{\Phi}_{\boldsymbol{s}} = \boldsymbol{\Phi}_{\boldsymbol{g}} = \boldsymbol{\Phi}_{\boldsymbol{e}} = \frac{14}{5} \cdot \mathbb{I}_{d \times d}$.

For comparison, the spatial component was removed from the model above (*cf.* Equation 46) producing the model

$$\boldsymbol{y}_{tprlg} = \boldsymbol{\mu}_t + \boldsymbol{loc}_{tl} + \boldsymbol{field}_{tr(l)} + \boldsymbol{g}_{tg} + \boldsymbol{e}_{tprlg} \tag{50}$$

with independent inverse-Wishart distributions

$$\begin{aligned} \mathbf{V}_{\boldsymbol{loc}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{loc}}, \nu_{\boldsymbol{loc}}) \\ \mathbf{V}_{\boldsymbol{field}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{field}}, \nu_{\boldsymbol{field}}) \\ \mathbf{V}_{\boldsymbol{g}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{g}}, \nu_{\boldsymbol{g}}) \\ \mathbf{V}_{\boldsymbol{e}} &\sim \mathcal{W}^{-1}(\boldsymbol{\Phi}_{\boldsymbol{e}}, \nu_{\boldsymbol{e}}) \end{aligned} \tag{51}$$

and $\nu_{\boldsymbol{loc}} = \nu_{\boldsymbol{field}} = \nu_{\boldsymbol{g}} = \nu_{\boldsymbol{e}} = 8$ and $\boldsymbol{\Phi}_{\boldsymbol{loc}} = \boldsymbol{\Phi}_{\boldsymbol{field}} = \boldsymbol{\Phi}_{\boldsymbol{g}} = \boldsymbol{\Phi}_{\boldsymbol{e}} = \frac{14}{4} \cdot \mathbb{I}_{d \times d}$. Both models were implemented using *GIBBS1F90* (see software availability for details).

4.3 Summary and overview

Separation of micro-environmental trait covariances from macro-environmental factors was explored by adding an additional random component describing the spatial correlations between plots in the fields to the random component evaluating influence of location. The spatial random component is defined by a Gaussian kernel matrix, whose coefficients depend on an unknown bandwidth parameter and on physical distances between plots.

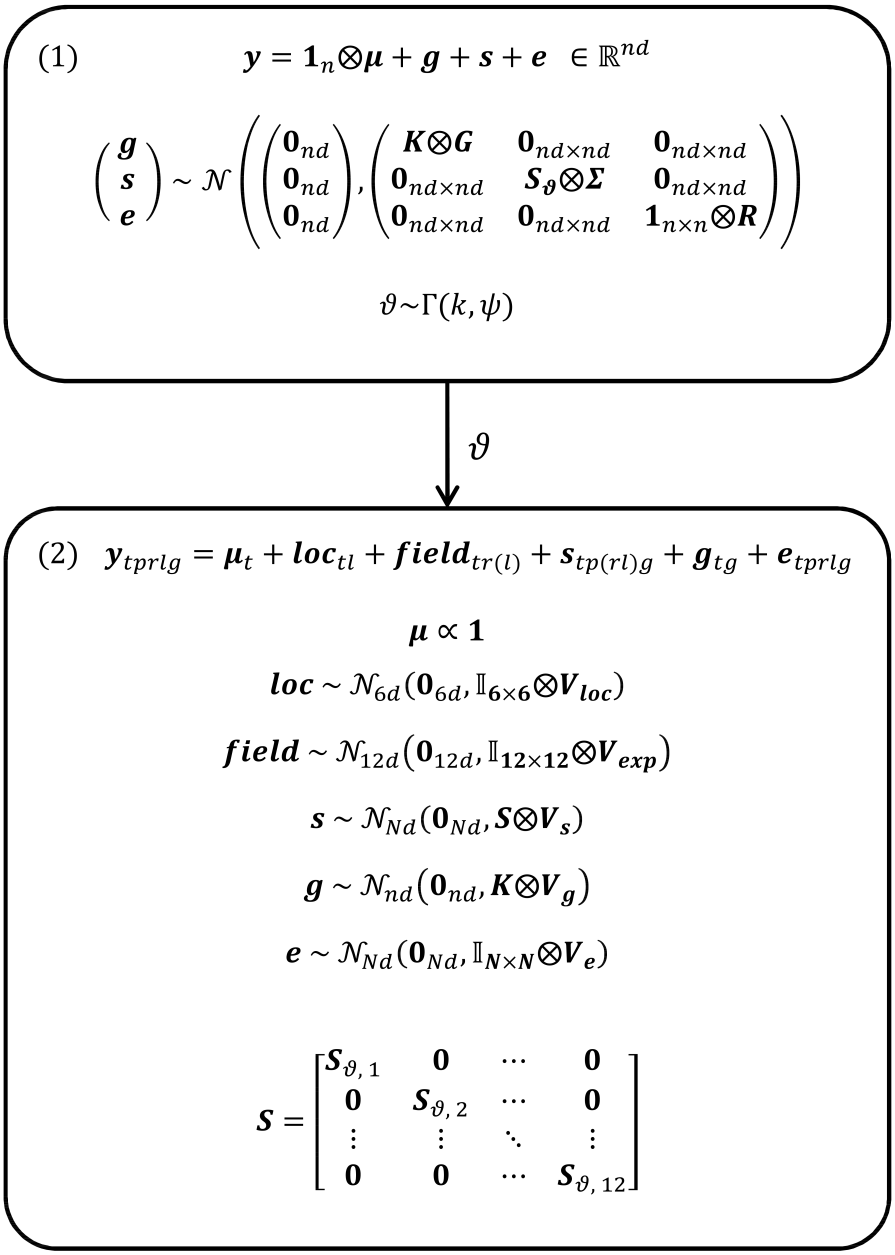


Figure 5: Refining phenotypic covariance decomposition with respect to environment: (1) For each field, a multiple-trait Gaussian model including a spatial random component (*cf.* Equation 33) was fitted to the phenotypic records of its plots. The posterior means for the bandwidth parameters of each field were thereby estimated separately. (2) Afterward, a one-stage model (*cf.* Equation 46) including genomic, micro-, macro-environmental, and residual components was fitted to all fields jointly, and the respective trait covariances were extracted.

5 Methods for model evaluation and software availability

Predictive ability and model fit of MTM *vs.* SEM (*cf.* Equations 5 and 28) were used to assess if the structures found in BN analysis were a better representation of trait connections than a fully recursive structure. The deviance information criterion (DIC) (Spiegelhalter et al. 2002), the logarithm of the Bayesian marginal likelihood (logL), and the predictive ability from 10 replicates of a fivefold cross-validation (CV) were employed for this purpose. The same 50 training- and test-set combinations were used for all models, including the single-trait model. Predictive ability was measured as the correlation between the adjusted means $\boldsymbol{y}$ and their predicted values.

Usefulness of the micro-environmental component in the spatial models (Equation 30 *vs.* Equation 33, and Equation 46 *vs.* Equation 50) was assessed by DIC. Fit of the analyses by field were also investigated using the logarithm of the likelihood at posterior mean in addition to DIC.

5.1 Deviance information criterion and likelihood

The "logarithm of likelihood at posterior mean" is the logarithm of the likelihood of the data (D) given the parameters (Θ), estimated as mean from their MCMC samples:

$$\log L(\overline{\Theta}) = \log p(D \mid \overline{\Theta}) \tag{52}$$

$$\overline{\Theta} = \frac{1}{S}\sum_{l=1}^{S}(\Theta^{(l)}) \tag{53}$$

where S is the number of MCMC samples and $\Theta^{(l)}$ is the l^{th} MCMC sample of the parameters. In contrast, the "logarithm of the mean likelihood" is the mean over all logarithms of likelihoods of the data given the current parameter sample:

$$\log \overline{L(\Theta)} = \frac{1}{S}\sum_{l=1}^{S}\log p(D \mid \Theta^{(l)}) \tag{54}$$

According to Spiegelhalter et al. (2002), the effective number of parameters can then be derived as

$$\mathrm{p_D} = \overline{D(\Theta)} - D(\overline{\Theta}) = -2\log \overline{L(\Theta)} + 2\log L(\overline{\Theta}) \tag{55}$$

where $D(\Theta)$ is called the Bayesian deviance. Spiegelhalter et al. (2002) then proposed the DIC as the sum of the Bayesian deviance at the posterior mean and twice the effective number of parameters:

$$\mathrm{DIC} = D(\overline{\Theta}) + 2p_D = \overline{D(\Theta)} + p_D \tag{56}$$

The marginal likelihood (ML) is the probability of the data given the model (M):

$$\text{ML} = p(D \mid M) = \int p(D \mid \Theta,\ M) p(\Theta \mid M)\, d\Theta \tag{57}$$

Using Bayes' theorem, one can state that

$$\frac{p(\Theta \mid M)}{ML} = \frac{p(\Theta \mid D,\ M)}{p(D \mid \Theta,\ M)} \tag{58}$$

Integrating both sides over Θ yields

$$\frac{1}{ML} \int p(\Theta \mid M)\, d\Theta = \int \frac{1}{p(D \mid \Theta,\ M)} p(\Theta \mid D,\ M)\, d\Theta \tag{59}$$

If $p(\Theta \mid D,\ M)$ is a proper prior, the integral over the prior distribution on the left-hand side equals one. It follows that

$$\frac{1}{ML} = \mathbb{E}_{\Theta \mid D,\, M} \left[\frac{1}{p(D \mid \Theta,\ M)} \right] \tag{60}$$

Newton and Raftery (1994) then proposed the harmonic mean of the MCMC likelihood values as an estimator of the marginal likelihood, that is

$$\widehat{\text{ML}} = \left[\frac{1}{S} \sum_{l=1}^{S} \frac{1}{p(D \mid \Theta^{(l)},\ M)} \right]^{-1} \tag{61}$$

This explanation can be found in Sorensen and Gianola (2002). They also describe a numerically stable computation method following Equation 61 to compute the marginal likelihood that was used in this thesis.

5.2 Software availability

Analyses were performed with the statistical programming tool R (R Core Team 2014) unless stated otherwise. Single-trait prediction was implemented with the *BLR* function (de los Campos and Pérez-Rodríguez 2012). Structure learning, multiple-trait mixed models and SEM were implemented using the R package *bnlearn* (Scutari 2010; Nagarajan et al. 2013) and the R function *MTM*, an extension of the *BGLR* package (de los Campos and Pérez-Rodríguez 2014; Lehermeier et al. 2015) which is available on Github (https://github.com/QuantGen/MTM). *MTM* was modified for the spatial models with unknown bandwidth parameters, see Appendix for the R code. The one-stage models were implemented using *GIBBS1F90*, a Fortran 90 program from the *BLUPF90* family (Misztal et al. 2016).

6 Results

6.1 Genomic and residual trait connections

6.1.1 Trait correlations and Bayesian networks

Phenotypic, genomic, and residual correlations between the five traits were estimated with the MTM (*cf.* Equation 5) for the Flint and Dent panels separately (Table 4). In both panels, there was a strong positive genomic correlation between flowering traits DtTAS and DtSILK, and between the yield-related traits DMY and PH. Genomic correlations were positive for trait combinations not including DMC, whereas DMC was negatively correlated with all other traits. The magnitude of genomic correlations differed between the two panels; for example, the genomic correlation between DMC and DMY was −0.29 in the Dent panel, and −0.64 in the Flint panel. Posterior SDs of correlation estimates varied between 0.01 and 0.10. Single-trait predictive abilities (Table 4) were high and similar to those found in Lehermeier et al. (2014), where a detailed single-trait analysis of the phenotypic and genotypic data can be found.

As the genomic values in the MTM (*cf.* Equation 5) were correlated between genotypes due to kinship, estimates were transformed to meet the model assumptions of the BN algorithms.The transformation modified the relationship between estimated genomic values as expected (see Figure 6).

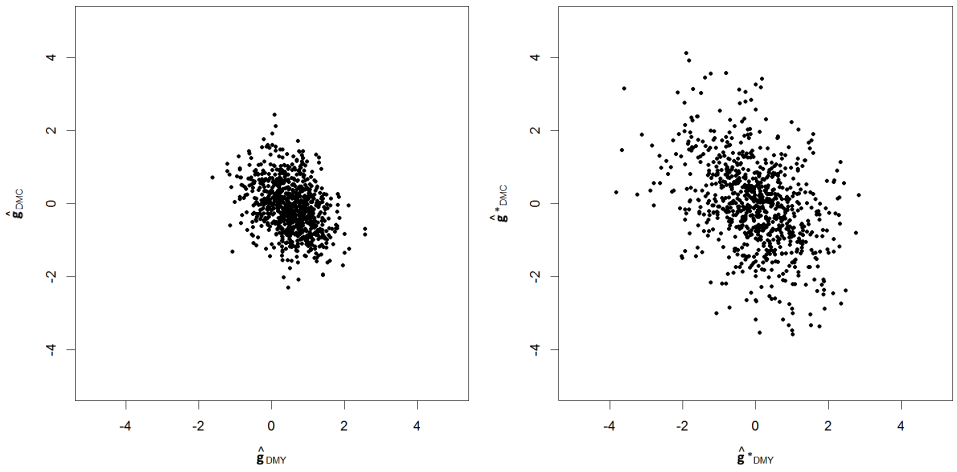


Figure 6: Transformation of the genomic component: relationships between estimated genomic values of DMY ($\hat{\boldsymbol{g}}_{\text{DMY}}$) and DMC ($\hat{\boldsymbol{g}}_{\text{DMC}}$) from the Dent panel and their counterparts ($\hat{\boldsymbol{g}}^*_{\text{DMY}}$ and $\hat{\boldsymbol{g}}^*_{\text{DMC}}$) after transformation.

BN inference was sensitive to the choice of algorithm and test (for detailed results see Figure A1, Figure A2, Figure A3, and Figure A4). The inferred networks varied in up to two edges and/or three directions. In general, score-based algorithms tended to choose more edges than constraint-based algorithms, and the networks of residual components were sparser than those of genomic components. Within the constraint-based or score-based approaches, inference was more consistent than between them.

Table 4: Phenotypic, genomic, and residual correlations between five traits in Dent (lower triangular) and Flint (upper triangular) with posterior SDs in parentheses where appropriate as well as predictive ability

Dent\Flint	DMY	DMC	PH	DtTAS	DtSILK
Phenotypic correlations					
DMY		−0.36	0.68	0.60	0.56
DMC	−0.13		−0.50	−0.59	−0.64
PH	0.59	−0.27		0.66	0.66
DtTAS	0.39	−0.51	0.46		0.91
DtSILK	0.32	−0.57	0.45	0.80	
Genomic correlations					
DMY		−0.64 (0.07)	0.82 (0.04)	0.75 (0.05)	0.73 (0.05)
DMC	−0.29 (0.10)		−0.65 (0.06)	−0.71 (0.05)	−0.77 (0.04)
PH	0.79 (0.04)	−0.27 (0.08)		0.76 (0.04)	0.75 (0.04)
DtTAS	0.60 (0.07)	−0.66 (0.06)	0.60 (0.06)		0.95 (0.01)
DtSILK	0.46 (0.08)	−0.67 (0.06)	0.52 (0.07)	0.88 (0.02)	
Residual correlations					
DMY		0.14 (0.05)	0.41 (0.04)	0.19 (0.05)	0.14 (0.06)
DMC	0.06 (0.05)		−0.10 (0.06)	−0.26 (0.05)	−0.25 (0.05)
PH	0.33 (0.05)	−0.22 (0.05)		0.35 (0.05)	0.36 (0.05)
DtTAS	0.20 (0.05)	−0.26 (0.05)	0.25 (0.05)		0.68 (0.03)
DtSILK	0.17 (0.05)	−0.35 (0.05)	0.34 (0.05)	0.63 (0.03)	
Single-trait predictive ability[a]					
Flint	0.63 (0.04)	0.66 (0.05)	0.69 (0.04)	0.74 (0.04)	0.76 (0.04)
Dent	0.51 (0.05)	0.64 (0.04)	0.69 (0.04)	0.61 (0.04)	0.68 (0.03)
Multiple-trait predictive ability[a]					
Flint	0.64 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.04)	0.76 (0.04)
Dent	0.52 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.04)	0.68 (0.04)

Traits: biomass dry matter yield (DMY) (dt/ha), biomass dry matter content (DMC) (%), plant height (PH) (cm), days to tasseling (DtTAS) (days), days to silking (DtSILK) (days).

[a] Average of 10 random 5-fold cross-validations.

BN inference of residual components was very similar in the Dent and Flint panels, irrespective of the algorithmic settings (Figure A2 and Figure A4). Connections that appeared in both panels were those between the flowering traits, from DtSILK to PH, from DMY to DMC, and between DMY and PH. Specific to the Flint panel was the edge from DtTAS to DMC, while the Dent panel showed a corresponding connection from DtSILK to DMC. An additional edge in the Dent panel not detected in the Flint panel was between DMC and PH. One network in the Flint panel showed an edge from PH to DtTAS, but it was only supported by 56% of the bootstrap samples in the respective network algorithm.

BN inference of genomic components was more variable than that of residual components within and across panels (Figure A1 and Figure A3). In the Flint panel, algorithms recognized edges from DtSILK to DtTAS, from DtTAS to PH, from DtSILK to DMC; and between DMY and DMC, between DMY and DtTAS, between DMY and PH, and between DM and DtTAS. In the Dent panel, only the connection between DMC and DMY was never inferred by any algorithm. A fully recursive network (*i.e.*, including all possible arcs) was never inferred, neither on the genomic nor on the residual components.

6.1.2 Structural equation model analyses

The structures inferred by the different BN algorithms were integrated into SEM (*cf.* Equation 28) and compared with the full structure by using model-fit criteria and predictive abilities of the respective SEM and of the standard multiple-trait model (MTM; *cf.* Equation 5) representing the full structure (Table 5, Table 6, Table A3, and Table A4). When two or more BN algorithms inferred the same network for the same component (genomic or residual), SEM analysis was only performed once for this structure. In the SEM, genomic and residual networks were assessed separately ($\mathbf{G}$ and $\mathbf{E}^*$ or $\mathbf{G}^*$ and $\mathbf{E}$) (Table 5 and Table A3) and jointly ($\mathbf{G}^*$ and $\mathbf{E}^*$) (Table 6 and Table A4).

The number of connections in the networks and the resulting number of nonnull entries in $\mathbf{G}^*$ and $\mathbf{E}^*$ were derived (Table 5 and Table 6). These illustrate the variable reduction attained by application of the structures to the residual or genomic components. For each SEM, the sum of the number of connections in the networks and the sum of nonnull entries in the covariance matrices were split into the genomic and the residual contributions. In general, residual networks were sparser than genomic networks, and SEM on residual structures were more parsimonious in the Flint panel than in the Dent panel.

Table 5: Single-structure evaluation: model fit for the multiple-trait model (MTM) and for structural equation models (SEM) including a genomic ($\mathbf{\Lambda}_{g^*}$) or residual ($\mathbf{\Lambda}_{e}$) trait structure denoted by the BN algorithm it originates from

BN Giving $\mathbf{\Lambda}_{g^*}$	BN Giving $\mathbf{\Lambda}_{e}$	DIC[a]	p_D[b]	logL[c]	No. Connections[d]	No. Nonnull Parameters[e]
Dent						
—	GS 3	−87.0	+61.6	+66.0	10 + 6 = 16	15 + 13 = 28
—	GS 1, 2, 4	−78.5	+65.7	+72.1	10 + 5 = 15	15 + 12 = 27
—	TABU 1, 2	−62.7	+45.1	+23.3	10 + 6 = 16	15 + 15 = 30
TABU 1	—	−0.7	−12.6	+12.1	8 + 10 = 18	15 + 15 = 30
—	—	0.0	1024.9[f]	0.0	20	30
TABU 2	—	+1.0	−14.1	−7.3	9 + 10 = 19	15 + 15 = 30
GS 1, 2, 3, 4	—	+57.0	−16.9	−32.4	7 + 10 = 17	15 + 15 = 30
Flint						
—	TABU 1	−124.1	+64.0	+105.2	10 + 5 = 15	15 + 13 = 28
—	GS 1, 2, 3, 4	−122.3	+63.3	+126.3	10 + 5 = 15	15 + 13 = 28
—	TABU 2	−117.7	+61.0	+131.1	10 + 6 = 16	15 + 14 = 29
—	—	0.0	1086.0[f]	0.0	20	30
TABU 1, 2	—	+35.8	+23.1	−3.2	7 + 10 = 17	14 + 15 = 29
GS 1, 2, 3, 4	—	+72.1	+4.2	−57.1	6 + 10 = 16	12 + 15 = 27

For notation of BN algorithms see Table 2.

[a] Deviance information criterion: DIC of SEM minus DIC of MTM.

[b] Effective number of parameters: p_D of SEM minus p_D of MTM.

[c] Logarithm of Bayesian marginal likelihood: logL of SEM minus logL of MTM.

[d] Number of connections in the genomic network plus number of connections in the residual network.

[e] Number of nonnull parameters in the genomic covariance matrix plus number of nonnull parameters in the residual covariance matrix.

[f] Absolute value of p_D.

Table 6: Double-structure evaluation: model fit for the multiple-trait model (MTM) and for structural equation models (SEM) including a genomic ($\mathbf{\Lambda}_{g^*}$) and residual ($\mathbf{\Lambda}_e$) trait structure denoted by the BN algorithm it originates from

BN Giving $\mathbf{\Lambda}_{g^*}$	BN Giving $\mathbf{\Lambda}_e$	DIC[a]	p_D[b]	logL[c]	No. Connections[d]	No. Nonnull Parameters[e]
Dent						
TABU 1	GS 3	−87.0	+47.6	+63.8	8 + 6 =14	15 + 13 = 28
TABU 2	GS 3	−86.0	+47.4	+66.1	9 + 6 =15	15 + 13 = 30
TABU 1	GS 1, 2, 4	−75.1	+48.0	+35.5	8 + 5 =13	15 + 12 = 27
TABU 2	GS 1, 2, 4	−74.4	+47.8	+52.7	9 + 5 =14	15 + 12 = 27
TABU 1	TABU 1, 2	−64.3	+34.6	+66.0	8 + 6 =14	15 + 15 = 30
TABU 2	TABU 1, 2	−63.9	+34.5	+54.1	8 + 6 =14	15 + 15 = 30
GS 1, 2, 3, 4	GS 3	−31.2	+47.0	+41.3	7 + 6 =13	15 + 13 = 28
GS 1, 2, 3, 4	GS 1, 2, 4	−26.6	+51.3	+21.1	7 + 5 =12	15 + 12 = 27
GS 1, 2, 3, 4	TABU 1, 2	−9.2	+33.2	+23.0	7 + 6 =13	15 + 15 = 30
—	—	0	1024.9[f]	0	20	30
Flint						
TABU 1, 2	TABU 1	−17.2	+108.1	+52.0	7 + 5 =12	14 + 13 = 27
GS 1, 2, 3, 4	TABU 2	−15.6	+77.7	+40.4	6 + 6 =12	12 + 14 = 26
TABU 1, 2	GS 1, 2, 3, 4	−14.8	+105.5	+65.9	7 + 5 =12	14 + 13 = 27
TABU 1, 2	TABU 2	−14.0	+103.5	+65.3	7 + 6 =13	14 + 14 = 28
—	—	0	1086.0[f]	0	20	30
GS 1, 2, 3, 4	TABU 1	+30.4	+76.5	+34.8	6 + 5 =11	12 + 13 = 25
GS 1, 2, 3, 4	GS 1, 2, 3, 4	+33.1	+74.2	+22.9	6 + 5 =11	12 + 13 = 25

For notation of BN algorithms see Table 2.

[a] Deviance information criterion: DIC of SEM minus DIC of MTM.

[b] Effective number of parameters: p_D of SEM minus p_D of MTM.

[c] Logarithm of Bayesian marginal likelihood: logL of SEM minus logL of MTM.

[d] Number of connections in the genomic network plus number of connections in the residual network.

[e] Number of nonnull parameters in the genomic covariance matrix plus number of nonnull parameters in the residual covariance matrix.

[f] Absolute value of p_D.

Model fit. Model fit was assessed with the DIC and with the logarithm of the Bayesian marginal likelihood (logL) (Table 5 and Table 6). logL evaluates the fit of the model to the data and prior assumptions, and DIC takes into account and penalizes the number of parameters in the model. Models were ranked according to their DIC, which differed from their logL ranking. In Dent (Flint), eight (four) SEM with both a structured residual and structured genomic component had a lower DIC than the multiple-trait model (Table 6). In Flint, no SEM with only a genomic structure had a lower DIC than the multiple-trait model (Table 5). In general, DIC and logL varied more in Flint than in Dent (Table 5 and Table 6).

Predictive ability. Predictive abilities and their SDs were derived from 10 replicates of a fivefold CV with random sampling of training and test sets within each panel (Table A3 and Table A4). Averaged over traits and for each trait individually, predictive abilities were similar for all models (single-trait, MTM, and SEM) and differences between models were smaller than 1 SD and not significant. The magnitude of predictive-ability estimates in the single-structure SEM was consistent with predictive-ability estimates in the double-structure SEM in both panels. Since all SEM with a genomic structure in the Flint panel had higher DIC (Table 5) and slightly lower predictive ability estimates (Table A3) than the MTM, genomic structures might not have been identified correctly, which might also have affected the double-structure SEM.

Choice of best network. Considering all model performance criteria jointly, the best-performing networks in the SEM analysis are given in Figure 7. In the Dent panel, the SEM with the structure derived from the residuals with GS 3 and the SEM derived from the genomic component with TABU 1 performed best considering both the double-structure and single-structure settings.

In the Flint panel, the SEM with the structure derived from the genomic component with TABU 1 and 2 had the best fit among SEM with a predefined structure, but performed worse than the MTM in the single-structure setting. Regarding the residual structure in the Flint panel, the structure derived with GS 1, 2, 3, and 4 was selected because it had a considerably larger logL than the structure derived with TABU 1 when at the same time DIC and effective number of parameters were very close.

Consequently, Figure 7 shows the structures derived with TABU 1 for the genomic components and the structure derived with GS 3 for the residual components for both panels.

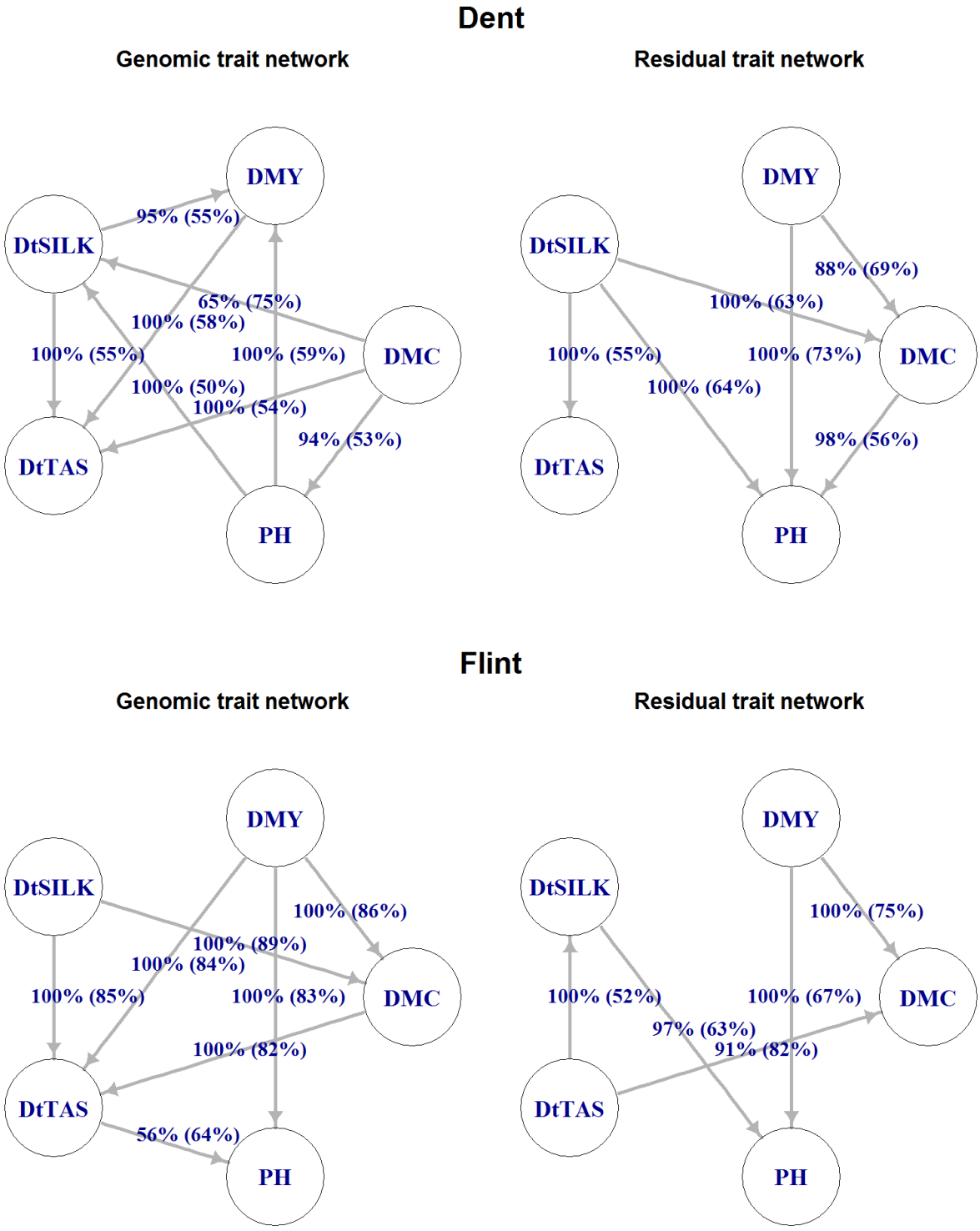


Figure 7: BN for genomic values and residuals in Dent and Flint: graphs of the genomic component from the tabu-search algorithm using BIC as score. Structure learning was performed with 500 bootstrap samples. Graphs of the residuals from the GS algorithm with mutual information as test statistic, also with 500 bootstrap samples and a significance level of $\alpha = 0.01$. Labels of edges indicate the proportion of bootstrap samples supporting the edge and (in parentheses) the proportion having the direction shown. Edges that were not significant in the averaging process due to a network-internal empirical test on the arc's strength are not shown.

6.2 Genomic, micro-, macro-environmental, and residual trait correlations

Runtime of a model with a random spatial component including the MH algorithm for the bandwidth parameter was almost as fast as without this additional MH step using the R code implemented here (see Appendix for code, and Algorithms A1 and 1 for details): Spatial models with two kernels (Equations 33) ran twice as long as models with only one, genomic, kernel (Equations 30) on a common Windows-7 machine (64-bit operating system).

Table 7: Posterior mean of bandwidth parameter for several prior and proposal settings (Flint panel)

Prior Assumption	$\vartheta \sim \Gamma(2,2)$			$\vartheta \sim \Gamma(4,0.15)$		
Starting Value ($\vartheta^{(t=0)}$)	0.10	1.00	2.00	0.10	1.00	2.00
SD Proposal (σ^2_ϑ)	0.05	0.10	0.10	0.10	0.05	0.10
Field						
INRA 1	0.531	0.532	0.531	0.520	0.518	0.520
KWS 1	0.555	0.559	0.557	0.547	0.544	0.545
SYN 1	1.092[bc]	1.231[bc]	1.412[bc]	0.802[c]	0.802[c]	0.804[c]
TUM 1	0.559	0.561	0.563	0.548	0.545	0.546
CIAM 1[a]	1.246[c]	1.249[c]	1.247[c]	1.186[c]	1.187[c]	1.188[c]
HOH 1	0.393	0.391	0.392	0.390	0.391	0.390
INRA 2	0.257	0.258	0.261	0.260	0.259	0.261
KWS 2	0.453	0.455	0.453	0.450	0.449	0.451
SYN 2[a]	0.782[c]	0.782[c]	0.810[bc]	0.712[c]	0.711[c]	0.711[c]
TUM 2	1.593[bc]	1.809[bc]	1.866[bc]	0.884[c]	0.882[c]	0.883[c]
CIAM 2[a]	0.888	0.893	0.893	0.872	0.874	0.873
HOH 2[a]	0.619	0.619	0.621	0.598	0.598	0.598

Ploudaniel 2011 (INRA), Einbeck 2011 (KWS), Wadersloh 2011 (SYN), Roggenstein 2011 (TUM), La Coruna 2011 (CIAM)

[a] Low prior sensitivity.

[b] MCMC not converged.

[c] SD of posterior mean larger than 0.1.

There were small differences among the posterior mean of the bandwidth parameter for different starting values within settings (Table 7). In three out of twelve fields, convergence of the bandwidth parameter's distribution was slow and perhaps not attained when the the less informative prior was used, with trace plots of MCMC samples of ϑ indicating convergence problems. Sensitivity to choice of hyper-parameters was mild in four fields and not appreciable in six fields, see differences between the two settings in Table 7. In summary, low prior sensitivity and good mixing indicated identifiability of and a reasonable information on the bandwidth parameter ϑ. Convergence to its posterior distribution was satisfactory.

Similarities of model fit (Table A5) and posterior means of ϑ (Table 7) were similar across settings, and settings with converged chains of the bandwidth parameter were favored over those with convergence problems (visual inspection of sampling chain) by model fit. Adding the spatial component improved model fit (Equations 33 *vs.* Equations 30) in each of the fields (Table A6), and, consis-

tently, adding the spatial component to the one-stage model (*cf.* Equation 46) also improved DIC and reduced the determinant of the residual covariance matrix (Equation 46 *vs.* Equation 50).

Correlations between predicted genomic values of this one-stage model with predicted genomic values of a comparable two-stage model were 0.98 for DMY, 0.99 for DMC, 0.98 for PH, 0.99 for DtTAS, and 0.99 for DtSILK. The two-stage model (MTM according to Equation 5) was recalculated for this comparison with a kinship matrix calculated according to Astle and Balding (2009) and without a population correction for the adjusted means.

Trait correlations and SD for all components are given in Table 8. Recall that in the two-stage approach (*cf.* Equations 1 and 5) there was a correction for family effects, *i.e.*, the adjusted means were mean-centered in each family. The one-stage model (*cf.* Equation 46) has no correction for family effects, *i.e.*, no fixed effects for the family means were included. So the genomic correlations in the one-stage model represent family relatedness in addition to allelic effects in contrast to the genomic correlations in Table 4. As a result, genomic correlations between DMY and DtTAS, between DMC and DtSILK, and between DtTAS and DtSILK were lower by more than 0.05.

Micro-environmental trait correlations ranged from −0.29 to 0.57 with SD smaller than 0.05. Macro-environmental (*i.e.*, from the location component) and field trait correlations had a large SD, but it is to be expected that variance components from six locations and twelve fields are necessarily imprecise.

Table 8: Genomic, micro-, macro-environmental, field, and residual trait correlations from the one-stage model (*cf.* Equation 46) with posterior SDs in parentheses (Flint panel)

	DMC	PH	DtTAS	DtSILK
Genomic correlations				
DMY	−0.60 (0.04)	0.81 (0.02)	0.68 (0.03)	0.68 (0.03)
DMC		−0.63 (0.04)	−0.69 (0.03)	−0.71 (0.03)
PH			0.74 (0.03)	0.75 (0.03)
DtTAS				0.86 (0.01)
Micro-environmental correlations (spatial component)				
DMY	0.28 (0.03)	0.57 (0.03)	−0.29 (0.04)	−0.29 (0.04)
DMC		−0.05 (0.03)	−0.17 (0.03)	−0.20 (0.03)
PH			−0.21 (0.03)	0.17 (0.03)
DtTAS				0.56 (0.02)
Macro-environmental correlations (location component)				
DMY	0.11 (0.30)	0.25 (0.30)	−0.15 (0.32)	−0.32 (0.33)
DMC		0.02 (0.30)	−0.27 (0.30)	−0.29 (0.32)
PH			0.00 (0.37)	0.08 (0.38)
DtTAS				0.48 (0.28)
Field correlations				
DMY	−0.30 (0.35)	0.32 (0.33)	−0.38 (0.33)	−0.37 (0.36)
DMC		−0.27 (0.32)	0.32 (0.33)	0.31 (0.33)
PH			−0.19 (0.37)	−0.35 (0.40)
DtTAS				0.40 (0.27)
Residual correlations[a]				
DMY	0.49 (0.03)	−0.08 (0.04)	0.13 (0.03)	−0.03 (0.03)
DMC		−0.02 (0.04)	0.03 (0.03)	−0.06 (0.03)
PH			0.21 (0.04)	0.24 (0.04)
DtTAS				0.49 (0.02)

Traits: biomass dry matter yield (DMY) (dt/ha), biomass dry matter content (DMC) (%), plant height (PH) (cm), days to tasseling (DtTAS) (days), days to silking (DtSILK) (days).
[a] The determinant of the residual trait covariance matrix is smaller than 10^{-7}.

7 Discussion

7.1 Learning genomic and residual trait networks

Scope of Bayesian network and structural equation model methodology. This thesis showed that structures derived from BN algorithms conveyed more information than mere trait-association values because many correlations are suggested to arise from indirect associations that are mediated by one or more variables. For example, the relatively strong genomic correlations between PH and DMC and between DtSILK and DMY in the Flint panel, and the relatively strong genomic correlation between DtSILK and DMC in the Dent panel, did not correspond to an edge in the genomic networks (Figure 7). As illustrated by Valente et al. (2015), a wrong interpretation or use of trait associations can lead to erroneous selection decisions or interventions. Therefore, knowledge on network structures among traits is useful on top of only trait correlations with respect to indirect selection strategies.

Genomic correlations, networks, and model performance criteria varied between the Dent and Flint panels, and incorporating genomic networks into prediction models improved model fit more in Dent than in Flint. As the DH lines in the Dent panel were genetically more diverse than in the Flint panel (Lehermeier et al. 2014), this could imply that formulating an SEM is more advantageous when dealing with a diverse set of genotypes.

It is noteworthy that the DIC in the single-structure SEM was more variable in Flint than in Dent. Variable selection through the network algorithms was more stringent in Flint and, therefore, the SEM differed more from each other and from the MTM than in Dent (Table 5). Based on the comparison of the two data sets, the impact of genetic structure on network inference could not be shown conclusively. In combination with an investigation of methods for control of variable selection intensity (*i.e.*, propensity of an algorithm, test, or score to a sparser network), this topic warrants further research.

In both panels, the marginal logL was higher for many SEM than for the multiple-trait model, even though the network structure on the covariance matrices implied variable selection. This shows that restrictions on the covariance structures among traits were generally supported by the data. Following the principle of parsimony, a fully recursive structure might not be the best representation of connections among traits.

Choice of method for network construction. The results suggest that the choice of method for network learning is crucial, and that a thorough assessment of network structures is necessary when dealing with real-life data. Resulting networks might not explain or fit the data better than a full network as seen here (especially for the genomic component in Flint). Assessment of the different network structures was done by considering several model quality criteria (DIC, logL, number of parameters, and cross-validated predictive ability) together, and by evaluating if these criteria ranked networks consistently.

Marginal likelihoods are used routinely to evaluate the relative plausibility of prior assumptions (*e.g.*, Spiegelhalter et al. 2002) from the observed data. After having learned the BN, they were translated into prior assumptions for SEM, and then the various SEM were assessed for their priors via marginal likelihoods. DIC conveys information on number of effective parameters in model comparisons. Different prior assumptions translate into different numbers of effective parameters in the SEM, *i.e.*, model complexity varies over networks. As DIC reflects model complexity, it is a natural companion to the marginal likelihood for SEM ranking. While model-fit criteria evaluate the quality of data description by a network, predictive ability reflects the generalization of structure estimation from one subset of the data (training set) to another subset (test set). However, true structures of traits remain unknown and validating inferred connections in experiments with a broad range of genetic material after hypothesis generation with BN and SEM is crucial.

SEM analysis suggested that score-based approaches were better for learning the structure of the genomic networks, while constraint-based approaches were better for the residual networks. This might have resulted from constraint-based algorithms having been more restrictive (given the chosen significance level) and from residual networks being sparser than genomic networks. Evidence for the significance of edges (*i.e.*, support of the edge in more than 70% of the bootstrap samples) was generally higher in the constraint-based networks than in the score-based networks. Influence of the significance level in the constraint-based algorithms warrants further investigation. However, it might be expected that a larger significance level α for learning BN from the residuals would result in more complex networks, and, at some point, the additional complexity would result in lower predictive ability and/or overfitting. Advanced network construction algorithms, such as max-min hill-climbing (Tsamardinos et al. 2006), combine a constraint-based edge search with a score-based directing of the edges. Such combined approaches should be investigated in future research because they allow variation of the significance level in combination with a score-based network search (*i.e.*, directing arcs).

Interpretation of network structure. If variables A and B are directly associated with each other, and B and C are alike, then an association between A and C would be a logical consequence, even if there is no direct association between A and C (*cf.* part A of Figure 1). When a BN is inferred, such associations that can be explained by a chain of other associations are not depicted as connections. This means that a connection in a network is more reliable than a significant association between two variables because BN eliminate connections that are unlikely to be direct.

Therefore, BN connections provide more information than pairwise associations between traits. If there were a causal connection between a child and a parent variable, then the child variable would change with a change of the parent variable. For example, if silking time in Dent material is changed, *e.g.*, by early seeding, it is likely that PH at harvest changes too, according to the residual network in the Dent panel. If the genotypic value of PH is changed in the Dent panel, *e.g.*, by selection, this change is likely to affect the genotypic value of DMY, according to the genomic network in the Dent panel. Nonetheless, the interpretation of BN connections as causal effects is delicate and needs fur-

ther assumptions than those made here, especially the assumption of absence of additional variables influencing those already included in the study. Thus, interpretation of networks is suggested as an overall association among more than two entities, *i.e.*, values of the child variable are associated with values of the parent variable.

Residual networks followed some temporal order as traits measured at harvest (PH, DMY, and DMC) depended on traits measured during the vegetation period (DtTAS and DtSILK). Additionally, the inference of direction between DtTAS and DtSILK in the residual networks was relatively uncertain with 52 and 55% of bootstrap samples supporting the direction. This might be because both flowering traits are determined by (the time for the) transition from vegetative to reproductive growth (TVR) and no direct dependence between them truly exists, as both traits depend on TVR. The connection between DtTAS and DtSILK would then be an example of an induced connection by an unobserved confounder in a network.

7.2 Investigation of micro-environmental covariance by a spatial component

Evaluation of spatial model. Formulation of the spatial component with a Gaussian kernel and a single bandwidth parameter with a Gamma prior was adapted from Pérez-Elizalde et al. (2015) for modeling spatial dependencies in field experiments. This approach modeled parsimoniously the two-dimensional field arrangements with a single random component, which was robust with respect to sensitivity to choice of hyper-parameters and convergence performance.

In contrast to grid search, Bayesian sampling of the bandwidth parameter facilitated investigation of model fit (likelihood at posterior mean) in a single model call. MCMC sampling allowed assessment of different priors and prior sensitivity and provided an uncertainty measure (SD of the posterior distribution) for the bandwidth parameter estimates.

Results in this thesis (*cf.* Table 7) showed that convergence issues or sensitivity to hyper-parameter choice can indicate lack of statistical information, strong posterior inter-correlation structure, or non-random spatial trends such as in the second trial in Roggenstein (TUM 2), where a fertilizer gradient was observed in the field. It therefore can be assumed that when there is no spatial correlation in the data following a random Gaussian pattern, the bandwidth parameter presumable may not converge, or rather, converge to zero or very large values as these both states of the bandwidth parameter imply little values of spatial covariances.

Model-fit measures indicated that the spatial component was supported by the data, and although less data (Table 1 and Table 3) was used in the one-stage model (*cf.* Equation 46) than in the two-stage model (*cf.* Equations 1 and 5), estimates of genomic values were very similar in both approaches. In the one-stage model, phenotypic (co)variance could be effectively decomposed into genomic, macro- and micro-environmental components, thereby distinguishing between amendable genetic

and environmental trait connections and trait connections being induced by macro-environmental conditions.

Differences between one-stage and two-stage approach. Besides the spatial component, there were other differences between the one-stage (*cf.* Equation 46) and two-stage models (*cf.* Equations 1 and 5). Different kinship matrices were used for modeling the genomic component (simple matching in the MTM and Astle-Balding in the one-stage). Additionally, family effects were considered in the two-stage approach as the adjusted means were mean-centered per family. In contrast, family effects were not separately modeled in the one-stage approach (*cf.* Equation 46) and the genomic component was therefore modeling additive allelic and family effects together. Kinship following Astle and Balding (2009) and no extra fixed family effects were chosen for the one-stage model (*cf.* Equation 46) as it allowed faster computation. In addition, modeling family means as fixed effects did not change estimates of non-genomic covariances or the model-fit values appreciably in the one-stage model (results not shown).

Second and more important, all traits were considered jointly in the one-stage approach (*cf.* Equation 46). As a consequence, data of a plot was discarded for all traits if data was missing for a single trait (compare Table 1 and Table 3 for details). In the two-stage approach (*cf.* Equations 1 and 5), adjusted means were calculated for each trait separately and more data could therefore be used. Additionally, less data was available in the one-stage approach (*cf.* Equation 46) than in the two-stage approach (*cf.* Equations 1 and 5) since marker data was not available for some genotypes (6 in the Flint panel). Differences in predicted genomic values were, however, negligible (see Results).

Estimates of fixed effects are less sensitive to not identically distributed residuals than estimates of random components as the variance estimation of random components can be confounded if the error terms' variance assumptions are violated. The plots in the field experiments here had different measurement errors, *e.g.*, due to different number of available plants per plot at harvest, which conflicted with the assumption of identically distributed residuals in the phenotypic analysis (*cf.* Equation 1) as well as in the one-stage model (*cf.* Equation 46). The effect of this violated assumption is expected to be more prevalent in one-stage models and, therefore, one-stage models can be less robust than two-stage approaches.

This could also cause imbalances in CV schemes. The development and assessment of a suitable weighting to correct for the measurement errors was, however, out of the scope of this thesis and can also become challenging as it is not clear which factors are involved in the measurement errors, how they can be measured, and then translated into error term covariance assumptions. To overcome this challenge in future research, it might therefore be advisable to assess predictive ability of the one-stage approach with an independent validation set, which was unfortunately not available for the experimental data used in this thesis.

Scope of spatial model methodology. First, modeling spatial correlation using physical distance helps identifying traits that are regulated by joint micro-environmental factors. Second, this ap-

proach facilitates computation of a multiple-trait-one-stage model (*cf.* Equation 46) since spatial effects are modeled by a single random component and can be used even if a block design is not available.

For all traits, the same spatial correlation was assumed, which was a strong assumption that might be inadequate as physical distance might influence traits differently. However, this assumption improved identifiability of the bandwidth parameter, and is analogous to the corresponding approach for genomic components, where bandwidth parameters are estimated for the genomic component (Pérez-Elizalde et al. 2015).

Predefining a single Gaussian random dependence structure in experimental data is an inflexible approach because it does not account for different spatial dependencies in different directions in trials, or for different distributions at different positions in the field (*e.g.*, Gilmour et al. 1997, Bernal-Vasquez et al. 2014, or Velazco et al. 2017). If nonrandom spatial effects or strong gradients are present, such spatial patterns would need to be modeled in addition to the spatial component proposed in this thesis as the micro-environmental component here only models random effects following a Gaussian pattern. On the other hand, overfitting spatial patterns is expected to be less likely when using a single predefined random component with a single free parameter. It is also assumed to be computationally easier and statistically better identified than using several components and many more parameters that need to be estimated. Whether or not spatial modeling is helpful for genomic prediction, and when which approach is the most beneficial is problem dependent and warrants further research.

Convergence of the MCMC sampling including MH usually depends heavily on the variance of the proposal distribution chosen for the respective parameter (σ^2_ϑ). If the variance is chosen too small, the MCMC chain mixes badly, *i.e.*, the acceptance rate will be high, but successive samples will move around the space slowly, and many samples are needed in the burn-in period. In addition, MCMC samples will be strongly auto-correlated after convergence. If the variance is chosen too large, the acceptance rate is very low, which directly implies slow convergence (*e.g.*, Green et al. 2015). Choosing an optimal variance for the proposal distribution therefore requires intensive trial and error, and is problem dependent.

If data from a plot is missing, estimation of the proposed spatial component is expected to be more robust to such missing data than estimation in a simple row-column adjustment because both directions are combined and, therefore, a gap has less weight per coefficient of the spatial component. A similar reasoning applies to spatial adjustment at field borders, which could be useful when investigating large plantations or natural populations, *e.g.*, of trees.

The implementation of a Gibbs Sampler with an embedded MH step provided in this thesis laid ground for further research on epistasis. Jiang and Reif (2015) show that a genome-enabled prediction model including epistatic effects (where the kernel is modeled as the Hadamard product of the additive relationship matrix) is equivalent to a reproducing kernel Hilbert space (RKHS) prediction

model. For this equivalence to hold, the RKHS model needs to be formulated with a Gaussian kernel (*cf.* Equation 35) and a bandwidth parameter chosen accordingly. So the code developed in this thesis can be used there and facilitate research in posterior distributions of epistatic effects and their variances, respectively.

8 Conclusion

The main contributions and conclusions from investigating trait covariances and their structures are:

- The novel approach allowed to determine trait connections separately on the genomic and residual levels. This is a major development in BN trait connection studies as it aligns them with (non-structural) multiple-trait genome-enabled prediction where genetic and residual factors are assumed to have independent distributions.
- A suitable transformation for adapting the genomic component to assumptions of the BN learning methodologies could be found in this thesis, and genomic and residual effects were used for BN learning instead of their variances. BN learning becomes thus presumably more robust to noise in the data since effects are expected to be more accurately estimated than covariance matrices in mixed models.
- An integration of the separate genomic and residual trait networks into SEM was presented and used as trait structure evaluation tool. Consequently, genomic and residual networks could then be assessed by SEM separately and jointly. SEM were used herein for the first time as evaluation of found BN trait connections in experimental data, and the found BN could be shown to fit data better than the standard model.
- Spurious connections in maize data could be identified on the genomic and the residual levels in this study. Identification of spurious genomic connections can be useful for multiple-trait prediction and indirect selection.
- Bootstrapping was used for the first time to identify potential confounders, and an example (TVR) for such a confounder was found and discussed.
- A computationally feasible one-stage model was implemented in this thesis and allowed a refined decomposition of phenotypic covariance into genomic, micro- and macro-environmental components.
- In the respective R implementation, computation of a Gaussian kernel with a single bandwidth parameter was facilitated by sampling this bandwidth parameter in an MH algorithm embedded in a Gibbs sampler. The implementation had a reasonable runtime and is of general usefulness, *e.g.*, regarding spatial genomic kernels and modeling epistasis.
- Convergence and low sensitivity with respect to hyper-parameters indicated identifiability of the bandwidth parameter when modeling spatial covariance in the experimental data examined.
- Using a single random component for both field dimensions as proposed here in modeling spatial covariance enables future investigation of micro-environmental trait connections via BN.

9 Summary

Relationships among traits were investigated on the genomic and residual levels using novel methodology developed in this thesis. This included inference on these relationships via Bayesian networks and an assessment of the networks with structural equation models. The methodology employed three steps. First, a Bayesian multiple-trait Gaussian model was fitted to the data to decompose phenotypic values into their genomic and residual components. Second, genomic and residual network structures among traits were learned from estimates of these two components. Network learning was performed using six different algorithmic settings for comparison, of which two were score-based and four were constraint-based approaches. Third, structural equation model analyses ranked the networks in terms of model fit and predictive ability, and compared them with the standard multiple-trait fully recursive network. For juxtaposing macro-environmental to micro-environmental trait connections, it was proposed to include a component modeling covariance due to spatial arrangement of plants in experimental trials as a representative of shared micro-environmental influences. The spatial random component added to the model was defined by a Gaussian kernel matrix, of which the coefficients depend on an unknown bandwidth parameter and on physical distances between plots. The model was implemented in a Bayesian framework and the bandwidth parameter was sampled within a Metropolis-Hastings step. Macro-environmental connections were modeled via the location.

The developed methods were applied to experimental data from the European heterotic maize pools Dent and Flint (*Zea mays* L.). Inferences on genomic and residual trait connections were depicted separately as directed acyclic graphs. These graphs provide information beyond mere pairwise genetic or residual associations between traits, illustrating conditional independencies and hinting at potential causal links among traits. Network analysis suggested some genetic correlations as potentially spurious. Micro-environmental covariances were estimated for each field at all locations separately. Convergence of the Monte Carlo Markov chains and identifiability of the bandwidth parameter were suggested by consistency of estimates across various hyper-parameters and Metropolis-Hastings specifications, *i.e.*, various starting values and standard deviation values of the Metropolis-Hastings proposal density were examined. Including the spatial components into the models was supported by increased likelihood. Data from all fields were integrated into a one-stage model, from which genomic, micro-, macro-environmental, and residual contributions to the phenotype could be calculated across the whole experiment. The implementation of the Gibbs sampler with a Metropolis-Hastings step for the bandwidth parameter might also be useful for genomic analysis, *e.g.*, for estimating epistatic values.

10 Zusammenfassung

Mit einer neuen Methodik wurden Merkmalszusammenhänge getrennt nach genomischem und anderweitigem Ursprung untersucht. Dazu wurden die Beziehungen zwischen den Merkmalen mit Bayesschen Netzen bestimmt und danach mit Strukturgleichungsmodellen evaluiert. Die Methodik bestand aus drei Schritten. Zuerst wurden die phänotypischen Werte mit einem gemischten Modell in ihre genomischen und anderweitigen Anteile zerlegt. Von beiden Anteilen wurden dann im zweiten Schritt Bayessche Netze gelernt. Das Lernen der Netze wurde durch sechs verschiedene algorithmische Szenarien realisiert, von denen zwei ergebnisoptimierend und vier testbasiert waren. Im dritten Schritt wurden die damit gefundenen Netze mit Strukturgleichungen nach ihrer Anpassungsgüte und ihrer Vorhersagekraft geordnet und so untereinander und mit dem Standardmodell für mehrere Merkmale verglichen. Um Effekte der Mikro- von denen der Makroumwelt trennen zu können, wurde ein Zufallsfaktor dem Modell hinzugefügt, der sich die räumliche Anordnung der Parzellen im Experiment zur Modellierung der Mikroumwelt zunutze macht. Dieser neue Zufallsfaktor wurde durch einen Gauß-Kern definiert, dessen Koeffizienten von einem unbekannten Parameter und dem physikalischen Abstand zwischen den Parzellen abhängen. Das Modell wurde nach Bayes implementiert und der unbekannte Parameter darin in einem Metropolis-Hastings-Algorithmus gezogen. Die Makroumwelt wurde durch den Standort im Modell repräsentiert.

Die Methodiken wurden auf experimentelle Daten angewendet, die aus den zwei heterotischen europäischen Maisgruppen (*Zea mays* L.), Dent und Flint, stammten. Die von den genomischen und anderweitigen Anteilen getrennt hergeleiteten Bayesschen Netze wurden als gerichtete, azyklische Graphen abgebildet. Diese Abbildungen bieten genauere Einblicke als paarweise Assoziationsmaße zwischen Merkmalen, da sie bedingte Unabhängigkeiten abbilden und potentielle kausale Zusammenhänge aufzeigen. Die Netze legten nahe, dass manchen genetischen Korrelationen keine Kausalität zugrunde liegt. Effekte der Mikroumwelt wurden zunächst für jede experimentelle Einheit an allen Orten getrennt berechnet. Die konsistente Schätzung des Parameters über verschiedene Hyperparameter und Einstellungen des Metropolis-Hastings-Algorithmus (wie Startwert und Standardabweichung der Vorschlagsdichte) ist ein gutes Indiz dafür, dass die Monte-Carlo-Markov-Kette konvergiert und der Parameter selbst identifiziert ist. Zum Schluss wurden alle experimentellen Einheiten gemeinsam in ein Ein-Schritt-Modell zusammengefasst, aus dem sich für alle Orte des Experiments gemeinsam die Anteile des Phänotyps gemäß genomischen Faktoren, Mikro- und Makroumwelt sowie anderer Einflussfaktoren ableiten lassen. Die Implementierung des Gibbs-Sampling mit dem Metropolis-Hastings-Algorithmus für den unbekannten Parameter kann daneben auch für die Schätzung von Epistasie nützlich sein.

11 References

Aliferis C F, Statnikov A, Tsamardinos I, Mani S, and Koutsoukos X D (2010a), Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. The Journal of Machine Learning Research 11:171–234

Aliferis C F, Statnikov A, Tsamardinos I, Mani S, and Koutsoukos X D (2010b), Local causal and Markov blanket induction for causal discovery and feature selection for classification part II: Analysis and extensions. The Journal of Machine Learning Research 11:235–284

Astle W, and Balding D J (2009), Population structure and cryptic relatedness in genetic association studies. Statistical Science 24(4):451–471

Aten J E, Fuller T F, Lusis A J, and Horvath S (2008), Using genetic markers to orient the edges in quantitative trait networks: the NEO software. BMC Systems Biology 2(1):1–21

Bauer E, Falque M, Walter H, Bauland C, Camisan C, Campo L, Meyer N, Ranc N, Rincent R, Schipprack W, Altmann T, Flament P, Melchinger A E, Menz M, Moreno-González J, Ouzunova M, Revilla P, Charcosset A, Martin O C, and Schön C C (2013), Intraspecific variation of recombination rate in maize. Genome Biology 14(9):R103

Bernal-Vasquez A M, Möhring J, Schmidt M, Schönleben M, Schön C C, and Piepho H P (2014), The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. BMC Genomics 15(1):646

Bouckaert R R (1995), Bayesian belief networks: from construction to inference. Ph.D. thesis, Faculteit Wiskunde en Informatica, Utrecht University, Utrecht, The Netherlands

Butler D, Cullis B R, Gilmour A, and Gogel B (2009), ASReml-R reference manual. Department of Primary Industries and Fisheries, Brisbane, Australia

Chickering D M (1995), A transformational characterization of equivalent Bayesian network structures. Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, pages 87–98, Morgan Kaufmann Publishers Inc., Montréal, Qué, Canada

Cover T M, and Thomas J A (2006), Elements of Information Theory. John Wiley and Sons Inc., Hoboken, New Jersey, USA, 2 edition

Daly R, and Shen Q (2007), Methods to accelerate the learning of Bayesian network structures. Proceedings of the 2007 UK workshop on computational intelligence, Citeseer, Imperial College, London, UK

De Faveri J, Verbyla A P, Cullis B R, Pitchford W S, and Thompson R (2017), Residual variance–covariance modelling in analysis of multivariate data from variety selection trials. Journal of Agricultural, Biological and Environmental Statistics 22(1):1–22

de los Campos G (2015), Technical Note: Multi-Trait Models Based on Eigenvalues and Eigenvectors. Technical report, Department of Epidemiology and Biostatistics and Department of Statistics, Michigan State University, East Lansing, MI, USA. Unpublished, available from Plant Breeding, Technical University of Munich, Germany.

de los Campos G, and Pérez-Rodríguez P (2012), BLR: Bayesian Linear Regression. R package version 1.3. http://CRAN.R-project.org/package=BLR

de los Campos G, and Pérez-Rodríguez P (2014), BGLR: Bayesian Generalized Linear Regression. R package version 1.0.3. http://CRAN.R-project.org/package=BGLR

de Maturana E L, de los Campos G, Wu X L, Gianola D, Weigel K A, and Rosa G J M (2010), Modeling relationships between calving traits: a comparison between standard and recursive mixed models. Genetics Selection Evolution 42(1):1–9

de Maturana E L, Wu X L, Gianola D, Weigel K A, and Rosa G J M (2009), Exploring biological relationships between calving traits in primiparous cattle with a Bayesian recursive model. Genetics 181(1):277–287

Edwards D I (2000), Introduction to graphical modelling. Springer, New York, NY, USA, 2 edition

Falconer D S (1952), The problem of environment and selection. The American Naturalist 86(830):293–298

Falconer D S, and Mackay T F C (1996), Introduction to Quantitative Genetics. Longmans Green, Essex, UK, 4 edition

Felipe V P S, Silva M A, Valente B D, and Rosa G J M (2015), Using multiple regression, Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes. Poultry Science 94(4):772–780

Fisher R A (1918), The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52:399–433

Ganal M W, Durstewitz G, Polley A, Bérard A, Buckler E S, Charcosset A, Clarke J D, Graner E M, Hansen M, Joets J, Le Paslier M C, McMullen M D, Montalent P, Rose M, Schön C C, Sun Q, Walter H, Martin O C, and Falque M (2011), A large maize (Zea mays L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6(12):e28334

Geiger D, and Heckerman D (1994), Learning Gaussian networks: a unification for discrete and Gaussian domains. Technical Report MSR-TR-94-10, Morgan Kaufmann Publishers, Microsoft Research, Redmond, Washington, USA

Geiger D, and Heckerman D (1995), Learning Gaussian networks: a unification for discrete and Gaussian domains. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, volume 1, pages 274–284

Gianola D, and Sorensen D (2004), Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. Genetics 167(3):1407–1424

Gilmour A R, Cullis B R, and Verbyla A P (1997), Accounting for natural and extraneous variation in the analysis of field experiments. Journal of Agricultural, Biological, and Environmental Statistics 2(3):269–293

Good P I (2005), Permutation, Parametric and Bootstrap Tests of Hypotheses. Springer Science and Business Media, Inc., New York, NY, USA, 3 edition

Green P J, Łatuszyński K, Pereyra M, and Robert C P (2015), Bayesian computation: a summary of the current state, and samples backwards and forwards. Statistics and Computing 25(4):835–862

Grubbs F E (1950), Sample Criteria for Testing Outlying Observations. The Annals of Mathematical Statistics 21(1):27–58

Guo G, Zhao F, Wang Y, Zhang Y, Du L, and Su G (2014), Comparison of single-trait and multiple-trait genomic prediction models. BMC Genetics 15(1):1–7

Hageman R S, Leduc M S, Korstanje R, Paigen B, and Churchill G A (2011), A Bayesian framework for inference of the genotype–phenotype map for segregating populations. Genetics 187(4):1163–1170

Hazel L N (1943), The genetic basis for constructing selection indexes. Genetics 28(6):476–490

Henderson C (1977), Best linear unbiased prediction of breeding values not in the model for records. Journal of Dairy Science 60(5):783–787

Jia Y, and Jannink J L (2012), Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192(4):1513–1522

Jiang J, Zhang Q, Ma L, Li J, Wang Z, and Liu J F (2015), Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity 115(1):29–36

Jiang Y, and Reif J C (2015), Modeling epistasis in genomic selection. Genetics 201(2):759–768

Korb K B, and Nicholson A E (2010), Bayesian Artificial Intelligence. Chapman and Hall Ltd., Boca Raton, FL, USA, 2 edition

Kullback S (1959), Information Theory and Statistics. John Wiley and Sons Inc., Hoboken, NY, USA

Lam W, and Bacchus F (1994), Learning Bayesian belief networks: an approach based on the MDL principle. Computational Intelligence 10(3):269–293

Legendre P (2000), Comparison of permutation methods for the partial correlation and partial Mantel tests. Journal of Statistical Computation and Simulation 67(1):37–73

Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger A E, Menz M, Meyer N, Moreau L, Moreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack W, Schönleben M, Walter H, Charcosset A, and Schön C C (2014), Usefulness of multiparental populations of maize (Zea mays L.) for genome-based prediction. Genetics 198(1):3–16

Lehermeier C, Schön C C, and de los Campos G (2015), Assessment of genetic heterogeneity in structured plant populations using multivariate whole-genome regression models. Genetics 201(1):323–337

Li R, Tsaih S W, Shockley K, Stylianou I M, Wergedal J, Paigen B, and Churchill G A (2006), Structural model analysis of multiple quantitative traits. PLoS Genetics 2(7):e114

Lynch M, and Walsh B (1998), Genetics and Analysis of Quantitative Traits, volume 1. Sinauer Associates, Sunderland, MA, USA

Maier R, Moser G, Chen G B, and Ripke S (2015), Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. The American Journal of Human Genetics 96(2):283–294

Malik H N, Malik S I, Hussain M, Chughtai S U R, and Javed H I (2005), Genetic correlation among various quantitative characters in maize (Zea mays L.) hybrids. Journal of Agriculture and Social Sciences 1:262–265

Margaritis D (2003), Learning Bayesian network model structure from data. Ph.D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

Möhring J, and Piepho H P (2009), Comparison of weighting in two-stage analysis of plant breeding trials. Crop Science 49(6):1977–1988

Mir C, Zerjal T, Combes V, Dumas F, Madur D, Bedoya C, Dreisigacker S, Franco J, Grudloyma P, Hao P X, Hearne S, Jampatong C, Laloë D, Muthamia Z, Nguyen T, Prasanna B M, Taba S, Xie C X, Yunus M, Zhang S, Warburton M L, and Charcosset A (2013), Out of America: tracing the genetic footprints of the global diffusion of maize. Theoretical and Applied Genetics 126(11):2671–2682

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, and Vitezica Z (2016), Manual for BLUPF90 family of programs. University of Georgia, Athens, USA

Morota G, Valente B, Rosa G, Weigel K, and Gianola D (2012), An assessment of linkage disequilibrium in Holstein cattle using a Bayesian network. Journal of Animal Breeding and Genetics 129(6):474–487

Nagarajan R, Scutari M, and Lèbre S (2013), Bayesian Networks in R with Applications in Systems Biology. Springer, New York, NY, USA, 2013 edition

Nazarian A, and Gezan S A (2016), GenoMatrix: a software package for pedigree-based and genomic prediction analyses on complex traits. Journal of Heredity 107(4):372–379

Neto E C, Ferrara C T, Attie A D, and Yandell B S (2008), Inferring causal phenotype networks from segregating populations. Genetics 179(2):1089–1100

Neto E C, Keller M P, Attie A D, and Yandell B S (2010), Causal graphical models in systems genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. The Annals of Applied Statistics 4(1):320–339

Newton M A, and Raftery A E (1994), Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). Journal of the Royal Statistical Society Series B (Methodological) 56(1):3–48

Pearl J (1988), Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., Los Altos, CA, USA, 1 edition

Pearl J (2000), Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge, UK

Peñagaricano F, Valente B D, Steibel J P, Bates R O, Ernst C W, Khatib H, and Rosa G J M (2015), Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. BMC Systems Biology 9(1):1–9

Pérez-Elizalde S, Cuevas J, Pérez-Rodríguez P, and Crossa J (2015), Selection of the bandwidth parameter in a Bayesian kernel regression model for genomic-enabled prediction. Journal of Agricultural, Biological, and Environmental Statistics 20(4):512–532

Piepho H P, Richter C, and Williams E (2008), Nearest Neighbour Adjustment and Linear Variance Models in Plant Breeding Trials. Biometrical Journal 50(2):164–189

Porth I, Klápště J, Skyba O, Friedmann M C, Hannemann J, Ehlting J, El-Kassaby Y A, Mansfield S D, and Douglas C J (2013), Network analysis reveals the relationship among wood properties, gene expression levels and genotypes of natural Populus trichocarpa accessions. New Phytologist 200(3):727–742

Pszczola M, Veerkamp R F, de Haas Y, Wall E, Strabel T, and Calus M P L (2013), Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. Animal 7(11):1759–1768

R Core Team (2014), R: a language and environment for statistical computing

Rissanen J (1978), Modeling by shortest data description. Automatica 14(5):465–471

Robertson A (1959), The sampling variance of the genetic correlation coefficient. Biometrics 15(4):469–485

Rockman M V (2008), Reverse engineering the genotype–phenotype map with natural genetic variation. Nature 456(7223):738–744

Rodríguez-Álvarez M X, Boer M P, van Eeuwijk F A, and Eilers P H (2016), Spatial models for field trials. arXiv preprint arXiv:1607.08255

Roff D A (1995), The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. Heredity 74(5):481–490

Rosa G J M, Valente B D, de los Campos G, Wu X L, Gianola D, and Silva M A (2011), Inferring causal phenotype networks using structural equation models. Genetics Selection Evolution 43(6):1–13

Schadt E E, Lamb J, Yang X, Zhu J, Edwards S, GuhaThakurta D, Sieberts S K, Monks S, Reitman M, and Zhang C (2005), An integrative genomics approach to infer causal associations between gene expression and disease. Nature Genetics 37(7):710–717

Scutari M (2010), Learning Bayesian networks with the bnlearn R package. Journal of Statistical Software 35(3):1–22

Scutari M, and Brogini A (2012), Bayesian network structure learning with permutation tests. Communications in Statistics - Theory and Methods 41(16-17):3233–3243

Scutari M, Howell P, Balding D J, and Mackay I (2014), Multiple quantitative trait analysis using Bayesian networks. Genetics 198(1):129–137

Scutari M, Mackay I, and Balding D (2013), Improving the efficiency of genomic selection. Statistical applications in genetics and molecular biology 12(4):517–527

Scutari M, and Nagarajan R (2011), On identifying significant edges in graphical models. Proceedings of workshop on probabilistic problem solving in biomedicine, pages 15–27, Springer, Bled, Slovenia

Scutari M, and Nagarajan R (2013), Identifying significant edges in graphical models of molecular networks. Artificial Intelligence in Medicine 57(3):207–217

Searle S R (1961), Phenotypic, genetic and environmental correlations. Biometrics 17(3):474–480

Sneath P H A, and Sokal R R (1973), Numerical Taxonomy. The Principles and Practice of Numerical Classification. W. H. Freeman and Company Ltd., San Francisco, CA, USA

Sorensen D, and Gianola D (2002), Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health, Springer, New York, NY, USA

Spiegelhalter D J, Best N G, Carlin B P, and Van Der Linde A (2002), Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(4):583–639

Tenaillon M I, and Charcosset A (2011), A European perspective on maize history. Comptes Rendus Biologies 334(3):221–228, on the trail of domestications, migrations and invasions in agriculture

Töpner K, Rosa G J M, Gianola D, and Schön C C (2017), Bayesian networks illustrate genomic and residual trait connections in maize (*Zea mays* L.). G3: Genes, Genomes, Genetics 7(8):2779–2789

Tsamardinos I, Brown L E, and Aliferis C F (2006), The max-min hill-climbing Bayesian network structure learning algorithm. Machine learning 65(1):31–78

Valente B D, Morota G, Peñagaricano F, Gianola D, Weigel K, and Rosa G J M (2015), The causal meaning of genomic predictors and how it affects construction and comparison of genome-enabled selection models. Genetics 200(2):483–494

Valente B D, Rosa G J M, de los Campos G, Gianola D, and Silva M A (2010), Searching for recursive causal structures in multivariate quantitative genetics mixed models. Genetics 185(2):633–644

Valente B D, Rosa G J M, Gianola D, Wu X L, and Weigel K (2013), Is structural equation modeling advantageous for the genetic improvement of multiple traits? Genetics 194(3):561–572

Velazco J G, Rodríguez-Álvarez M X, Boer M P, Jordan D R, Eilers P H C, Malosetti M, and van Eeuwijk F A (2017), Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. Theoretical and Applied Genetics 130(7):1375–1392

Verma T S, and Pearl J (1990), Equivalence and synthesis of causal models. Uncertainty in Artificial Intelligence 6, pages 255–268, Elsevier Science Publisher B.V., Amsterdam, The Netherlands

Vázquez A I, Bates D M, Rosa G J M, Gianola D, and Weigel K A (2010), Technical note: an R package for fitting generalized linear mixed models in animal breeding. Journal of Animal Science 88(2):497–504

Wang H, and van Eeuwijk F A (2014), A new method to infer causal phenotype networks using QTL and phenotypic information. PLoS ONE 9(8):e103997

Williams E, Piepho H P, and Whitaker D (2011), Augmented p-rep designs. Biometrical Journal 53(1):19–27

Winrow C J, Williams D L, Kasarskis A, Millstein J, Laposky A D, Yang H S, Mrazek K, Zhou L, Owens J R, and Radzicki D (2009), Uncovering the genetic landscape for multiple sleep-wake traits. PLoS One 4(4):e5161

12 Appendix

Reduced representation of the fully conditional posterior distribution of the bandwidth parameter

First, a suitable definition of $\boldsymbol{s}^*$ is derived, and, then, the equivalence in Equation 45 is shown. As $\boldsymbol{\Sigma}$ is assumed to follow an inverse-Wishart distribution, it is positive definite and can be decomposed into its Cholesky factors

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' \tag{A1}$$

where $\mathbf{L}$ is a $(d \times d)$-dimensional lower triangular matrix. Define $\mathbf{M} = \mathbb{I}_{n\times n} \otimes \mathbf{L}$, so that $\mathbf{M}^{-1} = \mathbb{I}_{n\times n} \otimes \mathbf{L}^{-1}$ and $(\mathbb{I}_{n\times n} \otimes \mathbf{L})(\mathbb{I}_{n\times n} \otimes \mathbf{L}^{-1}) = \mathbb{I}_{nd\times nd}$. Defining

$$\boldsymbol{s}^* = \mathbf{M}^{-1}\boldsymbol{s} \tag{A2}$$

yields

$$\begin{aligned} \text{Var}(\boldsymbol{s}^*) &= \text{Var}(\mathbf{M}^{-1}\boldsymbol{s}) = \mathbf{M}^{-1}\text{Var}(\boldsymbol{s})\left(\mathbf{M}^{-1}\right)' = \mathbf{M}^{-1}(\mathbf{S}_\vartheta \otimes \boldsymbol{\Sigma})(\mathbf{M}^{-1})' \\ &= (\mathbb{I}_{n\times n} \otimes \mathbf{L}^{-1})(\mathbf{S}_\vartheta \otimes \boldsymbol{\Sigma})(\mathbb{I}_{n\times n} \otimes \mathbf{L}^{-1})' = \mathbf{S}_\vartheta \otimes \mathbb{I}_{d\times d} \end{aligned} \tag{A3}$$

because $\mathbf{L}^{-1}\boldsymbol{\Sigma}(\mathbf{L}^{-1})' = \mathbb{I}_{d\times d}$, which follows directly from Equation A1. Hence, $\boldsymbol{s}^* \mid \vartheta \sim \mathcal{N}_{nd}(\mathbf{0}, \mathbf{S}_\vartheta \otimes \mathbb{I}_{d\times d})$. A similar transformation and the resulting covariance structure was used by Vázquez et al. (2010).

For Equation 45 to hold, it suffices to show that

$$\text{p}(\boldsymbol{s} \mid \vartheta) = \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} \cdot \text{p}(\boldsymbol{s}^* \mid \vartheta) \tag{A4}$$

This can be easily seen by re-formulating the multivariate normal distribution of $\boldsymbol{s}$:

$$\begin{aligned} \text{p}(\boldsymbol{s} \mid \vartheta) &= \frac{1}{\sqrt{(2\pi)^{nd}\det(\mathbf{S}_\vartheta \otimes \boldsymbol{\Sigma})}} \exp\left(-0.5\boldsymbol{s}'(\mathbf{S}_\vartheta \otimes \boldsymbol{\Sigma})^{-1}\boldsymbol{s}\right) \\ &= \frac{1}{\sqrt{(2\pi)^{nd}\det(\mathbf{S}_\vartheta)^d\det(\boldsymbol{\Sigma})^n}} \exp\left(-0.5\boldsymbol{s}'(\mathbf{S}_\vartheta{}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\boldsymbol{s}\right) \\ &= \frac{1}{\sqrt{(2\pi)^{nd}\det(\mathbf{S}_\vartheta)^d}} \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})^n}} \exp\left(-0.5\boldsymbol{s}'\left[\mathbf{S}_\vartheta{}^{-1} \otimes (\mathbf{L}^{-1})'(\mathbf{L}^{-1})\right]\boldsymbol{s}\right) \\ &= \frac{1}{\sqrt{(2\pi)^{nd}\det(\mathbf{S}_\vartheta)^d}} \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} \exp\left(-0.5\left[\boldsymbol{s}'(\mathbb{I}_{n\times n} \otimes \mathbf{L}^{-1})'\right](\mathbf{S}_\vartheta{}^{-1} \otimes \mathbf{L}^{-1})\boldsymbol{s}\right) \\ &= \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} \frac{1}{\sqrt{(2\pi)^{nd}\det(\mathbf{S}_\vartheta)^d}} \exp\left(-0.5\boldsymbol{s}^{*\prime}(\mathbf{S}_\vartheta{}^{-1} \otimes \mathbb{I}_{d\times d})\left[(\mathbb{I}_{n\times n} \otimes \mathbf{L}^{-1})\boldsymbol{s}\right]\right) \end{aligned}$$

$$= \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} \frac{1}{\sqrt{(2\pi)^{nd} \det(\mathbf{S}_{\vartheta} \otimes \mathbb{I}_{d\times d})}} \exp\left(-0.5 \boldsymbol{s}^{*\prime} (\mathbf{S}_{\vartheta} \otimes \mathbb{I}_{d\times d})^{-1} \boldsymbol{s}^{*}\right) = \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} \mathrm{p}(\boldsymbol{s}^{*} \mid \vartheta) \tag{A5}$$

Table A1: Full-sib families in the Flint and Dent panel

Name	Parents	Number of DH lines
Dent (10 families)		**831**
CFD02	F353 × B73	64
CFD03	F353 × D06	97
CFD04	F353 × D09	100
CFD05	F353 × EC169	64
CFD06	F353 × F252	92
CFD07	F353 × F618	104
CFD09	F353 × Mo17	53
CFD10	F353 × UH250	94
CFD11	F353 × UH304	81
CFD12	F353 × W117	82
Flint (11 families)		**805**
CFF03	UH007 × D152	72
CFF04	UH007 × EC49A	29
CFF05	UH007 × EP44	17
CFF06	UH007 × EZ5	26
CFF07	UH007 × F03802	129
CFF08	UH007 × F2	54
CFF09	UH007 × F283	133
CFF10	UH007 × F64	62
CFF12	UH007 × UH006	92
CFF13	UH007 × UH009	98
CFF15	UH007 × DK105	93

F353 is the central Dent line and UH007 is the central Flint line.

Table A2: Number of analyzed plots per field in 2011 (Flint)

Field	Number of analyzed plots
INRA 1	528
KWS 1	531
SYN 1	525
TUM 1	529
CIAM 1	528
HOH 1	528
INRA 2	528
KWS 2	526
SYN 2	520
TUM 2	523
CIAM 2	528
HOH 2	529

Ploudaniel (INRA), Einbeck (KWS), Wadersloh (SYN), Roggenstein (TUM), La Coruna (CIAM). Total plot count was 6323.

Table A3: Single-structure evaluation: Predictive ability[a] with SD for the MTM (*cf.* Equation 5) and for SEM (*cf.* Equations 28) including a genomic ($\mathbf{\Lambda}_{g^*}$) or residual ($\mathbf{\Lambda}_e$) trait structure denoted by the BN algorithm it originates from

BN Giving $\mathbf{\Lambda}_{g^*}$	BN Giving $\mathbf{\Lambda}_e$	DMY	DMC	PH	DtTAS	DtSILK
Dent						
—	GS 3	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
—	GS 1, 2, 4	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.05)	0.68 (0.04)
—	TABU 1, 2	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
TABU 1	—	0.52 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.05)	0.68 (0.04)
—	—	0.52 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.04)	0.68 (0.04)
TABU 2	—	0.52 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.04)	0.68 (0.04)
GS 1, 2, 3, 4	—	0.52 (0.05)	0.63 (0.04)	0.69 (0.04)	0.62 (0.04)	0.67 (0.03)
Flint						
—	TABU 1	0.64 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.04)	0.76 (0.04)
—	GS 1, 2, 3, 4	0.64 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.04)	0.76 (0.04)
—	TABU 2	0.64 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.04)	0.76 (0.04)
—	—	0.64 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.04)	0.76 (0.04)
TABU 1, 2	—	0.62 (0.04)	0.67 (0.05)	0.69 (0.05)	0.74 (0.03)	0.75 (0.03)
GS 1, 2, 3, 4	—	0.63 (0.04)	0.66 (0.05)	0.67 (0.05)	0.73 (0.04)	0.75 (0.04)

Traits: biomass dry matter yield (DMY) (dt/ha), biomass dry matter content (DMC) (%), plant height (PH) (cm), days to tasseling (DtTAS) (days), days to silking (DtSILK) (days).
For notation of BN algorithms see Table 2.

[a] Average of 10 random 5-fold cross-validations.

Table A4: Double-structure evaluation: Predictive ability[a] with SD for the MTM (*cf.* Equation 5) and for SEM (*cf.* Equations 28) including a genomic ($\mathbf{\Lambda}_{g^*}$) and residual ($\mathbf{\Lambda}_e$) trait structure denoted by the BN algorithm it originates from

BN Giving $\mathbf{\Lambda}_{g^*}$	BN Giving $\mathbf{\Lambda}_e$	DMY	DMC	PH	DtTAS	DtSILK
Dent						
TABU 1	GS 3	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
TABU 2	GS 3	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
TABU 1	GS 1, 2, 4	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
TABU 1	TABU 1, 2	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
TABU 2	TABU 1, 2	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.63 (0.05)	0.68 (0.04)
TABU 2	GS 1, 2, 4	0.53 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.05)	0.68 (0.04)
—	—	0.52 (0.05)	0.64 (0.04)	0.69 (0.04)	0.62 (0.04)	0.68 (0.04)
GS 1, 2, 3, 4	GS 3	0.53 (0.05)	0.63 (0.04)	0.69 (0.04)	0.62 (0.04)	0.67 (0.04)
GS 1, 2, 3, 4	GS 1, 2, 4	0.53 (0.05)	0.63 (0.04)	0.69 (0.04)	0.62 (0.04)	0.67 (0.04)
GS 1, 2, 3, 4	TABU 1, 2	0.53 (0.05)	0.63 (0.04)	0.69 (0.04)	0.62 (0.04)	0.67 (0.04)
Flint						
—	—	0.64 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.04)	0.76 (0.04)
TABU 1, 2	TABU 1	0.63 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.03)	0.76 (0.03)
TABU 1, 2	GS 1, 2, 3, 4	0.63 (0.04)	0.67 (0.05)	0.70 (0.04)	0.74 (0.03)	0.76 (0.03)
TABU 1, 2	TABU 2	0.63 (0.04)	0.67 (0.05)	0.70 (0.05)	0.74 (0.04)	0.76 (0.04)
GS 1, 2, 3, 4	TABU 2	0.64 (0.04)	0.67 (0.05)	0.68 (0.04)	0.73 (0.03)	0.75 (0.03)
GS1, 2, 3, 4	TABU 1	0.64 (0.04)	0.67 (0.05)	0.68 (0.05)	0.73 (0.04)	0.75 (0.04)
GS1, 2, 3, 4	GS 1, 2, 3, 4	0.64 (0.04)	0.67 (0.05)	0.68 (0.05)	0.73 (0.04)	0.75 (0.04)

Traits: biomass dry matter yield (DMY) (dt/ha), biomass dry matter content (DMC) (%), plant height (PH) (cm), days to tasseling (DtTAS) (days), days to silking (DtSILK) (days).
For notation of BN algorithms see Table 2.

[a] Average of 10 random 5-fold cross-validations.

Table A5: Model fit of genomic prediction models with spatial component for six prior and proposal settings in the Flint panel

Prior Assumption	$\vartheta \sim \Gamma(2,2)$			$\vartheta \sim \Gamma(4,0.15)$		
Starting Value ($\vartheta^{(t=0)}$)	0.10	1.00	2.00	0.10	1.00	2.00
SD Proposal (σ^2_ϑ)	0.05	0.10	0.10	0.10	0.05	0.10
Field						
Logarithm of the likelihood at posterior mean						
INRA 1	−703.7	−703.7	−704.2	−706.0	−707.4	−708.2
KWS 1	−328.0	−329.9	−328.1	−330.8	−330.9	−330.6
SYN 1	−772.3[b]	−772.7[b]	−770.6[b]	−781.5	−778.9	−780.3
TUM 1	−952.9	−953.1	−951.3	−955.9	−957.3	−955.8
CIAM 1[a]	−486.6	−486.3	−486.4	−501.9	−502.4	−501.5
HOH 1	−521.4	−524.7	−522.5	−524.5	−520.8	−523.9
INRA 2	−604.4	−604.5	−601.9	−600.4	−602.9	−600.5
KWS 2	−243.2	−242.4	−243.7	−243.4	−245.2	−242.5
SYN 2[a]	−770.3	−770.1	−769.8[b]	−770.9	−772.4	−773.1
TUM 2	−894.0[b]	−890.0[b]	−892.2[b]	−884.6	−885.1	−885.0
CIAM 2[a]	−323.0	−321.4	−321.1	−328.5	−326.1	−326.9
HOH 2[a]	−683.1	−684.8	−683.4	−687.0	−687.8	−686.0
DIC						
INRA 1	4490.8	4491.2	4491.6	4492.2	4494.4	4495.1
KWS 1	3833.5	3836.4	3834.1	3836.2	3836.3	3836.3
SYN 1	4551.6[b]	4551.1[b]	4548.6[b]	4560.8	4558.0	4559.7
TUM 1	4870.1	4870.4	4868.2	4871.9	4873.5	4872.8
CIAM 1[a]	4112.6	4112.0	4111.4	4127.3	4126.2	4126.0
HOH 1	4134.3	4137.3	4135.3	4137.0	4132.9	4137.2
INRA 2	4277.6	4278.3	4275.6	4273.7	4275.8	4274.9
KWS 2	3710.8	3710.6	3712.5	3711.2	3713.1	3710.1
SYN 2[a]	4564.4	4564.2	4563.2[b]	4563.3	4565.3	4566.4
TUM 2	4785.7[b]	4779.9[b]	4782.7[b]	4774.4	4774.0	4773.7
CIAM 2[a]	3770.2	3768.9	3768.7	3773.5	3771.1	3771.8
HOH 2[a]	4397.7	4399.4	4397.9	4399.9	4401.6	4399.5

Ploudaniel 2011 (INRA), Einbeck 2011 (KWS), Wadersloh 2011 (SYN), Roggenstein 2011 (TUM), La Coruna 2011 (CIAM)

[a] Low sensitivity to priors.

[b] MCMC not converged.

Table A6: Logarithm of the likelihood at posterior mean and DIC of genomic prediction models with and without spatial component (SC) (Flint panel)

Field	Logarithm of likelihood		DIC	
	with SC[a]	without SC	with SC[b]	without SC
INRA 1	−708.2	−2057.7	4490.8	5483.9
KWS 1	−330.9	−1568.7	3833.5	4657.0
SYN 1	−781.5	−2001.8	4548.6	5395.2
TUM 1	−957.3	−2301.8	4868.2	5867.0
CIAM 1	−502.4	−1951.8	4111.4	5181.1
HOH 1	−524.7	−1865.1	4132.9	5117.8
INRA 2	−604.5	−2060.8	4273.7	5455.2
KWS 2	−245.2	−1558.2	3710.1	4662.9
SYN 2	−773.1	−2020.7	4563.2	5465.9
TUM 2	−894.0	−2155.9	4773.7	5689.8
CIAM 2	−328.5	−1935.1	3768.7	5102.2
HOH 2	−687.8	−1968.1	4397.7	5272.0

Ploudaniel 2011 (INRA), Einbeck 2011 (KWS), Wadersloh 2011 (SYN), Roggenstein 2011 (TUM), La Coruna 2011 (CIAM)

[a] Minimum of all settings.

[b] Maximum of all settings.

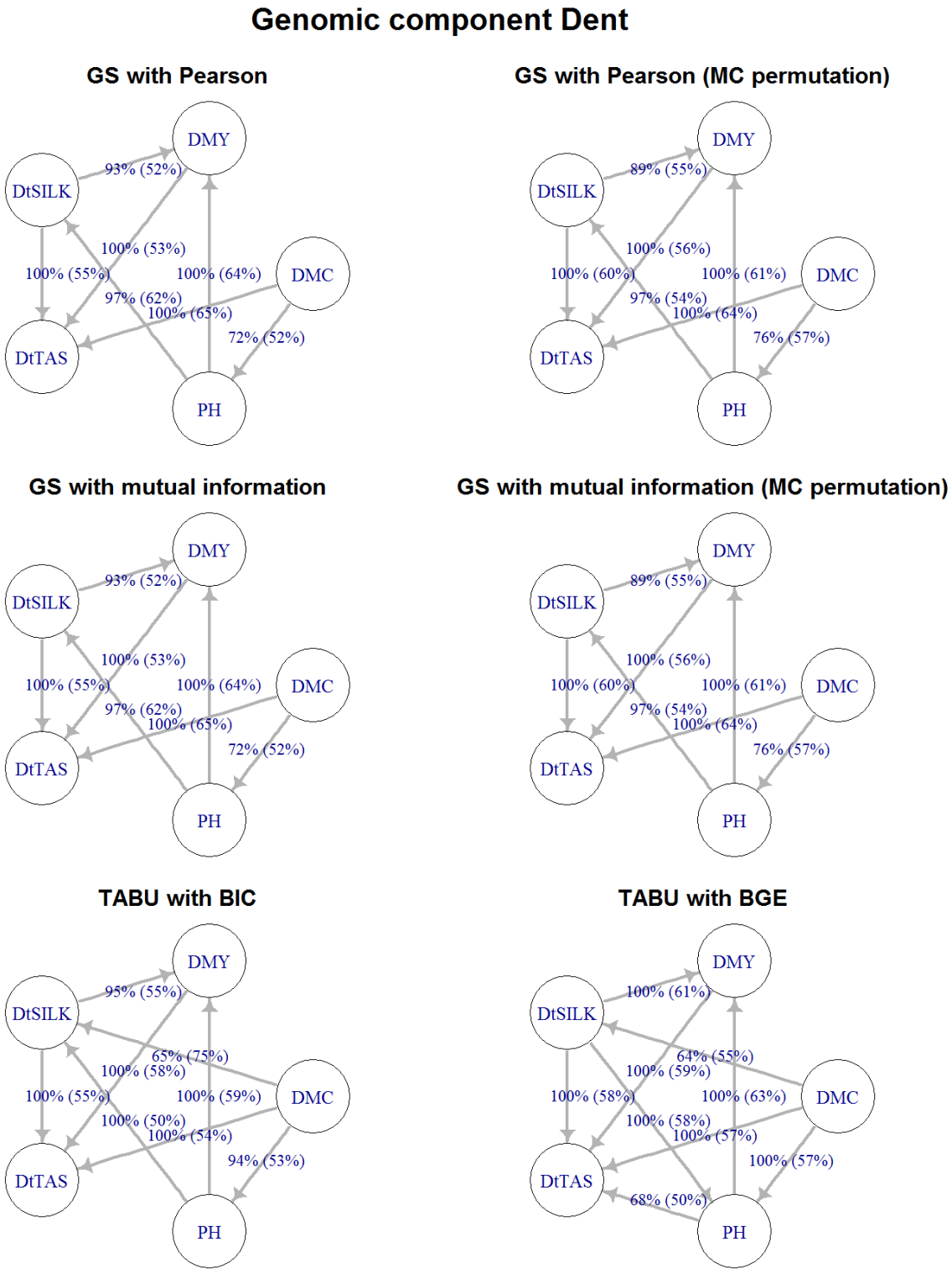


Figure A1: Networks of the genomic component in Dent: All networks identified the connections from DtSILK to DtTAS, from DtSILK to DMY, from DMC to PH, from PH to DMY, from DMY to DtTAS, from DMC to DtTAS, and between PH and DtSILK. Both score-based approaches (TABU 1 and 2) displayed an extra edge from DMC to DtSILK. The score-based approach with the BGE score (TABU 2) showed an additional edge from PH to DtTAS. The SEM favored the tabu-search with the BIC score (TABU 1) over all other settings. Labels of edges indicate the proportion of bootstrap samples supporting the edge and (in parentheses) the proportion having the direction shown. Edges that were not significant in the averaging process due to a network-internal empirical test on the arc's strength are not shown.

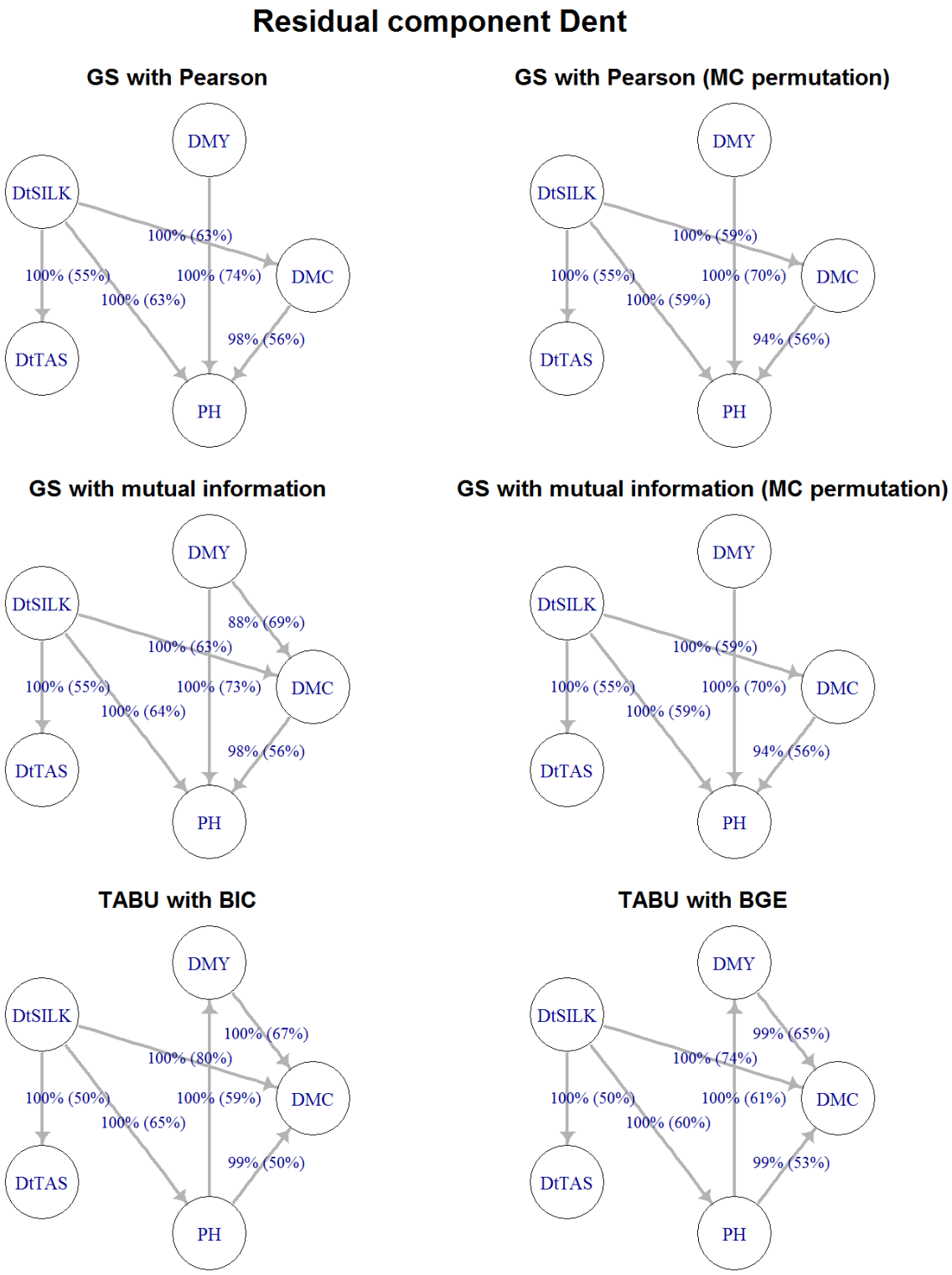


Figure A2: Networks of the residual component in Dent: All algorithms showed edges from DtSILK to DMC, from DtSILK to PH, from DtSILK to DtTAS, between DMY and PH, and between DMC and PH. One constraint-based approach (GS 3) and the score-based approaches (TABU 1, 2) identified an additional connection from DMY to DMC. The SEM favored the GS algorithm with the mutual information criterion (GS 3) over all other settings. Labels of edges indicate the proportion of bootstrap samples supporting the edge and (in parentheses) the proportion having the direction shown. Edges that were not significant in the averaging process due to a network-internal empirical test on the arc's strength are not shown.

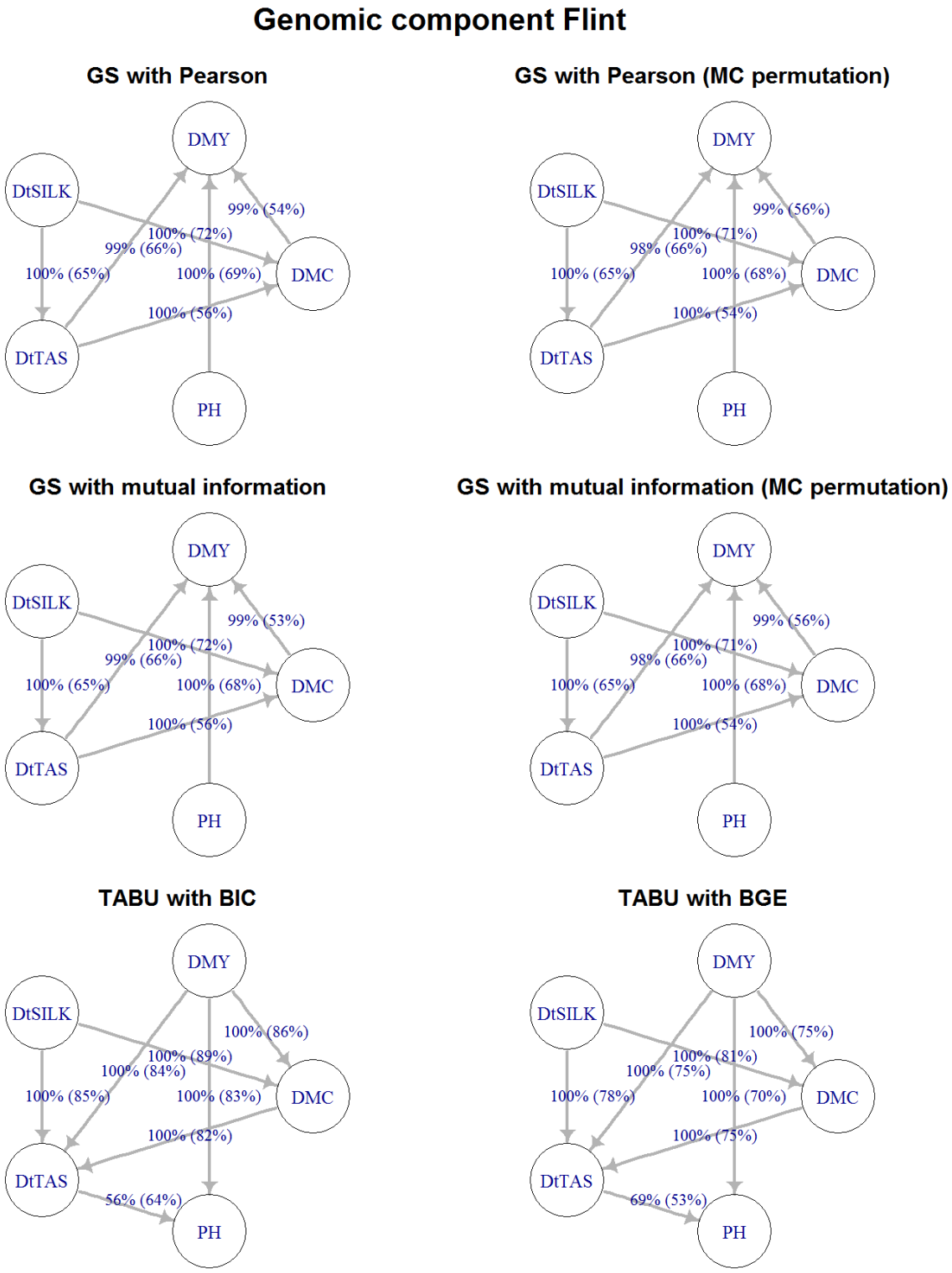


Figure A3: Networks of the genomic component in Flint. All algorithms identified the connections from DtSILK to DtTAS, from DtSILK to DMC, between DMY and DMC, between DMY and PH, between DMY and DtTAS, and between DtTAS and DMC. The score-based approaches (TABU 1, 2) oriented the connections with DMY and between DtTAS and DMC in the opposite direction than the constraint-based approaches (GS 1, 2, 3, 4). They also displayed an additional edge from DtTAS to PH. The SEM favored the tabu-search algorithms (TABU 1, 2) over all other settings. Labels of edges indicate the proportion of bootstrap samples supporting the edge and (in parentheses) the proportion having the direction shown. Edges that were not significant in the averaging process due to a network-internal empirical test on the arc's strength are not shown

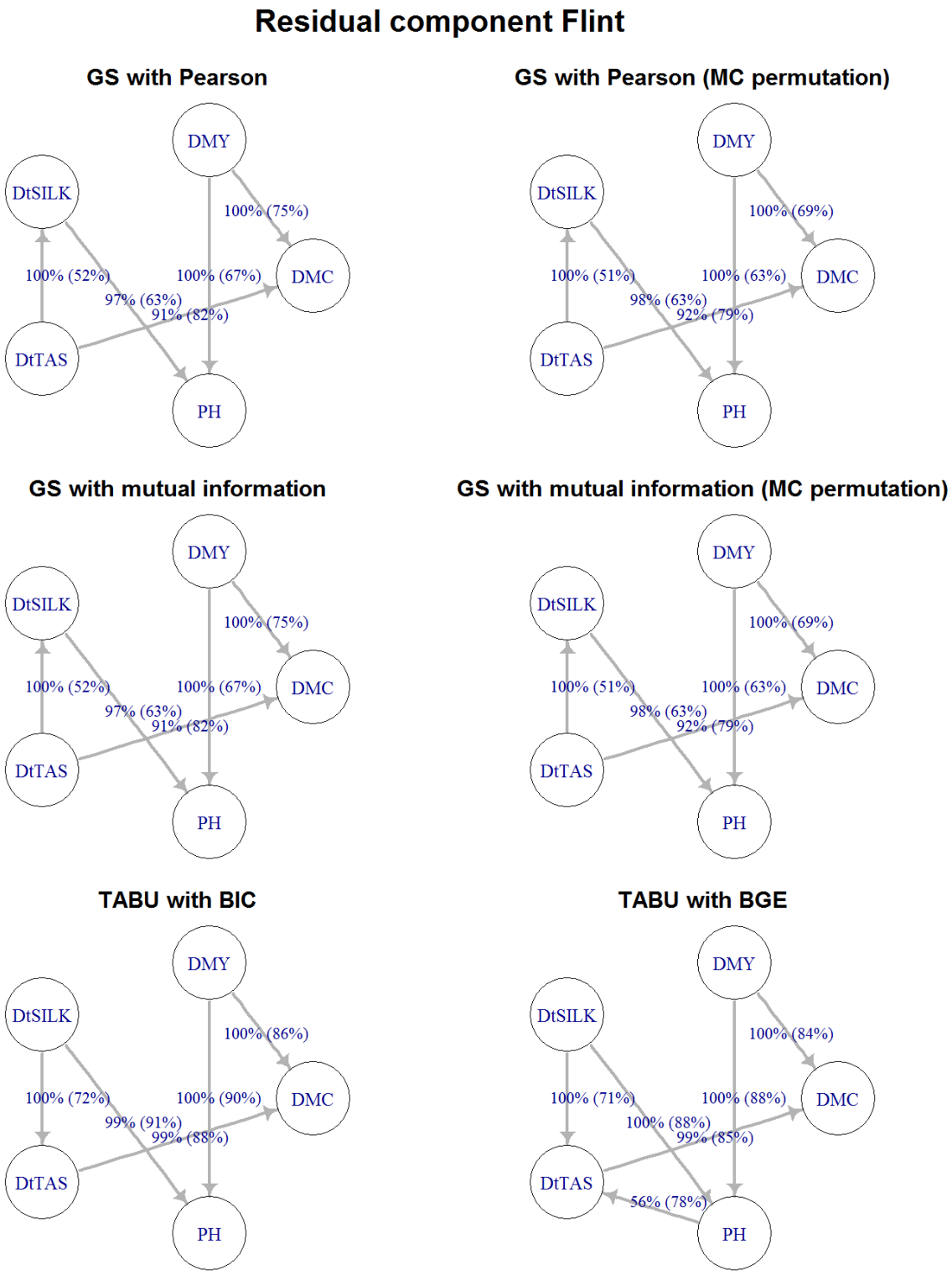


Figure A4: Networks of the residual component in Flint: All algorithms showed edges from DtSILK to PH, from DMY to PH, from DMY to DMC, from DtTAS to DMC, and between DtTAS and DtSILK. The score-based approach with the BGE (TABU 2) identified an additional connection from PH to DtTAS. The SEM favored the GS algorithm with the mutual information criterion (GS 3) over all other settings. Labels of edges indicate the proportion of bootstrap samples supporting the edge and (in parentheses) the proportion having the direction shown. Edges that were not significant in the averaging process due to a network-internal empirical test on the arc's strength are not shown.

Algorithm A1 Gibbs sampler for multivariate mixed models, *cf.* de los Campos (2015)

Let $\boldsymbol{\Lambda}_{\cdot i}$ be the i^{th} column of $\boldsymbol{\Lambda}$.

Initialize all MCMC chains with a sensible starting value.

for all random components $\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega} \otimes \mathbf{V})$ **do**

Compute quasi-orthogonal representation of $\boldsymbol{u}$ (see Equation 42).

Compute $\mathbf{A}_i = \left(\mathbf{E}^{-1} + (\boldsymbol{\Psi}^{-1})_{ii}\mathbf{V}^{-1}\right)^{-1}$ for all $i \in \{1, \ldots, m\}$.

Draw $\boldsymbol{\nu}_i = (\boldsymbol{\nu}_{i1}, \ldots, \boldsymbol{\nu}_{id}) \sim \mathcal{N}_d\left((\boldsymbol{\Lambda}_{\cdot i})'[\boldsymbol{y} - \mathbf{1}_n \otimes \boldsymbol{\mu} - \boldsymbol{u}_r]\mathbf{E}^{-1}\mathbf{A}_i, \mathbf{A}_i\right)$ for all $i \in \{1, \ldots, m\}$, with $\boldsymbol{u}_r$ being the sum of all other random components.

$\boldsymbol{u} \leftarrow (\boldsymbol{\Lambda} \otimes \mathbb{I}_{d \times d})\boldsymbol{\nu}$

$\boldsymbol{\nu}_0 \leftarrow (\boldsymbol{\Psi}^{-\frac{1}{2}} \otimes \mathbb{I}_{d \times d})\boldsymbol{\nu}$ ▷ $\boldsymbol{\nu}_0 \sim \mathcal{N}_{md}(\mathbf{0}, \mathbb{I}_{m \times m} \otimes \mathbf{V})$

Arrange $\boldsymbol{\nu}_0$ in $\mathbf{M}_{\boldsymbol{\nu}_0}$ such that rows contain traits and columns contain the m transformed eigenvector projections.

Draw $\mathbf{V} \sim \mathcal{W}^{-1}\left(\mathbf{M}_{\boldsymbol{\nu}_0}{}'\mathbf{M}_{\boldsymbol{\nu}_0} + \boldsymbol{\Phi}_{\mathbf{V}}, m + \nu_{\mathbf{V}}\right)$.

if $\boldsymbol{\Omega} = \mathbf{S}_\vartheta$ **then** do MH step for $\boldsymbol{\Omega} = \mathbf{S}_\vartheta$ and ϑ. ▷ see Algorithm 1

end for

Arrange $\boldsymbol{y} - \sum_{k=1}^{p} \boldsymbol{u}_k$ in $\mathbf{Y}^*$ such that rows contain traits and columns contain individuals.
Draw $\boldsymbol{\mu} \sim \mathcal{N}_d\left(\frac{\mathbf{Y}^*\mathbf{1}_n}{n}, \frac{\mathbf{E}}{n}\right)$.

Arrange the residuals in $\mathbf{M}_{\boldsymbol{e}}$ such that rows contain traits and columns contain individuals.
Draw $\mathbf{E} \sim \mathcal{W}^{-1}\left(\mathbf{M}_{\boldsymbol{e}}{}'\mathbf{M}_{\boldsymbol{e}} + \boldsymbol{\Phi}_{\mathbf{E}}, m + \nu_{\mathbf{E}}\right)$.

Algorithm A2 GS Markov Blanket, *cf.* Margaritis (2003)

Let Ω be the whole set of variables, **X** the node under investigation, **Y** any other node, and S the Markov blanket of **X** to be. Then

$S \leftarrow \emptyset$

while $\exists \mathbf{Y} \in \Omega \setminus \{\mathbf{X}\}$ such that **Y** is probabilistically dependent of **X** given S **do** $S \leftarrow S \cup \{\mathbf{Y}\}$
end while ▷ Growing phase

while $\exists \mathbf{Y} \in S$ such that **Y** is probabilistically independent of **X** given $S \setminus \{\mathbf{Y}\}$ **do** $S \leftarrow S \setminus \{\mathbf{Y}\}$
end while ▷ Shrinking phase

return S

Algorithm A3 GS (full version), *cf.* Margaritis (2003)

Compute the Markov blanket B(**X**) of all **X** according to Algorithm A2.

for all X and $\mathbf{Y} \in B(\mathbf{X})$ **do** determine if **Y** is a direct neighbor of **X** and if so, add **Y** to $N(\mathbf{X})$.
end for

for all X and $\mathbf{Y} \in N(\mathbf{X})$ **do** orient $\mathbf{Y} \rightarrow \mathbf{X}$ if $\exists \mathbf{Z} \in N(\mathbf{X}) \setminus N(\mathbf{Y}) \setminus \{\mathbf{Y}\}$ such that **Y** and **Z** are dependent given $S \cup \{\mathbf{X}\}$ $\forall S \subseteq T$, where T is the smaller of $B(\mathbf{Y}) \setminus \{\mathbf{X}, \mathbf{Z}\}$ and $B(\mathbf{Z}) \setminus \{\mathbf{X}, \mathbf{Y}\}$.
end for

while $\exists$ cycles **do**

Compute the set of edges C being part of a cycle.

Remove from the current graph the edge in C that is part of the greatest number of cycles and put it in R.

end while

Insert each edge from R in the graph in reverse order of removal in the step before, reversed.

while $\exists$**X** and $\mathbf{Y} \in N(\mathbf{X})$ with no direction between **X** an **Y** **do**
If there exists a directed path from **X** to **Y**, orient $\mathbf{X} \rightarrow \mathbf{Y}$.
end while

Algorithm A4 Tabu Search, *cf.* Bouckaert (1995)

Let STOP be any stopping criterion (*e.g.*, the maximal number of iterations) and n the length of the tabu-list. Let S be the set of the arrows of the current network, N(S) all networks that arise from S by changing one arc, and counter a running integer.

Initialize S (*e.g.*, an empty or full graph).
$S_{\text{best}} \leftarrow S$
counter $\leftarrow 0$
Initialize the tabu-list with all elements being in S.

repeat

Select an element $R \in N(S)$, $R \notin$ tabu-list, that maximizes the score.

$S \leftarrow R$

tabu-list[counter] $\leftarrow$ R

counter $\leftarrow$ (counter $+ 1$) modulo n

if the score of S exceeds the score of S_{best} **then** $S_{\text{best}} \leftarrow S$

until STOP

return S_{best}

Code for MTM function with spatial component

Example for a call

Let pheno be a matrix containing phenotypes, individuals in rows, traits in columns.
Let distances be a matrix containing physical distances between individuals, sorted as pheno.
Let kinship be a kinship matrix (*e.g.*, simple matching and Astle and Balding 2009), also sorted as pheno.

```
Y <- scale(pheno)

K <- list(

        list(  ## spatial kernel
        Dist = distances,
        theta = 0.1,        # start value for bandwidth parameter
        thetasd = 0.05,     # standard deviation for proposal truncated
                              normal for bandwidth parameter
        shape = 2,          # shape parameter of Gamma prior for bandwidth parameter
        scale = 2,          # scale parameter of Gamma prior for bandwidth parameter

        COV = list(     # prior for spatial trait covariance matrix
                type = 'spatial',
                df0 = 8,
                S0 = diag(14/3,5))
        ),

        list( ## genomic kernel
                K = kinship,
                COV = list(type = 'UN', df0 = 8, S0 = diag(14/3,5))
        )

)

resCov = list( ## residual kernel
        type = "UN",
        df0 = 8,
        S0 = diag(14/3, ncol(as.matrix(Y))))

model <- MTM(Y = Y, K = K, resCov = resCov, nIter = 1500, burnIn = 500, thin = 2)
```

Modified MTM function

```
### Original MTM code from Gustave de los Campos
### amended by Katrin Toepner
### Plant Breeding TUM, 2016
### katrin.toepner@tum.de
### Amendments between "NEW begin" and "NEW end".

### requires package msm and MASS
library(msm)
library(MASS)

#' Fits a (Bayesian) Multivariate Gaussian Mixed Effects Model using a Gibbs
#' Sampler.
#'
#' Data equation: Y = 1mu' + XB + U1 + ... + Uq + E \cr\cr where: \itemize{
#' \item{Y (numeric, nxp) is a matrix of phenotypes (individuals in rows, traits
#' in columns, NAs accepted),} \item{mu is a vector of (p) intercepts (included
#' by default),} \item{X (nxq) is an incidence matrix for q fixed effects,}
#' \item{B is a matrix of fixed effects,} \item{U1, ..., Uq are (nxq)
#' matrices of random effects with vec(Uj)~N(0, kronecker(Gj, Kj)), where Gj is a
#' pxp (unknown) co-variance matrix, Kj is an nxn user-defined covariance
#' matrix (kernel) which must by numeric, symmetric positive semi-definite,}
#' \item{E (nxp) is a matrix of model residuals, assumed to follow a MVN
#' distribution vec(E)~N(0, kronecker(R0, I)).} }
#'
#' Intercepts are included by default, fixed and random effects are optional.
#' The same fixed effects are applied to all traits. For each random effect the
#' user must provide a kernel (K). By default the residual co-variance matrix
#' (R0) and the co-variance matrices of random effects are un-structured;
#' however users can specify other models (e.g., DIAG=diagonal, FA=factor
#' analysis, and REC=recursive).
#'
#' @param Y Phenotype matrix (nxp numeric, traits (p) in columns, individuals
#'   (n) in rows).
#' @param K A 2-level list, 1st level defines random effects, inside each level
#'   a list is used to provide the kernel (K), a covariance structure (type,
#'   'UN', 'DIAG', 'FA', 'spatial' supported) and hyper-parameters (degree of freedom,
#'   df0, and scale, S0).
#' @param resCov A list used to define the co-variance matrix for model
#'   residuals (R0). Example: resCov=list(type='UN', df0=x, S0=V) specifies an
#'   un-structured covariance matrix, with an Inverse Whishart prior with degree
#'   of freedom df0 (scalar) and scale matrix (pxp) V.
#' @param nIter The number of iterations (integer).
#' @param thin Thining interval (integer).
#' @param burnIn The number of iterations to be discarded as burn-in (integer).
#' @param XF A numeric design matrix (nxq) for fixed effects. For factors use
#'   XF=as.matrix(model.matrix(~x+y...))[, -1].
#' @param saveAt A character path and a prefix used to define where to store
#'   samples (e.g., saveAt='c:/mtmFit/test_'. By default samples are saved in the
```

```
#'   current directory and filenames have no prefix.
#' @param tolD A numeric parameter used to define the minimum eigenvalue to be
#'   maintained in the model. Eigenvectors of kernels smaller than tolD are
#'   removed. The default value is tolD=1e-6.
#' @example examples/MTM.R
#' @return List containing estimated posterior means and estimated posterior
#'   standard deviations, including: $yHat.
#' @export

MTM <- function(Y, XF = NULL, K = NULL,
                resCov = list(type = "UN", df0 = 0, S0 = diag(0,ncol(as.matrix(Y)))),
                nIter = 110, burnIn = 10, thin = 2, saveAt = "", tolD = 1e-05) {

    if ((nIter - burnIn - thin) < 0) {
        stop("nIter must be greater than thin+burnIn")
    }
    iter <- 0

    Y <- as.matrix(Y)
    traits <- ncol(Y)
    n <- nrow(Y)
    hasXF <- !is.null(XF)
    hasK <- !is.null(K)

    ## Initializing overall mean
    mu <- colMeans(Y, na.rm = TRUE)
    post_mu <- rep(0, traits)
    post_YHat <- matrix(nrow = n, ncol = traits, 0)
    post_logLik <- 0

    ## Initializing missing values
    YStar <- Y
    whichNa <- list()
    whichNa$subjects <- which(apply(FUN = any, X = is.na(Y), MARGIN = 1))
    nNa <- length(whichNa$subjects)
    whichNa$traits <- list()
    if (nNa > 0) {
        for (k in 1:nNa) {
            whichNa$traits[[k]] <- which(is.na(Y[whichNa$subjects[[k]], ]))
            tmpSubject <- whichNa$subject[[k]]
            tmpTraits <- whichNa$traits[[k]]
            YStar[tmpSubject, tmpTraits] <- mu[tmpTraits]
        }
    }
    hasNa <- rowSums(is.na(Y)) > 0

    ## Initialization residuals
    E <- t(t(YStar) - mu)
```

```
## Initializing Fixed effects
if (hasXF) {
    XF <- as.matrix(XF)
    dimX.f <- ncol(XF)
    tmp <- eigen(crossprod(XF))
    if (any(tmp$values < 0)) {
        Stop("XF is not full-column rank")
    }
    Tb.f <- tmp$vectors %*% diag(1/sqrt(tmp$values))
    XTb.f <- XF %*% Tb.f
    B.f <- matrix(nrow = dimX.f, ncol = traits, 0)
    for (i in 1:traits) {
        B.f[, i] <- lm(E[, i] ~ XF - 1)$coef
    }
    post_B.f <- B.f
    E <- E - XF %*% B.f
}

## Initialization R0
resCov$R <- var(E)/2
resCov$L <- chol(resCov$R)
resCov$RInv <- chol2inv(resCov$L)
resCov$post_R <- matrix(nrow = traits, ncol = traits, 0)

if (resCov$type == "REC") {
    resCov$B <- matrix(nrow = traits, ncol = traits, 0)
    resCov$PSI <- diag(resCov$R)
    resCov$post_PSI <- rep(0, traits)
    resCov$post_B <- matrix(nrow = traits, ncol = traits, 0)
}

if (resCov$type == "FA") {
    resCov$nF <- ncol(resCov$M)
    sdU <- apply(FUN = sd, MARGIN = 2, X = E/2)
    FA <- factanal(E/2, factors = resCov$nF)
    resCov$B <- matrix(nrow = traits, ncol = resCov$nF, 0)
    resCov$B[resCov$M] <- (diag(sdU) %*% FA$loadings)[resCov$M]
    resCov$PSI <- (sdU^2) * FA$uniquenesses + 1e-04
    resCov$R <- tcrossprod(resCov$B) + diag(resCov$PSI)
    resCov$L <- chol(resCov$R)
    resCov$RInv <- chol2inv(resCov$L)
    resCov$F <- matrix(nrow = n, ncol = resCov$nF, 0)
    resCov$post_PSI <- rep(0, traits)
    resCov$post_B <- matrix(nrow = traits, ncol = resCov$nF, 0)
}

if (resCov$type == "DIAG") {
    resCov$R <- diag(apply(FUN = var, X = E, MARGIN = 2))/2
    resCov$RInv <- diag(1/diag(resCov$R))
```

```
        resCov$post_R <- matrix(nrow = traits, ncol = traits, 0)
    }

    ## Initializing Kernel-components
    if (hasK) {
        post_K <- list()
        nK <- length(K)

        for (k in 1:nK) {
            K[[k]]$G <- resCov$R/nK

            post_K[[k]] <- list()
            post_K[[k]]$G <- matrix(nrow = traits, ncol = traits, 0)
            post_K[[k]]$U <- matrix(nrow = n, ncol = traits, 0)

### NEW begin
                if (K[[k]]$COV$type == "spatial") {
                    K[[k]]$K <- dist2kernel(K[[k]]$Dist, K[[k]]$theta)
                    K[[k]]$logLike <- -Inf
                    theta <- K[[k]]$theta
                    K[[k]]$p_theta <- 0
                    post_K[[k]]$theta <- 0
                }
### NEW end

            if (is.null(K[[k]]$EVD)) {
                tmp <- eigen(K[[k]]$K)
            } else {
                tmp <- K[[k]]$EVD
            }

            K[[k]]$V <- tmp$vectors[, tmp$values > tolD]
            K[[k]]$d <- tmp$values[tmp$values > tolD]
            K[[k]]$nD <- length(K[[k]]$d)
            K[[k]]$U <- matrix(nrow = n, ncol = traits, 0)


            if (K[[k]]$COV$type == "REC") {
                K[[k]]$B <- diag(traits)
                K[[k]]$PSI <- diag(K[[k]]$G)
                post_K[[k]]$B <- matrix(nrow = traits, ncol = traits, 0)
                post_K[[k]]$PSI <- rep(0, traits)
            }

            if (K[[k]]$COV$type == "FA") {
                K[[k]]$COV$nF <- ncol(K[[k]]$COV$M)
                sdU <- sqrt(diag(K[[k]]$G))
                FA <- factanal(covmat = K[[k]]$G, factors = K[[k]]$COV$nF)
                K[[k]]$B <- matrix(nrow = traits, ncol = K[[k]]$COV$nF, 0)
```

```
            K[[k]]$B[K[[k]]$COV$M] <- (diag(sdU) %*% FA$loadings)[K[[k]]$COV$M]
            K[[k]]$PSI <- (sdU^2) * FA$uniquenesses + 1e-04
            K[[k]]$G <- tcrossprod(K[[k]]$B) + diag(K[[k]]$PSI)
            K[[k]]$F <- matrix(nrow = K[[k]]$nD, ncol = K[[k]]$COV$nF, 0)
            post_K[[k]]$PSI <- rep(0, traits)
            post_K[[k]]$B <- matrix(nrow = traits, ncol = K[[k]]$COV$nF, 0)
        }
    }
}

### Gibbs Sampler
time <- proc.time()[3]
for (i in 1:nIter) {

    logLik <- 0
    ## Fixed effects
    if (hasXF) {
        E <- E + XF %*% B.f
        B.f <- sampleBf(Y = E, XTb = XTb.f, Tb = Tb.f, R = resCov$R, L = resCov$L,
            RInv = resCov$RInv, traits = traits, dimX = dimX.f)
        E <- E - XF %*% B.f
    }

    ## Kernels
    if (hasK)
        {
            for (k in 1:nK) {
              E <- E + K[[k]]$U
              GInv <- chol2inv(chol(K[[k]]$G))
              tmp <- sampleU(Y = E, RInv = resCov$RInv, GInv = GInv, V = K[[k]]$V,
                d = K[[k]]$d, traits = traits, df0 = K[[k]]$df0, S0 = K[[k]]$S0)
              E <- E - tmp$U
              K[[k]]$U <- tmp$U


### NEW begin

              if (K[[k]]$COV$type == "spatial"){

                    tmp0 <- tmp$U0/sqrt(K[[k]]$d)
                    SS <- crossprod(tmp0) + K[[k]]$COV$S0
                    df <- K[[k]]$nD + K[[k]]$COV$df0
                    K[[k]]$G <- MCMCpack::riwish(S = SS, v = df)

                    LInv <- solve(t(chol(var(tmp$U))))
                    MInv <- kronecker(LInv, diag(1, n))
                    u_star <- matrix(crossprod(t(MInv), as.vector(tmp$U)), ncol=
                        traits)
```

```
                Kspat <- dist2kernel(K[[k]]$Dist, theta)
                tmp0 <- eigen(Kspat)
                Kspat_V <- tmp0$vectors[, tmp0$values > tolD]
                Kspat_d <- tmp0$values[tmp0$values > tolD]

                logLike    <- mvdnorm(data = t(u_star), sigma = Kspat,
                    sigmavectors = Kspat_V,  sigmavalues = Kspat_d)
                logLikeold <- mvdnorm(data = t(u_star), sigma = K[[k]]$K,
                    sigmavectors = K[[k]]$V, sigmavalues = K[[k]]$d)

                p_theta    <- log(dgamma(theta, shape = K[[k]]$shape, scale = K
                    [[k]]$scale))
                q_theta    <- log(dtnorm(theta, mean = K[[k]]$theta, sd = K[[k
                    ]]$thetasd, lower = 1e-6, upper = Inf))
                q_thetaold <- log(dtnorm(K[[k]]$theta, mean = theta, sd = K[[k
                    ]]$thetasd, lower = 1e-6, upper = Inf))
                alpha <- (logLike + p_theta + q_thetaold)-(logLikeold + K[[k]]
                    $p_theta + q_theta)

                if(log(runif(1)) < alpha) {
                        K[[k]]$K <- Kspat
                        K[[k]]$theta <- theta
                        K[[k]]$p_theta <- p_theta
                        tmp0 <- eigen(K[[k]]$K)
                        K[[k]]$V <- Kspat_V
                        K[[k]]$d <- Kspat_d
                        K[[k]]$nD <- length(K[[k]]$d)
                }

                theta <- rtnorm(1, mean = K[[k]]$theta, sd = K[[k]]$thetasd,
                    lower = 1e-6, upper = Inf) #truncated normal distribution

                }

### NEW end



            if (K[[k]]$COV$type == "UN") {
              tmp <- tmp$U0/sqrt(K[[k]]$d)
              SS <- crossprod(tmp) + K[[k]]$COV$S0
              df <- K[[k]]$nD + K[[k]]$COV$df0
              K[[k]]$G <- MCMCpack::riwish(S = SS, v = df)
            }

            if (K[[k]]$COV$type == "REC") {
              tmp <- tmp$U0/sqrt(K[[k]]$d)

              tmp <- sampleG.REC(U = tmp, M = K[[k]]$COV$M, PSI = K[[k]]$PSI,
```

```
            traits = traits, df0 = K[[k]]$COV$df0, S0 = K[[k]]$COV$S0,
            priorVar = K[[k]]$COV$var)
          K[[k]]$G <- tmp$G
          K[[k]]$B <- tmp$B
          K[[k]]$PSI <- tmp$PSI
        }

        if (K[[k]]$COV$type == "FA") {
          tmp <- tmp$U0/sqrt(K[[k]]$d)
          tmp <- sampleG.FA(U = tmp, F = K[[k]]$F, M = K[[k]]$COV$M, B = K[[k
              ]]$B,
            PSI = K[[k]]$PSI, G = K[[k]]$G, traits = traits, nF = K[[k]]
                $COV$nF,
            df0 = K[[k]]$COV$df0, S0 = K[[k]]$COV$S0, priorVar = K[[k]]
                $COV$var,
            n = K[[k]]$nD)
          K[[k]]$G <- tmp$G
          K[[k]]$PSI <- tmp$PSI
          K[[k]]$B <- tmp$B
          K[[k]]$F <- tmp$F
        }
        if (K[[k]]$COV$type == "DIAG") {
          tmp <- tmp$U0/sqrt(K[[k]]$d)
          K[[k]]$G <- sampleG.DIAG(U = tmp, traits = traits, df0 = K[[k]]
              $COV$df0,
            S0 = K[[k]]$COV$S0, n = n)
        }
      }

    } #ends has(K)

## Overal Mean
E <- t(t(E) + mu)
mu <- sampleMu(Y = E, L = resCov$L, n = n, traits = traits)
E <- t(t(E) - mu)

## Residual Variance
if (resCov$type == "UN") {
    SS <- crossprod(E) + resCov$S0
    df <- n + resCov$df0
    resCov$R <- MCMCpack::riwish(v = df, S = SS)
    resCov$L <- chol(resCov$R)
    resCov$RInv <- chol2inv(resCov$L)
}
if (resCov$type == "REC") {
    tmp <- sampleG.REC(U = E, M = resCov$M, PSI = resCov$PSI, traits = traits,
        df0 = resCov$df0, S0 = resCov$S0, priorVar = resCov$var)
    resCov$R <- tmp$G
    resCov$L <- chol(resCov$R)
```

```
    resCov$RInv <- chol2inv(resCov$L)
    resCov$B <- tmp$B
    resCov$PSI <- tmp$PSI
}
if (resCov$type == "FA") {

    tmp <- sampleG.FA(U = E, F = resCov$F, M = resCov$M, B = resCov$B, PSI =
        resCov$PSI,
        G = resCov$R, traits = traits, nF = resCov$nF, df0 = resCov$df0,
        S0 = resCov$S0, priorVar = resCov$var, n = n)
    resCov$R <- tmp$G
    resCov$L <- chol(resCov$R)
    resCov$RInv <- chol2inv(resCov$L)
    resCov$PSI <- tmp$PSI
    resCov$B <- tmp$B
    resCov$F <- tmp$F
}
if (resCov$type == "DIAG") {
    resCov$R <- sampleG.DIAG(U = E, traits = traits, df0 = resCov$df0, S0 =
        resCov$S0,
        n = n)
}
resCov$RInv <- chol2inv(chol(resCov$R))

## Imputing missing values
YHat <- YStar - E
if ((nNa > 0)) {
    for (j in 1:nNa) {
        subject <- whichNa$subject[[j]]
        missing <- whichNa$traits[[j]]
        observed <- (1:traits)[-missing]
        tmp <- sampleY(R = resCov$R, y = Y[subject, ], yHat = YHat[subject,
          ], e = E[subject, ], missing = missing, observed = observed, traits =
              traits)
        YStar[subject, ] <- tmp$y
        E[subject, ] <- YStar[subject, ] - YHat[subject, ]
        logLik <- logLik + tmp$logLik
    }
}


## Completing the logLik computation
if (nNa < n) {
    logLik <- logLik + dMVNorm(X = E[!hasNa, ], Sigma = resCov$R, mu = rep(0,
        traits), log = TRUE)
}

## Running means
if ((i > burnIn) & (i%%thin == 0)) {
```

```
            iter <- iter + 1
            k <- (iter - 1)/(iter)


            post_mu <- post_mu * k + mu/iter
            resCov$post_R <- resCov$post_R * k + resCov$R/iter
            post_YHat <- post_YHat * k + YHat/iter
            post_logLik <- post_logLik * k + logLik/iter

            if (resCov$type %in% c("REC", "FA")) {
                resCov$post_B <- resCov$post_B * k + resCov$B/iter
                resCov$post_PSI <- resCov$post_PSI * k + resCov$PSI/iter
            }

            if (hasXF) {
                post_B.f <- post_B.f * k + B.f/iter
            }

            if (hasK) {
                for (j in 1:nK) {
                  post_K[[j]]$U <- post_K[[j]]$U * k + K[[j]]$U/iter
                  post_K[[j]]$G <- post_K[[j]]$G * k + K[[j]]$G/iter

                  if (K[[j]]$COV$type %in% c("REC", "FA")) {
                    post_K[[j]]$B <- post_K[[j]]$B * k + K[[j]]$B/iter
                    post_K[[j]]$PSI <- post_K[[j]]$PSI * k + K[[j]]$PSI/iter
                  }
### NEW begin
                  if (K[[j]]$COV$type == "spatial") {
                     post_K[[j]]$theta <- post_K[[j]]$theta * k + K[[j]]$theta/iter
                  }
### NEW end
                }
            }

        }

        ## Saving Sammples
        tmp <- i%%thin == 0
        if ((tmp)) {

            tmp <- logLik
            fileName <- paste(saveAt, "logLik.dat", sep = "")
            write(tmp, ncol = length(tmp), file = fileName, append = T, sep = " ")


            tmp <- MCMCpack::vech(resCov$R)
            fileName <- paste(saveAt, "R.dat", sep = "")
            write(tmp, ncol = length(tmp), file = fileName, append = T, sep = " ")
```

```
            if (resCov$type %in% c("REC", "FA")) {
                if (sum(resCov$M) > 0) {
                  tmp <- resCov$B[resCov$M]
                  fileName <- paste(saveAt, "B_R.dat", sep = "")
                  write(tmp, ncol = length(tmp), file = fileName, append = T, sep = "
                      ")
                }
                tmp <- resCov$PSI
                fileName <- paste(saveAt, "PSI_R.dat", sep = "")
                write(tmp, ncol = length(tmp), file = fileName, append = T, sep = " ")

            }

            tmp <- c(mu)
            fileName <- paste(saveAt, "mu", ".dat", sep = "")
            write(tmp, ncol = length(tmp), file = fileName, append = T, sep = " ")

            if (hasXF) {
                tmp <- as.numeric(B.f)
                fileName <- paste(saveAt, "mu", ".dat", sep = "")
                write(tmp, ncol = length(tmp), file = fileName, append = T, sep = " ")
            }

            if (hasK) {
                for (k in 1:nK) {

                  tmp <- MCMCpack::vech(K[[k]]$G)
                  fileName <- paste(saveAt, "G_", k, ".dat", sep = "")
                  write(tmp, ncol = length(tmp), file = fileName, append = T, sep = "
                      ")
### NEW begin
                  if ((K[[k]]$COV$type == "spatial")) {
                    tmp <- K[[k]]$theta
                    fileName <- paste(saveAt, "theta_", k, ".dat", sep = "")
                    write(tmp, ncol = length(tmp), file = fileName, append = T, sep = "
                        ")
                  }
### NEW end
                  if ((K[[k]]$COV$type %in% c("REC", "FA"))) {
                    tmp <- K[[k]]$PSI
                    fileName <- paste(saveAt, "PSI_G_", k, ".dat", sep = "")
                    write(tmp, ncol = length(tmp), file = fileName, append = T, sep = "
                        ")
                    if (sum(resCov$M) > 0) {
                      tmp <- t(K[[k]]$B)[t(K[[k]]$COV$M)]
                      fileName <- paste(saveAt, "B_G_", k, ".dat", sep = "")
                      write(tmp, ncol = length(tmp), file = fileName, append = T,
                        sep = " ")
```

```
                    }
                  }
                }
            }
        }

        tmp <- proc.time()[3]
        cat(paste("Iter: ", i, "time: ", (round(tmp - time, 4))))
        cat("\n")
        cat("\n")
        time <- tmp
    } #end i in nIter

    tmp <- list()
    tmp$R <- resCov$post_R
    if (resCov$type %in% c("REC", "FA")) {
        tmp$B <- resCov$post_B
        tmp$PSI <- resCov$post_PSI
    }
    out <- list(mu = post_mu, YHat = post_YHat, resCov = tmp)

    if (hasXF) {
        out$B.f = post_B.f
    }

    if (hasK) {
        out$K <- post_K
    }

    out$DIC <- getDIC(Y = Y, YHat = post_YHat, R = resCov$post_R, meanLogLik =
        post_logLik)

    return(out)
}


### FIXED EFFECTS
### ##############################################################################
sampleBf <- function(Y, XTb, Tb, R, L, RInv, traits, dimX) {
    SOL <- crossprod(XTb, Y)
    Z <- matrix(nrow = dimX, ncol = traits, rnorm(dimX * traits))
    E <- tcrossprod(Z, L)
    B <- SOL + E
    B <- Tb %*% B
    return(B)
}

sampleMu <- function(Y, L, n, traits) {
    sol <- colMeans(Y)
```

```
    L <- L/sqrt(n)
    mu <- as.numeric(crossprod(L, rnorm(traits)) + sol)
    return(mu)
}

### MISSING RECORDS
### ##########################################################################
sampleY <- function(R, y, yHat, e, missing, observed, traits) {
    if (length(missing) == traits) {
        e <- crossprod(chol(R), rnorm(traits))
        y <- yHat + e
        logLik <- 0
    } else {
        Roo <- matrix(R[observed, observed], nrow = length(observed), ncol = length(
            observed))
        RooInv <- chol2inv(chol(Roo))
        Rmm <- matrix(R[missing, missing], nrow = length(missing), ncol = length(
            missing))
        Rom <- matrix(R[observed, missing], nrow = length(observed), ncol = length(
            missing))

        Bmo <- crossprod(Rom, RooInv)

        yHat2 <- as.numeric(Bmo %*% e[observed])   ##### %*% austauschen!
        CondVar <- Rmm - Bmo %*% Rom    ##### %*% austauschen!
        L <- chol(CondVar)
        e <- crossprod(L, rnorm(length(missing)))
        y[missing] <- yHat[missing] + yHat2 + e
        tmp <- (y - yHat)[observed]
        logLik <- dMVNorm_i(x_i = tmp, SigmaInv = RooInv, mu = rep(0, length(observed))
            ,
            log = TRUE)
    }
    out <- list(y = y, logLik = logLik)
    return(out)
}


### Sample U
### ##########################################################################
sampleUj <- function(j, Y, RInv, GInv, V, d, traits) {
    CInv <- chol2inv(chol(RInv + GInv/d[j]))
    T <- crossprod(t(RInv), CInv)    ##### %*% -> crossprod
    YStar <- crossprod(t(Y), T)      ##### %*% -> crossprod
    sol <- as.numeric(crossprod(V[, j], YStar))
    L <- chol(CInv)
    uj <- as.numeric(crossprod(L, rnorm(traits))) + sol
    return(uj)
}
```

```
sampleU <- function(Y, RInv, GInv, V, d, traits, df0 = 0, S0 = 0) {
    tmp <- matrix(unlist(lapply(FUN = sampleUj, Y = Y, RInv = RInv, GInv = GInv,
        V = V, d = d, traits = traits, X = 1:length(d))), byrow = TRUE, ncol = traits)

    U <- crossprod(t(V), tmp)          ##### %*% -> crossprod
    return(list(U = U, U0 = tmp))
}



### Sample G
### ##############################################################################


sampleG.REC <- function(U, M, PSI, traits, priorVar = 100, df0 = rep(0, traits),
    S0 = rep(0, traits)) {
    ### Model: U=UB+D (ONLY RECURSIVE ALLOWED!)  Current sample of random effects
    ### ('data') T a pxp matrix with TRUE/FALSE indicating possition of non-null
    ### recursive effects (FALSE in diagonal!)  PSI px1 the variance of the orthogonal
    ### shocks ...
    B <- matrix(nrow = traits, ncol = traits, 0)
    for (i in 1:traits) {

        dimX <- sum(M[i, ])

        if (dimX > 0) {
            tmpX <- U[, M[i, ]]
            tmpY <- U[, i]

            C <- crossprod(tmpX)/PSI[i] + 1/priorVar
            CInv <- chol2inv(chol(C))
            rhs <- crossprod(tmpX, tmpY)/PSI[i]

            sol <- crossprod(CInv, rhs)
            L <- chol(CInv)
            shock <- crossprod(L, rnorm(dimX))
            tmpB <- as.numeric(sol + shock)
            B[i, M[i, ]] <- tmpB
            uStar <- tmpY - matrix(tmpX, ncol = dimX) %*% (tmpB)
            SS <- as.numeric(crossprod(uStar)) + S0[i]
            df <- nrow(U) + df0[i]
            PSI[i] <- SS/rchisq(n = 1, df = df)
        } else {
            SS <- as.numeric(crossprod(U[, i])) + S0[i]
            df <- nrow(U) + df0
            PSI[i] <- SS/rchisq(n = 1, df = df)
        }
    }
```

```
    tmp <- solve(diag(traits) - B)
    G <- tmp %*% diag(PSI) %*% t(tmp)

    out <- list(B = B, PSI = PSI, G = G)

    return(out)
}


##############################################################################################

sampleG.FA <- function(U, F, M, B, PSI, G, traits, nF, n, df0 = rep(1, traits), S0 =
    rep(1/100,
    traits), priorVar = 100) {
    ### Gibbs sampler for FA model Model: U=BF+D

    ## sampling common factors LOOP OVER FACTORS
    for (i in 1:nF) {
        tmpY <- U - F[, -i] %*% matrix((B[, -i]), ncol = traits)
        rhs <- tmpY %*% matrix(B[, i]/PSI, ncol = 1)
        CInv <- 1/(sum((B[, i]^2)/PSI) + 1)
        sol <- CInv * rhs
        SD <- sqrt(CInv)
        F[, i] <- rnorm(n = n, sd = SD, mean = sol)
    }

    # sampling loadings LOOP OVER TRAITS LOOP OVER FACTORS
    for (i in 1:traits) {
        for (j in 1:nF) {
            if (M[i, j]) {
                tmpY <- U[, i] - F[, -j] %*% matrix(B[i, -j], ncol = 1)
                CInv <- 1/as.numeric(crossprod(F[, j])/PSI[i] + 1/priorVar)
                rhs <- as.numeric(crossprod(F[, j], tmpY)/PSI[i])
                sol <- CInv * rhs
                SD <- sqrt(CInv)
                B[i, j] <- rnorm(n = 1, mean = sol, sd = SD)
            }
        }
        D <- U[, i] - F %*% B[i, ]
        df <- df0[i] + n
        SS <- S0[i] + crossprod(D)
        PSI[i] <- SS/rchisq(df = df, n = 1)
    }
    if (nF > 1) {
        B <- varimax(B)$loadings[]
    }
    G <- tcrossprod(B) + diag(PSI)
```

```
    out <- list(F = F, PSI = PSI, B = B, G = G)
}

################################################################################
sampleG.DIAG <- function(U, traits = ncol(U), n = nrow(U), df0 = rep(0, traits),
    S0 = diag(0, traits)) {
    ### Gibbs sampler for DIAG covaraince matrices
    G <- matrix(nrow = traits, ncol = traits, 0)
    ## sampling common factors LOOP OVER FACTORS
    for (i in 1:traits) {
        tmp_SS <- sum(U[, i]^2) + S0[i]
        tmp_df <- n + df0[i]
        G[i, i] <- tmp_SS/rchisq(df = tmp_df, n = 1)
    }

    return(G)
}


if (FALSE) {
    B <- cbind(rnorm(1000, sd = 1), rnorm(1000, sd = 2))
    sampleG.DIAG(B)

}



################################################################################

### NEW begin

 dist2kernel <- function(Dist, theta){
     fun=function(a){return(exp(-theta*a^2))}
     return(fun(Dist))
 }

my.inv <- function(X, eigenvectors, eigenvalues, tol = sqrt(.Machine$double.eps)){
  ## Generalized Inverse of a semi-positive definite matrix, inspired by ginv

     Positive <- eigenvalues > max(tol * eigenvalues[1L], 0)
    if (all(Positive))
         crossprod(t(eigenvectors), (1/eigenvalues * t(eigenvectors)))
    else if (!any(Positive))
         array(0, dim(X)[2L:1L])
    else crossprod(t(eigenvectors[, Positive, drop = FALSE]),
         ((1/eigenvalues[Positive]) * t(eigenvectors[, Positive, drop = FALSE])))
 }

 mvdnorm <- function(data, sigma, sigmavectors, sigmavalues, log=TRUE){
```

```
    inv <- my.inv(sigma, sigmavectors, sigmavalues)
    det <- sum(log(sigmavalues))
    dim <- length(sigmavalues)

    p <- 0
    for (i in 1:nrow(data)){
      p <- p - (0.5*det) - (dim/2 * log(2*pi)) -  (0.5*crossprod(data[i,], crossprod(
          t(inv), data[i,])))
    }

  if (!log) {
      p <- exp(p)
  }
  return(p)

}

### NEW end

dScaleInvChisq <- function(x, df, S, log = FALSE) {
  y <- S/x
  out <- dchisq(y, df = df, log = TRUE) + log(S/(x^2))
  if (!log) {
      out <- exp(out)
  }
  return(out)
}

dMVNorm <- function(X, Sigma, mu, log = TRUE) {
  ## X may be a matrix, rows are IID draws

  TInv <- solve(chol(Sigma))
  Z <- crossprod((t(X) - mu), TInv)
  out <- sum(dnorm(x = as.vector(Z), log = TRUE)) + nrow(X) * sum(log(diag(TInv)))
  if (!log) {
      out <- exp(out)
  }
  return(out)

}


dMVNorm_i <- function(x_i, SigmaInv, mu, log = TRUE) {

  ## works for a single random draw and requires SigmaInv

  e <- as.matrix(x_i - mu)
  out <- -(length(e)/2 * log(2 * pi)) + log(det(SigmaInv))/2 - (crossprod(e, SigmaInv
```

```
        ) %*% e)/2
    if (!log) {
        out <- exp(out)
    }
    return(out)
}



getDIC <- function(Y, YHat, R, meanLogLik = NULL, logLik = NULL) {
    #### Returns Deviance Information Criterion and Effective Number of Parameters Y
    #### (nxp) the data matrix (NA's for misssing) YHat the posterior mean of the
    #### conditional expectation logLik the samples of the log-Likelihood generated by
    #### MIM() meanLogLik the posterior mean of logLik
    if (!is.null(logLik)) {
        meanLogLik <- mean(logLik)
    }
    E <- Y - YHat
    logLikAtPostMean <- 0
    for (i in 1:nrow(Y)) {
        observed <- !is.na(Y[i, ])
        e <- matrix(nrow = 1, E[i, observed])
        mu <- rep(0, sum(observed))
        Rtmp <- R[observed, observed]
        if (length(Rtmp) > 0) {
            logLikAtPostMean <- logLikAtPostMean + dMVNorm(X = e, Sigma = Rtmp, mu = mu
                ,
                log = TRUE)
        }
    }

    pD <- -2 * (meanLogLik - logLikAtPostMean)
    DIC <- pD - 2 * meanLogLik
    return(list(meanLogLik = meanLogLik, logLikAtPostMean = logLikAtPostMean, DIC = DIC
        ,
        pD = pD))

}
```

13 Publication out of this thesis

The following article has been published in advance out of this thesis:

Töpner K, Rosa G J M, Gianola D, and Schön CC (2017), Bayesian networks illustrate genomic and residual trait connections in maize (*Zea mays* L.). G3: Genes, Genomes, Genetics 7(8):2779–2789

The publication, including supporting material, can be accessed via the following link:
https://doi.org/10.1534/g3.117.044263

The following Sections include parts of this paper:
1, 2, 3.2, 3.3, 3.4, 3.5, 5, 6.1, 7.1, 8, 9, 10, 11, and 12

Candidate's contribution: developing application of methodology, in particular facilitation of BN analysis on the genomic values by applying a suitable transformation, use of bootstrapping to measure uncertainty of connections, separation of genomic from residual structures, formulation of SEM including two structured components instead of one, and use of SEM as a network and algorithm evaluation tool; statistical data analysis, interpretation and discussion of results, composition of graphs and tables, writing and revisions of the manuscript.

14 Acknowledgement

First of all, I thank Prof. Dr. Chris-Carolin Schön for giving me the opportunity to pursue this PhD and for supporting two stays at the University of Wisconsin-Madison as well as to attend numerous conferences. Thank you for providing feedback, advice, and leadership.

I also thank Prof. Daniel Gianola (Ph. D.) for co-supervising my PhD, for his inspiring ideas, and for serving as examiner on my graduate committee. Thank you very much for welcoming me at your department at the University of Wisconsin-Madison.

Many thanks go to Prof. Guilherme J. M. Rosa (Ph. D.) for introducing me to Bayesian networks, for laying the basis for both of my PhD projects, and for all constant scientific and personal advice, support, and encouragement. I also thank Bruno Valente (Ph. D.) for discussing and developing ideas based on his paper published in 2010 and for constant interest in my work.

Thanks a lot also to Prof. Dr. Aurélien Tellier for chairing my graduate committee.

Special thanks go to two anonymous reviewers at Genetics, who offered exceptionally detailed and dedicated reviews.

I thank all members of plant breeding at the Life Science School of the Technical University of Munich for their continuous scientific and personal support, especially Dr. Eva Bauer, Dr. Nicole Krämer, Dr. Christina Lehermeier, Dr. Sebastian Steinemann, Dr. Sandra Unterseer, Hans-Jürgen Auinger, Wiltrud Erath, and Manfred Schönleben.

I am also very grateful to all members of the Animal Sciences department at the University of Wisconsin-Madison for their warm welcome and interesting interdisciplinary and international discussions.

The first two years of my position at the Technical University of Munich were funded by its Institute of Advanced Study, which also offered an inspiring atmosphere and interesting events. I am grateful for having received the Laura Bassi Award by the Technical University of Munich, which included a financial contribution to my last year.

Last but not least, I am deeply grateful for being surrounded by my friends and family, whose love and care constantly contribute to all my projects and achievements, be it by giving any kind of support or by being the sense behind my work.