Expert Systems With Applications 93 (2018) 156–168

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Combination of expert decision and learned based Bayesian Networks for multi-scale mechanical analysis of timber elements

Hélder S. Sousa[a,*], Francisco Prieto-Castrillo[b,c,d], José C. Matos[a], Jorge M. Branco[a], Paulo B. Lourenço[a]

[a] ISISE, Department of Civil Engineering, University of Minho, Guimarães 4800-058, Portugal
[b] BISITE Research Group, University of Salamanca, Edificio I+D+i, Salamanca 37008, Spain
[c] Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139–4307, USA
[d] Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 1 May 2017
Revised 23 September 2017
Accepted 28 September 2017
Available online 3 October 2017

Keywords:
Bayesian Networks
Timber
Multi-scale analysis
Expert systems
Learning algorithms
Ranking

ABSTRACT

The use of Bayesian Networks allows to organize and correlate information gathered from different sources and its optimization may incorporate restrictions adjusting the network based on expert knowledge and network operativeness, in such a way that it may satisfactorily represent a given domain. The main goal of this paper is to study if an optimized learned Bayesian Network may be used as a prior structure for an expert based network of an engineering structural material analysis. The methodology is applied to a database of results from an experimental campaign that focused on the mechanical characterization of timber elements recovered from an early 20th century building. To that study case it is evidenced that through a suitable combination of model averaging and supervision steps it is possible to achieve robust and reliable models to underpin the causal structure of a typical multi-scale timber analysis.

1. Introduction

In the engineering field, the analysis of a system is often a complex problem when the objective is to predict a result variable based on information from heterogeneous sources, which by their own nature are necessarily uncertain to some extent. The analysis of these systems requires that dependencies, or correlations, are defined between different source variables and the result variable. In the field of timber engineering, one of the main motivations for probabilistic hierarchical modelling is to understand how properties, composition and structure at lower scale levels may influence and be used to predict the material properties on a macroscopic and structural engineering scales. For instance, different models were proposed to hierarchically define stiffness and strength of timber elements, by considering the presence of weak sections separated by segments of clear wood (Isaksson, 1999; Machado & Palma, 2011; Riberholt & Madsen, 1979). In this case, it is known that significant improvement of the model can be achieved by including expert knowledge in its building process (Constantinou, Fenton, & Neil, 2016; Fenton & Neil, 2013; Pearl, 2009; Rebonato, 2010; Spiegelhalter, Abrams, & Myles, 2004; Zhou, Fenton, & Neil, 2014). Furthermore, the collaboration between humans and machines (i.e., social machines) is gaining more and more attention in the AI (artificial intelligence) and expert systems community (Smart, Simperl, & Shadbolt, 2014). A key factor for this assessment is the correct integration of the supervision steps in the elicitation workflow. What type of information the expert is exposed to and when it is made available can substantially improve the final model performance. On the other hand, expert knowledge must be consistently integrated with data. This means that expert constraints should be coherent with respect to the conditional independencies found in data. Finally, the knowledge base has to be continuously updated with new information to ensure learning. This poses the question of how to design human-machine interactive systems able to cope with these requirements. To that aim, Bayesian methods are prone to introduce new information and update prior beliefs within a predictive model (Armero & Conesa, 2006; Pendharkar, Subramanian, & Rodger, 2005).

Bayesian networks (BNs) have been used for inference on masonry and concrete elements using both quantitative and qualita-

* Corresponding author.
E-mail addresses: sousa.hms@gmail.com (H.S. Sousa), franciscop@usal.es (F. Prieto-Castrillo), jmatos@civil.uminho.pt (J.C. Matos), jbranco@civil.uminho.pt (J.M. Branco), pbl@civil.uminho.pt (P.B. Lourenço).

https://doi.org/10.1016/j.eswa.2017.09.060
0957-4174/

tive data obtained from destructive and non-destructive tests, as well as from engineering judgment, to obtain the material strength (Marsili, Croce, Klawonn, & Landi, 2017a; Marsili et al., 2017b). In timber engineering, BNs were used in Deublein, Schlosser, and Faber (2011) to describe the influence of different origins or dimensions of sawn structural timber on the relevant timber material properties conditioned from indicator values assessed by machine stress-grading, whereas inference on stiffness and strength of timber elements using BNs was discussed in Sousa, Branco, Lourenço, and Neves (2016b). In the case of Sousa et al. (2016b), the arrangement of nodes within the BN was solely derived from expert decision based on linear correlation between variables. Bayes updating methodology applied on timber engineering was further discussed in Feio and Machado (2015); Fink and Kohler (2015); Sousa, Machado, Branco, and Lourenço (2015). In this paper, information gathered in a multi-scale experimental campaign regarding timber element's mechanical characterization is used to obtain a probabilistic inference model where BNs are applied to combine in an effective way heterogeneous information sources with expert knowledge. As stressed, this knowledge must be consistently integrated with data, meaning that expert constraints should be coherent with respect to the conditional independencies found in data. The research goals of this work are to present a formal procedure to assert expert knowledge to probabilistic models, such as Bayesian Networks, in the context of timber engineering. Timber is a anisotropic material that poses several design challenges in civil engineering due to its variability, specially when dealing with existing structures. In that case, a combination of expert knowledge and non-destructive tests is often used to assess the mechanical properties of timber. However, there is no global method to measure the influence of expert knowledge and how it will influence the prediction power of the relevant mechanical properties of timber. The objective of this work, is therefore to create a semi-automated process that allows to build a Bayesian Network that must ensure the implementation of expert information but maintain a strong prediction power regarding the mechanical characterisation of existing timber elements. For this aim, a framework for the interactive elicitation for the inter-relationships of different metrics in the timber structural analysis domain is presented. The present work considers the same data used in Sousa et al. (2016b), but incorporating a semi-automated process to build a Bayesian network with expert information. In this case, the expert knowledge is incorporated along the elicitation workflow allowing to obtain a stronger prediction of each variable. The optimization process is verified by the prediction power and the operativeness of the network. By using this framework, expert knowledge can be implemented in databases where the relation between variables are not prior known and also for networks with a large number of nodes, as the initial framework for the network is given by an automated process. Finally, the procedure allows to obtain a set of semi-automated models that may be compared to operative models (based only on expert decision), thus allowing to suggest possible network arrangements that allow to update the prior beliefs of the expert aiming at a stronger prediction of the variables in study.

2. Materials and methods

2.1. Bayesian methods

In Bayesian methods, probabilities are taken as the best possible expression of the degree of belief in the occurrence of an event or conjugation of events. This approach does not stipulate that probabilities are direct and unbiased predictors, but if the analysis is carried out methodically, the probabilities will be correct if averaged over a sufficient large number of decision situations (Vrouwenvelder, 2002). To fulfil that premise it is required that the experts subjective and intuitive part is neither continuously over confident nor over conservative. To that aim, calibration to common practice and to empirical data is considered as an adequate procedure. The JCSS Probabilistic Model Code (JCSS, 2000) concludes that, compared to the classic frequentist interpretation, the Bayesian approach is more adequate, as it allows to update distributions when having more statistical data. Reasoning between the correlation of source variables and experts interpretation has long been a matter of discussion. Singh and Valtorta (1995) proposed an algorithm that uses conditional independence test based methods to generate an ordering on the nodes in a BN. However, the same authors note that the reasonability of use of the proposed model relies on the premise that it must have a much higher probability that the next most probable BN. More recently, Constantinou et al. (2016) focused on the problem whereby the distribution of some continuous variable in a BN is known from data, but where it is intended to explicitly model the impact of some additional expert variable, for which there is expert judgment but no data.

BNs are probabilistic graphical models that combine the principles from probability and graph theory to examine the relationship among variables in a domain. In essence, a BN is represented by a pair (G, P) where G is a directed acyclic graph (DAG) over the set V of random variables and P is a joint probability distribution of V. G links encode the variable dependencies and their cause-effect relationship. This constitutes the qualitative part of the BN. BNs have also a quantitative part where local conditional probabilities are mapped into the network nodes allowing to factorize the joint probability distribution P in a very effective way (Spirtes, Glymour, & Scheines, 1993). More technically a BN is a minimal map of independencies (or minimal I-map) (Pearl, 1997). This is equivalent to say that BNs follow the Markov Condition (i.e., each variable $X \in V$, is conditionally independent of the set of all its non-descendants given the set of all its parents). These definitions allow to implement a BN in terms of independencies rather than in terms of dependencies. However, it would be preferable to use the BN to encode dependencies in the sense that for instance, the parents of a node are taken to be direct causes of that variable. Unfortunately, having two connected variables in G does not necessarily imply that these variables are dependent (Pearl, 1997). This is only true under two assumptions: causal sufficiency and faithfulness (Spirtes et al., 1993). Causal sufficiency means that there are not unobserved, latent, variables that are parent of one or more observed variables of the domain. The faithfulness assumption means that the BN is a dependency (or D-map) of P; every triple of conditionally dependent variables (X, Y, Z) is d-connected in G. Whereas faithfulness is an assumption usually made, causal sufficiency is not likely to hold in real applications. Accounting for the aforementioned conditions, a BN can be constructed manually, (semi-)automatically from data, or through a combination of a manual and a data driven process. The data driven approach is accomplished by means of structure learning algorithms which reveal the potential causal pathways among variables. This strategy has its pros and cons (see Margaritis (2003) for a detailed analysis of structure learning from data). On one hand, new interesting interconnections among variables can be discovered. On the other hand, the causal sufficiency assumption can fail and not all the resulting directed edges in the form $X \rightarrow Y$ (interpreted as "X causes Y") are reliable. In particular, reversing the arrow to $Y \rightarrow X$ might encode a causal relationship which is more reliable and this operation can still be consistent with the conditional independencies present in the data. This indeterminacy needs to be tackled by external (supervision) support. Hence, different supervision steps are needed to: (1) impose causal information and (2) rule out equivalent structures by selecting the most appropriate

according to the domain semantics. As stressed, this expert interaction is known to report significant improvements in model robustness and reliability. In this case, a network is considered robust if small errors in the independence tests do not lead to large variations in the output results. Within this context, since conditional independence tests can fail to capture the real independencies (specially for small databases),models having incorrect edge directions can be disregarded by means of one ore more supervision steps. This way, an external supervision can help in correcting these inconsistencies, thus improving the robustness of the network.

In this work, although coding variable associations through undirected links is often considered the most appropriate procedure to implement relationships, only causal bonds will be considered, as specific cause-effect relations are aimed at. The idea is to create a network that can be applied in real life scenarios where only a direction sense is considered useful or physically possible. For example, the mechanical properties of timber should be obtained through the results of the tests and not the other way around. Therefore, in our current domain, the expert is forced to provide directed bonds at his interventions. There are several types of structure learning algorithms. These can be classified into three groups: constraint-based (CS), score-based (SB) and hybrid. The most popular CS algorithms are perhaps PC (Spirtes et al., 1993), Grow-Shrink (GS) (Margaritis, 2003), Incremental Association Markov Blanket (IAMB) (Tsamardinos, Aliferis, & Statnikov, 2003) and Semi-Interleaved SI-HITON-PC (Aliferis, Statnikov, Tsamardinos, Subramani, & Koutsoukos, 2010) algorithms. PC orders the independencies present in data from small to large to achieve larger efficiency while GS, IAMB and SI-HITON-PC make structure discovery faster by using the Markov Blanket concept (Spirtes et al., 1993). The MB of a node is defined as the set of nodes that make that node conditionally independent of any other node in the network. Hence MB can be understood as shields for independency (given the MB any other information is irrelevant for that node).

Variations of these structure learning algorithms can be found in literature (see Tsamardinos et al. (2003) for a comparison). Constraint based (CS) models use conditional independence X^2, information based (e.g., mutual information MI) or other statistical tests (Margaritis, 2003; Pearl, 1997; Spirtes et al., 1993)). The resulting networks are sparse and fairly interpretable. However, CS algorithms have clear disadvantages. In particular, they are not robust, since small errors in the independence tests can lead to large variations in the output networks (Margaritis, 2003). This is a major issue when dealing with empirical distributions because most surely theoretical independencies will not be preserved in the sample. On the other hand, score based methods (SB) consist of two basic components: a scoring metric and a search procedure. The scoring metric reflects the goodness-of-fit of the data and can be the log-likelihood (LL), Minimum Description Length (MDL) (Bouckaert, 1993), Bayesian Information Criterion (BIC) (Kass & Wasseman, 1995), Akaike Information Criterion (AIC) (Akaike, 1973) or more sophisticated metrics. However, it is widely known that the LL score is guaranteed to overfit, fostering more complex networks since the addition of new edges tends to improve the metric (Koller & Friedman, 2011). In this regard, MDL, BIC and AIC seek a trade-off between fit to data and model complexity. The Bayesian Dirichlet (BD) and its variant BDE scoring functions (Cooper & Herskovits, 1992; de Campos & Ji, 2010; Heckerman, Geiger, & Chickering, 1995) penalizes complex graphs less than BIC and is also a common criteria for evaluating BN structures. The searching procedure can be any heuristic, like the popular Hill-Climbing (HC) algorithm or the Tabu search (TABU) (see for instance Russell & Norvig (2003)) that generates networks for evaluation by the scoring metric. As it is known, learning a BN structure by brute-force search is NP-Complete (Maxwell, 1996). This fact has made the heuristic methods popular in the BN community. The drawback of this is the sub-optimality of the solutions when the algorithm gets stuck in local maxima. Finally, hybrid algorithms combine CS and BS to seek an optimum balance between both approaches. An example of hybrid method is the Max-Min Hill-Climbing MMHC (Tsamardinos et al., 2003).

2.2. Database

In this work, the results of a multi-scale experimental campaign regarding twenty chestnut timber elements was considered as initial database. The testing phase sequence, with three different scales regarding the size of the timber elements is described in Sousa, Branco, and Lourenço (2016a). From one to the next scale, the timber elements were sawn into smaller sizes in order to clearly define the influence of defects and differentiate between clear wood segments and segments with defects (each beam was sawn into three boards). In each phase, bending modulus of elasticity was measured according to the test principles of EN 480 (CEN, 2010). Furthermore, non-destructive tests were also made, namely visual inspection as to categorize the different levels of defects (size and distribution of knots and grain misalignment) and ultrasound tests as its results (i.e., dynamic modulus of elasticity) are a common parameter correlated with the elements stiffness (e.g., static bending modulus of elasticity). To create the BN, the parameters of the experimental campaign and its results were divided in three categories: (1) geometry; (2) onsite non-destructive tests (NDT); (3) mechanical tests. The first category is related to location of each segment within the element and its measurement scale. In that category, the following nodes are defined: (a) element, (b) section, (c) position and (d) scale. Element considers the identification of the twenty timber beams as separate and independent elements. Section defines the segment position in the horizontal direction (divided in seven segments), while position defines the segment position in the vertical direction (top, middle or bottom boards). Scale defines the phase of the experimental campaign starting with the elements on their initial onsite conditions (old beams), after as sawn beams and finally as sawn boards. The second category, related to onsite NDT, include two nodes corresponding to visual inspection results (according to UNI 11119 (UNI, 2004)) and to ultrasound results (in terms of propagation velocity). The last category, related to mechanical tests, include two nodes corresponding to the global and local bending moduli of elasticity according to EN 408 (CEN, 2010). The division of these properties into different categories also expresses different levels of knowledge in a common mechanical characterization of a structural elements. The first category includes information that is easily available with a geometrical survey, while the second and third category require measurement of the element's defects and testing. Globally, the dataset consists of 755 observations of 8 variables: "*Element*", "*Position*", "*Section*", "*Scale*", "*Visual inspection*", "*Ultrasound*", "*Local bending modulus of elasticity*", "*Global bending modulus of elasticity*". Here, "*Element*", "*Position*", "*Section*", "*Scale*", "*Visual inspection*", are categorical whereas "*Ultrasound*", "*Local bending modulus of elasticity*", and "*Global bending modulus of elasticity*" are numerical. "*Element*" has 20 possible classes labeled with capital letters from "A" to "T", "*Position*" has: "bottom", "middle" and "top" levels. "Scale" splits into "sawn beam" and "sawn board" classes. "*Visual inspection*" has 4 classes: "I", "II", "III" and "non-classifiable – NC". As a pre-processing step the numerical variables were binned into 4 levels according to their quartiles. This ensures that there are approximately the same number of observations per class. The variables included in the database are summarized in (Table 1)

Table 1
Variables used in the database.

Variable	Symbol	Type	Classification	Range of values	Description
Element	E	Geometry	Categorical	A to T	20 timber elements retrieved from a construction site
Position	P	Geometry	Categorical	Bottom, middle, top	Vertical location of measurements, each corresponding to one third of the height of the element
Section	Se	Geometry	Categorical	1 to 7	Horizontal location of measurements, each corresponding to a 40 cm segment
Scale	Sc	Geometry	Categorical	Beams or boards	Size of the elements corresponding to beams with 7x15 cm cross-section or boards with 7x4 cm cross-section
Visual inspection	VI	Non-destructive test	Categorical	I, II, III, NC	Visual grading obtained according to (UNI, 2004)
Ultrasound	Us	Non-destructive test	Numerical	1880–5630 m/s	Indirect ultrasound measurements, considering the propagation velocity in 40 cm segments
Local modulus of elascticity	EL	Mechanical test	Numerical	2590–23385 MPa	Modulus of elasticity measured between loading points according to (CEN, 2010), in a 4-point bending test
Global modulus of elasticity	EG	Mechanical test	Numerical	1120–17740 MPa	Modulus of elasticity measured between support points according to (CEN, 2010), in a 4-point bending test

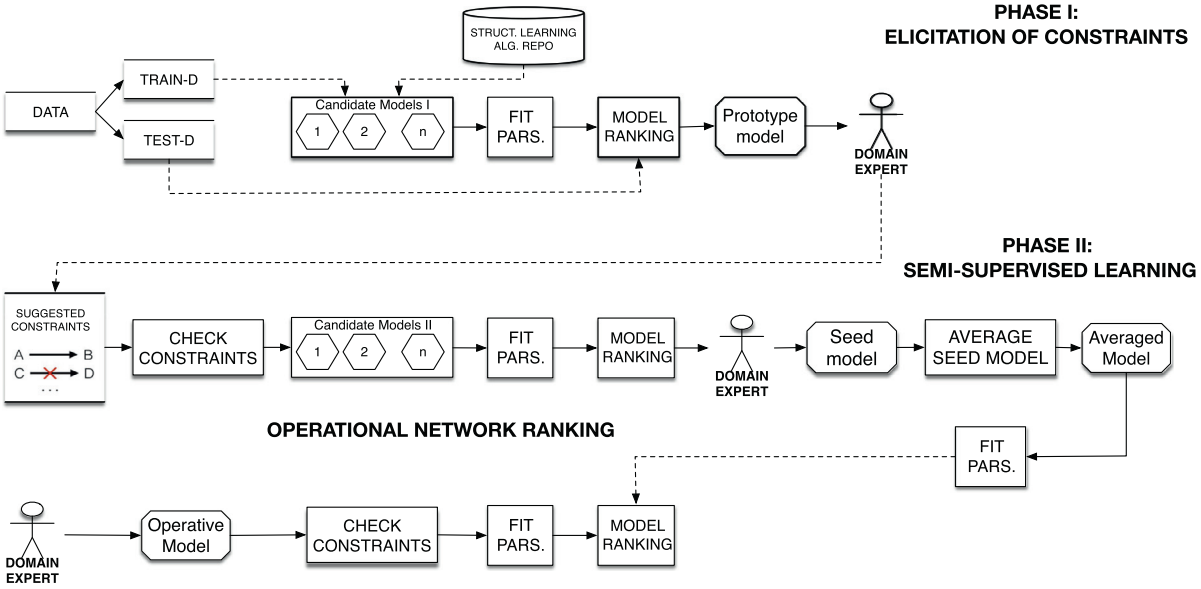

Fig. 1. The implemented workflow consists on two phases: restriction elicitation and model-refinement. In both cases the automatic learning is aided by expert intervention at different stages, whether there is supervision or not, a set of candidate models is generated and evaluated. In phase 1 the winner models are used to discover expert constraints. In phase 2 this information is used to obtain a seed learning strategy which is then averaged over a set of equivalent replicates. Finally the expert can evaluate the performance of his operative model by using the rankings created in phases 1 and 2.

2.3. Model construction

As stressed above, a suitable combination of the expert knowledge with the inferred models is the most promising approach to achieve a reliable and robust solution when different sources of information are available for a multi-scale analysis. In this work, the approach was implemented in two phases (Fig. 1).[1] In phase I, the expert is confronted with the learned models without any constraints. This is made to avoid any bias in the expert's ranking of the model. In phase II, the network is refined by implementing, when possible, the suggested constraints taking into account the predictive power before and after the supervision steps. In both phases, the learning process of the network structure is made through different algorithms that include or not supervision constraints.

The dataset is relatively small (<10 variables and $<10^3$ records) and the consequent computing times are consequently reasonably low. This would encourage the use of CS methods which in the worst case are exponential (Margaritis, 2003). However, as highlighted, both CS and SB have their pros and cons, being the poor robustness of CS the most limiting factor. This motivates to design a hyper-heuristic layer with different CS, BS and hybrid learning algorithms that generate a pool of candidate seed models. The same algorithm produces a different output when a different conditional independence test or scoring function is used. Hence, a set of structure learning methods was created, by combining these algorithms with the X^2 and MI tests and with the LL, BIC, AIC and BDE scores (Table 2).

When this repository of learning objects is available, the data is split into train and test subsets (Fig. 1) with a proportion of

[1] The respective software codes are available at: https://franciscoprietocastrillo@bitbucket.org/franciscoprietocastrillo/timberms.git

Table 2
Structure learning algorithms, tests and metrics used in this study.

Algorithm	Test/score	References
Constraint based		
GS	X^2	Margaritis (2003)
GS	MI	Margaritis (2003)
IAMB	X^2	Tsamardinos et al. (2003)
IAMB	MI	Tsamardinos et al. (2003)
SI-HITON-PC	X^2	Aliferis et al. (2010)
SI-HITON-PC	MI	Aliferis et al. (2010)
Score based		
HC	LL	
HC	BIC	Russell and Norvig (2003), Kass and Wasseman (1995)
HC	AIC	Russell and Norvig (2003), Akaike (1973)
HC	BDE	Russell and Norvig (2003), Cooper and Herskovits (1992), Heckerman et al. (1995)
TABU	LL	
TABU	BIC	Russell and Norvig (2003), Kass and Wasseman (1995)
TABU	AIC	Russell and Norvig (2003), Akaike (1973)
TABU	BDE	Russell and Norvig (2003), Cooper and Herskovits (1992), Heckerman et al. (1995)
Hybrid		
MMHC	X^2	Tsamardinos et al. (2003)
MMHC	MI	Tsamardinos et al. (2003)

(67/33%). In phase I, once the models are learned from the train set, they are extended to DAGs in case they are not directed graphs. This is achieved by generating directions which are consistent with the conditional independencies in data and also avoid cycles. Then, for each model, the parameters (local probability density functions) are fitted with the train data. Since the existing dataset is small, the parameters were estimated in a Bayesian setting (see e.g., Koller & Friedman (2011)) rather than through the Maximum Likelihood Estimation (MLE) method. A uniform prior distribution was used with a minimum smoothing provided by an imaginary sample size of 1 as described in Scutari and Denis (2014). Next, a scale for model performance was created. For every fitted model the Kullback–Leibler Divergence (KLD) and the prediction error were computed. Since KLDs diverge at null data observation bins, the test dataset was smoothed by adding a small quantity of 10^{-10}. Then the KLD between the fitted joint and the smoothed observed frequencies were obtained. Prediction errors are simply the differences between the empirical joint in the test data and the fitted joint in the model averaged over the nodes in the model. To improve the reliability of this measure, instead of using a single test sample, a 10 test subsamples out from our whole data of size 33% of the total was created. Then, for each subsample, the model parameters were learned from the remaining 67% of data as before and the prediction error was computed. The final measure is the averaged error over the 10 subsamples. The method is implemented with the "hold-out" method of the *bn.cv* function available in the *bnlearn* R package (see Scutari and Denis (2014) for a general description about Bayesian Network analysis with *bnlearn*, and Scutari (2017) for more details on the method). To further illustrate how this method works for the present case, consider to have 10 iterations. At iteration k, dataset $test_k$ is built by randomly selecting 33% of all data, while $train_k$ is the remaining 67% of the data. Then the network is trained with $train_k$ and the prediction error for $test_k$ is computed. This gives $predErr_k$. Then the process is repeated for $k = 1,...,10$ and after the average of $predErr_k$ is taken. Next, the ranks for both prediction error and KLD were created with "1" meaning the best performing model under that score, "2" the second best performing model and so forth up to "n" existing models. Finally, both ranks are added in a total score and the minimum of that quantity is subtracted as to normalize the scale. Although the same scoring functions used in the learning step should not be used in the evaluation step, the respective rankings of LL, BIC, AIC and BDE were also included for completeness of their analysis.

Phase I ends with the presentation of the best ranked models from the CS, BS and hybrid learning algorithm families to the expert (prototype models). This helps the expert to better define the causal constraints and other restrictions he wants to implement.

In phase II the expert's feedback is encoded in terms of: (a) node constraints and (b) edge constraints. Node constraints refer to the elimination or not of a node. The reason for including this is that although the dataset is assumed to contain all the relevant features, once exposed to the relationships among variables, the expert might suggest that some variables are unnecessary given the observed structure. This is the case, for instance, when the dataset includes auxiliary variables whose relevance for the model is unknown beforehand. The consistency of the suggested constraints is then checked. As stressed, reversing an edge direction may introduce loops in the DAG, violate some structural constraints or render an intolerable ranking for the BN. Hence, only some edges are allowed to switch their directions. Therefore, Pearson's conditional independence test is used to check if every suggested independence can be supported by the data. To this end the null hypothesis for A,B being independent is ruled out if the associated $p - value$ for the X^2 statistic is lower or equal to 0.05. If constraints are consistent, new candidate models are generated as in phase I but with the constraints already implemented. Then, the same steps as in phase I are followed to reach the three best representative for CS, BS and hybrid models. At this point a new supervision step is introduced to select the most reliable model. This allows to rule out models holding arc directions which make no sense from the practitioner's perspective. The most reliable model (seed model in Fig. 1) is then submitted to a non-parametric bootstrap procedure. This works as follows. First, 10,000 replicas of the training data are created. Each replica is built by uniformly extracting observations with replacement a number of times equal to the size of the training data. Then, a graph instance from each replica is learned by using the same learning algorithm and scoring function as those in the seed model. For each pair of nodes A, B the proportion of models having edge $A - B$ is counted; if this is above a threshold (50% for this case) that edge is kept. Finally, for each surviving edge its direction is looked into. Conditioning on $A - B$ being present, the probability of $A \rightarrow B$ and $B \rightarrow A$ is computed. The directed edge which is above 50% is then kept. This is basically a simple majority vote implementation to improve model robustness (Efron, 1982; Sinha, 1986). The parameters of the resulting averaged model are then learned from the training dataset and the model performance is evaluated as in phase I.

3. Results

3.1. Phase I: Unsupervised

In Fig. 2 the resulting models of the algorithms in Table 2 are presented. It can be noticed how the complexity of the models in terms of edges largely varies across the different strategies and metrics. In particular, the LL score being the most complex since, as previously stressed, LL favors the creation of edges. The score values are shown in Table 3. As expected, the correlation between the number of arcs and the score functions is very high as it is among the score functions themselves. Also, it is noticed that the highest scores according to LL (most complex models) correspond to the lowest values in BIC and AIC but not for BDE (the SI-HITON-PC based networks resulted in semi-directed graphs that could not be converted into DAGs and could not be included in this ranking). The sum of the BIC, AIC and BDE ranks (total rank) sorts the al-

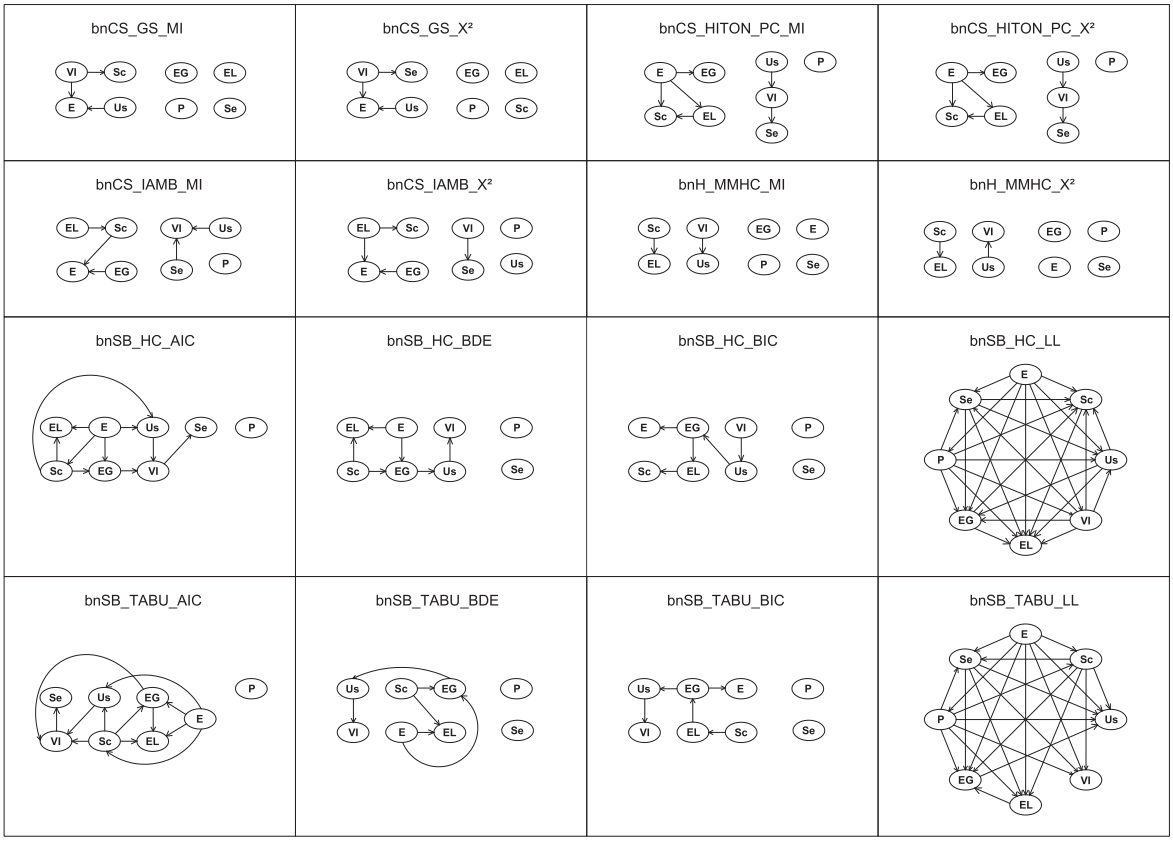

Fig. 2. Unsupervised models generated with the algorithms provided in Table 2.

Table 3
Ranking of models in the unsupervised phase. For the prediction error, KLD, LL, BIC, AIC, and BDE metrics the rank is obtained for each model according to the values obtained for each metric. The last column is computed by adding the prediction error and KLD ranks and then subtracting the minimum of that value.

Algorithm	Arcs	Error	KLD	LL	BIC	AIC	BDE	Rank
bnSB-HC-BDE	6	1	2	4	4	2	1	1
bnSB-HC-AIC	10	2	1	2	8	4	5	1
bnSB-TABU-BDE	6	2	2	4	4	2	1	2
bnSB-TABU-AIC	10	2	3	3	6	3	4	3
bnCS-HITON-PC-MI	5	2	5	5	3	5	3	5
bnSB-HC-BIC	5	3	4	6	1	1	2	5
bnSB-TABU-BIC	5	4	4	6	1	1	2	6
bnCS-IAMB-MI	5	6	6	7	5	6	6	10
bnCS-IAMB-X^2	4	5	7	8	7	8	7	10
bnH-MMHC-MI	2	7	8	11	2	7	8	13
bnCS-GS-MI	3	7	9	10	9	10	9	14
bnCS-GS-X^2	3	7	9	9	10	9	10	14
bnH-MMHC-X^2	2	8	8	11	2	7	8	14
bnSB-HC-LL	28	9	10	1	12	12	11	17
bnSB-TABU-LL	24	10	10	1	11	11	12	18

gorithms in a way that evidences the clusters (distances) among rankings; HC and TABU search heuristics group in the first positions closely. Finally, we also notice that constraint based and hybrid algorithms in general perform poorly compared with the score based algorithms (with the exception of LL scores).

In Table 4 we show a summary of the in/out degree and size of both the neighbour and Markov Blanket size of each node. The quantities shown are the mean values obtained over the whole set of models. The aim of this statistic is to evaluate the role of the variables in the different networks and see if the cause/effect relationships can be elucidated. Nodes which are likely to get more head connections (high *kin*) are better candidates for being interpreted as effects rather than causes. However, since causality can not be concluded from data unless there is a collider structure, we look into nodes with in-degree ≥ 2. This way, a high mean in-degree is interpreted as a proxy of the likelihood of a node for being a collider, which in turn is a proxy for the likelihood of that node being an effect.[2] On the other hand the neighbour size and the Markov Blanket size provide an estimate of the *mean influence* of that node in the network in terms of connectivity (high *nbrsize*) and conditional independencies (large *mbsize*)

[2] Collider descendant probability is neglected since this is less likely because a very small amount of nodes is present.

Table 4
Node statistics for unsupervised models. Measures: kin = in-degree. kout = out-degree, mbsize = Markov blanket size, nbrsize = number of neighbours. The quantities are the mean values obtained from the whole set of models for each node and measure.

Measure	E	P	Se	Sc	*VI*	Us	EL	EG
kin ≥ 2	4.000	0.000	2.000	2.000	5.000	4.000	6.000	6.000
kin	0.625	0.125	0.500	0.687	1.000	1.125	1.437	1.250
kout	1.562	0.750	0.562	1.250	0.687	0.812	0.250	0.875
mbsize	2.562	0.875	1.375	2.500	2.250	2.375	2.125	2.562
nbrsize	2.437	0.875	1.312	2.312	2.125	2.062	2.062	2.312

It is noticed that both local and global bending modulus (*EL* and *EG*) are the nodes more likely to be an effect according to their high *kin* ≥ 2 measure. This makes sense as these quantities are intrinsic to the material and the other variables are commonly used for measuring them. For instance, this is an evidence that the mechanical properties are being measured by use non-destructive tests, and not the other way around. Also, *mbsize* and *nbrsize* are highly correlated as expected. It is also noted that Position is the node with lowest connectivity.

From these results it is concluded that bnSB-HC-BDE, bnCS-HITON-PC-MI and bnH-MMHC-MI are the best candidates to represent score based, constraint based and hybrid algorithms.

The correlations showed in phase I must be considered with caution since they are unsupervised and correspond to the arrangement of nodes that allow to a higher prediction of the variables. Nevertheless, some indications for the next step are already noticeable. Firstly, it is noticed that the node *P* is independent from any other variable in the three better scored models which is consistent with the experimental results. When performing the tests at each position (bottom, middle or top boards) the results are made for each segment and no significant difference or variation of results was found between those positions.

It is also noticed that it is common to find a direct correlation between *EG* and *E* in the better ranked models, whereas *EL* is directly correlated with *Sc*. In line with the experimental results, this evidences that *EL* is a parameter more related to the material itself and thus scale dependent within the variation on each element, whereas the *EG* is influenced by the variation between different elements. In other words, there are significant different values of *EL* when changing the scale within the same element, whereas there are significant different values of *EG* between elements. Nevertheless, it is noted that *EL* and *EG* may also be indirectly correlated through the node *E*. This corroborates the known correlation between these two parameters evidencing that within a same element there is a significant correlation between local and global modulus of elasticity. It is also observed that the two non-destructive tests, visual inspection and ultrasounds, are closely related. This occurs because both measured the same segments at the same scale and for both the existence and size of defects (mainly knots) are a parameter that directly influences the results. On one hand, knots are a defect that influences the choice of a given visual grade and also knots are an obstacle to the ultrasound wave thus influencing the propagation velocity. In this case, these two nodes are closely related because of a parameter that influences directly both these variables.

3.2. Phase II: Supervision and model refinement

At this point, the model is refined by adding information provided by expert decision. Regarding this specific experimental campaign, the *Se* node was removed since it only is representative of the location of a segment and does not have any physical representation with any of the tests. Also, the *E* and *Sc* nodes were defined as root nodes as often they are known parameters of the analysis. Implementing these constraints is consistent with the workflow and the results are shown in Fig. 3

Table 5
Arc strengths in the boot model measured as proportions of edge/direction presence in 10,000 samples.

From	To	Strength	Direction
Sc	EL	1.00	1.00
E	EG	1.00	1.00
E	Us	1.00	1.00
Sc	EG	1.00	1.00
Sc	Us	0.97	1.00
E	EL	0.81	1.00
Us	VI	0.80	0.96
EG	VI	0.70	1.00

At this point, a new supervision step is introduced to check if the model with the highest performance is reliable. In this case, the bnSB-HC-AIC had the highest score and this model is approved because, in terms of physical interpretation, the colliders modulus of elascticity (both local and global) evidence that there is a significant importance on the variability between elements (due to the element node) and between scales of the same element (due to the node scale). This is found in literature where the variability within and between elements (even from the same species and origin) is proved to be significant.

As before, BDE scored with HC heuristic had a high score, thus following the workflow, the averaged (bootstraped) version of bnSB-HC-BDE is shown in Fig. 4. The arcs' strength for the booting process are shown is Table 5 as proportions of edge/direction presence in 10,000 samples.

In Table 6 the final rankings of the supervised models are presented, including the averaged network. This model (dirAvgNet) scores almost as good as the best algorithms in Table 6. Its prediction power is only one position below the seed model. The quantitative difference between both prediction errors is nevertheless small: 0.52 for dirAvgNet and 0.51 for bnSB-HC-AIC (1%). Also, both models have the best KLD ranking. This finding suggests that the averaged model quality is as least as good as the best ranked model and that perhaps its small loss of accuracy compared to bnSB-HC-AIC is because this latter slightly overfits.

At this point, both stages of the BN construction process are completed. Initially, the framework of dependencies is established and then the network is refined with expert decision. The network that results from this process is more robust and allows to perform inference on the parameters related to timber's mechanical properties. However, the use of the network is yet to be defined, as it depends on the scope of the research. In this case, the obtained BN may directly be used to ascertain different ranges for a visual grading according to the information of the visual inspection, since each class of inspection results from information obtained by the ultrasound parameter and by stiffness parameters. On the other hand, it is more common that the desirable outputs (effect nodes) of the BN are those related to the mechanical properties (in this case *EL* and *EG*) since the other parameters (cause

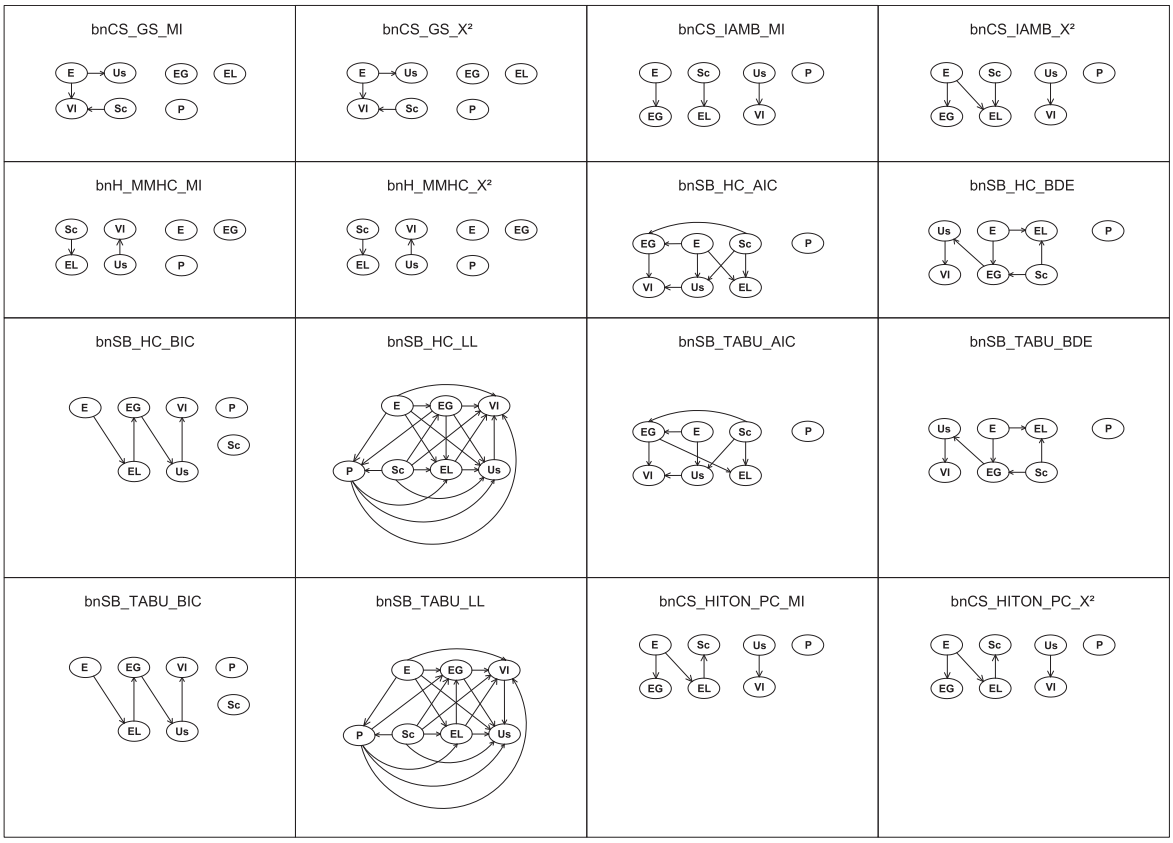

Fig. 3. Supervised models generated with the algorithms provided in Table 2.

Table 6
Ranking of models in the supervised phase. Ranks are computed as in Table 3.

Algorithm	Arcs	Error	KLD	LL	BIC	AIC	BDE	Rank
bnSB-HC-AIC	8	1	1	2	9	3	5	1
dirAvgNet	8	2	1	2	9	3	5	2
bnSB-HC-BDE	6	2	2	4	6	1	1	3
bnSB-TABU-BDE	6	2	2	4	6	1	1	3
bnSB-TABU-AIC	8	3	3	3	7	2	4	5
bnCS-HITON-PC-MI	4	3	5	6	3	4	3	7
bnCS-HITON-PC-X^2	4	3	5	6	3	4	3	7
bnCS-IAMB-X^2	4	5	4	5	5	5	2	8
bnSB-HC-BIC	4	4	6	7	2	6	6	9
bnSB-TABU-BIC	4	4	6	7	2	6	6	9
bnCS-IAMB-MI	3	8	7	8	1	7	7	14
bnSB-HC-LL	20	6	10	1	10	10	10	15
bnSB-TABU-LL	20	7	10	1	10	10	11	16
bnCS-GS-MI	3	9	9	9	8	9	9	17
bnCS-GS-X^2	3	9	9	9	8	9	9	17
bnH-MMHC-MI	2	10	8	10	4	8	8	17
bnH-MMHC-X^2	2	11	8	10	4	8	8	18

nodes used as input information) may be obtained more easily by onsite inspection.

4. Operative network

4.1. Basis for the operative network

During a safety assessment of an existing timber structure, it is required that the mechanical properties of the elements are determined. However, often it is not possible to test all elements or to remove them from the building to perform laboratory tests. Therefore, NDTs are used, but the information retrieved is only an indication of the real mechanical properties of the element. In order to obtain reliable information, the data obtained in the multi-scale experimental campaign was used to construct an operative network that infers on the stiffness of the elements (bending modulus of elasticity). This operative model is fully proposed by the authors regarding their past experience on the assessment of timber structures (Sousa et al., 2016a; Sousa et al., 2016b) without a learning process. The operative network was constructed considering three main premises: (1) geometrical parameters would be considered as root nodes, since these are parameters easily obtainable for timber elements onsite and only depend on the size of the element; (2) in the acyclic graph, NDTs would be parent nodes to the mechanical properties, since the main objectives of using these NDTs are to provide information about the conservation state of the element and to correlate that information with the mechanical

Table 7
Ranking of models including the operative network shown in Fig. 5. Ranks are computed as in Table 3.

Algorithm	Arcs	Error	KLD	LL	BIC	AIC	BDE	Rank
bnSB-HC-AIC	8	1	1	2	10	3	5	1
dirAvgNet	8	2	1	2	10	3	5	2
bnSB-HC-BDE	6	2	2	4	7	1	1	3
bnSB-TABU-BDE	6	2	2	4	7	1	1	3
bnSB-TABU-AIC	8	3	3	3	8	2	4	5
bnCS-HITON-PC-MI	4	3	5	6	3	4	3	7
bnCS-HITON-PC-X^2	4	3	5	6	3	4	3	7
bnCS-IAMB-X^2	4	5	4	5	6	5	2	8
bnSB-HC-BIC	4	4	6	7	2	6	6	9
bnSB-TABU-BIC	4	4	6	7	2	6	6	9
bnCS-IAMB-MI	3	8	7	8	1	7	7	14
bnSB-HC-LL	20	6	11	1	11	11	11	16
operational	5	9	8	10	5	8	8	16
bnSB-TABU-LL	20	7	11	1	11	11	12	17
bnCS-GS-MI	3	10	10	9	9	10	10	19
bnCS-GS-X^2	3	10	10	9	9	10	10	19
bnH-MMHC-MI	2	11	9	11	4	9	9	19
bnH-MMHC-X^2	2	12	9	11	4	9	9	20

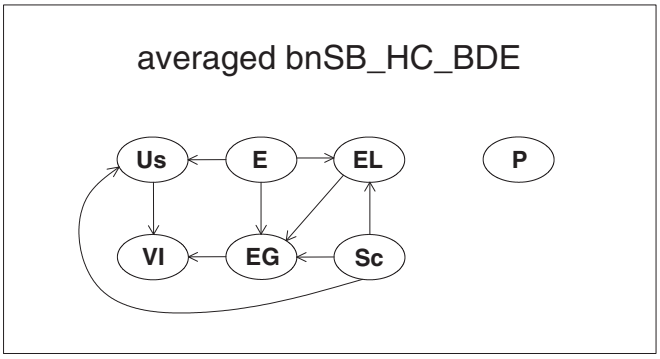

Fig. 4. Averaged network from the bnSB-HC-BDE seed model in the case of supervised training.

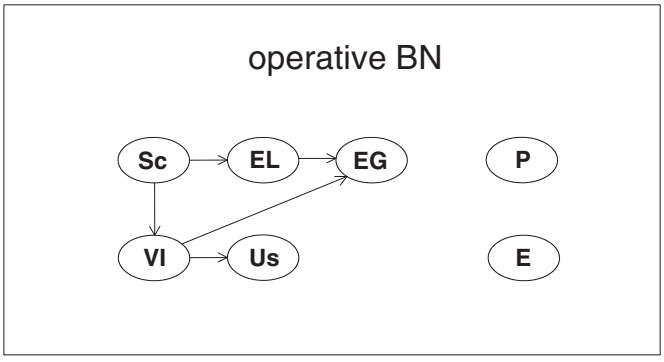

Fig. 5. Operative network proposed totally in a supervised way.

properties of the timber element without performing destructive tests; (3) the mechanical properties (bending stiffness) would be child nodes of the network, specifically *EG* that should take into account the information of local parameters (*EL*) and of global NDT (visual inspection). The main objective of the model is to predict the mechanical properties having as predictors the NDTs and scale without specific information of the element itself. The main objective of the model is to predict the mechanical properties having as predictors the NDTs and scale without specific information of the element itself. In Fig. 5 it is shown such model. From Table 7 it is noticed how this model actually outperforms the other models including the averaged network. This is only possible due to the knowledge of the expert on already well established correlations between variables. Moreover, the number of variables are small and thus a simple network is possible to be defined without significant multicollinearity constraints. In any case, it is observed that the scores are similar for the operational and the averaged network. This is a first indication that the averaged network, although being semi-supervised, can be used also for operative purposes.

4.2. Prediction model validation

In order to validate the results of the inference models, in both unsupervised and supervised phases (Phase I and II of the proposed workflow), with the results of the operative model (fully supervised), the conditional probabilities obtained for the node *EG* given its parents *EL* and *VI* are shown. The variable *EG* was considered because, in the analysis of existing timber elements, this is a mechanical property that depends on both the local material properties of timber, as well as on the number and intensity of defects. The conditional probabilities are computed by MLE fitting the local probabilities of each node of the networks. The results are presented in Fig. 6 when prior information is given for *VI* and in Fig. 7 when prior information is given for *EL*. The conditional probabilities computed through each model are used for comparison in order to compare both the feasibility of the networks, as well as the change of value for different sets of prior information. Measures of relevance, such as precision and recall, were not considered in this case study as the aim was to assess the feasibility of using the semi-automated networks independently of the strength of the connection between variables within the given database.

From the conditional probabilities considering *VI* as prior information it is clear that all models behave similarly to a change in visual grade. It is noticed that higher values of *EG* are more likely to happen for classes I and constantly decrease for the remaining classes. On the other hand, lower values of *EG* are more likely to happen for classes III and NC. It is also important to notice that the supervised and operative models have more similar results. For the case where *EL* is the prior information, a similar trend is observed. Lower values of *EL* result on higher conditional probabilities for lower values of *EG*, whereas higher values of *EL* result on higher conditional probabilities for higher values of *EG*. As seen before, *EL* result on higher conditional probabilities for lower values of *EG*. It should be noted that due to the arrangement of variables in the unsupervised and operative models, the values of *EG* given *EL* are equivalent for both models, therefore only the comparison to the supervised model is given in Fig. 7

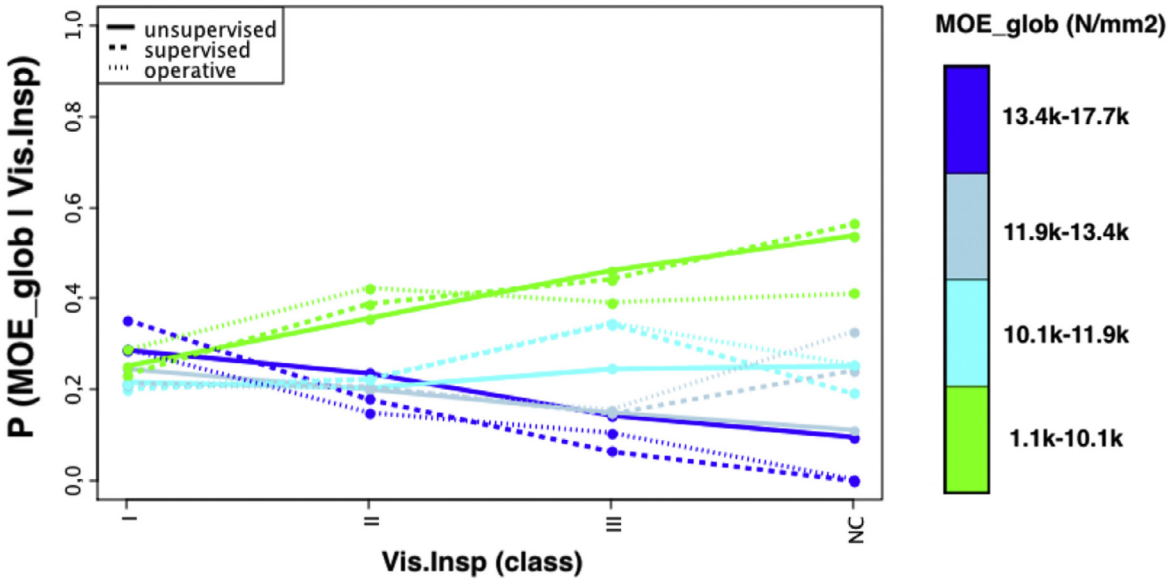

Fig. 6. Conditional probabilities of *EG* given *VI*.

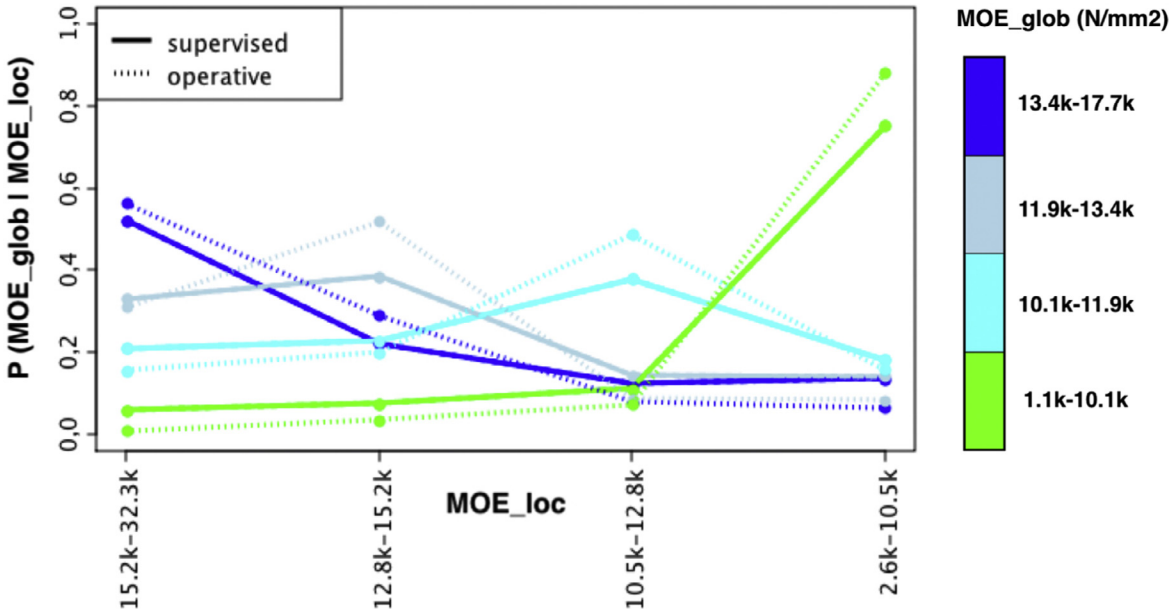

Fig. 7. Conditional probabilities of *EG* given *EL*.

The results evidence that a supervised network, based on a first structure given by an automated process, has a similar behaviour compared to a network based solely on expert knowledge, when different types of prior information are provided. Moreover, as seen before, this supervised network ranked better in terms of prediction power, thus it may be used in a case study application with a more accurate prediction of the target variables without contradicting the expert's expectations.

5. Concluding remarks

Engineering analysis of complex systems with multiple variables depend on the strength of correlation between those variables for a reliable inference of a specific target variable. When several variables are present in the analysis, optimized learned methods can be used in order to ascertain which models render the best combination of predictors. In this work, this concept was used together with Bayesian Networks as to infer on mechanical properties of timber. The proposed methodology was divided in two phases, unsupervised and supervised. Based on a ranking process a model was obtained and after compared to a model obtained exclusively from expert decision. The model obtained from the proposed methodology was able to outrank the prediction power of a network given fully by expert knowledge, still maintaining similar results for different scenarios adapted to common needs in the assessment of timber elements onsite. The comparison was made through conditional probabilities obtained with different prior information.

Accounting the proposed methodology, a comparison is made with work on the field of structure learning using Bayesian networks. For instance, in the interesting work of de Campos and Castellano (2007) the authors formalise the absence of a restriction by allowing undirected edges. This results in partially directed networks which are consistent with the constraints. The authors pro-

vide a self-consistency criterion that outperforms traditional conditional independence tests. In the case of this work, the analysis is restricted to totally directed constraints. The reason for doing this is to obtain a cause-effect situation as in this specific context, timber engineering, relationships among variables must be directed as to obtain a formal operative procedure that can be applied for the assessment of existing timber elements onsite, based on a specific database. Therefore, self-consistency conditions a), b) and c) from Proposition 1 in de Campos and Castellano (2007) are satisfied. However, even in the case of providing restriction consistency, the proposed approach relies on the supervision step of the best scored models. This allows to disregard models with meaningless arcs from the causal point of view. Therefore, this approach cannot be reduced to the self-consistency criterion, as self-consistent models in one step can be disregarded by the expert later on. Also, in Masegosa and Moral (2013) an interactive approach for integrating expert knowledge is proposed similarly to the present work. That method also relies on identifying the edges of the graph structure which are more unreliable considering the information present in the learning data and uses the expert knowledge at different stages to direct the edges accordingly to its assumptions. Similar procedure was also adopted by Chawla et al. (2016) where expert knowledge is used to prune or strengthen edges connecting variables.

In Flores, Nicholson, Brunskill, Korb, and Mascaro (2011), an algorithm based on adjacency matrices makes it clear that, in the context of a medical case study, the simplest kind of prior information was more effective in aiding model discovery than either using no prior information or using more sophisticated and detailed expert priors. In that case, a hybrid causal learning of Bayesian Networks was used for comparison of different expert information, whereas in the framework proposed in this work, the first phase is independent of expert information and it is updated in a second phase with the restrictions of the expert which already has the possibility to follow a suggested structure for its network. Also, in the field of medicine application, the work of Bouchaala, Masmoudi, Gargouri, and Rebai (2010) proposed a score based on implicit inference network. This score is then implemented in K2 and MWST algorithms. This is similar to the proposed method for this work, however in this case several structure learning algorithms were already considered and the method is not restricted to a set of implicit learning algorithms as new algorithms can be used and a ranking between models can be determined at the first stage of the procedure and after complemented with the expert information.

In contrast with Almeida et al. (2014), the first stage of this proposed method provides an initial framework that after can be updated with the experts knowledge and restrictions, nevertheless that work also indicates that Bayesian networks built by specialists can with minor perturbations yield better classifiers, while maintaining most of the interpretability of the original network. It is also noted that, in the present proposed method, the expert may choose to include or not his knowledge in the construction of the final network. This issue is also addressed in Constantinou et al. (2016) where it is aimed at preserving data-driven expectations when the expert variables remain unobserved, but also incorporating a method towards determining the accuracy of expert judgement in terms of the extent to which the variability of the revised empirical distribution is minimised. In this case, the method proposed in this work is dependent on the judgement of the expertise, however it also provides a ranking where the expert may decide to maintain or withdraw his information, with the aim of having a network with higher prediction power.

The case study analysed in the present work has a limited number of measurements which can be seen as a drawback to the analysis. Even if this is not limiting to the proposed method, an algorithm including Monte Carlo simulations, such as that proposed in Cano, Masegosa, and Moral (2013), can be further implemented if required without changing the main core of the learning structure and the ranking procedure. In the case of Cano et al. (2013), Monte Carlo simulations were used to decrease the cost of elicitation of informative prior probability distributions of the graph structures. Within the same concept, the work of Su, Borsuk, Andrew, and Karagas (2014) uses a Markov chain Monte Carlo based Bayesian Network structure learning algorithm to address the challenges of data containing missing values and incorporation of expert knowledge to improve computational efficiency. Although, in the present work the limitation regarding the use of both numerical and categorical data was undermined when categorising each variable prior to the construction of the networks, it should be noticed that, as mentioned by Chang, Zhou, Jiang, Li, and Zhang (2013), there is a combinatorial oversize issue when there are too many attributes and/or too many alternatives for each attribute in belief rule based algorithms.

Within the case study, it was found that similar trends of prediction of a global mechanical property of timber were obtained for both the supervised (learned network) and a network based on experts' knowledge, when different sets of prior information was given. The benefit of the use of the proposed method was, therefore, to obtain a network with a better prediction power without compromising its operativeness within a case study application. Even if the network provided fully by expert decision may seem quicker to obtain, this may not be the case for networks with high number of variables, where the semi-supervised procedure proposed in this work may be advantageous as it provides a first structure for the network. Moreover, the expert decision may lead to an operative model that is well fitted to its objectives (obtaining information on certain target variables), but may induce a decrease on the prediction power since it has more restrictions, thus the database may not be fully compatible with the experts initial assumptions. Therefore, although the number of samples and variables is low, it was evidenced that the proposed methodology may provide a reliable model consistent with the onsite evaluation of existing timber elements.

It should be noted that in the Bayesian approach there are many equivalent models which compete for representing knowledge and predict data in the best possible way. Furthermore, in general it is not possible to identify a single optimal causal Bayesian Network, but rather a small set of likely causal Bayesian Networks that fit the knowledge given by the data. Therefore, the advantage of the combination of expert supervision and automatic learning that was proposed in this work is that it allows to strengthen practitioner's confidence into a model. Moreover, the advantage of the proposed method is to allow the practitioner to have a framework for the construction of his network based on a ranking on prediction power. This is specially noted for networks with a high number of variables or when correlations between variables are not known in advance. In the case of a high number of variables, both a fully supervised approach followed by data verification, or rather an unsupervised approach completed by the expert intervention can be used. In the first case, the expert model will be compared in terms of ranking to other models and the expert may chose to continue with is model or update it, whereas, in the second case, the expert may intervene in the supervision of a network already proposed by the structure learning algorithm. The second case will allow for a more robust network based on its prediction power and will decrease the effort of the practitioner in determining the structure of the network.

It is stressed that even by using a very optimised approach to include expert information, as that provided in (de Campos & Castellano, 2007), the operativeness of the final model can be improved when more supervision steps are included in the workflow.

The drawbacks of this method are mainly related to the ranking procedure, since it depends on each type of structure learning algorithms which, if not chosen carefully, may be conflicting with each other. This issue is relevant since the metrics used to obtain the models should not be used to validate them. Moreover, the network must be created regarding a database that reliably represents the target variables and its correlations, which in the case of timber elements may depend on the wood species, state of conservation, among other parameters.

Acknowledgments

This work was partly financed by FEDER funds through the Competitivity Factors Operational Programme - COMPETE and by national funds through FCT Foundation for Science and Technology within the scope of the project POCI-01-0145-FEDER-007633. This research was partially supported by ProTimber research project (PTDC/ECM-EST/1072/2014) financed by the FCT.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the second international symposium on information theory, Armenia, USSR* (pp. 267–281).

Aliferis, F., Statnikov, A., Tsamardinos, I., Subramani, M., & Koutsoukos, X. (2010). Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. *Journal of Machine Learning Research, 11*, 171–234. doi:10.1016/S0031-8914(53)80099-6.

Almeida, E., Ferreira, P., Vinhoza, T. T. V., Dutra, I., Borges, P., Wu, Y., & Burnside, E. (2014). Expert Bayes: Automatically refining manually built Bayesian networks. In *Proceedings of the thirteenth international conference on machine learning and applications (ICMLA)*. Detroit, USA: IEEE.

Armero, C., & Conesa, D. (2006). Bayesian hierarchical models in manufacturing bulk service queues. *Journal of statistical planning and inference, 136*(2), 335–354. doi:10.1016/j.jspi.2004.07.007.

Bouchaala, L., Masmoudi, Gargouri, F., & Rebai, A. (2010). Improving algorithms for structure learning in Bayesian networks using a new implicit score. *Expert Systems with Applications, 37*(7), 5470–5475. doi:10.1016/j.eswa.2010.02.065.

Bouckaert, R. (1993). Probabilistic network construction using the minimum description length principle. In M. Clarke, R. Kruse, & S. Moral (Eds.), *Symbolic and quantitative approaches to reasoning and uncertainty*. Taylor & Francis Group. doi:10.1007/BFb0028180.

de Campos, L. M., & Castellano, J. G. (2007). Bayesian network learning algorithms using structural restrictions. *International Journal of Approximate Reasoning, 45*(2), 233–254. https://doi.org/10.1016/j.ijar.2006.06.009.

Cano, A., Masegosa, A. R., & Moral, S. (2013). A method for integrating expert knowledge when learning Bayesian networks from data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41*(5), 1382–11394. doi:10.1109/TSMCB.2011.2148197.

CEN (2010). *EN 408:2010 - Timber structures - Structural timber and glued laminated timber - Determination of some physical and mechanical properties*. Brussels: European Committee for Standardization.

Chang, L., Zhou, Y., Jiang, J., Li, M., & Zhang, X. (2013). Structure learning for belief rule base expert system: A comparative study. *Knowledge-Based Systems, 39*, 159–172. doi:10.1016/j.knosys.2012.10.016.

Chawla, V., Naik, H. S., Akintayo, A., Hayes, D., Schnable, P., Ganapathysubramanian, B., & Sarkar, S. (2016). A Bayesian network approach to county-level corn yield prediction using historical data and expert knowledge. In *Proceedings of the twenty-second ACM SIGKDD workshop on data science for food, energy and water, San Francisco, USA*.

Constantinou, A., Fenton, M., & Neil, M. (2016). Integrating expert knowledge with data in Bayesian networks: Preserving data-driven expectations when the expert variables remain unobserved. *Expert Systems with Applications, 56*, 197–208. doi:10.1016/j.eswa.2016.02.050.

Cooper, G., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning, 9*, 309–347. doi:10.1007/BF00994110.

de Campos, C., & Ji, Q. (2010). Properties of Bayesian Dirichlet scores to learn Bayesian network structure. In *Proceedings of the twenty-fourth AAAI conference on artificial intelligence* (pp. 431–436).

Deublein, M., Schlosser, M., & Faber, M. (2011). Hierarchical modelling of structural timber material properties by means of Bayesian probabilistic networks. In Faber, Kohler, & Nishijima (Eds.), *Applications of statistics and probability in civil engineering* (pp. 1377–1385). Taylor & Francis Group.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. In *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, PA: Society for Industrial and Applied Mathematics. ISBN: 978-0-89871-179-0, doi: 10.1137/1.9781611970319.fm.

Feio, A., & Machado, J. (2015). In-situ assessment of timber structural members: Combining information from visual strength grading and NDT/SDT methods – A review. *Construction and Building Materials, 101*, 1157–1165. doi:10.1016/j.conbuildmat.2015.05.123.

Fenton, ., & Neil, M. (2013). *Risk assessment and decision analysis with Bayesian networks*. Taylor & Francis Group.

Fink, G., & Kohler, J. (2015). Quantification of different NDT/SDT methods in respect to estimate the load-bearing capacity. *Construction and Building Materials, 101*, 1181–1187. doi:10.1016/j.conbuildmat.2015.07.015.

Flores, M. J., Nicholson, A. E., Brunskill, A., Korb, K. B., & Mascaro, S. (2011). Incorporating expert knowledge when learning Bayesian network structure: A medical case study. *Artificial Intelligence in Medicine, 53*(3), 181–204. doi:10.1016/j.artmed.2011.08.004.

Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning, 20*(3), 197–243. doi:10.1007/BF00994016.

Isaksson, T. (1999). Modelling the variability of bending strength in structural timber. Report No. TVBK-1015, Department of Structural Engineering, Lund University.

JCSS (2000). *JCSS probabilistic model code, part 1: Basis of design. Probabilistic model code*. Joint Committee on Structural Safety. ISBN 978-3-909386-79-6

Kass, R., & Wasseman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association, 90*(431), 928–934. doi:10.1080/01621459.1995.10476592.

Koller, D., & Friedman, N. (2011). *Probabilistic graphical models: principles and techniques*. The MIT Press.

Machado, J., & Palma, P. (2011). Non-destructive evaluation of the bending behaviour of in-service pine timber structural elements. *Materials and Structures, 44*(5), 901–910. doi:10.1617/s11527-010-9674-9.

Margaritis, D. (2003). Learning Bayesian network model structure from data. (Ph.D. thesis). Carnegie Mellon University.

Marsili, F., Croce, P., Klawonn, F., & Landi, F. (2017a). A Bayesian network for the definition of probability models for compressive strength of concrete homogeneous population. In *Proceedings of the fourteenth international probabilistic workshop*. (pp. 269–283). Springer. doi:10.1007/978-3-319-47886-9_19.

Marsili, F., Croce, P., Klawonn, F., Vignoli, A., Boschi, S., & Landi, F. (2017b). A Bayesian network for the definition of probability models for masonry mechanical parameters. In *Proceedings of the fourteenth international probabilistic workshop*. (pp. 253–268). Springer. doi:10.1007/978-3-319-47886-9_18.

Masegosa, A. R., & Moral, S. (2013). An interactive approach for Bayesian network learning using domain/expert knowledge. *International Journal of Approximate Reasoning, 54*(8), 1168–1181. doi:10.1016/j.ijar.2013.03.009.

Maxwell, D. (1996). Learning Bayesian networks is np-complete. In *Learning from data: Artificial intelligence and statistics v* (pp. 121–130). New York: Springer. doi:10.1007/978-1-4612-2404-4_12.

Pearl, J. (1997). *Probabilistic reasoning in intelligent systems: networks of plausible inference* (2nd). Morgan Kaufmann.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd). Cambridge University Press.

Pendharkar, P., Subramanian, G., & Rodger, J. (2005). A probabilistic model for predicting software development effort. *IEEE Transactions on software engineering, 31*(7), 615–624. doi:10.1109/TSE.2005.75.

Rebonato, R. (2010). *Coherent stress testing: A Bayesian approach to the analysis of financial stress*. John Wiley & Sons.

Riberholt, H., & Madsen, P. (1979). Strength of timber structures, measured variation of the cross sectional strength of structural lumber. Report no. R 114, Struct. Research Lab., Technical University of Denmark.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd). Pearson Education.

Scutari, M. (2017). bnlearn – an R package for Bayesian network learning and inference – cross-validation for Bayesian networks. (Accessed September 13 2017). Available at: http://www.bnlearn.com/documentation/man/bn.cv.html.

Scutari, M., & Denis, J.-B. (2014). *Bayesian networks with examples in R*. Boca Raton: Chapman and Hall. ISBN 978-1-4822-2558-7, 978-1-4822-2560-0

Singh, M., & Valtorta, M. (1995). Construction of Bayesian network structures from data: a brief survey and an efficient algorithm. *International Journal of Approximate Reasoning, 12*(2), 111–131. doi:10.1016/0888-613X(94)00016-V.

Sinha, N. (1986). Robust bootstrap method for joint estimation of states and parameters of a linear system. *Journal of dynamic systems, measurement, and control, 108*(3), 255–263. doi:10.1115/1.3143775.

Smart, P., Simperl, E., & Shadbolt, N. (2014). A taxonomic framework for social machines. In D. Miorandi, V. Maltese, M. Rovastsos, A. Nijholt, & J. Stewart (Eds.), *Proceedings of the social collective intelligence: Combining the powers of humans and machines to build a smarter society* (pp. 51–58). Berlin, Germany: Springer. doi:10.1007/978-3-319-08681-1_3.

Sousa, H., Branco, J., & Lourenço, P. (2016a). Use of bending tests and visual inspection for multi-scale experimental evaluation of chestnut timber beams stiffness. *Journal of Civil Engineering and Management, 22*(6), 728–738. doi:10.3846/13923730.2014.914083.

Sousa, H., Branco, J., Lourenço, P., & Neves, L. (2016b). Inference on stiffness and strength of existing chestnut timber elements using hierarchical Bayesian probability networks. *Materials and Structures, 49*(10), 4013–4028. doi:10.1617/s11527-015-0770-8.

Sousa, H., Machado, J., Branco, J., & Lourenço, P. (2015). Onsite assessment of structural timber members by means of hierarchical models and probabilistic methods. *Construction and Building Materials, 101*, 1188–1196. doi:10.1016/j.conbuildmat.2015.05.127.

Spiegelhalter, D., Abrams, K., & Myles, J. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons.

Spirtes, P., Glymour, G., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Su, C., Borsuk, M. E., Andrew, A., & Karagas, M. (2014). Incorporating prior expert knowledge in learning Bayesian networks from genetic epidemiological data. In *Proceedings of the IEEE conference on computational intelligence in bioinformatics and computational biology*. Honolulu, USA: IEEE.

Tsamardinos, I., Aliferis, C., & Statnikov, A. (2003). Algorithms for large scale Markov blanket discovery. In *Proceedings of the sixteenth international Florida artificial intelligence research society conference* (pp. 376–381). AAAI Press.

UNI (2004). *EUNI 11119:2004 cultural heritage – Wooden artifacts – Load-bearing structures – Onsite inspections for the diagnosis of timber members*. Milan: UNI Ente Nazionale Italiano di Unificazione.

Vrouwenvelder, A. (2002). Developments towards full probabilistic design codes. *Structural Safety, 24*(2), 417–432. doi:10.1016/S0167-4730(02)00035-8.

Zhou, Y., Fenton, N., & Neil, M. (2014). An extended MPL-C model for Bayesian network parameter learning with exterior constraints. In L. van der Gaag, & A. Feelders (Eds.), *Probabilistic graphical models*. Springer International Publishing. doi:10.1007/978-3-319-11433-0_38.