Development of a framework to estimate the condition of mechanical, electrical and plumbing systems with Bayesian Networks.

An application to Air Handling Units in the Netherlands.

MSc. Construction Management and Engineering
Benjamin Ramousse

Delft University of Technology

Development of a framework to estimate the condition of mechanical, electrical and plumbing systems with Bayesian Networks: An application to Air Handling Units in the Netherlands.

Graduation thesis

University	Delft University of Technology
Faculty	Civil Engineering and Geosciences (CiTG)
Curriculum	MSc Construction Management and Engineering

Author

Name	Benjamin Ramousse
Student number	5526337
E-mail address	b.ramousse@student.tudelft.nl

Graduation committee

Chairman, TU Delft	Dr.ir. Oswaldo Morales Napoles
1st supervisor, TU Delft	Dr.ir. Ad Straub
2nd supervisor, TU Delft	MSc. Miguel A. Mendoza Lugo
Company supervisor, DEMO Consultants B.V.	ir. Menno van den Broeke

In partial fulfilment of the Master of Science in **Construction Management and Engineering**.

September 21, 2023

Preface

Ending this deliverable is at the same time touching and relieving. Similarly to my international experience in Delft, the completion of this master's thesis has been a long journey, lonely at times, but which thankfully allowed to discover new parts of myself and others. The last eight months have fed my growing interest in statistics, programming, 'numbers', somewhat helped me realize where my strengths resided - and where they certainly did not.

I am excited to share this deliverable with my committee members, who I'd like to thank for their guidance all the way throughout. Menno and Ad, thank you for enriching my non-existent knowledge of buildings, your critical support and for answering sometimes stupid questions. Oswaldo and Miguel, thank you for your enthusiasm for Bayesian networks (and somewhat for my thesis), which made the process much cooler and interesting, as well as your respective expertise and confidence.

Thank you to my friends, here in Delft and back home, for the support and fun we had together. Having a drink and a laugh was probably the best possible motivation to keep studying when the mind did not feel like it. In particular, I'd like to share my gratitude to the thesis (mental) support group, whose remarkable insights have strongly contributed to my success. Get yourself a support group, you won't regret it.

Lastly, thank you to my family. I have and will keep finding strength in the pride you have in me, so I owe that degree to you too !

Enjoy the read, it's a long one !

Benjamin Ramousse,
Delft, 18/09/2023

Executive summary

As the Western European building stock ages, attention is increasingly allocated to the maintenance of building components, particularly mechanical, electrical and plumbing (MEP) systems. Although the latter are essential in ensuring the correct operation of a building and the safety of its occupants, they remain the crafts where the most defects are observed, resulting in significant material costs. This phenomenon partly finds explanation in the shortcomings of current condition assessment methods for MEP systems, which often poorly describe the actual state of the components.

As a result, novel approaches to estimate the condition of these building elements are investigated by industry participants. Among them, Bayesian Networks (BNs) are probabilistic models that progressively gain momentum for real-life applications. In the context of the present research, their relevance is twofold: (i) their graphical structure allows to visually model influence between large sets of variables, and (ii) they robustly handle missing data. Unfortunately, like most probabilistic models, their quantification requires extensive amounts of empirical data which is extremely sparse for MEP systems. Therefore, this thesis attempts to answer the following question:

> **How can Bayesian Networks be applied to estimate the condition of mechanical, electrical, and plumbing systems in the absence of empirical data?**

Methods

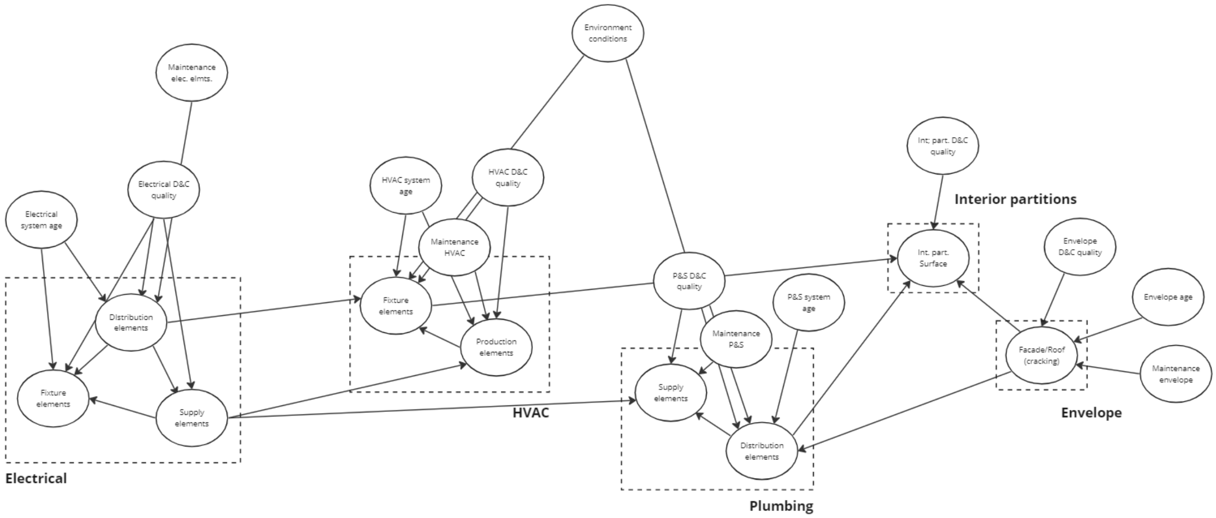
In their 'traditional' discrete form, BNs have a limited range of applications. First, they do not allow the integration of continuous variables, which for numerous physical problems is a major drawback. Second, the number of parameters to quantify discrete networks quickly becomes intractable as the number of states and parents increases, again limiting their implementation for complex systems. Therefore, Non-Parametric Bayesian Networks (NPBNs) are adopted in this research, whose formulation is based on (conditional) rank correlations (dependence) and marginal distributions associated to each of the network's variables.

To overcome the challenge imposed by the limited availability of empirical data, several studies have investigated the use of field experts' judgments for the quantification of BNs. While the elicitation of univariate distributions has been thoroughly studied, the assessment of dependence remains an emerging topic in structured expert judgment (SEJ) literature. Consequently, this thesis focuses on the development of a method for the assessment of rank correlations by field experts, whereas a lesser effort is allocated to the elicitation of the marginal distributions.

Existing research has delved into the use of two approaches for the elicitation of dependence: statistical and conditional fractile estimates. Here, the suitability of probabilities of concordance, a third type of probabilistic assessment, is investigated. Under the normal copula assumption, common in the context of SEJ, unconditional rank correlations can be retrieved from probabilities of concordance using a set of closed-form relations. Then, the individual experts' opinions are aggregated using dependence-calibration, a performance-based aggregation method gaining momentum for NPBNs. The application of these approaches to MEP systems in discussed later in this summary.

Findings

The first step in the creation of a BN is the definition of a graph. Therefore, a classification of the MEP systems is developed and constitutes the foundation of the network. Subsequently, the factors influencing the condition of the sub-systems classified previously are identified. The literature reviewed suggests a distinction between two types of relationships: those between exogenous variables (e.g. maintenance or environmental conditions) and building components, and those between components themselves. Following the identification of these relationships, a ‘global’ graph encompassing all MEP systems arises:



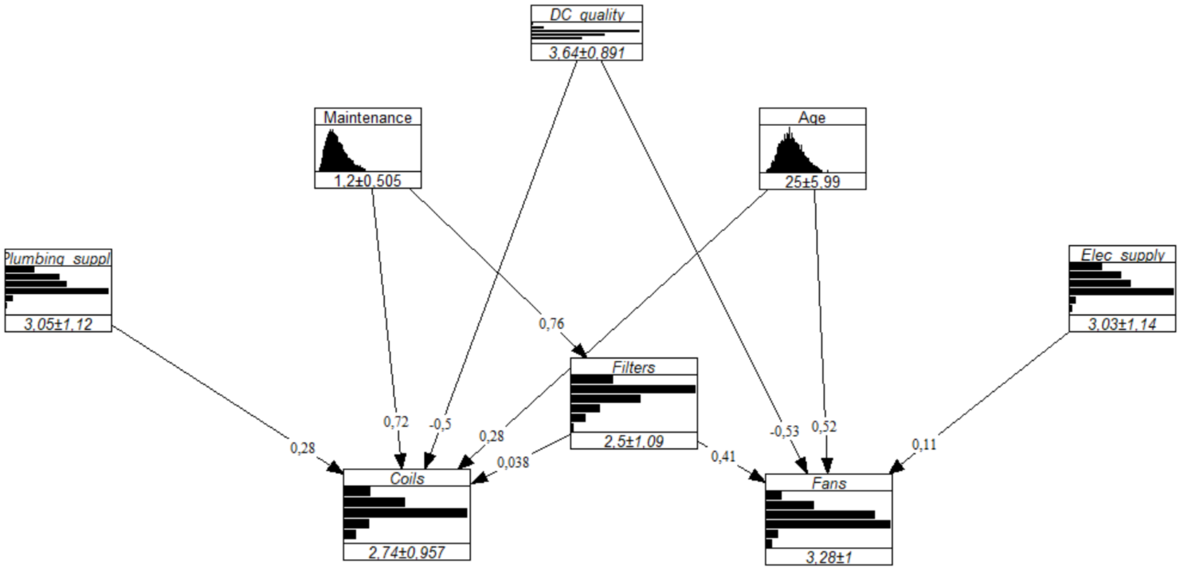
Before engaging in the quantification of this network, it is crucial to interrogate its feasibility given the time span of this research and the absence of empirical data. With 23 variables and over 30 edges, the assessment of all correlations (leave alone of the marginal distributions) is practically unrealistic solely based on experts’ judgments. Therefore, the remainder of the report presents a case study on air handling units (AHUs), for which the elicitation method is implemented. The graph defined for AHUs is illustrated in the figure below.

Questions for the assessment of probabilities of concordance related to the newly created graph are then formulated, taking a similar form as follows:

> “Two buildings A and B are randomly selected among all non-residential buildings in the Netherlands. Given that the AHU in building A is maintained more regularly than in building B ($y_A \leq y_B$), what is the probability that the coils are in better condition in building A than building B ($x_A \leq x_B$) ?”

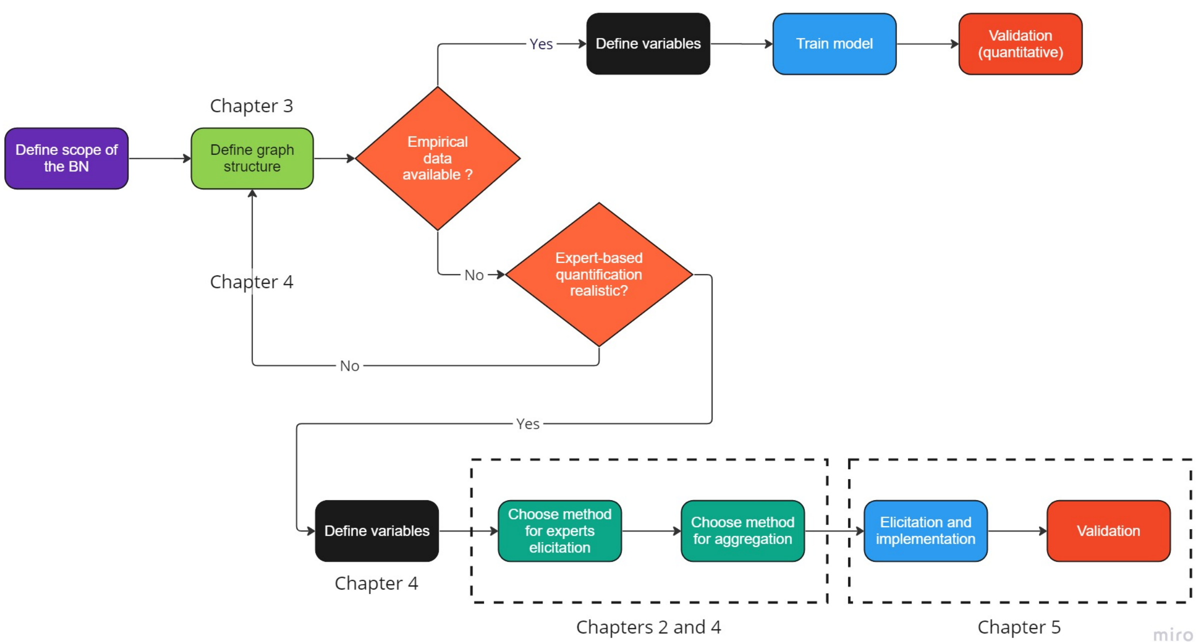
Similar questions were created for each of the network’s edges and presented to a panel of five experts, resulting in five individual correlation matrices. As mentioned previously, the experts were then evaluated using seed questions in the context of dependence-calibration. Additional questions were thus asked to the participants with regards to precipitation in the Netherlands, a choice motivated by the absence of data related to AHUs and mechanical systems for calibration. The respondents’ calibration scores were then calculated using their assessments on the seed questions and the correlation matrix retrieved from empirical data. Finally, a combination of the experts’ dependence structures was built using their calibration scores in a weighted average, resulting in a unique set of correlations which were implemented in the NPBN.

Lastly, two of the five experts consulted previously participated in the elicitation of the marginal distributions, either by the direct provision of the distribution or through answers to qualitative statements. The resulting model, including both marginal distributions and (conditional) rank correlations, is illustrated below.



To conclude, the NPBN is validated. While the lack of empirical data prevents the quantification of the model's predictive validity, a scenario analysis is performed to observe its output under different input combinations. It reveals that the exclusion of the environmental conditions from the network results in unrealistic outcomes, thus refuting an assumption made earlier in this research. Moreover, a global sensitivity analysis is conducted based on Sobol's method, which demonstrates the high contributions of all inputs to the outputs' variances. Consequently, evidence on any of the inputs substantially reduces the uncertainty in the output distributions, a comforting conclusion on the relevance of the chosen factors.

The final result of this thesis is a flowchart illustrating the construction process of a Non-Parametric Bayesian Network. It provides academics and practitioners with a foundational framework for the creation of Bayesian Networks, irrespective of the quantification method selected. While this thesis proposes the implementation of a particular expert-based elicitation method, the most suitable approach should be chosen with regards to the system modelled.



Contents

List of Figures

List of Tables

Nomenclature

Abbreviation	Definition
AHU	Air Handling Unit
BCA	Building Condition Assessment
BIS	Building Inspection System
BN	Bayesian Network
CBM	Condition-Based Maintenance
CDF	Cumulative Distribution Function
CM	Corrective Maintenance
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DM	Decision-Maker
GSA	Global Sensitivity Analysis
HVAC	Heating, Ventilation and Air Conditioning
MEP	Mechanical, Electrical and Plumbing
NPBN	Non-Parametric Bayesian Network
P&S	Plumbing & Sanitary
PdM	Predictive Maintenance
PM	Preventive Maintenance

Chapter 1

Introduction

1.1 Background

In most Western European countries, the second half of the XXth century saw a steep increase in real estate production, both in the housing and the non-residential markets. Research in that period essentially focused on developing new techniques and expertise in the construction process, while little attention was given to the later stages of buildings' life cycle. Nevertheless, the ageing of the building stock in the region sparks the industry's interest in maintenance and future developments in the field.

Maintenance programs have kept evolving over the last hundred years to meet the industry's needs, and despite the lack of consensus on maintenance taxonomy, three distinctive movements are identifiable, an overview of which is presented in Figure 1.1. The first and historic type is reactive or corrective maintenance (CM), whereby works are performed after the occurrence of a failure in order to bring a component back into a state where it can perform its intended functions (CEN, 2010; Sullivan, Pugh, Melendez, & Hunt, 2010). However, the strategy's efficiency heavily relies of occupants' disposition to request repairs, especially for indoor spaces (Straub, 2012).

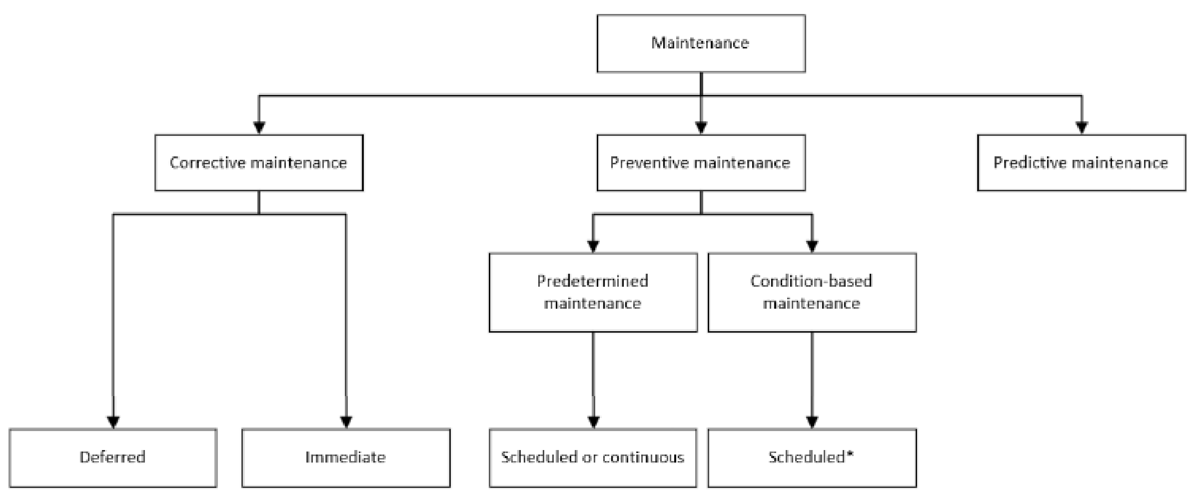


Figure 1.1: Overview of the maintenance types.

To overcome the shortcomings of CM, asset managers progressively adopted predetermined preventive maintenance (PM) for which interventions are scheduled in fixed-time intervals (Straub, 2012; Sullivan et al., 2010; Endrenyi, Anders, & Leite da Silva, 1998) to reduce or eliminate deterioration of building components

(Endrenyi et al., 2001; Lee & Cha, 2016) without failure actually happening (Lind & Muyingo, 2012). The shift to PM marked a turning point in research on (building) maintenance, bringing to the fore the relevance of maintenance planning optimization. Barlow and Hunter (1960) pioneered maintenance theory by applying reliability theory to compare PM policies and determine optimal time intervals between interventions. Lower (repair) expenses and increased up-times allowed PM to prevail on reactive maintenance, with such benefits being accessible without an holistic understanding of buildings (Nakagawa, 2005; Sullivan et al., 2010).

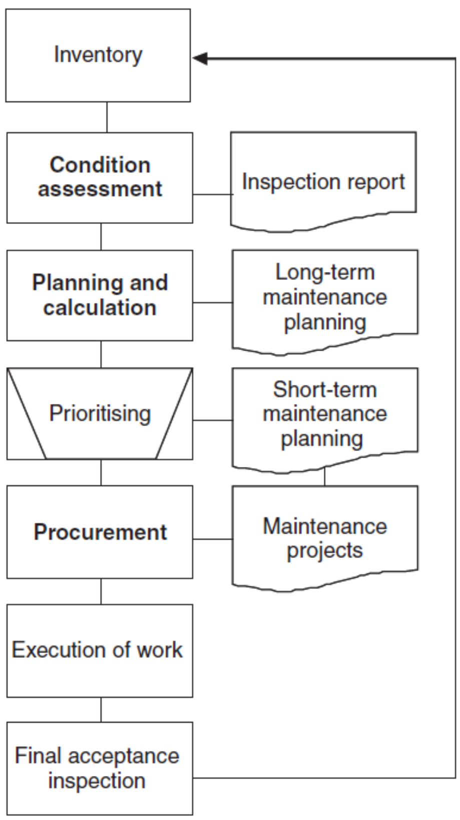


Figure 1.2: Condition-based maintenance process (Straub, 2012).

Rapidly though, scholars and practitioners came to realize that certain components were replaced despite being in good condition, thus incurring unnecessary expenses. For that reason, condition-based maintenance (CBM) gained momentum; in contrast to predetermined PM, it is planned based on the observed condition of the assets under supervision, which is assessed during periodic inspections (Endrenyi et al., 2001; Lind & Muyingo, 2012) - a process illustrated in Figure 1.2. Progressively, several national standards arose to facilitate the implementation of CBM on a larger scale, a review of which is presented in chapter 3. In practice, however, the assessment of some components' condition is lacking and performed using theoretical lifespan curves - which use age as the sole variable - or simply left out of the multi-year maintenance planning. CBM and its setbacks are further discussed in subsection 3.1.2.

The third and most recent type of maintenance is predictive maintenance (PdM), which builds on condition-based maintenance to forecast the deterioration of building components over time. Using (real-time) data and a set of predictive models, asset managers are able to estimate failure risks and organize interventions accordingly (Santiago, Antunes, Barraca, Gomes, & Aguiar, 2019; PwC, 2017) . This newer approach admittedly decreases the likelihood of catastrophic failures and improves cost effectiveness since works are performed only when needed (Sullivan et al., 2010). Nonetheless, several challenges must be addressed to anchor predictive maintenance in practice, including the important upfront investment required to implement sensors (Grussing & Liu, 2014; PwC, 2017) and the difficulties related to the analysis of large data sets (Paolanti et al., 2018; Santiago et al., 2019).

As the previous paragraphs highlight, the creation of an appropriate maintenance program encompasses numerous factors, raising the question of its impact on operations. First, proper maintenance increases end-user satisfaction by diminishing downtime and increasing mean time to failure (Endrenyi et al., 2001) - for which repairs can bear high costs. Second, implementing simple and inexpensive measures can result in significant improvements in buildings' energy efficiency (up to 20%, see Sullivan et al., 2010). This is of the utmost importance when considering the impact of buildings on global energy consumption; in the European Union, the sector consumes around 40% of all energy and contributes to the same extent to greenhouse gas emissions (GHG) (Hosamo, Nielsen, Kraniotis, Svennevig, & Svidt, 2023).

1.2 Problem statement

In the context presented previously, mechanical, electrical and plumbing (MEP) systems and their maintenance are under practitioners' scrutiny. However, the condition assessment of these components is difficult and often based on trivial models or the expertise of third parties, resulting in a poor integration in the overall maintenance strategies. Although the introduction of sensors and Building Information Modelling (BIM) in newer installations has allowed to build faults detection and prediction models (e.g. Hosamo et al., 2023), a large share of the existing Western European building stock does not have access to this resource. Therefore, **there is a need for a method to estimate the condition of mechanical, electrical and plumbing systems based on easily retrievable information**, i.e. with little to no additional measurement.

1.3 Research objective

To tackle the problem formulated above, this research aims to develop a model to estimate the technical condition of the mechanical (Heating, Ventilation and Air Conditioning, HVAC), plumbing and electrical systems. In contrast with other existing methods, it must allow to perform inference using data that asset managers and owners can retrieve with minimum effort, such as a system's age or the environmental conditions to which it is exposed. To that end, Bayesian Networks (BNs) are deemed relevant; their growing use in a range of practical areas showcases their pertinence when modelling influence between large sets of variables (see chapter 2 for a detailed introduction).

1.4 Research questions

In line with the research objective, this thesis will attempt to answer the following question:

> **How can Bayesian Networks be applied to estimate the condition of mechanical, electrical, and plumbing systems in the absence of empirical data?**

which can further be decomposed in four sub-questions:

RQ1: What are the barriers and enablers driving the integration of Bayesian Networks in building condition assessment?
Despite their numerous strengths, Bayesian Networks are not limitless. Therefore, a review of existing research is first conducted to understand the capacities of BNs and design the model accordingly. The literature investigated includes books, academic papers as well as scientific reports related to Bayesian networks and their applications.

RQ2: Which factors affect the condition of mechanical, electrical and plumbing systems?
The kernel of Bayesian networks is the creation of a comprehensive graph structure. Therefore, the main mechanical, electrical and plumbing (MEP) sub-systems that ought to be modelled are first identified. Then, the factors influencing the condition of these sub-systems are investigated, including both physical variables (e.g. age, environmental conditions) and interdependencies between building components. Because of the complexity of MEP systems, the works of Bortolini and Forcada (2018, 2020) form the foundational framework of this endeavour, complemented by additional building pathology literature.

RQ3: How can the model be populated in the absence of empirical data?
Bayesian networks' structure is characterized by their nodes and edges, respectively necessitating the specification of their marginal distributions and rank correlations. However, building condition assessment data is scarce and strongly scattered. As a result, a novel expert-based approach for the elicitation of rank correlations, based on probabilities on concordance and dependence-calibration, is applied to a case study on air handling units while marginal distributions are retrieved directly from consultations with experts.

RQ4: To what extent can the proposed model estimate the condition of mechanical systems?
Lastly, the reliability of the model is investigated and its relevance with regards to the previously defined objectives is assessed. Although the absence of empirical data hampers the quantification of the model's predictive validity, a scenario analysis and a sensitivity analysis are applied to determine the validity of the network qualitatively.

The methods applied to address those questions are introduced in chapter 2. Because it involves qualitative and quantitative methods, this thesis' methodology is a **mixed-method**.

1.5 Relevance

Before moving on to the rest of this report, this research's relevance, both on practical and theoretical grounds, is briefly presented.

1.5.1 Practical

In spite of the large contribution of MEP systems to the overall building's performance (Eleftheriadis & Hamdy, 2017), and therefore occupants' comfort and satisfaction (Waddicor et al., 2016; Bortolini & Forcada, 2019), they remain the crafts where the most defects are observed (Weeks & Leite, 2021). The lack of attention given to the maintenance of MEP systems is problematic for at least three reasons: (i) their malfunctions significantly impact occupants' well-being and perception of building quality (Zalejska & Hungria, 2019; Olanrewaju, Khamidi, & Idrus, 2010), (ii) the replacement or repairs of defective systems entails substantial material costs (Islam, Nazifa, & Mohamed, 2019; Weeks & Leite, 2021) and (iii) disruptions can significantly affect business operations. Being able to estimate the condition of these elements more accurately is therefore paramount for asset managers to allocate their resources appropriately and increase their installations' uptime. To that end, the framework developed in the present research aims to provide managers with a reliable and accessible solution for the estimation of MEP systems' condition.

1.5.2 Theoretical

In addition to the practical considerations aforementioned, this thesis contributes to two scientific fields. First, it contributes to the research on probabilistic modelling for civil engineering applications, and more particularly buildings. While various works (articles, theses) have focused on coupling infrastructure with probabilistic models, little attention has been given to buildings. Secondly, it presents a novel methodology for the elicitation of experts' judgments in Non-Parametric Bayesian Networks, involving both probabilities of concordance for individual assessments and d-calibration for aggregation (see chapter 2). In particular, the use of probabilities of concordance provides an alternative to the already widespread conditional probabilities of exceedance, as discussed later in this report.

Chapter 2

Methods

A set of methods and tools were applied to reach the objectives defined in the previous chapter. After introducing (Non-Parametric) Bayesian Networks, the procedure designed for the elicitation of experts' judgments is presented, supported by state-of-the-art literature.

2.1 Bayesian Networks

Before engaging in the construction of the network, it is essential to understand the concepts on which Bayesian Networks (BNs) are built. Therefore, this section introduces BNs and Non-Parametric Bayesian Networks (NPBNs), which were adopted in this research. Then, the barriers and drivers of the implementation of BNs in various industries are discussed to gain a grasp of the expected capacities and limitations of the model developed in this study.

2.1.1 Introduction

Introduced in the 1970s in cognitive sciences, Bayesian Networks were theorized by Judea Pearl in his 1988 book *"Probabilistic Reasoning in Intelligent Systems"* (Pearl, 1988). In their 'traditional' discrete form, they contain two types of information: a directed acyclic graph (DAG) and Conditional Probability Tables (CPTs), as illustrated in Figure 2.1. In graph theory, a graph $\mathcal{G}$ is defined by a pair $\mathcal{G} = (V, E)$ where V is a set of *vertices* (or nodes) and E a set of *edges* (or links) between elements of V (Cowell, 1999). It is called directed when the edges are directed, explicitly describing influence between nodes and acyclic because the edges do not form any cycle (Cowell, 1999), i.e. there is no path starting and ending at the same node. Furthermore, to each vertex with at least one parent is associated a conditional probability table which translates the *strength* of its parents' influence.

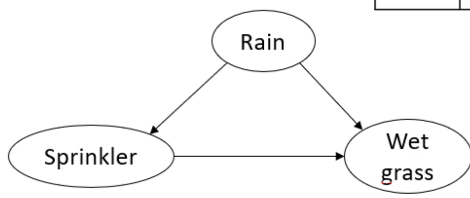
A fundamental assumption in Bayesian modelling is that of conditional independence. Let $(x_1, ..., x_k, ..., x_n)$ be an enumeration of all the nodes of a BN, and $pa(k)$ the set of parents of a node x_k. Then x_k is independent of all $x_i \notin pa(k)$ given $x_j \in pa(k)$, which mathematically translates to (N. L. Zhang & Poole, 1996):

$$P(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} P(x_i|pa(i)). \tag{2.1}$$

The combination of the probability tables and the conditional independence property allows to perform inference efficiently, i.e. to determine the distribution of a variable X given a vector of observations $\mathbf{Y} = \mathbf{y_0}$. Although this conditional distribution could be retrieved from the joint distribution $P(X, Y)$ and the total probability law, this approach is practically unrealistic as it involves an exponential number of additions (N. L. Zhang & Poole, 1996). Therefore, efficient inference relies on the notion of *factorization* (N. L. Zhang & Poole, 1996; Pearl & Russell, 2000): using the example in Figure 2.1, we aim to determine the joint distribution $P(rain, sprinkler, wet\,grass)$. Given the structure of the graph, it can be factorized as follows:

$$P(rain, sprinkler, wet\,grass) = P(rain)P(sprinkler|rain)P(wet\,grass|rain, sprinkler). \tag{2.2}$$

Rain	
T	F
0.2	0.8



	Sprinkler	
Rain	T	F
F	0.4	0.6
T	0.01	0.99

		Wet grass	
Sprinkler	Rain	T	F
F	F	0	1
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Figure 2.1: Example of a discrete Bayesian network and conditional probability tables.

From this joint distribution and using Bayes' rule, one can compute a range of marginal and conditional distributions as a function of the factors in Equation 2.2, which are stored in the CPTs. For instance, by applying the total probability law and given that the sprinkler is activated ($Sprinkler = T$), the probability that the grass is wet is:

$$\begin{aligned} P(wet\ grass = T|sprinkler = T) &= P(wet\ grass = T|sprinkler = T, rain = T)P(rain = T) \\ &\quad + P(wet\ grass = T|sprinkler = T, rain = F)P(rain = F) \\ &= 0.9 \times 0.8 + 0.99 \times 0.2 \\ &= 0.918 \end{aligned}$$

Additionally, Bayesian networks do not only allow top-down (prior-to-posterior) inference; in fact, as illustrated in Figure 2.2, inference can be performed posterior-to-prior (Cowell, 1999), for instance to determine the probability of occurrence of a disease given the observation of certain symptoms.

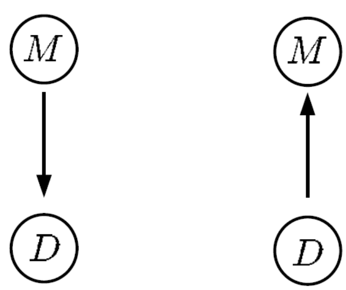


Figure 2.2: Prior-to-posterior (left) and posterior-to-prior (right) Bayesian inference (Cowell, 1999).

2.1.2 Non-Parametric Bayesian Networks

The previous sub-section primarily addressed discrete BNs, which have limited real-life applications. For complex systems, the use of CPTs quickly becomes intractable due to the exponential increase in their size with the number of possible states and parents. In contrast, Non-Parametric Bayesian Networks (NPBNs) approach dependence from another angle: dependence between variables is associated with (conditional) rank correlations, whose values depend on the non-unique ordering of each variable's parents, and bivariate copulas. While copula-based models were explored in Clemen, Fischer, and Winkler (2000), Non-Parametric BNs were first formalized in Kurowicka and Cooke (2005) to expand the spectrum of applications of discrete BNs. The following paragraphs introduce theoretical foundations on NPBNs and related mathematical concepts.

Copulas were introduced in Sklar (1959) and the associated Sklar's theorem, which states that:

Theorem 1. *Given a joint cumulative distribution function (CDF) $F(x_1, ..., x_n)$ for random variables $X_1, ..., X_n$ with marginal CDFs $F_1(x_1), ..., F_n(x_n)$, F can be written as a function of its marginals:*

$$F(x_1, ..., x_n) = C_\theta(F_1(x_1), ..., F_n(x_n)),$$

where $C_\theta(u_1, ..., u_n)$ is a joint distribution function with uniform marginals. Moreover, if each F_i is continuous, then C_θ is unique, and if each F_i is discrete, then C_θ is unique on $Ran(F_1) \times ... \times Ran(F_n)$, where $Ran(F_i)$ is the range of F_i. C_θ is called copula with parameters θ.

While the most prominent measure of dependence in copulas is Pearson's product moment correlation (ρ), NPBNs instead assess bivariate dependence using Spearman's rank correlation (r). Let X and Y be two random variables with finite expectations $E(X)$, $E(Y)$, finite standard deviations σ_X, σ_Y and cumulative distributions functions F_X, F_Y. Then, their product moment correlation and rank correlation are:

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}, \tag{2.3}$$

$$r(X, Y) = \frac{E(F_X F_Y) - E(F_X)E(F_Y)}{\sigma(F_X)\sigma(F_Y)} = \rho(F_X, F_Y). \tag{2.4}$$

When unambiguous, the notations $\rho_{X,Y}$ (for $\rho(X, Y)$) and $r_{X,Y}$ (for $r(X, Y)$) are used in the remainder of this research. Likewise, conditional rank correlations $r(X_i, X_j|X_k, ..., X_z)$ are noted $r_{X_i,X_j|X_k,...,X_z}$ when possible. In contrast with product moment correlations ρ which assess linear dependence between two variables, rank correlations provide a more general measure of monotonic dependence, independent of the marginal distributions (A. M. Hanea, Morales-Nápoles, & Ababei, 2015; Morales-Nápoles, Kurowicka, & Roelen, 2008) - hence their designation as non-parametric.

In NPBNs, each edge is associated to a (conditional) rank correlation and a one-parameter copula; for each term i with parents $\{i_1, ..., i_k\}$, the rank correlation associated with the edge $i_{k-j} \to i$ is:

$$\begin{cases} r(i, i_k) & j = 0, \\ r(i, i_{k-j}|i_k, ..., i_{k-j+1}) & 1 \le j \le k-1. \end{cases} \tag{2.5}$$

The assignment is vacuous if $pa(x_i) = \varnothing$. Then, NPBNs' main result, demonstrated in A. M. Hanea, Kurowicka, and Cooke (2006), states that:

Theorem 2. *Given the following conditions, the joint distribution of the n variables of a network is uniquely determined:*

1. *A directed acyclic graph (DAG) with n nodes specifying conditional independence relationships in a BBN;*
2. *n variables, assigned to the nodes, with continuous invertible distribution functions;*
3. *The specification 2.5, i = 1,...,n, of conditional rank correlations on the arcs of the BBN;*
4. *A copula realizing all correlations [-1, 1] for which correlation 0 entails independence.*

and the conditional rank correlations 2.5 are algebraically independent.

Using the same example as in the previous sub-section, Figure 2.3 illustrates the NPBN formulation and the (non-unique) assignment of rank correlations to each edge. For instance, $r_{R,WG|S}$ is the conditional rank correlation between R (*Rain*) and WG (*Wet grass*) given S (*Sprinkler*). An advantage of Non-Parametric Bayesian networks is their flexibility; the copula formulation is independent of the marginal distributions, which can be changed to the liking of the user/researcher. Moreover, the addition of new variables only

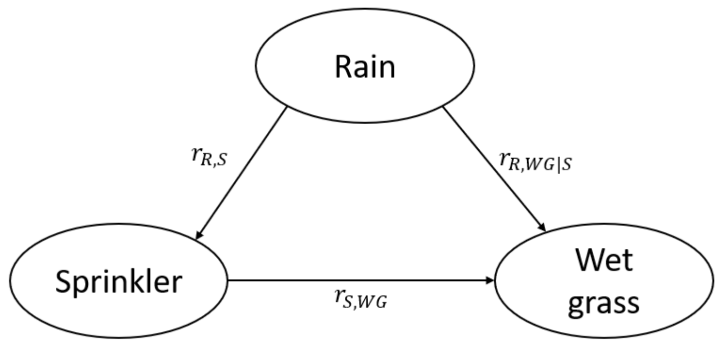


Figure 2.3: Example of a Non-Parametric Bayesian Network.

involves the elicitation of the (conditional) rank correlations assigned to the new edges rather than a complete review of the CPTs. For a thorough presentation of NPBNs, the reader is invited to consult A. M. Hanea (2008) and A. M. Hanea et al. (2015).

Among the copulas with the zero-independence property, the bivariate Gaussian copula offers several advantages and is thus often selected. It is defined as:

$$C_\rho(u_1, u_2) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \tag{2.6}$$

where Φ_ρ is the bivariate standard normal cumulative distribution function (CDF) with product moment correlation ρ and Φ^{-1} the inverse univariate standard normal CDF. The normal copula allows significantly faster sampling of joint distributions (A. M. Hanea et al., 2015; Morales-Nápoles & Steenbergen, 2015) because of one of its intrinsic properties: for multivariate normal distributions, all conditional distributions are also normal. Additionally, closed-form relations between different measures of dependence exist under this assumption: Pearson's correlation ρ, Spearman's rank correlation r and Kendall's tau τ. Such relations are particularly pertinent when attempting to compute rank correlations from other statistical quantities, such as probabilities of concordance - see subsection 2.2.1. Therefore, all (conditional) copulas in this research were considered normal.

Finally, a support for the implementation of the network was chosen. Despite the wide range of software for the implementation of Bayesian networks (Netica, Bayesia ...), only a few are suitable for Non-Parametric BNs. We opted for a combination of UniNet Academic[1] and pyBanshee (Koot et al., 2023), a Python-based open-source implementation of the MATLAB toolbox BANSHEE (Mendoza-Lugo & Morales-Nápoles, 2023). Whereas the former's agreeable Graphical User Interface was useful when interacting with external stakeholders, pyBanshee and Python offer more flexibility and power when performing analyses, as highlighted by its existing applications (Paprotny et al., 2021; Mendoza-Lugo, Morales-Nápoles, & Delgado-Hernández, 2022).

2.1.3 Barriers

The first contact with Bayesian networks and the related literature can be daunting and discourage the one without supervision or support. In addition to the different declinations, of which the reader already knows at least two, there seems to be two levels of detail: papers which describe the application of BNs to their field with little detail on the process of building the network, and papers discussing algorithmic/optimization which are of little use for the beginner. This trend is discussed in Kabir and Papadopoulos (2019) where the authors underline that there are no formal semantic guidelines for developing BNs, increasing the volatility in models' efficiency.

Similarly to most probabilistic models, a particularly sensitive aspect of Bayesian Networks is the collection of relevant data to quantify the model. Quantifying joint distributions (either through CPTs or marginal distributions and correlations) heavily relies on data, of which empirical data and experts' judgments are the two main sources, both presenting significant challenges. First, historical data is scarce, scattered, and

[1]Software in closed-access, see TU Delft's related webpage.

of extremely varying quality (Hänninen, 2014; Kyrimi, Dube, et al., 2021; G. Zhang & Thai, 2016), making the data gathering and processing tedious. Despite the implementation of condition assessment standards in several countries, this problem remains relevant in the building industry in which numerous asset owners store inspection data in a range of different databases and rating systems.

Likewise, selecting and eliciting experts' judgments is a complex task. First, one must gather individuals with extensive knowledge of the system under study (Kabir & Papadopoulos, 2019; Kaikkonen, Parviainen, Rahikainen, Uusitalo, & Lehikoinen, 2021), an increasingly challenging task as experts are decreasingly approachable. Then, the selection of an appropriate elicitation method must be made. Methods based on consensus, such as the Delphi method, bear an important behavioural bias (G. Zhang & Thai, 2016), inducing researchers to develop new mathematical aggregation methods (Hänninen, 2014) - for instance Cooke's method (Rongen, Morales-Nápoles, & Kok, 2022). Finally, the formulation of sensible questions to collect information is challenging given that probabilistic reasoning is all but intuitive, particularly for events of rare occurrence (Uusitalo, 2007) such as catastrophic building defects. Although limiting the number of parents per node is recommended to obtain reliable information from experts (Uusitalo, 2007; G. Zhang & Thai, 2016), it constrains the possibilities of creating a complex network. All in all, the overwhelming predominance of data-driven models over knowledge-driven ones in the medical field highlights the difficulties associated with experts' judgments collection (Kyrimi, McLachlan, et al., 2021).

For real life applications, which often involve continuous or timestamped data, Bayesian networks can be implemented only under certain (limiting) assumptions. When using discrete BNs, continuous variables must be discretized (i.e. decomposed in a finite quantity of bins), thus leading to a loss of information (Kyrimi, Dube, et al., 2021; Uusitalo, 2007; Hu, Xiong, Zhang, & Wang, 2022; Hart & Pollino, 2009). The topic of discretization is actively researched, as determining the appropriate number of intervals and the method (e.g. equal interval, equal quantile) involves several aspects such as the complexity of the network and the available computing power (Uusitalo, 2007; Nojavan A., Qian, & Stow, 2017). Moreover, integrating time in BNs is only feasible under certain assumptions, e.g. time invariance of the graph structure and the CPTs (Cui, Du, & Sun, 2023). For these reasons, applications of dynamic Bayesian Networks and continuous/hybrid BNs are still limited (Phan, Smart, Capon, Hadwen, & Sahin, 2016; Kaikkonen et al., 2021; S. H. Chen & Pollino, 2012; Kabir & Papadopoulos, 2019).

Finally, the implementation of BNs encounters resistance from industry participants themselves. In more conservative industries, such as healthcare and construction, reluctance arises from the lack of proven impact of these models on the decision making process (Kyrimi, Dube, et al., 2021). Therefore, methods to quantify the impact of these models must be developed and thoroughly applied to newly created tools (S. H. Chen & Pollino, 2012).

2.1.4 Drivers

Despite the limitations listed above, Bayesian Networks are a powerful tool which can help address several problems in a range of industries, from healthcare (Kyrimi, Dube, et al., 2021; Kyrimi, McLachlan, et al., 2021) to maritime engineering (Hänninen, 2014; G. Zhang & Thai, 2016; Phan et al., 2016). A review of the benefits identified in academia is presented below.

Their flexibility allows BNs to be extremely versatile with regards to both their scope and usage (Hart & Pollino, 2009). Because they do not necessarily have a unique output, they can be used for evidential reasoning (Pearl & Russell, 2000), i.e. to infer the state of any of its nodes, or as decision models (Kabir & Papadopoulos, 2019; Kyrimi, Dube, et al., 2021; Hänninen, 2014). In the latter case, the impact of local changes on the level of the system can be estimated by sensitivity analysis (Hänninen, 2014; Kaikkonen et al., 2021). Furthermore, both numerical and categorical variables can be modelled. Such feature is particularly relevant as the demand for complex environmental (Kaikkonen et al., 2021) and socio-economic (Hänninen, 2014; Zerrouki, Estrada-Lugo, Smadi, & Patelli, 2019; S. H. Chen & Pollino, 2012) models rises.

Moreover, applications of BNs to continuous and/or dynamic problems are sharply increasing. Dynamic BNs are the subject of extensive research, and great progress has been made with regards to modelling, learning and inference techniques (Shiguihara, Lopes, & Mauricio, 2021), enabling the scaling of deterioration models on the level of the building (Morato, Papakonstantinou, Andriotis, Nielsen, & Rigo, 2022). In certain fields, such as the chemical and process industries, research on dynamic BNs has even outweighed the one on static BNs (Zerrouki et al., 2019). The development of Non-Parametric Bayesian Networks for continuous and hybrid BNs (A. M. Hanea et al., 2006) and their application to civil engineering problems has eased the burden of modelling continuous variables, shifting scholars' attention on experts' judgments elicitation and other quantification methods.

In certain fields, data is increasingly available, changing the role of experts in the modelling process. Several algorithms, e.g. K2 and greedy-hill (Doguc & Ramirez-Marquez, 2009; Scutari, Graafland, & Gutiérrez, 2019; Kitson, Constantinou, Guo, Liu, & Chobtham, 2023), make use of empirical data to create a (provisional) network; experts are therefore not consulted to build a graph from scratch, but rather to adjust the algorithms' output to reach higher levels of accuracy and relevance. The continuous progress made in developing new algorithms and enhancing the existing ones provides a spectrum of methods, each more efficient in a given context (see Scutari et al., 2019).

Lastly, BNs are praised for their capacity to handle missing data (Uusitalo, 2007). As underlined in the Introduction, condition assessment data is often scattered between different stakeholders, and there is little consistency between assets due to varying equipment accessibility. Introducing BNs in this context could therefore allow to combine inspection data with other sources (Kaikkonen et al., 2021; Phan et al., 2016); in environmental engineering, for instance, BNs allow to model systems that cannot be observed holistically but for which the state of some elements can be inferred using existing knowledge (Uusitalo et al., 2018).

2.2 Experts' judgments

The previous section underlined the difficulty of quantifying Bayesian networks, particularly when empirical data is scarce. As presented in the introduction to Non-Parametric BNs (subsection 2.1.2), information on marginal distributions and (conditional) rank correlations were needed to complete the construction of the network. Because of the limited availability of inspection data for mechanical, electrical and plumbing (MEP) systems, that information was retrieved by consulting field experts. This section thus discusses state-of-the-art methods for the elicitation of experts' judgments for the assessment of **dependence** and presents the approach selected in this study.

2.2.1 Individual assessments

Assessing correlations between two variables can prove challenging for whomever is unfamiliar with statistical concepts. To that end, *indirect* elicitation methods (in opposition to *direct* ones) were developed to overcome the possible lack of confidence or intuition of experts when faced with probability reasoning (Renooij, 2001). Their application, however, is subjected to a number of biases which can outweigh their ease of use (Renooij, 2001) and result in unrealistic assessments. In contrast, some direct methods have shown great accuracy; although there is no consensus on an overall best method (G. Zhang & Thai, 2016; Werner, Bedford, Cooke, Hanea, & Morales-Nápoles, 2017), experimental results showed that directly asking experts for correlations often provides excellent results, which can be further improved by training interviewees (e.g. illustrating correlations using scatter plots) (Clemen et al., 2000).

Direct methods can themselves take various forms, but are commonly classified in three approaches: (i) statistical approaches, (ii) conditional fractile estimates and (iii) probabilities of concordance (Clemen & Reilly, 1999; Morales-Nápoles et al., 2008; Werner et al., 2017). In the first, experts directly provide rank correlations estimates or related quantities such as ratios of rank correlations (Morales-Nápoles, Hanea, & Worm, 2014; Morales-Nápoles, Delgado-Hernández, De-León-Escobedo, & Arteaga-Arcos, 2014). In the second, experts provide conditional probabilities of exceedance, answering questions such as: *"Suppose that variable X was observed above its q^{th} quantile, what is the probability that Y will also be observed above its q^{th} quantile?"*. From the results, the assessor can compute the associated (conditional) rank correlations, as described in Morales-Nápoles et al. (2008). Despite the popularity of this approach for quantifying NPBNs (Morales-Nápoles et al., 2008; Morales-Nápoles, Hanea, & Worm, 2014; D. Hanea, Jagtman, & Ale, 2012), computing rank correlations from exceedance probabilities requires knowledge of the marginal distributions and is suitable when working with continuous variables only, criteria that may restrain the research. As a result, the applicability of probabilities of concordance was investigated. A probability of concordance (P_c) is defined as follows: given a bivariate population (X, Y), two independent realizations (x_A, y_A) and (x_B, y_B) are considered. Then:

$$P_c = P((x_A - x_B)(y_A - y_B) > 0) = P(x_A \leq x_B | y_A \leq y_B).$$

To that date, no study has implemented probabilities of concordance for the elicitation of rank correlations for NPBNs. While their use is inadequate for rare events (Clemen & Reilly, 1999), it is highly relevant for problems that involve physically intelligible variables. For instance, take X the variable representing the weight of Dutch males between 18 and 50 years of age, and Y representing the height of the same population. $P_c(X, Y)$ is then obtained by answering the following question:

> "Two individuals A and B are randomly selected among Dutch males between 18 and 50 years old. Given that B is taller than A ($y_A \leq y_B$), what is the probability that B weighs more than A ($x_A \leq x_B$) ?"

which given little knowledge of the Dutch population and probability can be answered with relative ease. If a respondent believes that X and Y are completely positively (resp. negatively) correlated, then they should provide a value of $P_c = 1$ (resp. $P_c = 0$), while $P_c = 0.5$ indicates independence.

As outlined in 2.1.2, relations exist to retrieve rank correlations from P_c. First, P_c is linearly related to Kendall's tau τ (Clemen et al., 2000; Derumigny & Fermanian, 2019):

$$\tau = 2P_c - 1. \tag{2.7}$$

Given the normal copula assumption formulated earlier, closed-form relations exist between Kendall's τ, Pearson's ρ and Spearman's r (Fang, Fang, & Kotz, 2002; A. M. Hanea et al., 2015):

$$\rho = sin\left(\frac{\pi\tau}{2}\right), \tag{2.8}$$

$$r = \frac{6}{\pi} arcsin\left(\frac{\rho}{2}\right). \tag{2.9}$$

Figure 2.4 illustrates the non-linear relationship between P_c and r under the normal copula assumption. Interestingly, the function's derivative takes low values around $r = 0$, and conversely high values around the bounds of the definition domain. In practice, this observation indicates that rank correlations are more sensitive to variations, and thus errors, around $P_c = 0.5$. Therefore, experts' ability to nuance their probabilistic assessments is crucial in retrieving realistic correlation coefficients to populate the BN. To compare the relation in Figure 2.4 with the one between rank correlation and conditional probability of exceedance, Appendix B presents both function next to each other. Evidently, there is a similarity in the curves' shapes, and the calculation of the mean square error between these functions ($MSE \sim 10^{-10}$) supports their resemblance.

After retrieving unconditional rank correlations, *conditional* rank correlations are computed recursively using partial correlations and the ordering of each variable's parents. Indeed, under the normal copula assumption, partial and conditional correlations are equal, the former being defined as follows (Kurowicka & Cooke, 2006): if $X_1, \ldots, X_n$ are random variables, the partial correlation of X_1, X_2 given $X_3, \ldots, X_n$ is:

$$\rho_{12;3,\ldots,n} = \frac{\rho_{12;4,\ldots,n} - \rho_{13;4,\ldots,n}\rho_{23;4,\ldots,n}}{\sqrt{((1 - \rho^2_{13;4,\ldots,n})(1 - \rho^2_{23;4,\ldots,n}))}}. \tag{2.10}$$

While the process of computing the conditional rank correlations could be performed manually, it is already implemented in a software: *Matlatzinca*[2]. In addition to automating these operations, *Matlatzinca* indicates for each edge the range of mathematically acceptable unconditional rank correlations[3] which verify that all conditional correlations in Equation 2.10 are in the interval $[-1, 1]$. These elements are illustrated using the previous example of Rain, Sprinkler and Wet grass in Figure 2.5.

All in all, the protocol implemented to retrieve individual experts' opinions can be summarized in a set of elementary steps as follows:

1. The expert assesses the probability $P_c \in [0, 1]$;

[2]Software developed by researchers at the TU Delft and accessible on GitHub.
[3]Later in this study, this notion is referred to as the 'validity' of experts' assessments.

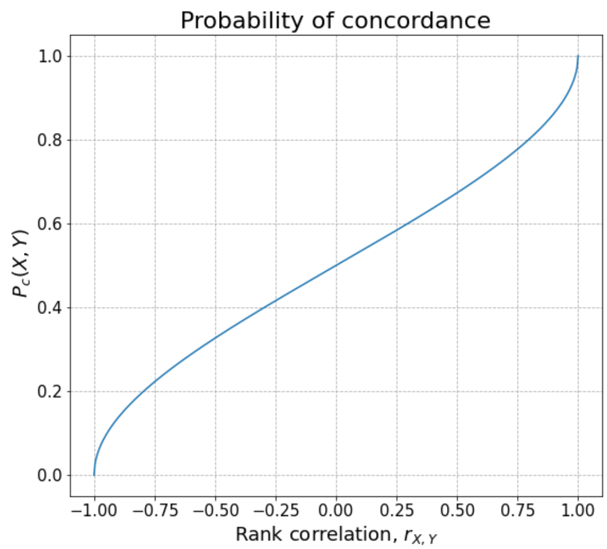


Figure 2.4: Probability of concordance as a function of the rank correlation.

2. It is converted to an unconditional rank correlation using Equation 2.7, Equation 2.8 and Equation 2.9;
3. The correlation coefficient is logged in *Matlatzinca*: if the respondent's answer is valid, move to the next question and go back to step 1;
4. Else, the expert is given the acceptable range for P_c. Because this range may be directly affected by their answers to the previous questions, the experts are also welcomed to review and modify them to their liking.

Given the nature of the questions and the topic of the research, the experts contacted must be familiar with HVAC systems and their deterioration. Forming a diverse group, both with regards to experience and positions (private/public), is believed to result in a greater accuracy of the elicited quantities (A. M. Hanea, Hemming, & Nane, 2022). Therefore, practitioners and scholars from the TU Delft as well as practitioners were welcomed to participate, regardless of their level of experience. Additionally, participants were required to have basic comprehension and expression skills in English given that the questionnaire/interviews were conducted in that language, a criterion that prove constraining for some (potential) respondents. Finally, despite the lack of consensus on the 'ideal' group size for experts elicitation, a number between four and ten shall suffice (A. M. Hanea, Hemming, & Nane, 2022).

2.2.2 Aggregation

After collecting the individual assessments, the latter must be aggregated in a unique correlation matrix. Two types of methods are found in literature: behavioral and mathematical (Clemen & Winkler, 1999). Behavioral methods attempt to reach a consensus between the experts, for instance through direct interaction (room discussion) or by integrating experts' answers throughout the interview rounds (Delphi method) (Clemen & Winkler, 1999). While behavioral methods are not discussed into details in this study, their critiques claim that agreement between experts is either impossible or leads to compromises that reflect none of the experts' opinions (Clemen & Winkler, 1999; French, 2011).

In contrast, mathematical methods attempt to overcome these behavioral biases aforementioned by combining individual assessments analytically. Most consist in weighting together the experts' judgments, but their complexity vary greatly, from arithmetic and geometric means to methods which account for experts' performance, such as the classical model (or Cooke's method, after Cooke (1991)) (Cooke & Goossens, 2008; French, 2011). In the latter, weights are defined using a set of calibration questions for which the assessor knows (or will soon know) the answer, and therefore reflect experts' informativeness and capacity to assess

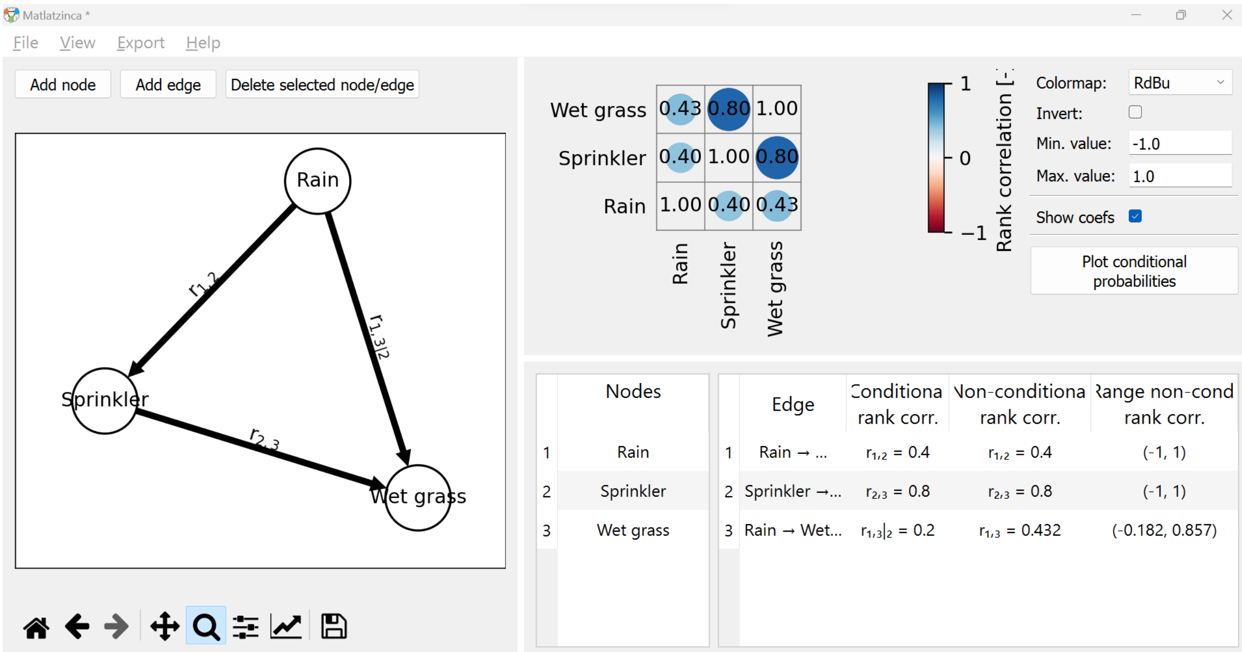


Figure 2.5: Graphical User Interface of *Matlatzinca*.

uncertainty (Cooke, 1991). Cooke's method has been applied to Bayesian modelling (and particularly to Non-Parametric BNs) in several practical problems (Morales-Nápoles, Hanea, & Worm, 2014; Morales-Nápoles, Delgado-Hernández, et al., 2014; Rongen et al., 2022; Cooke & Goossens, 2008).

Because Cooke's method is inadequate for scoring dependence assessments, another performance-based method was investigated: the dependence calibration (or d-calibration). This metric, first introduced in Morales-Nápoles, Hanea, and Worm (2014), has been applied in a handful of real-life problems (Nogal, Morales Nápoles, & O'Connor, 2019; Rongen, Morales-Nápoles, & Kok, 2023) and measures the distance between two correlation matrices: in the context of experts' judgments, we define the empirically observed correlation matrix R_m and an expert's estimation of the correlation matrix R_e. The d-calibration score $dCal_e$ is then defined as:

$$dCal_e = 1 - d_H(R_m, R_e) = 1 - \sqrt{1 - \frac{|R_m|^{\frac{1}{4}}|R_e|^{\frac{1}{4}}}{|\frac{1}{2}R_m + \frac{1}{2}R_e|^{\frac{1}{2}}}}, \tag{2.11}$$

where d_H is the Hellinger distance. The d-calibration score hence takes values between 0 and 1 (for $R_m = R_e$), with higher scores translating a statistically good estimation of the empirical correlation matrix by the expert. Implementing d-calibration, which is sometimes considered as an extension of Cooke's method for dependence assessment (Nogal et al., 2019), unfortunately carries the same burden of identifying relevant seed variables, elaborated on in chapter 4.

Chapter 3

Building condition

Building defects and technical condition are tightly intertwined. As defects arise, building components' ability to perform their intended function deteriorates, eventually leading to a change in users' perceived state of an asset. This explains the focus of condition assessment methods on identifying (emerging) defects (NEN, 2006; Faqih & Zayed, 2021b) and their remediation. However, what makes for a defect? Should degradation induced by vandalism and wear and tear be assessed equally?

The purpose of this chapter is twofold. First, it introduces important definitions and the state-of-the-art on condition assessment, thus helping the reader grasp the key concepts underlying this thesis. Then, it presents the steps towards the creation of a graph structure: identification of the variables as well as the links between them, resulting in a 'global' graph encompassing HVAC, electrical and plumbing systems.

3.1 Background

3.1.1 Definitions

Defect, failure, fault, snag? Similarly to the wide range of terms employed in academia, definitions of "building defect" vary. A defect may consist in a shortcoming of a component/system's ability to operate according to its intended design (Olanrewaju et al., 2010; Ahzahar, Karim, Hassan, & Eman, 2011), a shortcoming in the function, performance, statutory or user requirements of a building (Ilozor, Okoroh, Egbu, & Archicentre, 2004; Watt, 2007; Das & Chew, 2011), or simply as an "imperfection" (Ahzahar et al., 2011; Bortolini & Forcada, 2018; Yacob, Ali, & Au-Yong, 2022). Nevertheless, it generally falls under one of three categories (Sommerville & McCosh, 2006): technical, whereby poor workmanship, materials or design of an element prevent its correct operation; omission, whereby part or all of an element are forgotten; and aesthetic, when the appearance of an element is altered (Sommerville & McCosh, 2006). This thesis solely focuses on technical defects.

Similarly, faults can emerge from a variety of sources, all of which are not accounted for in every study. On the one hand, most papers account for every types of defects, which can appear in the design and construction phases as well as during the entire building's life cycle (Josephson & Hammarlund, 1999; Ahzahar et al., 2011; Forcada et al., 2012, 2013; Forcada, Macarulla, Gangolells, & Casals, 2014); on the other hand, a handful of authors solely include faults resulting from wear and tear (Olanrewaju et al., 2010; Yan, Luh, & Pattipati, 2020; Yan, Cai, Li, Zhang, & Sun, 2021). Because the framework developed in this research focuses on the latter, it ought to be defined. The Cambridge Dictionary defines wear and tear as : "the damage that happens to an object in ordinary use during a period". The term "ageing" is sometimes employed, for instance in the ISO 15686-2 norm where ageing is associated to the "degradation due to long-term influence of agents related to use" (ISO, 2012). Finally, Yan et al. (2021) argue that gradual defects "cause gradual performance degradation of components or systems", a definition we choose to adopt for the remainder of the research. It is important to note that by focusing on gradual defects, breakdowns are not disregarded; instead, the study (and therefore the model) only investigates those breakdowns that are caused by the progressive deterioration of a (set of) system(s).

3.1.2 Building condition assessment

Building Condition Assessment (BCA) is paramount in asset management to properly allocate maintenance resources. Sometimes referred to as "performance evaluation" (ISO, 2014), it consists in technical inspections conducted by a competent assessor who reports the occurrence of defects, thus defining the needs for maintenance (Dejaco, Re Cecconi, & Maltese, 2017). However, the scope of BCA protocols highly varies across regions and sectors. Whereas some solely account for the technical condition of building components, others base their assessment on risks for occupants' hygiene or availability (Faqih & Zayed, 2021a).

Additionally, the rating systems can differ, Faqih and Zayed (2021a) and Lupășteanu, Lupășteanu, and Chingălată (2022) both offering reviews of practices worldwide. The standard in the Netherlands, NEN 2767, provides a method to grade technical building defects based on three criteria:

- **Criticality** reflects the impact of the defect on the component's function. Values range from 1 (minor defect) to 3 (severe defect);
- **Intensity** positions the current state on the scale of the degradation process. Values range from 1 (initial stage) to 3 (final stage), and can be determined using standardized lists of defects (e.g. NEN 2767-2);
- **Extent** measures the proportion of the component affected by the defect. Values range from 1 (<2%) to 5 ($\geq$ 70%) (Straub, 2002; NEN, 2006).

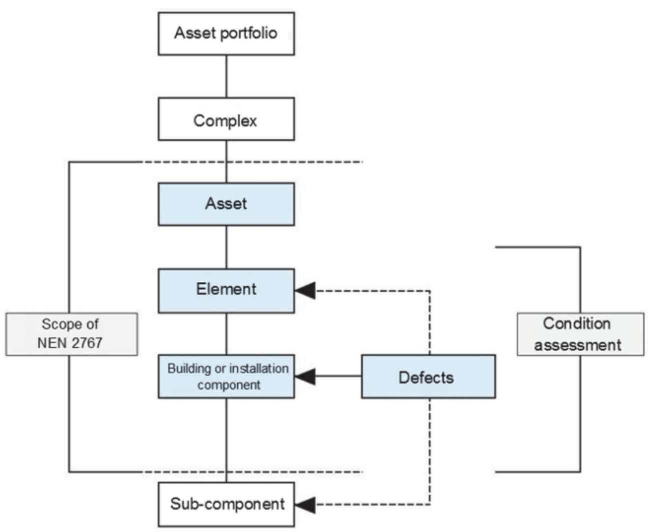


Figure 3.1: Building breakdown structure of NEN 2767 (NEN, 2006).

As illustrated in Figure 3.1, buildings are broken down in a set of components whose condition is evaluated with the previous criteria. Eventually, inspectors obtain a condition score from 1 (excellent) to 6 (very bad) following the protocol presented in Appendix A from which asset managers can build a maintenance plan[1]. Most inspections performed in the context of multi-year maintenance planning (*Meerjaren onderhoudsplanning, MJOP*) are sensory inspections, whose simplicity comes with limitations. First, they do not allow assessors to trace the origins of the reported defects, increasing the likelihood of reoccurrence of similar faults in the future. Moreover, they are intrinsically subjective, and despite the progress made since the implementation of the standard, their outcome remains sensitive to the inspector's experience and training (Faqih & Zayed, 2021a; Silva & de Brito, 2019). Lastly, some installations cannot be reached or observed without additional equipment (Lupășteanu et al., 2022): mechanical systems, such as HVAC, are commonly 'hidden' in suspended ceilings. Consequently, they are often left apart during BCA despite their key contribution to buildings' overall performance (Eleftheriadis & Hamdy, 2017) and the high costs their repairs incur (Islam et al., 2019).

Contrary to the trend observed in the overall building industry, the inspection of installations is often reported in the form of checklists which fail to precisely describe the deterioration of components. This lack of information is particularly problematic when attempting to model the behaviour of MEP systems, a process significantly made easier by the availability of relevant data (cf. section 2.1).

[1] The standard only provides a supporting framework for the assessment of a building's technical condition, and translating defects into maintenance activities is not part of the standard (NEN, 2006).

As presented in this section, building condition assessment strongly relies on the decomposition of an asset in a set of sub-systems and components. Therefore, such a decomposition shall be determined for the mechanical, electrical and plumbing (MEP) systems to design a consistent network.

3.2 Classification of MEP systems

Clearly, selecting the right variables for a Bayesian network is a challenging endeavour; accuracy and complexity are at both ends of a scale, and considering too many (or too little) variables could make it dangerously tilt on one side. As a result, this section includes a review of building pathology literature to determine a relevant classification of MEP systems, building on the inspection system presented in Bortolini and Forcada (2018).

3.2.1 Plumbing & Sanitary

Despite being relatively scarce (Chew, 2005), faults in plumbing systems are perceived as critical and severe (Zalejska & Hungria, 2019; Olanrewaju, Tan, & Soh, 2021). Three types of faults in particular are frequently cited in literature: pipes corrosion, water leakage and water supply problems (Bortolini & Forcada, 2018). Firstly, pipe corrosion is an inconspicuous yet severe defect which equally affects residential (Carretero-Ayuso, Moreno-Cansado, & García-Sanz-Calcedo, 2017) and non-residential (Chew, 2005) buildings. Second, water leakages are both common (Abdul-Rahman, Wang, Wood, & Khoo, 2014) and severe (Kian, 2001; Das & Chew, 2011; Carretero-Ayuso et al., 2017; Chong & Low, 2006; Yacob et al., 2022). Their occurrence negatively impacts the condition of the surrounding components and is hence highly undesirable. Third, water supply problems' - which encompass issues related to water temperature and pressure among others - impact on occupants' experience and the overall operation of the plumbing system is less thoroughly discussed in academia. Nonetheless, data collected by Chong and Low (2006) suggests that choked pipes and malfunctions of systems/accessories are somewhat frequent for different asset classes.

Because we are interested in classifying components (or sub-systems) rather than defects, the latter must be transformed accordingly. First, pipe corrosion and leakage both affect the water distribution elements and can thus be grouped under one umbrella variable: *Plumbing distribution elements*. Water supply problems, on the other hand, simply become *Plumbing supply elements*.

3.2.2 Heating, Ventilation and Air Conditioning

Among the building systems within the scope of this study, HVAC systems are undoubtedly the most researched. This is not innocuous: in addition to being central to occupants' indoor comfort (Alavi, Forcada, Bortolini, & Edwards, 2021; Hua, Göçer, & Göçer, 2014; Hosamo et al., 2023), they are incredibly complex and diverse (Taal & Itard, 2020b, 2022). This complexity hinders the creation of a classification that would include all HVAC defects, explaining the variety of research on fault detection and diagnosis models (Y. Chen, Wen, Chen, & Pradhan, 2018; Cheung & Braun, 2015; Hosamo, Svennevig, Svidt, Han, & Nielsen, 2022; Mirnaghi & Haghighat, 2020; Yu, Woradechjumroen, & Yu, 2014).

Bortolini and Forcada (2018) distinguish malfunctions of HVAC *production* elements (chiller/boiler malfunction, fan motor failure...) and *fixture* elements (thermostat malfunction, excessive vibration of air unit...). Similarly, the 4S3F framework (4 symptoms, 3 faults) developed in Taal and Itard (2020b) for fault detection and diagnosis categorizes faults as either component, control or model faults. Control faults, on the one hand, can be errors in controllers and actuators and hence correspond to fixture elements faults; component faults, on the other hand, relate to other production and distribution elements (Taal & Itard, 2020b).

Despite the limited evidence used to question the findings of Bortolini and Forcada (2018), a distinction between HVAC *production* and *fixture* elements is adopted.

3.2.3 Electrical installations

Although surveys suggest that defects related to electrical installations are less frequent than the ones aforementioned (Olanrewaju et al., 2021; Abdul-Rahman et al., 2014; Georgiou, Love, & Smith, 1999), owners and facility managers consider them critical (Bortolini & Forcada, 2018; Carretero-Ayuso et al., 2017). First, the good condition of lighting and other electrical fittings is paramount to provide a high visual comfort (Faqih, Zayed, & Soliman, 2020), thus making their dysfunction problematic (Das & Chew, 2011). Note, however, that the importance attached to visual comfort - and therefore lighting - widely varies across studies; in contrast with Das and Chew (2011), defective lighting ranks among the least critical defect in a study performed by Islam et al. (2019). Second, the faulty operation of distribution elements (e.g. cabling) can potentially trigger the occurrence of catastrophic events. For instance, exposed cabling considerably increases the likelihood of overheating, which can eventually burn the cables (Hassanain, Fatayer, & Al-Hammad, 2016) and create stress among occupants (Olanrewaju et al., 2010), thus making safe distribution elements essential to the overall building's performance (Abdul-Rahman et al., 2014).

After discussing the findings of Bortolini and Forcada (2018), adapting them to the present research and examining them based on literature, the resulting classification of MEP systems is presented in Table 3.1, and includes all seven sub-systems as well as the components part of each cluster.

Sub-system	**Components**	**References**
	Plumbing & Sanitary	
Water distribution elements	Pipes, joints, pipe penetrations	(Kian, 2001; Chew, 2005; Chong & Low, 2006; Das & Chew, 2011; Carretero-Ayuso et al., 2017; Bortolini & Forcada, 2018; Yacob et al., 2022)
Water supply elements	Valves, pumps	(Chong & Low, 2006; Bortolini & Forcada, 2018)
	Heating, Ventilation and Air Conditioning	
Fixture elements	Thermostat, heater, air extraction	(Bortolini & Forcada, 2018; Taal & Itard, 2020b)
Production elements	Chiller, boiler, fan motor	(Bortolini & Forcada, 2018; Taal & Itard, 2020a)
	Electrical systems	
Fixture elements	Lighting, switches, sockets	(Das & Chew, 2011; Bortolini & Forcada, 2018; Faqih et al., 2020)
Distribution elements	Cabling, meter	(Lai, 1993; Hassanain et al., 2016; Bortolini & Forcada, 2018; Olanrewaju et al., 2021)
Supply elements	Transformer, motor control centers	

Table 3.1: Overview of selected MEP sub-systems with supporting references.

3.3 Deterioration and influences

Whereas literature on building defects and their classification fairly extensive, research on their causes is relatively scarce. This phenomenon finds a simple explanation: after notification of a defect, an expert is mobilized to identify the defective component and, hypothetically, restore it to a state where it can perform its function (Straub, 2012). However, the cause is often difficult to apprehend (Josephson & Hammarlund, 1999) and can involve a range of environmental and human factors (Paton-Cole & Aibinu, 2021). Fortunately, some aspects of this topic have been covered in academia, and factors influencing the deterioration of MEP systems can be clustered into Environmental, Design & Construction (D&C) and Maintenance.

The influence of indoor and outdoor conditions on a building is well documented (Watt, 2007; Carretero-Ayuso, Rodríguez-Jiménez, Bienvenido-Huertas, & Moyano, 2021). Weather conditions (e.g. outdoor temperature, humidity) directly affect the usage of the mechanical systems (Zhao & Magoulès, 2012; Biswas, Robinson, & Fumo, 2016), which itself correlates with the efficiency of the said installations. The runtime of HVAC systems has proved to significantly impact their deterioration (De Silva, Setunge, & Tran, 2022): their efficiency allegedly decreases by 20% over 20 years on average (Eleftheriadis & Hamdy, 2017), although most authors simply acknowledge that building components deteriorate over time (Olubodun & Mole, 1999; Waddicor et al., 2016; Sui Pheng & Wee, 2001). Nevertheless, most studies were conducted on the asset level (Ahzahar et al., 2011; Olanrewaju et al., 2021) and therefore failed to identify causal schemes between climatic conditions and deterioration on the level of the component/sub-system. Additionally, observed correlations between climate-related variables and the condition of building components is not sufficient to draw conclusions on the existence of causality between them.

The design and construction (or installation) have an enormous influence on the condition of buildings from the handover (Forcada et al., 2012, 2013) and throughout the operation phases (Carretero-Ayuso et al., 2021; Hauashdh, Jailani, Abdul Rahman, & Al-Fadhali, 2022). Design errors, such as inadequate working drawing details (Faqih et al., 2020) and inappropriate design for maintainability (Watt, 2007; Ilozor et al., 2004; Asmone & Chew, 2020; Islam et al., 2021), tend to snowball in the execution phase as design requirements do not account for standard methods of construction (Watt, 2007), resulting in poor construction quality. Additionally, the selection of inappropriate construction materials can highlight other deficiencies and increase stress on the installations (Chew, 2005; Chong & Low, 2006; Ahzahar et al., 2011; Olanrewaju et al., 2021; Watt, 2007; Pan & Thomas, 2015). Errors committed on site equally impact the quality of the building and can be of different kinds, the most frequently cited being poor site management (Josephson & Hammarlund, 1999; Alencastro, Fuertes, & de Wilde, 2018) and poor workmanship (Josephson & Hammarlund, 1999; Ahzahar et al., 2011; Pan & Thomas, 2015; Ilozor et al., 2004; Olanrewaju et al., 2021; Chew, 2005; Faqih et al., 2020; Forcada et al., 2013). Clearly then, the pace at which building components deteriorate is strongly impacted by their *basic quality* (NEN, 2006; Ishak, Chohan, & Ramly, 2007).

Maintenance is paramount in slowing the deterioration of building elements. Several sources identified a clear relation between the implementation of an appropriate maintenance strategy and a reduction of building defects (Lai, 1993; Watt, 2007; Pan & Thomas, 2015; Ahzahar et al., 2011; Bortolini & Forcada, 2020; Sui Pheng & Wee, 2001; Waddicor et al., 2016). Nevertheless, studies suggest that budgets allocated to maintenance decline, entailing an increasing number of deteriorated buildings (Faqih et al., 2020). MEP systems are particularly affected by the industry's reluctance to adopt preventive maintenance (Bortolini & Forcada, 2018; Weeks & Leite, 2021), thus significantly increasing facility management costs (Islam et al., 2019; Weeks & Leite, 2021) and perceived discomfort (Bortolini & Forcada, 2019; Hosamo et al., 2023). Cuts in projects funds also translate in poor design-for-maintainability (Asmone & Chew, 2020), eventually complicating the operation of installations (Islam et al., 2019) and properties overall (Asmone & Chew, 2020).

Lastly, building components are interdependent insofar as the failure of one can affect the condition of another. Such dependencies, either spatial or operational, are for instance discussed in Bortolini and Forcada (2018) and Atef and Bristow (2019), and are modelled in Bortolini and Forcada (2020). Contrary to identifying spatial relationships, which given the availability of a BIM model is relatively straightforward, determining operational dependencies requires knowledge of the system of interest. For that reason, the existence of relationships (modelled by edges in Bayesian networks) can be assumed and reflected on using empirical data and/or experts' judgments.

3.4 'Global' network

As a result of the classification of all MEP sub-systems (3.2) and the identification of the factors influencing their deterioration (3.3), a graph encompassing all MEP systems emerges. It is presented in Figure 3.2, and additional hypotheses were made to reach this outcome.

First, variables such as Age, Design & Construction (D&C) quality and Maintenance are differentiated by cluster. In practice, however, it is impossible to define (for instance) an age common to all 'HVAC production elements' as it encompasses several distinct components. The following chapters illustrate how this limitation is overcome in the implementation of the BN. Second, merely considering the age of a component/sub-system can be misleading; repairs, partial or replacement of an element affect the condition of a sub-system, such that a component 20 years of age may have been repaired/improved several times during its lifetime (Straub, 2012). It is difficult, however, to model the impact of repairs on the (future) performance of a component (Endrenyi et al., 2001; Grussing & Liu, 2014), thus explaining our decision to solely consider the age. As demonstrated later in the report, additional variables can be added even after its quantification, and the seemingly simplistic structure created in this chapter is thus not definitive.

Despite our efforts to limit the number of variables in the network, quantifying the graph in Figure 3.2 seems unfeasible given its size ($\approx$ 25 variables, 30+ edges) and the limited resources available in this thesis. Consequently, the quantification will focus on a section of the network: air handling units. This key element of HVAC production systems, presented in the following chapter, is a great support to evaluate the pertinence of the elicitation method described in chapter 2.

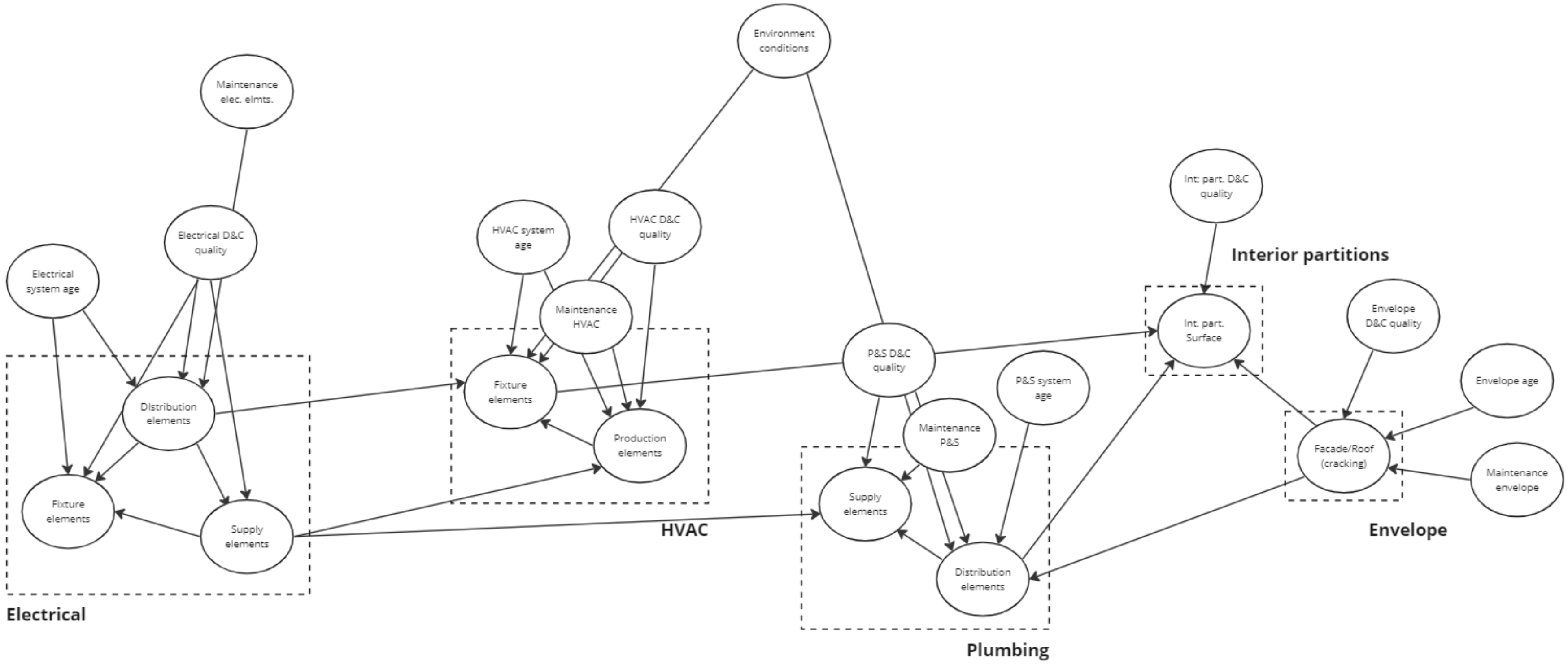


Figure 3.2: Network of all mechanical, electrical and plumbing systems.

Chapter 4

Case study: Air Handling Units

A wide array of air handling units (AHUs) is available on the market, all designed with a common purpose: maintain an acceptable indoor air quality. In single-family housing, split-systems are particularly popular due to their ease of installation. For other asset classes, however, central air handling units are commonly used and located on a building's roof. Figure 4.1 illustrates the process by which indoor air quality is preserved: outdoor air is filtered, conditioned by coils for heating or cooling, and distributed in the room(s) through ducts. Simultaneously, polluted indoor air is extracted and (partially) evacuated from the building. In this study, a system with air recirculation was selected; it is worth noting that according to one of the experts consulted for the elicitation, AHUs with thermal wheels are more common in the Netherlands.

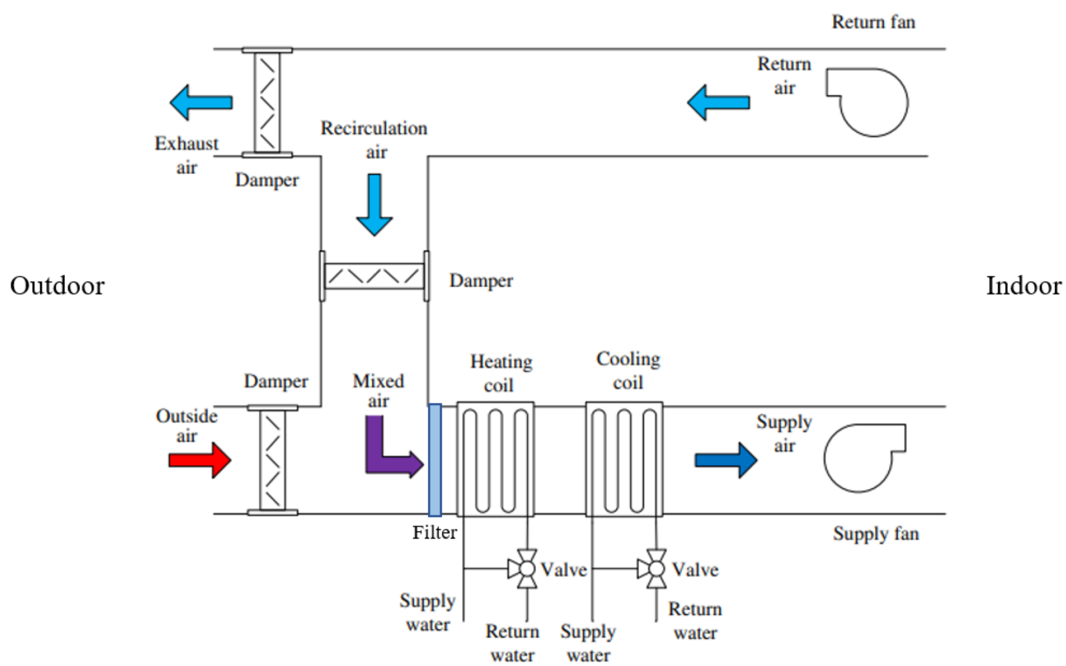


Figure 4.1: Air handling unit with air recirculation (adapted from Kusiak & Li, 2010).

As illustrated in Figure 4.2, the construction of the BN follows a process initiated with the definition of a graph structure. Due to the complexity of the 'global' graph created in the previous chapter, this chapter presents the definition of the network for air handling units (4.1) as well as the application of the experts elicitation method to this case study (4.2).

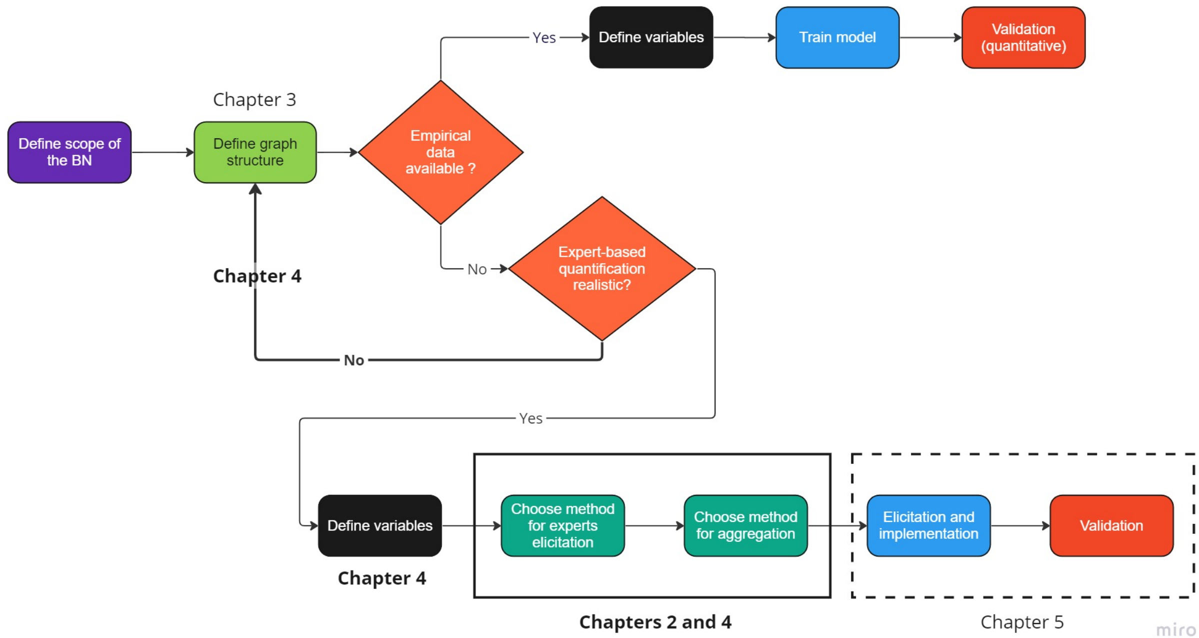


Figure 4.2: Construction process of the Non-Parametric Bayesian Network.

4.1 Graph structure

The main components of an AHU clearly stand out from Figure 4.1: supply and exhaust fans, heating and cooling coils, and the filter(s). In their study on gradual faults prediction, Yan et al. (2021) limited their effort to defects related to the supply fan and the cooling, obtaining satisfactory results. However, our research also investigated the relation between components. Consequently, we chose to include all the previously mentioned elements and group them as follows: **coils**, **fans**, and **filters**. Interestingly, as suggested by an expert, the deterioration of the heating and cooling coils is influenced by different factors: the latter, for instance, is more vulnerable to environmental perturbations (e.g. frost, pollution), thus resulting in stronger correlation. The decision to group components is knowingly oversimplistic and reflects the exploratory dimension of the research, whose focus is on the elicitation of experts' judgments rather than the creation of a complex and accurate model. For practical purposes, the condition of these components - and the associated variables - was defined in accordance with the 1-6 scale of NEN 2767, previously introduced in subsection 3.1.2.

To determine which factors affect these components' condition, we built on the results presented in section 3.3 and reassessed the relevance of each exogenous variable (excl. Age) in relation with air handling units.

While considering environmental can be crucial when examining components directly exposed to them, their impact on AHUs seems fairly even across the Netherlands. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) classifies regions based on climatic design conditions, derived from measurements that include temperature, humidity and precipitations. According to this standard, the country falls into two categories: 4A (mixed-humid, Southern half) and 5A (cool-humid), Northern half) (ANSI & ASHRAE, 2013). Because the distinction between these categories resides in differences in heating degree days, deemed insignificant for AHUs, environmental conditions were not included in this model.

The frequency of inspections and maintenance interventions has been shown to have a substantial impact on the operation of HVAC systems. Diminishing the time between inspections can allow to reduce equipment downtime and thus improve occupants' comfort and satisfaction (Au-Yong, Ali, & Ahmad, 2014). For instance, failing to replace filters regularly can result in reduced air flow or a total loss of their function. Similarly, cleaning coils is essential in preventing corrosion. Therefore, the variable 'Maintenance interval' was defined as the time period (in years) between two consecutive inspections.

Lastly, errors made during the design and construction (installation) phases are known to catalyze the deterioration of AHUs and HVAC systems as a whole. The findings of Carretero-Ayuso et al. (2017) indicate that inadequately placed joints and omissions in the ventilation systems account for a high percentage of the reported malfunctions. Additionally, several 'specialized' websites assert that poor ductwork design (incl. incorrect sizing and layout) and leaks, mostly due to joint failure, can considerably decrease the efficiency of the unit (Hoffmann, 2018; Rosone, 2023) and create stress on the fans and the coils. As a result, the variable 'Design & Construction quality' was defined as a measure of the unit's *basic quality*, including the quality of the materials, design and workmanship at the time of the installation. It takes values on a 1-5 scale defined as follows:

Very poor	**Poor**	**Medium**	**Good**	**Excellent**
1	2	3	4	5

In summary, these reflections enabled the creation of a graph that models the relationships between the AHU components and the variables above, shown in Figure 4.3. Several assumptions underlie this graph:

- Because of their comparatively short lifespan, the condition of the filters is exclusively affected by 'Maintenance interval'.
- The condition of the plumbing supply system only affects the coils as these elements are functionally interdependent: the warm or chilled water (or other fluid) from the plumbing system supplies the coils. Likewise, the electrical supply system exclusively interacts with the fans.
- Since the filters are responsible for reducing the number of particles entering the AHU, their failure allows the accumulation of particles on the coils and thus speeds up the deterioration by corrosion.
- The fans' condition can be impacted by the filters in at least two ways. First, polluted filters oblige the fans to exert more power to maintain the same perceived airflow. Secondly, particles that enter the AHU partially flow through the ducts where they accumulate, thus leading to reduced airflow and additional stress on the fans.

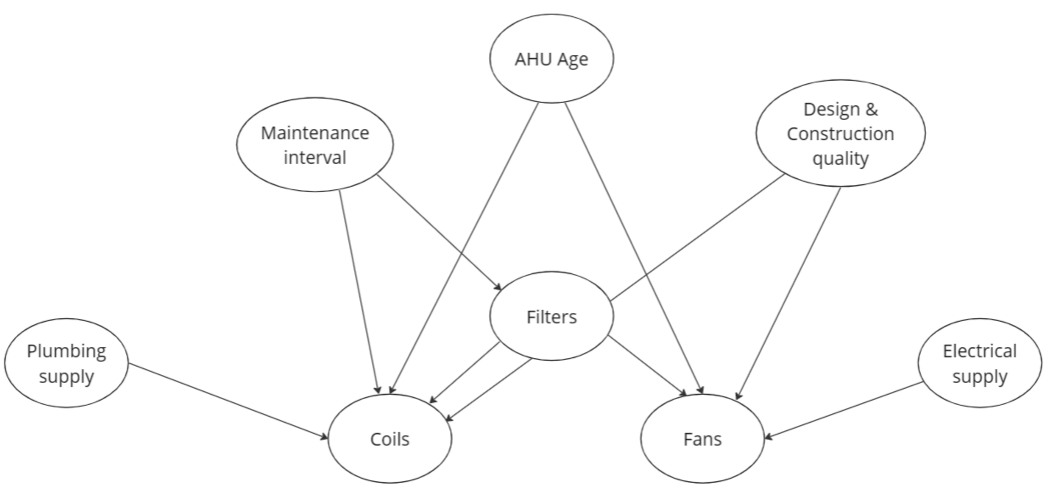


Figure 4.3: Graph structure for the air handling unit.

4.2 Quantification: experts' judgments

Section 2.2 introduced a novel framework for the elicitation of rank correlations in Non-Parametric Bayesian Networks (NPBNs). In this section, the application of these methods to the case study is presented including the list of participants, the questionnaire, and the seed variables used for dependence-calibration.

4.2.1 Individual assessments

Similarly to the case of Dutch males' weight, height and age discussed in subsection 2.2.1, the study of air handling units (and MEP systems as a whole) is based on physical quantities. Let X be the condition of the fans and Y the age of the AHU (in years). To retrieve the probability of concordance $P_c(X, Y)$, one needs the answer to the following question:

> "Two buildings A and B are randomly selected among all non-residential buildings in the Netherlands. Given that the air handling unit in building A is more recent than in building B ($x_A \leq x_B$), what is the probability that the fans are in better condition in building A than in building B ($y_A \leq y_B$) ?"

Because the graph in Figure 4.3 contains ten edges, the first and main section of the questionnaire - presented in Appendix C - included ten questions similar to the one formulated above.

The panel of participants consulted for the assessment of probabilities included five experts, whose details are laid in Appendix D. In addition to hypothetically result in more accurate assessments (A. M. Hanea, Hemming, & Nane, 2022), the diversity in experience and perspectives (practical/academic) offered an interesting opportunity to compare perceptions of the elicitation method used in this research. In the remainder of the thesis, the experts are referred as 'Expert A', 'Expert B', and so forth to ensure the unbiased interpretation of the results.

4.2.2 Aggregation

The formulation of relevant seed questions is a challenging task when empirical data is scarce or simply absent (Bolger & Rowe, 2015). As outlined previously, condition assessment data for MEP systems is not widely available. Therefore, we decided to select a seed variable familiar to the experts: precipitation[1].

Empirical data of hourly precipitation (*Dutch: uur som van de neerslag*) measured at three weather stations between the 01/01/2023 and 18/06/2023 was retrieved from the database of the Koninklijk Nederlands Meteorologisch Instituut (*Dutch Royal Institute of Meteorology - KNMI*). For reference, the location of the stations is illustrated in Figure 4.4 (left). Because of their geographical proximity, precipitations at these locations are likely correlated, an assumption supported by historical data. The (rank) correlation matrix of variables 'Gilze-Rijen', 'Rotterdam' and 'Eindhoven' was retrieved and is shown in the upcoming chapter. The second part of the questionnaire included the seed questions related to the graph in Figure 4.4 (right) and were formulated as follows:

[1] With over 100 rain days per year, rain is rooted in the Dutch culture. Source: https://www.statista.com/statistics/1012831/number-of-rain-days-in-the-netherlands/

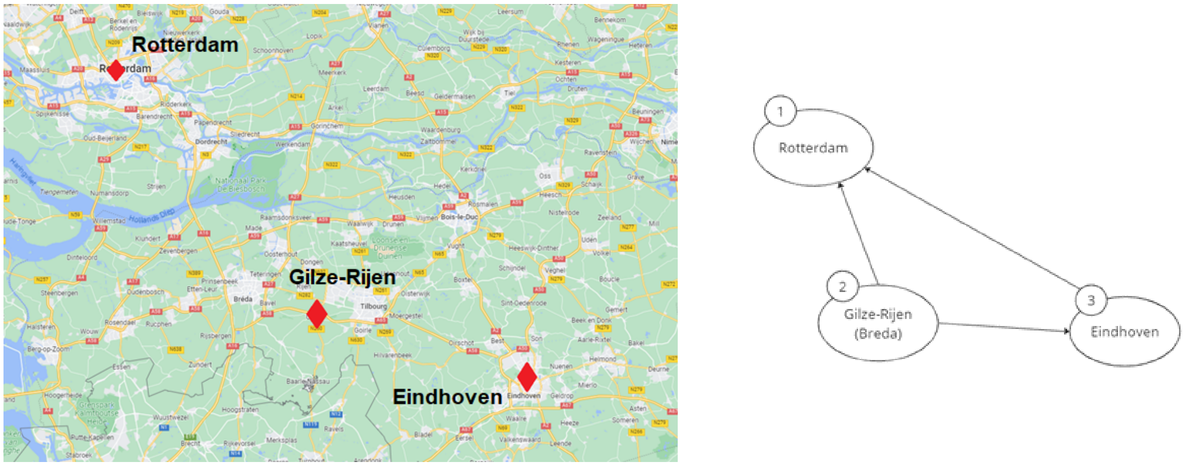


Figure 4.4: Location of the weather stations (left) and associated graph (right) used for the assessment of seed probabilities.

> "Two moments H1 and H2 (defined by the hour) are taken randomly between the 01/01/2023 and the 18/06/2023. Given that the hourly precipitation is higher at H2 than at H1 in Gilze-Rijen, what is the probability that the hourly precipitation is also higher at H2 than at H1 in Rotterdam ?"

The questions as presented to the experts are laid in Appendix C, and were answered following the same protocol as the 'main' questions, described in subsection 2.2.1. The resulting correlation matrices were then used to compute individual experts' d-calibration scores and design decision-makers, an endeavour presented in the next chapter.

4.2.3 Marginal distributions

Having dedicated extensive time and effort to the elicitation of the network's dependence structure, a less scientifically sound approach was adopted to determine the marginal distributions necessary to the completion of the model. Two of the questionnaire respondents, Marcel Klok and Frans Strik (cf. Appendix D), accepted to contribute by attempting to convert their experience into probability distributions. It is noteworthy that their assessments reflected the scope of their work, which is inextricably linked to the TU Delft and its campus. Therefore, attempts to implement the NPBN in different settings would certainly require the definition of new marginals.

Chapter 5

Results and Analysis

Following the completion of the elicitation process, we proceeded to analyze the experts' opinions retrieved from the questionnaire. The first part of this chapter presents and discusses in details the results, leading to a combination of the experts' assessments, or Decision-Maker (DM). Then, the model was subjected to a set of analytical procedures to assess its alignment with the research objectives.

5.1 Dependence structure

5.1.1 Individual assessments

Five correlation matrices were obtained based on each expert's responses to the 'main' section of the questionnaire (for numerical values, see Appendix E). As illustrated in Figure 5.1, experts A, D and E indicated the prevalence of specific relationships within the network. For instance, expert D suggested the existence of one or two main predictors of each component's condition, such as 'Age'/'Fans' ($r_{1,7} = 0.882$). However, the evaluation of high correlations raised problems during the elicitation as these experts were asked to review their responses multiple times to make them valid (cf. subsection 2.2.1). As a matter of fact, and despite understanding the mathematical concepts underlying the 'validity' of their answers, expert A claimed that the bounds limited his ability to reflect his experience numerically. Moreover, as observed in the next sub-section, the lack of nuance in some of the experts' assessments strongly penalized them in the d-calibration.

Furthermore, three out of five experts believed that the edges 'Plumbing supply elements' → 'Coils' and 'Electrical supply elements' → 'Fans' were irrelevant. These observations interrogate the pertinence of the global graph defined in Figure 3.2, these edges being the links between the AHU and the rest of the network. Nevertheless, no conclusion is drawn on whether these edges should be removed before BNs for the other MEP systems are created, as independent variables are not necessarily *conditionally* independent. This is a first limitation of the assessment of unconditional rank correlations rather than conditional ones..

In the next section, the experts' answers to the seed questions and their respective d-calibration scores are introduced.

5.1.2 Dependence-calibration

To obtain a unique set of rank correlations suitable for implementation in the Bayesian network, the individual correlation matrices presented earlier were aggregated. In this research, the d-calibration method was employed, which involved the definition of a weighted average of each expert's responses based on their performance on a predefined set of seed variables. As illustrated in Figure 5.2, the elicited results from all five experts reinforce the previous observations regarding the inclination of experts D and E to assess high correlations. As outlined in the previous sub-section, expert D's good understanding of probabilistic reasoning ($\rho_{1,2} \simeq \rho_{2,3} > \rho_{1,3}$) was penalized by his excessively large estimates. In contrast, experts B and C demonstrated their ability to provide moderate judgments, an important feature given the sensitivity of

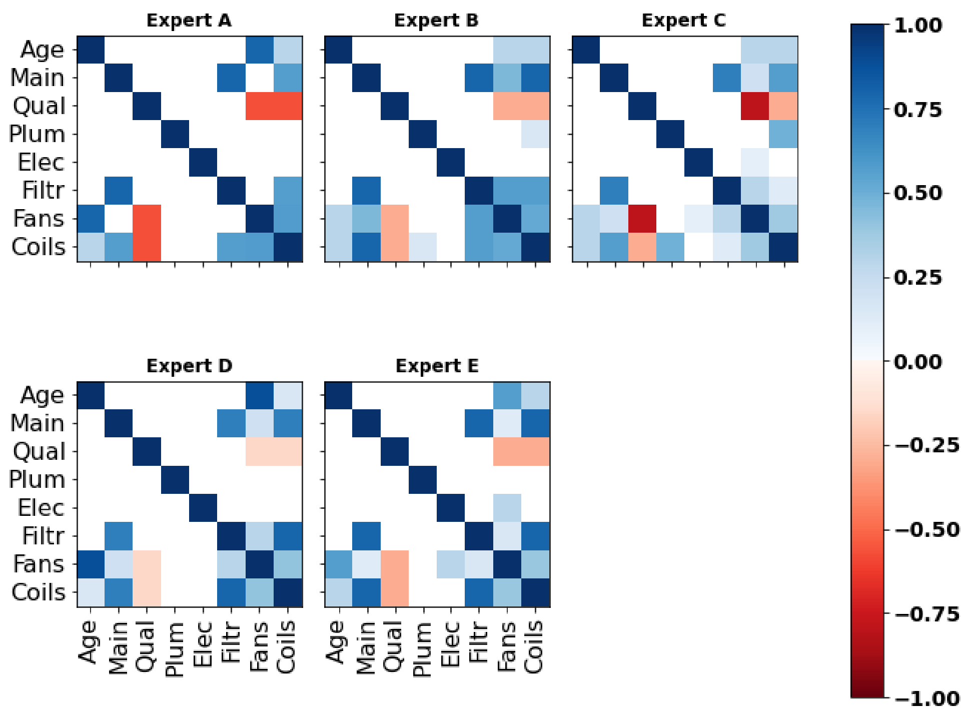


Figure 5.1: Correlation matrices retrieved from the ‘main section’ of the questionnaire.

the rank correlations for values of P_c around 0.5 (cf. subsection 2.2.1), thus resulting in higher calibration scores.

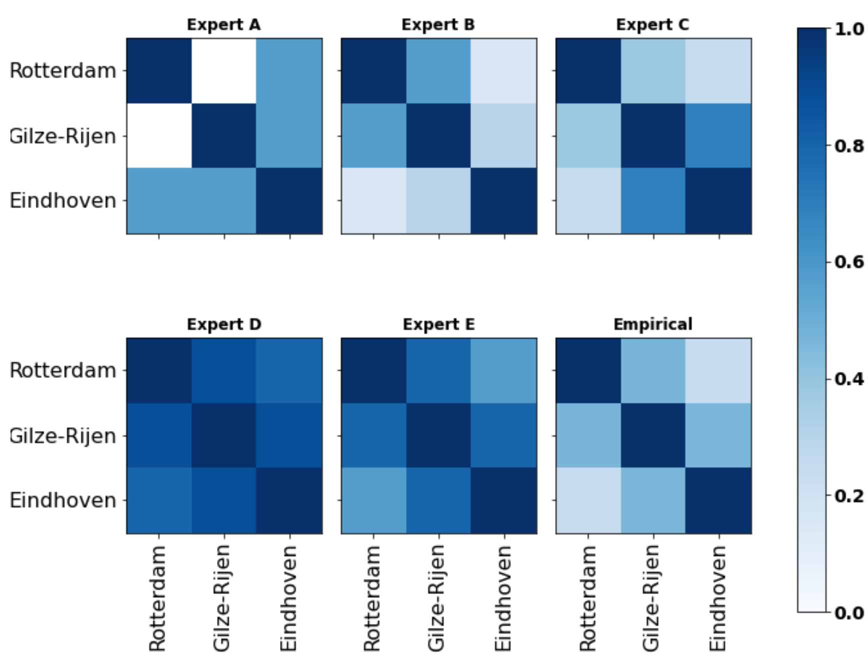


Figure 5.2: Correlation matrices from the seed questions and empirical correlation matrix.

Table 5.1 displays the d-calibration computed from the correlation matrices, where high values translate similarity between an expert's estimates and the dependence structure retrieved from empirical evidence. Clearly, two groups of experts arose: whereas experts B and C obtained excellent results, experts A, D and E performed less satisfactorily - although their scores remain acceptable. Interestingly, all experts were 'better' calibrated than in other studies using dependence-calibration (e.g. Nogal et al., 2019; Rongen et al., 2023); we write 'better' because the relatively small number of seed variables used, as well as their unrelatedness to the problem at hand, make probabilistic assessments simpler (Brooker, 2011; Nogal et al., 2019).

Decision-maker	D-calibration	Perceived comfort
Expert A	0.639	4
Expert B	0.907	4
Expert C	0.85	4
Expert D	0.516	2
Expert E	0.657	2
EWDM	0.869	-
GWDM	0.897	-
optDM	0.968	-

Table 5.1: Experts' dependence-calibration scores and perceived degree of comfort during the elicitation.

Experts were asked to evaluate the degree of comfort perceived in the assessment of probabilities. To that end, the following 1-5 Likert scale was used:

	Strongly disagree	**Disagree**	**Neither agree or disagree**	**Agree**	**Strongly agree**
I felt comfortable assessing probabilities.	1	2	3	4	5

whose answers are shown in Table 5.1. In line with the d-calibration scores, experts B and C demonstrated confidence in their assessments whereas experts D and E encountered difficulties translating their opinions into numerical values. While expert A appeared to express confidence in his estimates, he also voiced his discomfort during the session and (indirectly) regretted the use of *unconditional* probabilities. Both experts A and D perceived the questions as 'vague' and the use of unconditional probabilities of concordance inappropriate, since information about one variable does not allow to draw general conclusions about the state of others. Consequently, conditional probabilities of concordance may yield more consistent results, and are hence discussed in the Conclusion and Discussion.

5.1.3 Decision-makers

Expanding upon the previous analysis of d-calibration scores, this section aims to design and assess various combinations of the experts' judgments (or decision-makers, DMs) to gauge their reliability. Two distinct DMs were subjected to evaluation: the equal weights decision-maker (EWDM), defined as the average of the experts' correlation matrices, and the global weights decision-maker (GWDM), determined by a weighted average of the matrices. In the GWDM, each expert's weight corresponds to their respective (normalized) d-calibration scores. To effectively compare the performance of these decision-makers with the respondents', their calibration scores were computed and are presented in Table 5.1. Encouragingly, both decision-makers outperformed all but the highest scoring expert (C), whose score slightly surpassed that of the global weights decision-maker. Notably, the comparison between the GWDM and the EWDM did not exhibit a significant difference in performance, similar to findings in Rongen et al. (2023), due to the fairly high scores obtained by all experts and the absence of an outlier.

To observe whether the gap between equal and global weights decision-makers widens with the presence of an outlier, a poorly calibrated expert was added to the actual experts panel. This dummy expert's correlation matrix for the seed variables was:

$$R_{outlier} = \begin{bmatrix} 1 & 0.95 & 0.95 \\ 0.95 & 1 & 0.95 \\ 0.95 & 0.95 & 1 \end{bmatrix}$$

which is definite positive and performs significantly worse than the lowest-scoring expert (D): $dCal_{outlier} = 0.311$. The d-calibration scores of both decision-makers, computed using the new pool of experts, can be found in Table 5.2. The addition of an outlier notably affected the performance scores of both DMs; the EWDM's, however, decreased more than twice as much as the GWDM's. In the former, a minor weight is attributed to the new expert while the best performing experts (B and C) still predominently defined the correlation matrix, whereas in the latter the dummy's (poor) assessment highly influenced the outcome. These results underline the relevance of a scoring system whereby the influence of the least performing experts is minimized.

Decision-maker	**Without outlier**	**With outlier**
Expert D	0.516	-
Outlier	-	0.311
EW DM	0.869	0.818
GW DM	0.897	0.873
optDM	0.968	-

Table 5.2: Equal and global weights decision-makers' scores with and without outlier.

Next, the existence of a 'best' decision-maker was investigated, i.e. a combination of the experts that maximizes the calibration score. Given the gap between experts B, C, and the rest, it came with no surprise that the optimized DM (optDM in Table 5.1) is merely a weighted average of the former's correlation matrices. This new decision-maker was significantly better calibrated than the GWDM, with $dCal_{GWDM} = 0.897$ and $dCal_{optDM} = 0.968$. Nonetheless, we recall the limitations of the aggregation approach: the seed variables are completely unrelated to the research topic. Therefore, the d-calibration scores hereby assess the experts' familiarity with probability (normative expertise), but do not provide evidence on their substantive expertise (Bolger & Rowe, 2015). Defining a best DM based on this sole criterion would hence be litigious.

Previous applications of dependence-calibration in academia indicated that a larger weighing pool results in the definition of more consistent decision-makers (Rongen et al., 2023). To evaluate the robustness of the DMs constructed previously, we were interested in the spread in calibration scores across the different combinations of a given size, similarly to the analysis in Rongen et al. (2023). Figure 5.3 depicts a convergence of the d-calibration scores towards higher average values for a larger experts pool, a phenomenon accentuated by the presence of an outlier. These results confirm the concerns regarding the reliability of the optimized DM; the informativeness of experts B and C, despite their excellent scores, remains highly uncertain, thus defining the network's dependence structure only based on their estimates would increase the exposure to substantial errors between the expert-elicited matrix and the actual one. Additionally, the optimized DM excluded the most experienced experts, whose knowledge of air handling units is significant. As a result, the global weights decision-maker's dependence structure, shown in Figure 5.4, was used for the construction of the Non-Parametric Bayesian Network (NPBN).

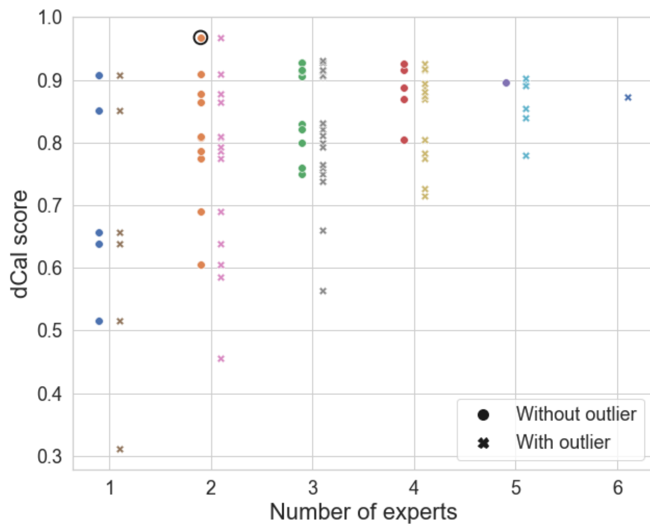


Figure 5.3: D-calibration scores of the global weights decision-maker for all combinations of experts, with and without outlier.

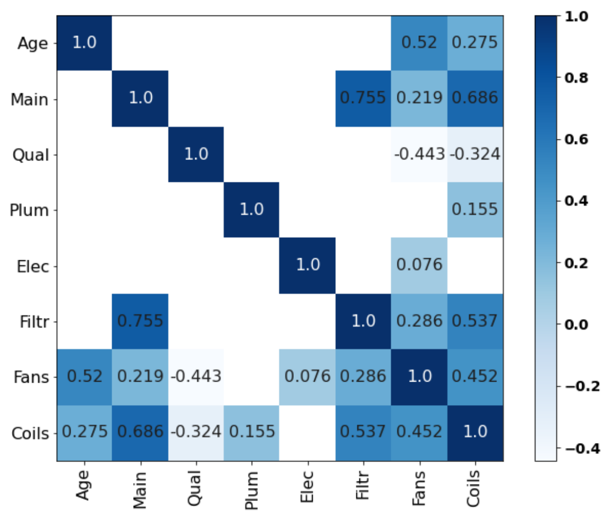


Figure 5.4: Correlation matrix implemented in the NPBN.

5.2 Marginal distributions

To complete the development of the network, the marginal distributions associated to each variable were defined. We recall that the model contains eight variables, presented in chapter 4:

- 'AHU Age': continuous. Defined on $\mathbb{R}_+^*$.
- 'Maintenance interval': continuous. Defined on $\mathbb{R}_+^*$.
- 'Design & Construction quality': discrete. Takes values between 1 (very poor) and 5 (excellent).

- ‘Filters’, ‘Fans’, ‘Coils’, ‘Plumbing supply elements’ and ‘Electrical supply elements’: discrete. Assessed on the 1-6 scale defined in NEN 2767 (Appendix A; NEN, 2006).

In accordance with chapter 4, the marginal distributions were determined by consulting individually two of the five respondents to the questionnaire. Using the visualization in UniNet, the experts’ opinions were progressively added to the software and modified if the distribution did not match their perception. Because of the hybrid nature of the network, the distributions were retrieved in two different ways:

- Discrete: experts provided the complete distribution, i.e. $P(X = i), i \in [\![1, 6]\!]$.
- Continuous: experts answered qualitative statements, including: “What is the mean (age) of the population of interest? Which proportion is (older/newer) that the mean value? How spread are the values around the mean?” A distribution was then defined in an attempt to fit the experts’ answers.

The experts both suggested that the continuous distributions are right-skewed, i.e. that the density function has a longer right tail. Therefore, two log-normal distributions were defined and are presented in Table 5.3 along with the other variables. Whereas information regarding the age was easily retrievable, the experts were less informed on industry-wide practices regarding frequency of maintenance, as the TU Delft’s policy strictly requires a yearly inspection of all installations. Concerning the discrete distributions, the filters and coils are on average in remarkably better condition than the fans: the filters are changed yearly, the coils are regularly cleaned whereas the fans’ components are more rarely replaced. This is because corrosion is the main source of deterioration of the coils whereas the fans’ is mechanical, the former being more easily attenuated.

Variable	**Distribution**	**(mean, std*)**
Age	$LN(\mu = 3.191, \sigma = 0.237)$	(24.98, 6.00)
Maintenance interval	$LN(\mu = 0.102, \sigma = 0.40)$	(1.20, 0.50)
D&C quality	$P(X = i)_{i \in [1,6]} =$ [0.01, 0.05, 0.44, 0.3, 0.2]	(3.63, 0.89)
Filters	$P(X = i)_{i \in [1,6]} =$ [0.15, 0.44, 0.25, 0.1, 0.05, 0.01]	(2.49, 1.08)
Fans	$P(X = i)_{i \in [1,6]} =$ [0.05, 0.15, 0.35, 0.39, 0.04, 0.02]	(3.28, 1.00)
Coils	$P(X = i)_{i \in [1,6]} =$ [0.1, 0.25, 0.5, 0.1, 0.05, 0]	(2.75, 0.94)
Plumbing supply elts	$P(X = i)_{i \in [1,6]} =$ [0.12, 0.2, 0.24, 0.4, 0.03, 0.01]	(3.05, 1.14)
Electrical supply elts	$P(X = i)_{i \in [1,6]} =$ [0.12, 0.2, 0.24, 0.4, 0.03, 0.01]	(3.05, 1.14)

Table 5.3: Marginal distributions of the variables. *std: standard deviation

All in all, a Bayesian Network including the newly determined marginal distributions arose and is displayed in Figure 5.5, with a summary of the associated (conditional) rank correlations presented in Table 5.4. To transfer the BN from UniNet to pyBanshee, and because the latter does not support the manual definition of discrete variables, the model was sampled using UniNet’s integrated sampling mode. The resulting data was then imported in pyBanshee in the form of ‘empirical data’. In order to validate the network quantified in this section, a set of qualitative and quantitative analyses meant to assess the model’s fidelity is presented hereafter.

5.3 Analyses

To go beyond the qualitative comments written in the previous section regarding the values elicited from the experts, the BN constructed previously was validated. In literature, validation often takes one of two forms: the model’s predictions are compared to empirical data (when available) or experts, who contributed or not to the model creation, are asked to assess the model’s output when subjected to a set of scenarios (Pollino, Woodberry, Nicholson, Korb, & Hart, 2007; S. H. Chen & Pollino, 2012; Pitchforth & Mengersen, 2013). Some authors argue that these tests are insufficient to assess the validity of a model, which is “the ability

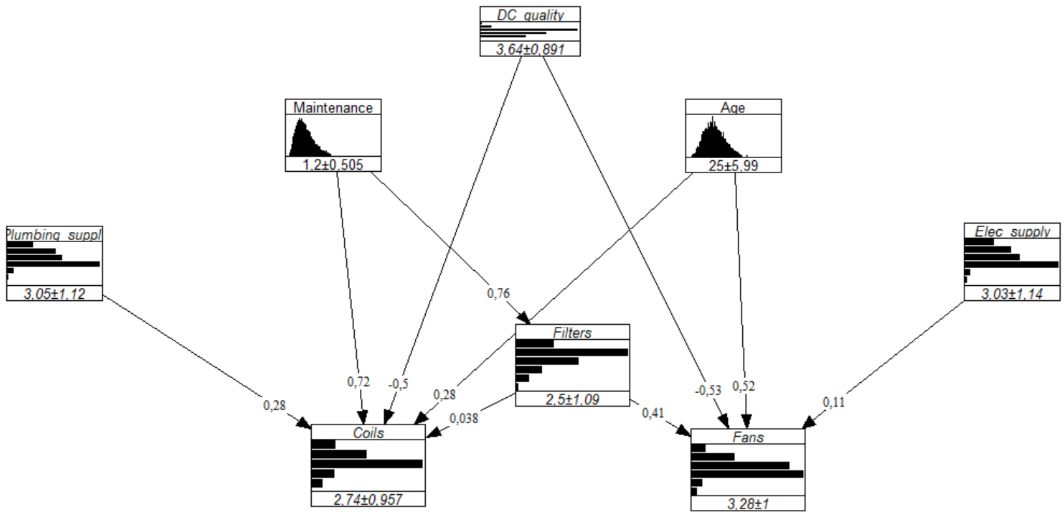


Figure 5.5: Visualization of the quantified NPBN in UniNet.

Rank correlation	**Value**
$r(Main, Filt)$	0.755
$r(Age, Fans)$	0.52
$r(D\&C, Fans \mid Age)$	-0.527
$r(Elec, Fans \mid Age, D\&C)$	0.108
$r(Filt, Fans \mid Age, D\&C, Elec)$	0.409
$r(Age, Coils)$	0.275
$r(Main, Coils \mid Age)$	0.718
$r(D\&C, Coils \mid Age, Main)$	-0.501
$r(Plum, Coils \mid Age, Main, D\&C)$	0.279
$r(Filt, Coils \mid Age, Main, D\&C, Plum)$	0.038

Table 5.4: Conditional rank correlations used in the NPBN in Figure 5.5.

of a model to describe the system that it is intended to describe both in the output and in the mechanism by which that output is generated" (Pitchforth & Mengersen, 2013). Clearly, the use of data was excluded in this research, simply because it was unavailable. Thus, two other techniques were employed: a scenario analysis and a sensitivity analysis.

5.3.1 Scenario analysis

Scenario - or 'what-if' - analysis is a common appproach for the validation of Bayesian networks, regardless of the availability of empirical data (S. H. Chen & Pollino, 2012; Hänninen, 2014; Saltelli et al., 2019; A. M. Hanea, Hilton, Knight, & P. Robinson, 2022). It allows to assess the logic of the model output for different sets of input variables as a complement to more advanced analyses, e.g. a sensitivity analysis. In essence, BNs have no input or output; evidence can be entered in any of the network's nodes, thus multiplying the number of scenarios required to thoroughly observe the model's behaviour. When validating the model, however, it is essential to account for its initial purpose (Hänninen, 2014); here, the estimation of air handling units components' condition. Clearly then, a set of output variables arose: 'Filters', 'Coils' and 'Fans'. Moreover, the experts' opinions indicated that among the five remaining variables, three have a substantially stronger influence on the model output: 'Age', 'Maintenance interval' and 'Design & Construction quality'. As a result, only the latter were considered as inputs; the other two variables ('Plumbing supply elements' and 'Electrical supply elements') were kept constant throughout the analyses at their mean values.

To the best of the author's knowledge, there is no formal guideline on the number of scenarios necessary to observe a model's behaviour holistically. Therefore, the following six scenarios were investigated:

- Scenario 1: old AHU, frequent maintenance,
- Scenario 2: recent AHU, unfrequent maintenance,
- Scenarios 3 and 4: very poor D&C quality, recent/old AHU,
- Scenarios 5 and 6: excellent D&C quality, recent/old AHU.

all of which are described in detail across the sub-section. In the absence of data, extreme conditions tests are particularly relevant because it subjects the model to scenarios for which the outcome is predictable and can be discussed qualitatively (Pitchforth & Mengersen, 2013), even without field expertise. Nevertheless, only combinations of inputs for which the output is not trivial were deemed insightful; simulating a recent AHU maintained consistently for instance would undoubtedly result in a system in excellent condition. For that reason, two variables with similar contributions to the output distributions were set at inversely extreme values to determine whether the model reacted realistically.

For each scenario, the (conditional) distribution of each output was determined using pyBanshee's inference method, with unconditional distributions displayed in grey and conditional distributions in blue. It is common practice to consider the variable's state with the highest probability as a BN's output (Cypko et al., 2017). In this thesis, however, another approach was adopted; the probability that the output is lower or equal than 3 was computed and compared between the unconditional ($P(Y \leq 3)$) and conditional ($P(Y \leq 3|\mathbf{X_i})$, with $\mathbf{X_i}$ the vector of observations in scenario i) cases. The relevance of this metric is explained by the physical meaning of the outputs: they represent the technical condition of three components used to plan multi-year maintenance. Although there is no absolute rule stating that an equipment shall be replaced or repaired beyond a condition level, informal discussions with practitioners suggested that a condition score of three is a reasonable threshold.

Scenario 1: old AHU and frequent maintenance.

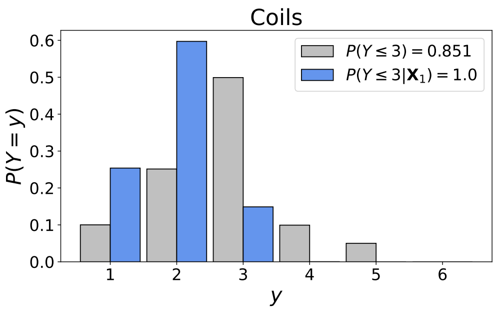
The first scenario consisted in the following configuration:

- 'AHU age': 40 years,
- 'Maintenance interval': 6 months,
- 'Design & Construction quality': 3.63 (mean value),

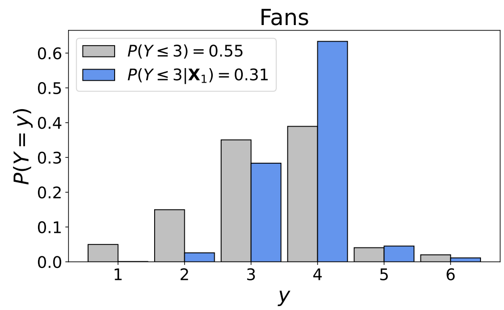
$\Rightarrow$ $\mathbf{X_1} = (X_0 = 40, X_1 = 0.5, X_2 = 3.63)$.

The dependence structure elicited from experts indicates that the age of the unit and the frequency at which it is maintained overwhelmingly affect its condition. The outcome of scenario 1 is shown in Figure 5.6. Unsurprisingly, the filters' condition has significantly improved due to its connection with maintenance. However this outcome, while consistent with our earlier assumptions, appears to be somewhat unrealistic from a physical standpoint. Expert D illustrated the relationship between 'Filters' and 'Maintenance' with the example of Schiphol airport, Netherlands' main international airport, where filters are replaced three to four times a year due to air pollution. In fact, experts almost unanimously (4/5) indicated that environmental condition should be included because of their impact on the filters' deterioration. Clearly then, this scenario showcases the model's disproportionate response as the probabilities associated to states 3 and above should not be null, as demonstrated by the example of Schiphol.

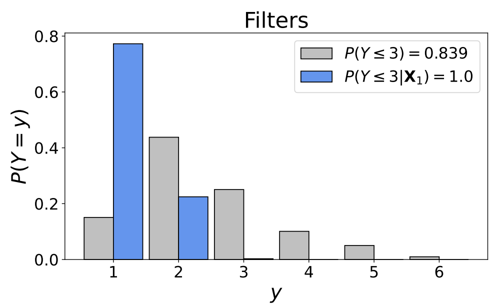
Similarly, the distribution of the variable 'Coils' shifted left, reflecting an improvement from the unconditional case. This finds explanation in the dependence structure of the BN, where the correlation between 'Coils' and 'Maintenance interval' is substantially higher than between 'Coils' and 'Age' (0.686 and 0.275, respectively). Because the main mode of deterioration of the coils is by corrosion, accelerated by frost and the accumulation of particles, consistently cleaning them allows to temper the phenomenon. Moreover, the shift in the distribution of 'Filters' also influences the one of 'Coils' since these variables are positively correlated. Interestingly, the probabilities of states 3, 4 and 5 are relatively low (0.15, 0 and 0, respectively) given the advanced age of the unit and the theoretical lifespan of the coils ($\sim$ 20-25 years). For the same reason, the probability that the coils are in condition 1 (0.26) is abnormally high, proof that the model's capacity to handle extreme cases is limited.



(a) Distribution of 'Coils'.



(b) Distribution of 'Fans'.



(c) Distribution of 'Filters'.

Figure 5.6: Unconditional and conditional distributions of the outputs (a) 'Coils', (b) 'Fans' and (c) 'Filters' under scenario 1.

Finally, the fans' condition was inversely impacted by the input values: the probability that the component is in condition 4 has dramatically increased (from 0.39 to 0.65), again in accordance with the correlation of 'Fans' with 'Age' being higher than with 'Maintenance interval' (0.52 and 0.219, respectively). Failure in the fans mainly involves mechanical malfunctions such as exhaustion of the motor or failure of the bearings, whose maintenance have limited impact on their lifespan. The conditional probability that the component is in reasonable condition or better seems high (0.31) but most of this density is in state 3, which is conform with the previous comments.

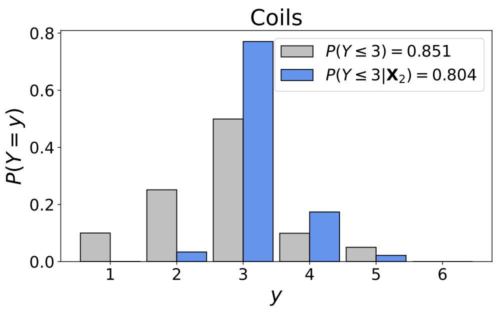
Scenario 2: recent AHU and unfrequent maintenance.
The second scenario consisted in the following configuration:

- 'AHU age': 10 years,
- 'Maintenance interval': 3 years,
- 'Design & Construction quality': 3.63 (mean value),

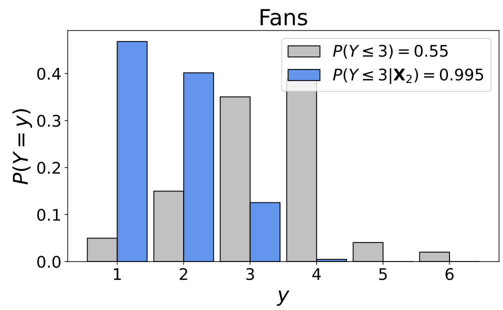
$\Rightarrow$ $\mathbf{X_2} = (X_0 = 10, X_1 = 3, X_2 = 3.63)$.

This scenario is assuredly paired with the previous, as 'Age' is now set at a low value and 'Maintenance interval' at an extremely high one. As expected, the variations observed in the last scenario are reversed; both the 'Coils' and 'Filters' distributions shifted right whereas the 'Fans' shifted left. The case of 'Coils' is particularly interesting. While $P(Y \leq 3|\mathbf{X_i})$ hardly varies between the conditional and unconditional cases, the distribution of probability between the states has completely changed: state 3 increased from 0.50 to 0.77 (!) whereas states 1 and 2 decreased to 0 and 0.04, respectively. These observations correctly reflect the unlikelihood of the component being in good condition due to the low maintenance, and simultaneously the unlikelihood of a bad condition given the young age of the unit, relative to its theoretical lifespan.

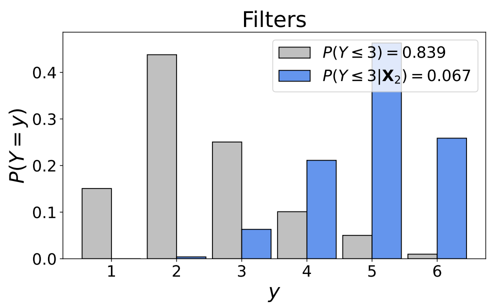
Moreover, the fans' young age results in a high probability $P(Y \leq 3|\mathbf{X_2}) = 0.995$, again aligning with the dependence structure of the BN. Nevertheless, $P(Y = 1|\mathbf{X_2}) = 0.47$ appears to be unrealistically high; given the age and the interpretation of the condition scores in NEN (2006)'s 'safety net', a condition score of 1 must be less likely than 2 (0.40).



(a) Distribution of 'Coils'.



(b) Distribution of 'Fans'.



(c) Distribution of 'Filters'.

Figure 5.7: Unconditional and conditional distributions of the outputs (a) 'Coils', (b) 'Fans' and (c) 'Filters' under scenario 2.

Scenarios 3/4: very poor D&C quality, recent/old AHU.

In the following four scenarios (3, 4, 5 and 6), the impact of the quality of the materials, workmanship and design on the condition at different stages of each component's life-cycle is evaluated. Because 'Filters' is uncorrelated to both 'Age' and 'D&C quality', only the distributions of 'Fans' and 'Coils' are discussed in the remainder of these scenarios.

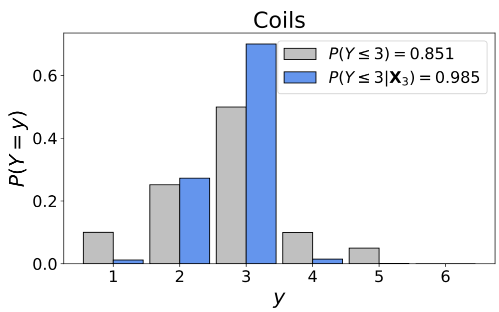
The third and fourth scenarios consisted in the following configurations:

- 'AHU age': 10 years (scen. 3)/ 40 years (scen. 4),
- 'Maintenance interval': 1.20 (mean value),
- 'Design & Construction quality': 1 (very poor),

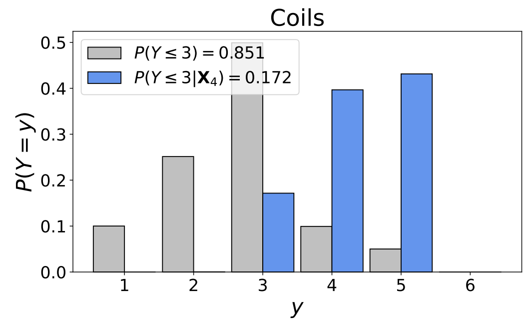
$\Rightarrow$ $\mathbf{X_3} = (X_0 = 10, X_1 = 1.2, X_2 = 1)$; $\mathbf{X_4} = (X_0 = 40, X_1 = 1.2, X_2 = 1)$.

Figure 5.8 shows the model output under scenarios 3 and 4. For a recent AHU, the quality of 'Coils' substantially influenced the distribution; despite a low age and a relatively frequent maintenance, the probability of the component being in state 1 was extremely low. Nevertheless, $P(Y \leq 3|\mathbf{X_3})$ remained very high (0.985) due to the high probabilities of states 2 and 3 (0.29 and 0.69, respectively). For an older AHU (scenario 4), a shift was observed: while state 3 was very likely ($\simeq 0.8$) in scenario 3, states 4 and 5 now share an important proportion of the total density. Clearly then, a very bad basic quality strongly influences the long-term degradation of the coils, echoing expert A's remark on the decrease of AHU components' lifespan due to an average deterioration of their basic quality.

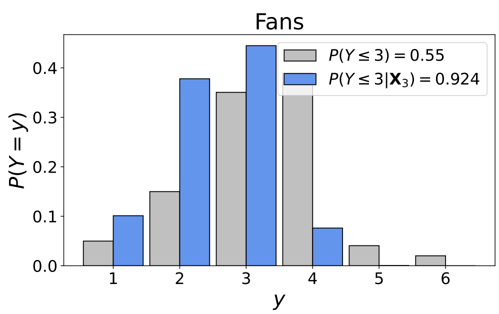
Similar observations hold for the fans. Comparing scenarios 2 and 3, the impact of a lower quality is minor, with a 'redistribution' of density from state 1 to state 3. However, a comparison between scenario 1 and scenario 4 highlights the steep increase in $P(Y = 5|\mathbf{X_i})$ and $P(Y = 6|\mathbf{X_i})$ (0.18 and 0.63, respectively) due to a reduction of the 'D&C quality' from 3.63 to 1, again emphasizing the importance of high quality installations in the advanced stages of their life cycle.



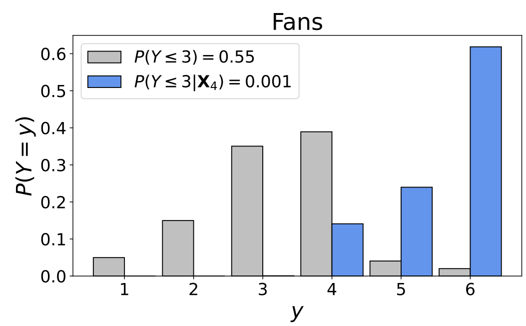
(a) Distribution of 'Coils' under scenario 3.



(b) Distribution of 'Coils' under scenario 4.



(c) Distribution of 'Fans' under scenario 3.



(d) Distribution of 'Fans' under scenario 4.

Figure 5.8: Unconditional and conditional distributions of 'Coils' (top) and 'Fans' (bottom) under scenario 3 (left) and scenario 4 (right).

Scenarios 5/6: excellent D&C quality, recent/old AHU.
The fifth and sixth scenarios consisted in the following configurations:

- 'AHU age': 10 years (scen. 5)/ 40 years (scen. 6),
- 'Maintenance interval': 1.20 (mean value),
- 'Design & Construction quality': 5 (excellent),

$\Rightarrow$ $\mathbf{X_5} = (X_0 = 10, X_1 = 1.2, X_2 = 5)$; $\mathbf{X_6} = (X_0 = 40, X_1 = 1.2, X_2 = 5)$.

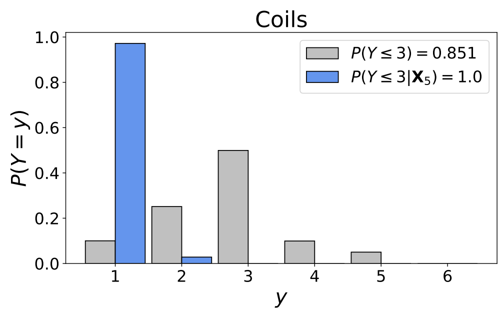
In contrast with scenarios 3 and 4, which investigated the influence of a poor basic quality on the states of the network's output variables, scenarios 5 and 6 aim to determine whether an investment in an excellent quality significantly affects the short- and long-term condition of the air handling unit.

Under similar operating conditions, components of greater quality theoretically have a longer life span than elements of poor quality, which were modelled in the previous scenarios. Therefore, the distributions of 'Coils' and 'Fans' in scenario 5 are unsurprisingly concentrated around low condition scores, resulting in values of $P(Y \leq 3|\mathbf{X_5})$ of 1 for both variables. However, the model's response to the increase in basic quality appears to be excessive, with $P(Y = 1|\mathbf{X_5})$ of 0.977 and 1 for 'Coils' and 'Fans', respectively. For instance, the degradation of 'Fans' also depends on the intensity at which the AHU is used, and while the network does not explicitly account for such factors, the output distributions should reflect the uncertainty due to such missing information.

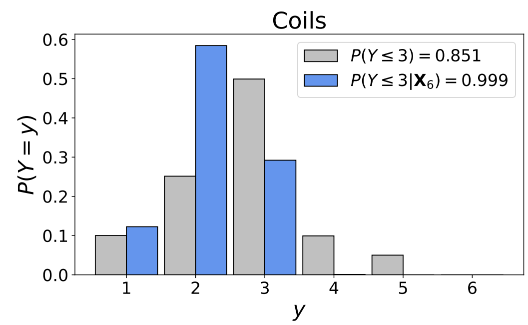
For older units, and in comparison with scenario 4, an increase of 'D&C quality' evidently results in a slower deterioration for both 'Coils' and 'Fans', with a substantial share of the distributions being below the threshold values: $P(Coils \leq 3|\mathbf{X_6}) = 0.999$ and $P(Fans \leq 3|\mathbf{X_6}) = 0.912$, against $P(Coils \leq 3|\mathbf{X_4}) = 0.172$

and $P(Fans \leq 3|\mathbf{X_4}) = 0.001$ in scenario 4. In alignment with the observations for scenario 5, the impact of an excellent quality on the components is too high. While it is logical to witness an improvement from scenario 4, the unit's age (40 years) *must* translate in medium-to-high likelihoods for states 4 and 5. Conversely, the probabilities of states 1 and 2 are too high as both components (almost) have reached their theoretical lifespan. Clearly then, scenarios 5 and 6 indicate that the rank correlations associated to the edges 'D&C quality' → 'Coils' (-0.324) and 'D&C quality' → 'Fans' (-0.443) are possibly too high (in absolute values), which may partly stem from expert C's strong assessment for D&C quality' → 'Fans' (-0.795). His assessment, which is substantially higher than the rest of the experts', strongly contributes to the final decision-maker because of his excellent calibration score.

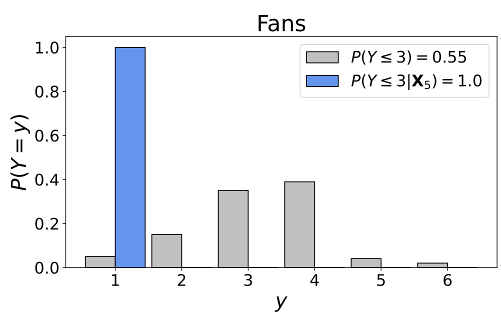
All in all, the influence of the basic quality of the air handling unit's components is correctly translated, even though some adjustments are required for the model to reduce the rank correlations and obtain realistic outputs. We notably observe that high-standard materials, design and construction can significantly extend the components' lifespan and have a greater impact as the system ages.



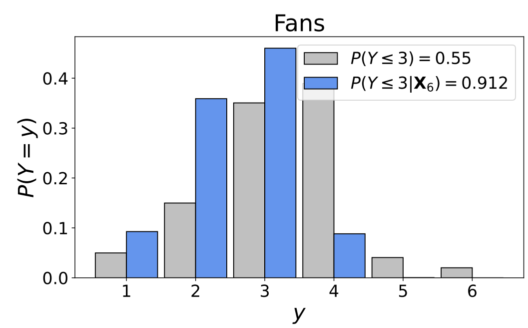
(a) Distribution of 'Coils' under scenario 5.



(b) Distribution of 'Coils' under scenario 6.



(c) Distribution of 'Fans' under scenario 5.



(d) Distribution of 'Fans' under scenario 6.

Figure 5.9: Unconditional and conditional distributions of 'Coils' (top) and 'Fans' (bottom) under scenario 5 (left) and scenario 6 (right).

A summary of the outcome of each scenario can be found in Table 5.5. In the absence of empirical data to quantitatively assess the predictive validity of the BN, the latter's validation relies on qualitative observations of the model's behaviour. While most comments formulated in the previous paragraphs simply indicate similarities between the outputs and the dependence structure implemented in the NPBN, the central finding of this analysis is the necessity to integrate environmental conditions in the network, supported by experts' feedback during the consultations.

As mentioned in the introduction to Non-Parametric BNs in chapter 2, the addition of new variables is facilitated by the modular nature of this type of model. To illustrate the effect the addition of environmental conditions would have on the outputs, let us consider an hypothetical network which includes this variable. The latter ('Environmental conditions') is defined on the following 1-5 scale:

Scenario	Configuration (Age, Maintenance, Quality)	$P(Y \leq 3 \vert \mathbf{X_i})$		
		Coils	**Fans**	**Filters**
1	(40, 0.5, 3.63)	1	0.31	1
2	(10, 3, 3.63)	0.804	0.995	0.067
3	(10, 1.2, 1)	0.985	0.924	0.908
4	(40, 1.2, 1)	0.172	0.001	0.908
5	(10, 1.2, 5)	1	1	0.905
6	(40, 1.2, 5)	0.999	0.912	0.905
7	(40, 0.5, 3.63, *Env*=1)	0.999	0.057	0.71

Table 5.5: Summary of the scenario analysis.

Very unfavorable	Unfavorable	Medium	Favorable	Very favorable
1	2	3	4	5

Because determining a realistic marginal distribution for the newly defined variable involves an expertise that was unavailable at this stage of the research, a Uniform distribution was chosen, i.e. 'Environmental conditions' $\sim \mathcal{U}(1, 5)$. Furthermore, consultations with the experts indicated that this factor mainly influences the deterioration of the filters, hence the creation of the edge 'Environmental conditions' $\rightarrow$ 'Filters'. Although this relationship is weaker than that between 'Maintenance interval' and 'Filters', the conditional correlation associated to the new edge shall be high; for two units maintained at the same frequency, the environmental conditions are a strong predictor for the conditions of their filters. Therefore, we considered $r(Env, Filt|Main) = -0.8$. The resulting model, used in the next scenario, is illustrated in Figure 5.10.

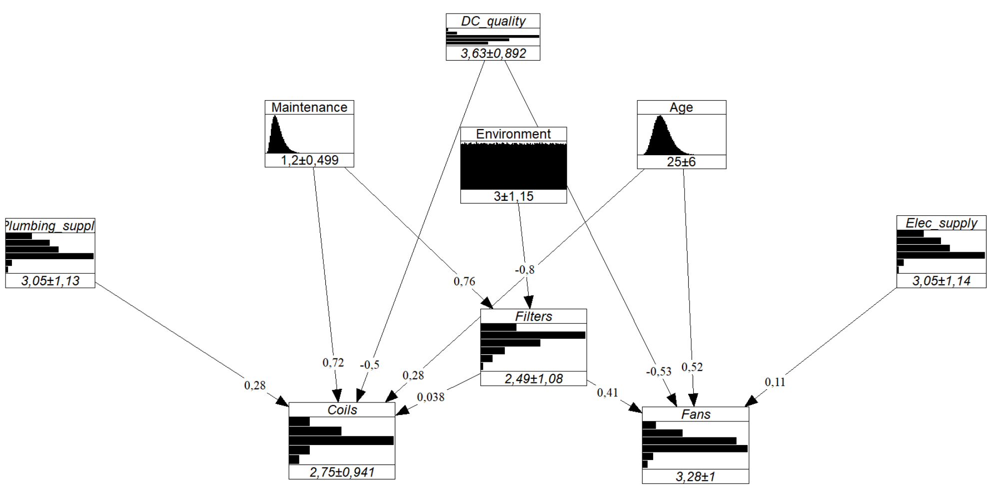


Figure 5.10: Hypothetical NPBN including the variable 'Environmental conditions'.

Scenario 7: old AHU, frequent maintenance and *very unfavorable environmental conditions.* The seventh scenario consisted in the following configuration:

- 'AHU age': 40 years,
- 'Maintenance interval': 6 months,
- 'Design & Construction quality': 3.63 (mean value),
- 'Environmental conditions': 1 (very unfavorable),

$\Rightarrow \mathbf{X_7} = (X_0 = 40, X_1 = 0.5, X_2 = 3.63, X_8 = 1)$.

Figure 5.11 illustrates the conditional distributions of 'Filters' obtained in scenarios 1 and 7. First, there is an evident change in the distribution. The discussion on scenario 1 underlined that the probabilities of states 3, 4 and 5 could not realistically be null without information on the environmental conditions. Here, evidence of very unfavorable climatic conditions clearly resulted in a concentration of the distribution around states 3 and 4, with probabilities of 0.646 and 0.268 respectively - aligning with the example of Schiphol airport presented previously. This brief discussion demonstrates that the addition of 'Environmental conditions', although not rigorous, was a fairly straightforward endeavour which yielded encouraging results. Clearly then, additions to the network can allow the continuous improvement of the model's predictive accuracy.

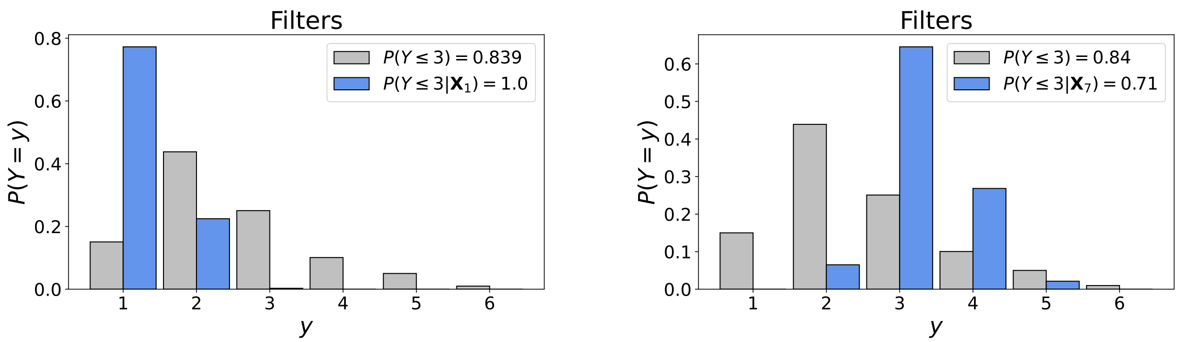


Figure 5.11: Unconditional and conditional distributions of 'Filters' under scenario 1 (left) and scenario 7 (right).

5.3.2 Sensitivity analysis

A rigorous sensitivity analysis is an essential ingredient to an exhaustive model validation. It consists in the assessment of a model's sensitivity to variations in its inputs (in a general sense) (S. H. Chen & Pollino, 2012; Pitchforth & Mengersen, 2013; Razavi et al., 2021) and serves two primary purposes: the identification of relationships between some of the variables as well as the quantification of the inputs' contributions to the uncertainty of the model output (Razavi & Gupta, 2015; Saltelli et al., 2019; Razavi et al., 2021). In Bayesian modelling, inputs cover a large panel of elements: graph structure, marginal distributions, or more commonly evidence and parameters (Pollino et al., 2007; Pitchforth & Mengersen, 2013) - rank correlations or CPTs.

A common misconception regarding sensitivity analyses is the belief that they are intrinsically local, meaning they allow to determine individual inputs' influence on the output separately. While such approach, known as Local Sensitivity Analysis (LSA), is widespread, it is in fact appropriate in a very limited range of applications (Saltelli et al., 2019). The literature raises three main criticisms of LSA: (i) it is not applicable to nonlinear models (of which BNs are part), (ii) it poorly covers the input space, and (iii) it does not evaluate the interactions between parameters (X.-Y. Zhang, Trame, Lesko, & Schmidt, 2015; Razavi & Gupta, 2015; Saltelli et al., 2019; Razavi et al., 2021).

These limitations of local sensitivity analyses have given rise to *global* methods. The latter aim to comprehensively explore the input domain, thus offering insights into a model's behaviour across all possible configurations. While an exhaustive review of global sensitivity analysis (GSA) is beyond the scope of this thesis - see X.-Y. Zhang et al. (2015) for an overview, this section introduces one of the most prominent technique: Sobol's method (X.-Y. Zhang et al., 2015; Razavi & Gupta, 2015; Saltelli et al., 2019; Ballester-Ripoll & Leonelli, 2022). Its essence lies in the decomposition of a model output's Y variance in a series of contributions of increasing dimensionality, encompassing not only individual 'direct' contributions (V_i) but also that of their interactions with other variables of the model (Saltelli et al., 2010; X.-Y. Zhang et al., 2015):

$$V(Y) = \sum_{i}^{d} V_i(Y) + \sum_{j>i}^{d} V_{ij}(Y) + ... + V_{12..d}(Y) \tag{5.1}$$

where d is the dimension of the input space. For a subset of indices $\{i_1, i_2, ..., i_k\} \subset [\![1, d]\!]$, we define Sobol sensitivity indices as:

$$S_{i_1 i_2 \dots i_k} = \frac{V_{i_1 i_2 \dots i_k}(Y)}{V(Y)} \tag{5.2}$$

S_i is called the first-order sensitivity index of variable X_i, S_{ij} the second-order sensitivity index of variables X_i and X_j, and so on. The first-order index is often expressed as follows:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X_{-i}}}(Y|X_i))}{V(Y)} \tag{5.3}$$

where $\mathbf{X_{-i}}$ is the vector of all factors but X_i and which is equivalent to the formulation in Equation 5.2 in dimension 1. Another commonly used measure is the total-order sensitivity index, which accounts for all of the contributions of variable X_i to the variance of the output. It is defined as:

$$S_{T_i} = S_i + \sum_j S_{ij} + \sum_{j,k} S_{ijk} + ... \tag{5.4}$$

$$= \frac{E_{\mathbf{X_{-i}}}(V_{X_i}(Y|\mathbf{X_{-i}}))}{V(Y)} \tag{5.5}$$

In the context of sensitivity analysis, both indices can be interpreted in terms of (reduction of) variance (Saltelli et al., 2010):

- $V_{X_i}(E_{\mathbf{X_{-i}}}(Y|X_i))$ corresponds to the expected reduction of variance in Y if X_i could be fixed. Therefore, a high value of S_i indicates that the first-order effect of X_i on the uncertainty of Y is high.
- $E_{\mathbf{X_{-i}}}(V_{X_i}(Y|\mathbf{X_{-i}}))$ is the expected variance in Y that would be left if all factors but X_i could be fixed. A high value of S_{T_i} thus suggests that the total effect, including first-order and interaction terms, of X_i on the output's uncertainty is high.

Despite the emerging popularity of Sobol's method for GSA, its suitability to Bayesian networks has been barely explored. First, computing Sobol's indices is expensive as it requires the approximation of multifold integrals and the construction of complex Monte Carlo simulations. The development of more efficient algorithms for GSA is an active research topic, and while new methods have emerged (latin hypercube, quasi-random MCS, surrogate models), random Monte Carlo remains the standard approach (Ballester-Ripoll & Leonelli, 2022). The convergence of such algorithm is extremely slow ($O(\sqrt{n})$), thus requiring an extensive amount of samples to obtain consistent results. Second, the presence of correlated inputs, common in BNs, is ignored by most implementations based on MCS and other sampling techniques. Although it is not the case in this research, the lack of standardized approach for correlated inputs arguably hinders the development of Sobol's method and GSA for BNs. To the best of the author's knowledge, Sobol's indices were only applied to BNs on two occasions, in Li and Mahadevan (2018) and Ballester-Ripoll and Leonelli (2022). Their approaches, however, are inappropriate for our implementation of NPBNs. First, Li and Mahadevan (2018)'s study solely focuses on the calculation of the first-order sensitivity index, thus only covering half of our objectives, and relies on a stratified sampling technique not adapted to hybrid BNs. Second, Ballester-Ripoll and Leonelli (2022)'s work directly applies to discrete BNs, and while they suggest alternatives to implement their approach to continuous variables, it involves their discretization which is undesirable in the current context. Notwithstanding their limitations, both studies are solid alternatives to 'raw' MCS for networks with dozens of nodes; here, due to the limited size of the network, the algorithms for S_i and S_{T_i} were manually implemented on Python. The scripts for both methods can be found in Appendix F.

As previously discussed, the presence of two loops strongly decreases the efficiency of the algorithms; as Equation 5.3 and Equation 5.5 illustrate, the computation of S_i (S_{T_i}) requires the iteration over values of X_{T_i} (X_i) for *all* values of X_i (X_{T_i}). As a result, the execution times of these algorithms for a sample size of 1000 (for iterations) are substantial: 30 minutes for S_i, 48 minutes for S_{T_i}[1]. Moreover, variations are still

[1]On a machine equipped with Intel Core i5-8265 CPU @1.60 GHz.

observed in the algorithms' results for 1000×1000 samples, which do not affect the interpretability of the results but illustrate the slow convergence of the algorithm.

Output	**Input**		
	Age	Maintenance	D&C quality
	First-order index S_i		
Filters	0.001	0.522	0.001
Fans	0.254	0.045	0.530
Coils	0.075	0.446	0.248
	Total-order index S_{T_i}		
Filters	0.464	0.667	0.460
Fans	0.656	0.441	0.587
Coils	0.413	0.523	0.445

Table 5.6: First- and total-order Sobol indices of the input variables for n=1000.

The result of the simulations are shown in Table 5.6. Due to the normal copula assumption, the conditional variance of the distribution $X_i|X_j$ is directly linked to its unconditional variance and the product moment correlation. In the context of NPBNs, two elements then affect the conditional variance of a variable: the variance of its marginal distribution, and the strength of the influence of its observed parents. Because the first-order sensitivity index quantifies a *reduction* of variance due to one parent's direct contribution, higher values of S_i are unsurprisingly obtained for strongly correlated variables.

Interestingly, the first-order contribution of 'D&C quality' to 'Fans' is greater than that of 'Age', a surprising observation considering the higher correlation between the latter two variables (-0.443 and 0.52, respectively). In contrast, 'AHU Age' and 'Maintenance interval' substantially contribute to the variance of 'Fans' through interaction terms. This is because both of these variables exert a strong influence on 'Coils' (with correlations of $r(Age, Coils) = 0.275$, $r(Main, Coils) = 0.686$), itself positively correlated to 'Fans' ($r(Filtr, Fans) = 0.286$). Moreover, the direct impact of maintenance on the filters increases the second-order contribution of 'Maintenance interval' to the variance of 'Fans', resulting in a notably high value of S_{T_i}.

Similar observations arise for the variable 'Coils'. Its correlations with input variables align with their first-order contributions, with values of $S_i = 0.075 < 0.248 < 0.446$ and $|r| = 0.275 < 0.324 < 0.686$ for 'AHU Age', 'D&C quality' and 'Maintenance interval', respectively. Comparing the results in Table 5.6 and the network in Figure 5.5, we unsurprisingly notice that the total-order indices are inextricably linked with the rank correlations. However, even for pairs of uncorrelated variables, a significant value of S_{T_i} is obtained, e.g. 'Age'/'Filters' for which $S_{T_i} = 0.464$, which finds explanation in the small dimension of the BN and thus the importance of the interaction terms. Consequently, all input variables are relevant and cannot be fixed, thus underlining the importance of eliciting all rank correlations rigorously.

Chapter 6

Conclusion and Discussion

To conclude the thesis, the following sections assess the extent to which the findings discussed in the previous chapter answer the problem identified in the Introduction. First, the results obtained throughout the study are aggregated to answer the research questions formulated earlier in this deliverable. Then, the findings are nuanced by investigating the limitations encountered during the research with regards to the methodology and the resources at hand. Finally, practical and academic recommendations are proposed to build on the conclusions of this study.

6.1 Conclusion

The opening discussion on the state-of-the-art of building condition assessment brought to the fore the relevance of a new method to estimate the condition of mechanical, electrical and plumbing (MEP) systems based on easily retrievable information. To address this problem, this thesis investigated the applicability of (Non-Parametric) Bayesian Networks (BNs) in this context, characterised by the limited availability of empirical data. This section sequentially delves into the four sub-research questions and, at last, the main question defined in section 1.4.

RQ1: What are the barriers and enablers driving the integration of Bayesian Networks in building condition assessment?

The review conducted in this research highlighted the various arguments in favor and against the implementation of Bayesian networks in certain fields. This section solely discusses the ones that crucially affected the approach adopted in this thesis.

First, barriers rapidly emerged with regards to the quantification of BNs. Similarly to most probabilistic models, it requires an extensive amount of data, a significant challenge given the scarcity of condition assessment data for mechanical, electrical and plumbing systems. Interactions with practitioners from both residential and non-residential real estate organizations underlined this absence (or lack of integration) of empirical data, thus raising the question of alternative quantification methods.

As suggested in literature, the involvement of experts is common practice in the quantification of BNs as data scarcity is a phenomenon that affects various industries. However, the elicitation of experts' judgments remains a challenging endeavour. In addition to gathering a panel of knowledgeable and willing experts, the task of retrieving probabilities or other statistical quantities is complicated, as probabilistic reasoning is all but intuitive. While a range of techniques have been developed by scholars to overcome these barriers, new issues can emerge during their implementation. Moreover, the construction of a complex network - with a high number of nodes and edges - requires the assessment of an exponential amount of information, further exacerbating the limitations of expert-based BNs.

Nevertheless, the emergence of new formulations for BNs, such as Non-Parametric Bayesian Networks (NPBNs), facilitates the creation of such model in data-sparse environments. By building on marginal distributions and bivariate copulas rather than probability tables, NPBNs enable the quantification of complex networks with a lesser number of parameters, thus facilitating the integration of experts' opinions. Additionally, their formulation allows the inclusion of new variables without the complete redefinition of their children's probability tables, allowing the gradual complexification of the model. Lastly, the graphical structure of BNs assists in the interaction with experts who can visually grasp the purpose of the model.

RQ2: Which factors affect the condition of mechanical, electrical and plumbing systems?

The construction of the NPBN was initiated by the creation of the graph structure modelling all MEP sub-systems and the factors influencing their deterioration. As mentioned previously, restricting the number of variables included in the network is essential in keeping its quantification realistically possible. Therefore, a classification of the relevant MEP sub-systems was defined building on the results of Bortolini and Forcada (2018) and complemented by literature on building science.

Then, the factors affecting the condition of the sub-systems aforementioned were investigated. Existing building pathology research suggest that two types of influences exist: influence of exogenous variables (e.g. environment, maintenance) on building components, and influence between components themselves. The literature review conducted indicates that three major (exogenous) factors can be distinguished: Design and Construction quality, Maintenance and the Environmental conditions - all of which were integrated in the graph. Furthermore, the (geographical and operational) relationships between building elements were determined using the findings of Bortolini and Forcada (2020) and other studies and included in the structure.

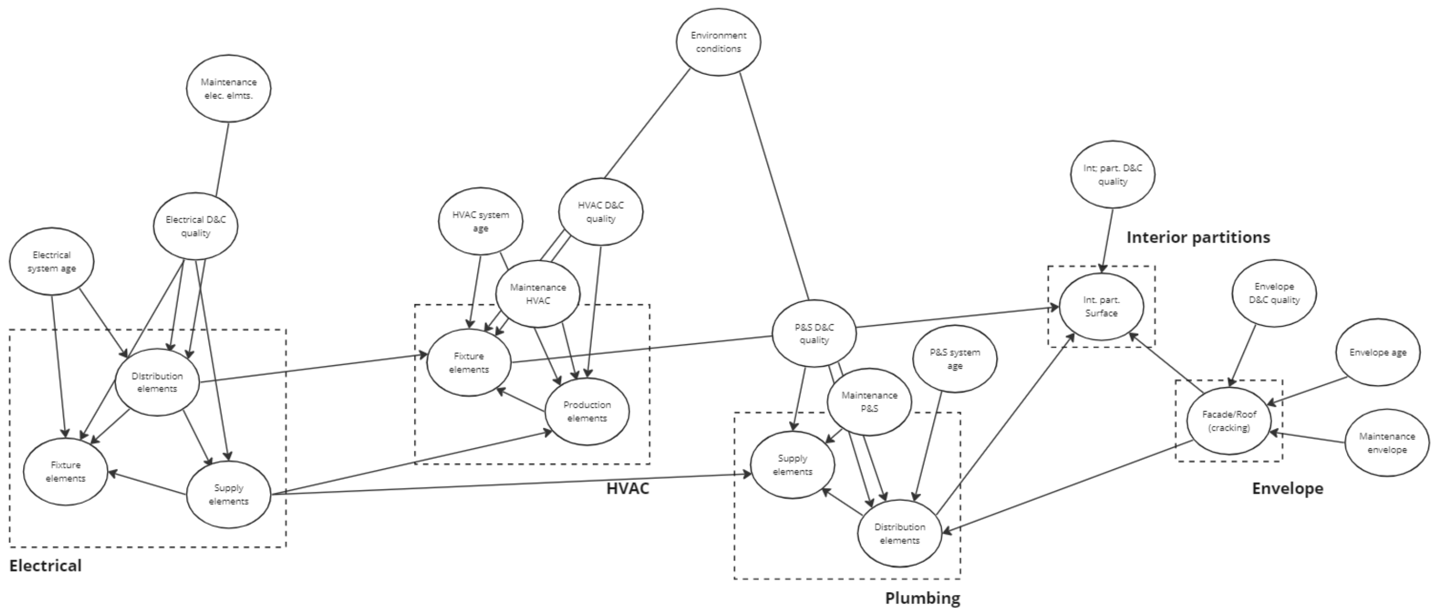


Figure 6.1: Graph structure encompassing all mechanical, electrical and plumbing systems.

The resulting graph encompassing all MEP systems, referred to as 'global' graph, is shown in Figure 6.1. Before engaging in the quantification of this graph, the feasibility of populating a network with 23 variables and over 30 edges with experts' judgments as a sole source information was interrogated. Given the difficulties enunciated earlier in this section and the time span of this thesis, the remainder of the study focused on the construction (and quantification) of a section of the graph presented in Figure 6.1. The case study selected delved into air handling units (AHUs), a sub-system of the 'HVAC production elements', for which a graph was created. The latter is shown in Figure 6.2.

RQ3: How can the model be populated in the absence of empirical data?

In order to fully quantify the newly defined graph for AHUs, two elements were to be determined: the rank correlations, associated to each edge, and the marginal distributions associated to each node. While the topic of the elicitation of univariate distributions with experts' judgments has been widely investigated in academia, the assessment of dependence is a comparatively emerging field. Consequently, a substantial share of the elicitation effort in this research was dedicated to the definition of the NPBN's dependence structure, for which a novel approach was developed.

First, a method to retrieve the experts' individual assessments was designed. While previous studies investigated the use of statistical and conditional fractile estimate approaches, the use of probabilities of concordance was here formalized. Similarly to conditional exceedance probabilities, concordance probabilities (noted P_c) allow to compute rank correlations given a set of hypotheses including the normal copula assumption. In contrast with exceedance probabilities, however, prior knowledge of the marginal distributions is not required to retrieve the rank correlations, thus providing additional flexibility to the assessor. A panel of 5 experts participated in the elicitation, which consisted in ten questions (= 10 edges) based on probabilities of concordance, resulting in five separate correlation matrices.

Secondly, the individual matrices were aggregated to obtain a unique dependence structure implementable in the NPBN. Dependence-calibration is a method that consists in combining experts' assessments through a weighted average, with the weights reflecting the experts' performance on a set of seed questions. While the seed questions ideally align with the topic of the elicitation (here AHUs), the data scarcity problem mentioned previously hindered the formulation of relevant seed questions. As a result, precipitation in the Netherlands was chosen due to the familiarity of the experts with the topic. Three seed questions were added to the questionnaire presented to the experts, allowing to compute the experts' d-calibration scores (noted $dCal_e$). Following the elicitation, two observations arose. First, all experts obtained satisfactory scores, the lowest being $dCal = 0.516$ (expert D); second, two of the five experts performed significantly better than their peers, with d-calibration scores of 0.907 and 0.85 (experts B and C). The emergence of two groups aligns with the confidence displayed by the participants during the consultations and reinforces the relevance of a robust aggregation method. Finally, the d-calibration scores were used to define a 'decision-maker' as the weighted average of the experts' correlation matrices, which reassuringly performed better than all but one expert ($dCal(GWDM) = 0.897$).

Thirdly, the marginal distributions were defined in collaboration with two of the five participants of the elicitation. Contrary to the assessment of dependence, which relied on a mathematically sound approach, the univariate distributions were determined by directly querying the experts (discrete variables) or through qualitative statements (continuous variables). Consequently, while the resulting distributions were somewhat defined based on field expertise, they do not constitute a solid foundation on which the network may be further developed, and future applications of the model shall include the redefinition of the marginal distributions. Following the elicitation of the marginal distributions, the complete version of the Non-Parametric Bayesian Network was built. It is shown in Figure 6.2.

RQ4: To what extent can the proposed model estimate the condition of mechanical systems?

Eventually, the newly constructed NPBN was validated. Although goodness-of-fit, a prominent validation approach for BNs, could not be assessed due to the lack of data, two analyses were conducted: a scenario analysis and a sensitivity analysis. Because these analyses require the selection of inputs and outputs, the latter were chosen as follows:

- **Input variables:** 'AHU Age', 'Maintenance interval' and 'Design & Construction quality';
- **Output variables:** 'Filters', 'Coils' and 'Fans'.

This distinction aligns with the model's objective, where the inputs are the "the easily retrievable information" and the outputs the air handling units components.

The scenario analysis aimed to observe the model's behaviour when subjected to extreme case values. The six scenarios were defined by considering different combinations of (opposite) input values, for instance a high value of 'AHU Age' and a small value of 'Maintenance interval'. Unsurprisingly, the output distributions reflected the rank correlation coefficients elicited by the experts. Interestingly though, some configurations indicated inconsistencies in the model's parameters, such as the presence of excessively high correlations associated to 'Design & Construction quality'. Additionally, the analysis demonstrated that the exclusion of the environmental conditions from the graph in Figure 6.2 leads to unrealistic outcomes. Therefore, an hypothetical BN including the variable 'Environmental conditions' was created and subjected to a seventh scenario, illustrating the ease with which variables can be integrated in the existing network. Despite

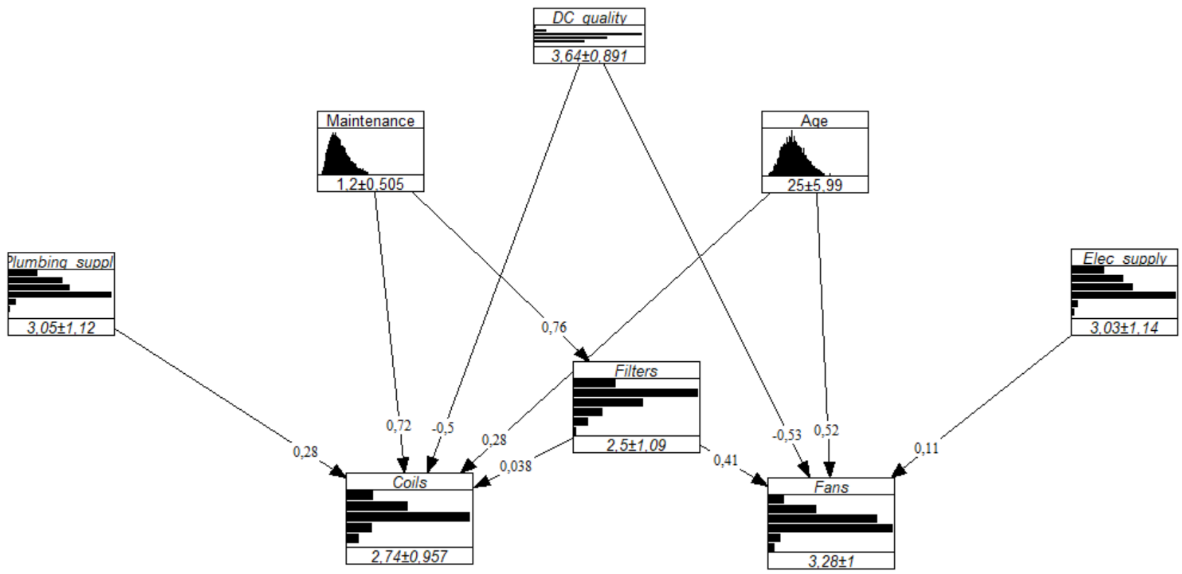


Figure 6.2: Quantified NPBN for the estimate of air handling units' condition.

being realized without field expertise, this addition yielded promising results which further refute the initial assumption of ignoring environmental conditions in the study of AHUs.

A global sensitivity analysis was then conducted to quantify the model outputs' sensitivity to variations of its inputs. This thesis presented an implementation of Sobol's first- (S_i) and total-order (S_{T_i}) indices, a prominent method still hardly explored in Bayesian modelling. These indices assess the contribution of an input variable to the variance of a given output variable, with high values indicating a strong contribution. The analysis suggested that (i) the first-order contributions of the input variables to an output's variance are highly influenced by their correlation with the latter, and (ii) all input variables have a significant total-order contribution to all outputs' variances ($\min_i S_{T_i} = 0.413$), thus making factor fixing absurd.

All in all, the answers to the four sub-research questions provided above allow to address the main research question tackled in this thesis:

> **How can Bayesian Networks be applied to estimate the condition of mechanical, electrical, and plumbing systems in the absence of empirical data ?**

After identifying the foreseeable obstacles in the construction of the NPBN, this research investigated and implemented a (partially) novel method to quantify the model in the complete absence of empirical data. The process followed through this deliverable provides a framework for practitioners and academics to reproduce a similar effort, regardless of the context at stake. The flowchart in Figure 6.3 illustrates the 'recipe' of building a Non-Parametric Bayesian Network, from the definition of the scope to the evaluation of the model.

6.2 Evaluation

It is essential to critically review the methods and results presented in this research to assess the extent to which its findings can be generalized. Therefore, this section investigates the relevance of the assumptions and methods adopted throughout the study.

The first key result is the creation of the 'global' graph encompassing all mechanical, electrical and plumbing (MEP) systems which served as the basis for the development of the NPBN. The classification of the MEP sub-systems, however, lacks specificity; each variable is composed of several sub-systems (as

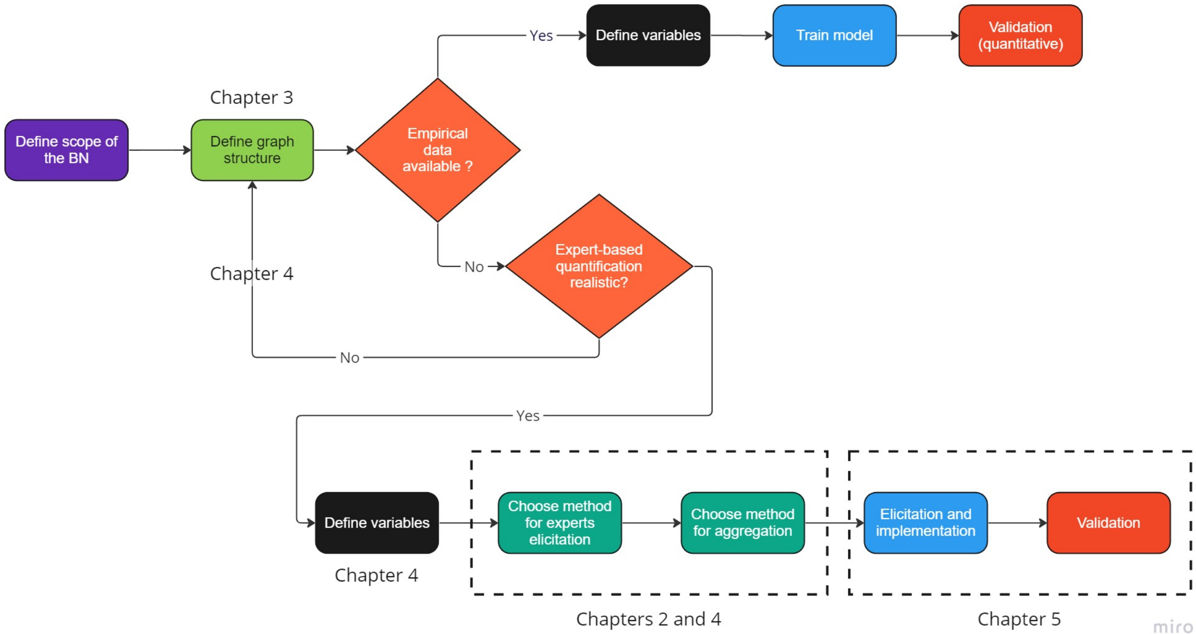


Figure 6.3: Construction process of the Non-Parametric Bayesian Network and associated chapters.

illustrated by the case study on AHUs), making the decomposition hardly usable as such. This issue may have strong repercussions on the development/quantification on a higher level: in Figure 6.2, for instance, the components of the AHU are linked with the sub-systems 'Plumbing supply elements' and 'Electrical supply elements'. For the related edges, rank correlations were assessed using experts' judgments as previously explained. A detailed study of these sub-systems, however, may reveal the existence of several sub-(sub-)systems that must be modelled individually, then voiding the effort put in the elicitation of these rank correlations.

This thesis was to date the first application of probabilities of concordance for the assessment of dependence in expert-based Non-Parametric BNs. While the experts' feedback indicated that this method is relevant and accessible, two elements may influence the fidelity of the elicited values. First, the closed-form equations used to compute the rank correlations from the concordance probabilities require the normal copula assumption, which is rarely verified despite being common for expert-elicited NPBNs. Second, some experts expressed difficulty assessing *unconditional* concordance probabilities, i.e. accounting for the uncertainty given that evidence on only one (parent) variable is available. Instead, the use of *conditional* concordance probabilities may help reduce the 'vagueness' perceived by some of the respondents, and is discussed in the Recommendations.

Furthermore, the present implementation of dependence-calibration for the aggregation of experts' individual assessments is also subject to criticisms. Although the participants' scores indicate a solid performance, the seed variables (unrelated to AHUs and buildings overall) merely allowed to quantify the experts' normative expertise, i.e. their familiarity with probabilistic assessments. As a result, the d-calibration scores - and thus the final decision-maker - completely disregard the experts' substantive expertise, a limitation highlighted by the largely superior scores obtained by the two least experienced respondents. The challenges faced in the selection of relevant seed variables reinforces the reluctance of certain authors to implement performance-based aggregation methods when empirical data is unavailable.

The structure of the NPBN constructed and displayed in Figure 6.2 is purposely oversimplistic due to the exploratory character of the research. The absence of the environmental conditions, for instance, illustrates the effort to limit the number of correlations to be elicited by experts at the expense of fidelity. While the addition of this variable proved to be trivial, other (structural) changes may have substantial effects on the previously determined *unconditional* correlations. Moreover, the elicitation of the marginal distributions was attributed significantly less resources than that of the dependence structure, and possibly includes a bias as both experts' perspectives are strongly intertwined with the TU Delft and its campus.

Lastly, the scale chosen for the assessment of the components' condition is highly specific and might align with only a few grading systems for building condition assessment. As a result, the model's outputs may lack of interpretability for stakeholders unfamiliar with NEN 2767 and its philosophy, thus further limiting the generalisability of the network. Nevertheless, the approach adopted for the quantification of the NPBN remains relevant, not only for the estimation of building components' conditions but also for any system whose condition/state is assessed on an ordinal scale. Examples include the condition of infrastructure works, education (grades) and customer satisfaction, which can be evaluated on an ordinal scale.

6.3 Recommendations

Following the discussion on the research's findings and their limitations, this section presents suggestions for future academic endeavours as well as recommendations regarding the continuation of the effort initiated in this thesis.

6.3.1 Probabilities of concordance for dependence assessment

The implementation of a novel method for the elicitation of dependence assessments, based on probabilities of concordance, presents an alternative to existing approaches based on conditional probabilities of exceedance and direct assessments of rank correlations. Surprisingly, this study is to the best of the author's knowledge the first implementation of this approach for Bayesian Networks, which yielded promising results. However, some axes of improvement arose, particularly regarding the use of *conditional* concordance probabilities.

Let $\mathbf{Z}$ be a vector of covariates, then for each $\mathbf{z} \in \mathbb{R}^p$, the concordance probability between two random variables X and Y is:

$$P_c(X, Y|\mathbf{Z} = \mathbf{z}) = P(x_1 \leq x_2 | y_1 \leq y_2, \mathbf{Z} = \mathbf{z}).$$

with (x_1, y_1) and (x_2, y_2) two random draws of variables X and Y. To illustrate the practical impact of this modification, let us consider the edge between 'Maintenance interval' and 'Coils' (Figure 6.2). To assess $P(Maint, Coils|Age)$, an expert would be presented the following question:

> "Two buildings A and B are randomly selected among all non-residential buildings in the Netherlands. Given that the AHUs in **buildings A and B are both z years old**, and that the AHU in building A is maintained more regularly than in building B ($y_A \leq y_B$), what is the probability that the coils are in better condition in building A than building B ($x_A \leq x_B$) ?"

However, for conditional concordance probabilities to be relevant, additional research should investigate the extent to which their use facilitates the elicitation and whether the protocol used to retrieve rank correlations from unconditional concordance probabilities still applies. The latter is crucial as the validity of the closed-form formulas used to retrieve rank correlations (subsection 2.2.1) are not trivial in the conditional case, and is demonstrated in Appendix G. Moreover, the dependence in $\mathbf{z}$ can be eliminated by assuming that $P_c(X, Y|\mathbf{Z} = \mathbf{z})$ is constant in $\mathbf{z}$, similarly to conditional exceedance probabilities. All in all, the implementation of conditional probabilities of concordance could enhance the intepretability of the questionnaire presented to the experts and therefore the quality of the collected assessments.

6.3.2 Dependence-calibration in data-sparse environments

The application of dependence-calibration in this research highlighted the challenge of selecting appropriate seed variables when little to no empirical data is available. Past studies relied on the wide availability of data in their field (e.g. Nogal et al., 2019, with traffic data) or knowledge of the 'true' dependence structure (e.g. Rongen et al., 2023). However, due to the emerging nature of the method, there is no guideline in data-sparse environments, sometimes constraining assessors to use equal weights decision-makers (Wang, Li, Dong, & Ding, 2019) or unrelated seed variables, as was the case in this study. Therefore, future research could focus on assessing the potential loss of accuracy between a BN quantified with field data and another with 'general knowledge' information, such as precipitation or physiological data.

6.3.3 Classification of MEP systems and graph

The previous section underlined the lack of specificity of the classification developed in this thesis for all MEP systems, which formed the basis of the 'global' network. Because of its foundational role in the development of the BN, the classification must be refined and a new graph structure should hence be created. In particular, the identification of relationships between building components is arguably a challenging task, as methods based on empirically observed correlations (e.g. Bortolini & Forcada, 2020) are limited by the availability of condition data. Therefore, researchers eager to identify such relationships may resort on field expertise and additional literature.

6.3.4 Is it worth it?

This thesis, conducted in eight months, resulted in the creation of a Non-Parametric Bayesian Network to estimate the condition of air handling units. As discussed previously, this case study aims to be the first step in the development of a network that encompasses all MEP systems. Yet, it is essential to assess the extent to which continuing this study's work may profit for industry participants. The experts interrogated during the elicitation unanimously indicated that, given accurate predictions, the creation of such network would be highly relevant to their practice. In addition to the prediction feature, the model can allow decision-makers to assess the influence of certain policy changes (maintenance interval, investment in high-quality components) on the condition of the system and its evolution over time.

Nonetheless, the amount of effort dedicated to the construction of the BN - for a yet undetermined predictive accuracy - raises skepticism. For asset owners/managers without access to empirical data, the construction of a NPBN similar to this study's may seem illogical. Despite the limitations of current condition assessment practices mentioned throughout the report, their sole purpose is the long-term planning of maintenance, which is reevaluated with detailed inspections of the different building components. Therefore, a (potential) marginal improvement to the current 'handmade' estimates might be of limited added value given the required upfront investment. Conversely, stakeholders with access to empirical data and/or internal field expertise are encouraged to build on this research's findings and continue developing an increasingly accurate model.

References

Abdul-Rahman, H., Wang, C., Wood, L. C., & Khoo, Y. M. (2014, April). Defects in Affordable Housing Projects in Klang Valley, Malaysia. *Journal of Performance of Constructed Facilities*, *28*(2), 272–285. Retrieved 2023-03-14, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000413 doi: 10.1061/(ASCE)CF.1943-5509.0000413

Ahzahar, N., Karim, N. A., Hassan, S. H., & Eman, J. (2011, January). A Study of Contribution Factors to Building Failures and Defects in Construction Industry. *Procedia Engineering*, *20*, 249–255. Retrieved 2023-03-27, from https://www.sciencedirect.com/science/article/pii/S1877705811029717 doi: 10.1016/j.proeng.2011.11.162

Alavi, H., Forcada, N., Bortolini, R., & Edwards, D. J. (2021, March). Enhancing occupants' comfort through BIM-based probabilistic approach. *Automation in Construction*, *123*, 103528. Retrieved 2023-03-08, from https://linkinghub.elsevier.com/retrieve/pii/S0926580520311080 doi: 10.1016/j.autcon.2020.103528

Alencastro, J., Fuertes, A., & de Wilde, P. (2018, January). The relationship between quality defects and the thermal performance of buildings. *Renewable and Sustainable Energy Reviews*, *81*, 883–894. Retrieved 2023-03-27, from https://linkinghub.elsevier.com/retrieve/pii/S1364032117311735 doi: 10.1016/j.rser.2017.08.029

ANSI, & ASHRAE. (2013). *Climatic Data for Building Design Standards.*

Asmone, A. S., & Chew, M. Y. L. (2020, October). Development of a design-for-maintainability assessment of building systems in the tropics. *Building and Environment*, *184*, 107245. Retrieved 2023-03-12, from https://linkinghub.elsevier.com/retrieve/pii/S0360132320306168 doi: 10.1016/j.buildenv.2020.107245

Atef, A., & Bristow, D. (2019, December). Capturing the spatial and operational interdependencies among building systems using building information modelling. *Structure and Infrastructure Engineering*, *15*(12), 1613–1629. Retrieved 2023-04-17, from https://doi.org/10.1080/15732479.2019.1636285 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15732479.2019.1636285) doi: 10.1080/15732479.2019.1636285

Au-Yong, C. P., Ali, A. S., & Ahmad, F. (2014, July). Improving occupants' satisfaction with effective maintenance management of HVAC system in office buildings. *Automation in Construction*, *43*, 31–37. Retrieved 2023-06-12, from https://www.sciencedirect.com/science/article/pii/S092658051400065X doi: 10.1016/j.autcon.2014.03.013

Ballester-Ripoll, R., & Leonelli, M. (2022). Computing Sobol indices in probabilistic graphical models. *Reliability Engineering and System Safety*, *225*. doi: 10.1016/j.ress.2022.108573

Barlow, R., & Hunter, L. (1960, February). Optimum Preventive Maintenance Policies. *Operations Research*, *8*(1), 90–100. Retrieved 2022-12-30, from http://pubsonline.informs.org/doi/10.1287/opre.8.1.90 doi: 10.1287/opre.8.1.90

Biswas, M. A. R., Robinson, M. D., & Fumo, N. (2016, December). Prediction of residential building energy consumption: A neural network approach. *Energy*, *117*, 84–92. Retrieved 2023-07-04, from https://www.sciencedirect.com/science/article/pii/S0360544216315006 doi: 10.1016/j.energy.2016.10.066

Bolger, F., & Rowe, G. (2015). The Aggregation of Expert Judgment: Do Good Things Come to Those Who Weight? *Risk Analysis*, *35*(1), 5–11. Retrieved 2023-07-18, from https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.12272 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.12272) doi: 10.1111/risa.12272

Bortolini, R., & Forcada, N. (2018, October). Building Inspection System for Evaluating the Technical Performance of Existing Buildings. *Journal of Performance of Constructed Facilities*, *32*(5), 04018073. Retrieved 2023-01-03, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0001220 doi: 10.1061/(ASCE)CF.1943-5509.0001220

Bortolini, R., & Forcada, N. (2019, November). A probabilistic-based approach to support the comfort performance assessment of existing buildings. *Journal of Cleaner Production*, *237*, 117720. Retrieved 2023-01-27, from https://linkinghub.elsevier.com/retrieve/pii/S0959652619325764 doi: 10.1016/j.jclepro.2019.117720

Bortolini, R., & Forcada, N. (2020, November). A probabilistic performance evaluation for buildings and constructed assets. *Building Research & Information*, *48*(8), 838–855. Retrieved 2022-12-16, from https://www.tandfonline.com/doi/full/10.1080/09613218.2019.1704208 doi: 10.1080/09613218.2019.1704208

Brooker, P. (2011, October). Experts, Bayesian Belief Networks, rare events and aviation risk estimates. *Safety Science*, *49*(8), 1142–1155. Retrieved 2023-03-07, from https://www.sciencedirect.com/science/article/pii/S0925753511000762 doi: 10.1016/j.ssci.2011.03.006

Carretero-Ayuso, M. J., Moreno-Cansado, A., & García-Sanz-Calcedo, J. (2017, July). An analysis of technical facilities failures in modern Spanish houses. *Building Services Engineering Research and Technology*, *38*(4), 490–498. Retrieved 2023-03-14, from https://doi.org/10.1177/0143624417701802 (Publisher: SAGE Publications Ltd STM) doi: 10.1177/0143624417701802

Carretero-Ayuso, M. J., Rodríguez-Jiménez, C., Bienvenido-Huertas, D., & Moyano, J. J. (2021, July). Interrelations between the types of damages and their original causes in the envelope of buildings. *Journal of Building Engineering*, *39*, 102235. Retrieved 2023-02-27, from https://linkinghub.elsevier.com/retrieve/pii/S2352710221000917 doi: 10.1016/j.jobe.2021.102235

CEN. (2010). *Maintenance. Maintenance terminology* (Withdrawn ed.). (OCLC: 9521555828)

Chen, S. H., & Pollino, C. A. (2012, November). Good practice in Bayesian network modelling. *Environmental Modelling & Software*, *37*, 134–145. Retrieved 2023-02-05, from https://linkinghub.elsevier.com/retrieve/pii/S1364815212001041 doi: 10.1016/j.envsoft.2012.03.012

Chen, Y., Wen, J., Chen, T., & Pradhan, O. (2018, July). Bayesian Networks for Whole Building Level Fault Diagnosis and Isolation. *International High Performance Buildings Conference*. Retrieved from https://docs.lib.purdue.edu/ihpbc/266

Cheung, H., & Braun, J. E. (2015, December). *Development of Fault Models for Hybrid Fault Detection and Diagnostics Algorithm: October 1, 2014 – May 5, 2015* (Tech. Rep. No. NREL/SR-5500-65030). National Renewable Energy Lab. (NREL), Golden, CO (United States). Retrieved 2023-04-12, from https://www.osti.gov/biblio/1235409 doi: 10.2172/1235409

Chew, M. (2005, April). Defect analysis in wet areas of buildings. *Construction and Building Materials*, *19*(3), 165–173. Retrieved 2023-03-14, from https://linkinghub.elsevier.com/retrieve/pii/S0950061804001527 doi: 10.1016/j.conbuildmat.2004.07.005

Chong, W.-K., & Low, S.-P. (2006, August). Latent Building Defects: Causes and Design Strategies to Prevent Them. *Journal of Performance of Constructed Facilities*, *20*(3), 213–221. Retrieved 2023-03-14, from https://ascelibrary.org/doi/10.1061/%28ASCE%290887-3828%282006%2920%3A3%28213%29 doi: 10.1061/(ASCE)0887-3828(2006)20:3(213)

Clemen, R. T., Fischer, G. W., & Winkler, R. L. (2000, August). Assessing Dependence: Some Experimental Results. *Management Science*, *46*(8), 1100–1115. Retrieved 2023-05-01, from https://pubsonline.informs.org/doi/10.1287/mnsc.46.8.1100.12023 doi: 10.1287/mnsc.46.8.1100.12023

Clemen, R. T., & Reilly, T. (1999, February). Correlations and Copulas for Decision and Risk Analysis. *Management Science*, *45*(2), 208–224. Retrieved 2023-06-05, from https://pubsonline.informs.org/doi/10.1287/mnsc.45.2.208 doi: 10.1287/mnsc.45.2.208

Clemen, R. T., & Winkler, R. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, *19*(2), 187–203. doi: 10.1023/A:1006917509560

Cooke, R. M. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. New York, NY, US: Oxford University Press. (Pages: xii, 321)

Cooke, R. M., & Goossens, L. L. (2008, May). TU Delft expert judgment data base. *Reliability Engineering & System Safety*, *93*(5), 657–674. Retrieved 2023-03-22, from https://linkinghub.elsevier.com/retrieve/pii/S0951832007000944 doi: 10.1016/j.ress.2007.03.005

Cowell, R. G. (Ed.). (1999). *Probabilistic networks and expert systems*. New York: Springer.

Cui, L., Du, Y.-M., & Sun, C. (2023, June). On system reliability for time-varying structure. *Reliability Engineering & System Safety*, *234*, 109146. Retrieved 2023-03-01, from https://linkinghub.elsevier.com/retrieve/pii/S0951832023000613 doi: 10.1016/j.ress.2023.109146

Cypko, M. A., Stoehr, M., Kozniewski, M., Druzdzel, M. J., Dietz, A., Berliner, L., & Lemke, H. U. (2017, November). Validation workflow for a clinical Bayesian network model in multidisciplinary decision making in head and neck oncology treatment. *International Journal of Computer Assisted Radiology and Surgery*, *12*(11), 1959–1970. Retrieved 2023-08-16, from http://link.springer.com/10.1007/s11548-017-1531-7 doi: 10.1007/s11548-017-1531-7

Das, S., & Chew, M. Y. L. (2011, December). Generic Method of Grading Building Defects Using FMECA to Improve Maintainability Decisions. *Journal of Performance of Constructed Facilities*, *25*(6), 522–533. Retrieved 2022-12-16, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000206 doi: 10.1061/(ASCE)CF.1943-5509.0000206

Dejaco, M. C., Re Cecconi, F., & Maltese, S. (2017, January). Key Performance Indicators for Building Condition Assessment. *Journal of Building Engineering*, *9*, 17–28. Retrieved 2023-05-23, from https://linkinghub.elsevier.com/retrieve/pii/S2352710216302893 doi: 10.1016/j.jobe.2016.11.004

Derumigny, A., & Fermanian, J.-D. (2019, July). A classification point-of-view about conditional Kendall's tau. *Computational Statistics & Data Analysis*, *135*, 70–94. Retrieved 2023-07-18, from https://www.sciencedirect.com/science/article/pii/S0167947319300271 doi: 10.1016/j.csda.2019.01.013

De Silva, D. T., Setunge, S., & Tran, H. (2022, March). Effect of Runtime on the Deterioration of HVAC Components in Building Services. *Journal of Infrastructure Systems*, *28*(1), 04021049. Retrieved 2023-03-14, from https://ascelibrary.org/doi/10.1061/%28ASCE%29IS.1943-555X.0000656 (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)IS.1943-555X.0000656

Doguc, O., & Ramirez-Marquez, J. E. (2009, February). A generic method for estimating system reliability using Bayesian networks. *Reliability Engineering & System Safety*, *94*(2), 542–550. Retrieved 2023-01-13, from https://linkinghub.elsevier.com/retrieve/pii/S0951832008001804 doi: 10.1016/j.ress.2008.06.009

Eleftheriadis, G., & Hamdy, M. (2017). Impact of building envelope and mechanical component degradation on the whole building performance: A review paper. In (Vol. 132, pp. 321–326). (ISSN: 1876-6102) doi: 10.1016/j.egypro.2017.09.739

Endrenyi, J., Aboresheid, S., Allan, R., Anders, G., Asgarpoor, S., Billinton, R., … Singh, C. (2001, November). The present status of maintenance strategies and the impact of maintenance on reliability. *IEEE Transactions on Power Systems*, *16*(4), 638–646. Retrieved 2022-12-16, from http://ieeexplore.ieee.org/document/962408/ doi: 10.1109/59.962408

Endrenyi, J., Anders, G., & Leite da Silva, A. (1998, May). Probabilistic evaluation of the effect of maintenance on reliability. An application [to power systems]. *IEEE Transactions on Power Systems*, *13*(2), 576–583. Retrieved 2022-12-16, from http://ieeexplore.ieee.org/document/667385/ doi: 10.1109/59.667385

Fang, H.-B., Fang, K.-T., & Kotz, S. (2002, July). The Meta-elliptical Distributions with Given Marginals. *Journal of Multivariate Analysis*, *82*(1), 1–16. Retrieved 2023-06-13, from https://www.sciencedirect.com/science/article/pii/S0047259X01920172 doi: 10.1006/jmva.2001.2017

Faqih, F., & Zayed, T. (2021a, January). A comparative review of building component rating systems. *Journal of Building Engineering*, *33*, 101588. Retrieved 2023-06-29, from https://www.sciencedirect.com/science/article/pii/S2352710220301406 doi: 10.1016/j.jobe.2020.101588

Faqih, F., & Zayed, T. (2021b, March). Defect-based building condition assessment. *Building and Environment*, *191*, 107575. Retrieved 2023-03-08, from https://linkinghub.elsevier.com/retrieve/pii/S0360132320309422 doi: 10.1016/j.buildenv.2020.107575

Faqih, F., Zayed, T., & Soliman, E. (2020, December). Factors and defects analysis of physical and environmental condition of buildings. *Journal of Building Pathology and Rehabilitation*, *5*(1), 19. Retrieved 2023-03-08, from https://link.springer.com/10.1007/s41024-020-00084-0 doi: 10.1007/s41024-020-00084-0

Forcada, N., Macarulla, M., Fuertes, A., Casals, M., Gangolells, M., & Roca, X. (2012, August). Influence of Building Type on Post-Handover Defects in Housing. *Journal of Performance of Constructed Facilities*, *26*(4), 433–440. Retrieved 2023-03-14, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF

.1943-5509.0000225 doi: 10.1061/(ASCE)CF.1943-5509.0000225

Forcada, N., Macarulla, M., Gangolells, M., & Casals, M. (2014, September). Assessment of construction defects in residential buildings in Spain. *Building Research & Information*, *42*(5), 629–640. Retrieved 2023-03-12, from http://www.tandfonline.com/doi/abs/10.1080/09613218.2014.922266 doi: 10.1080/09613218.2014.922266

Forcada, N., Macarulla, M., Gangolells, M., Casals, M., Fuertes, A., & Roca, X. (2013, December). Posthandover Housing Defects: Sources and Origins. *Journal of Performance of Constructed Facilities*, *27*(6), 756–762. Retrieved 2023-03-14, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000368 doi: 10.1061/(ASCE)CF.1943-5509.0000368

French, S. (2011, March). Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, *105*(1), 181–206. Retrieved 2023-07-20, from http://link.springer.com/10.1007/s13398-011-0018-6 doi: 10.1007/s13398-011-0018-6

Georgiou, J., Love, P., & Smith, J. (1999, January). A comparison of defects in houses constructed by owners and registered builders in the Australian State of Victoria. *Structural Survey*, *17*(3), 160–169. Retrieved 2023-03-20, from https://doi.org/10.1108/02630809910291343 (Publisher: MCB UP Ltd) doi: 10.1108/02630809910291343

Grussing, M. N., & Liu, L. Y. (2014, June). Knowledge-Based Optimization of Building Maintenance, Repair, and Renovation Activities to Improve Facility Life Cycle Investments. *Journal of Performance of Constructed Facilities*, *28*(3), 539–548. Retrieved 2023-01-11, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000449 doi: 10.1061/(ASCE)CF.1943-5509.0000449

Hanea, A. M. (2008). *Algorithms for non-parametric Bayesian belief nets* (Doctoral dissertation, Technische Universiteit Delft). Retrieved 2023-06-17, from https://repository.tudelft.nl/islandora/object/uuid%3A2cd65be4-1708-4967-8f09-902a40386090

Hanea, A. M., Hemming, V., & Nane, G. F. (2022). Uncertainty Quantification with Experts: Present Status and Research Needs. *Risk Analysis*, *42*(2), 254–263. Retrieved 2023-07-15, from https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.13718 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.13718) doi: 10.1111/risa.13718

Hanea, A. M., Hilton, Z., Knight, B., & P. Robinson, A. (2022). Co-designing and building an expert-elicited non-parametric Bayesian network model: demonstrating a methodology using a Bonamia Ostreae spread risk case study. *Risk Analysis*, *42*(6), 1235–1254. Retrieved 2023-08-16, from https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.13904 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.13904) doi: 10.1111/risa.13904

Hanea, A. M., Kurowicka, D., & Cooke, R. M. (2006). Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets. *Quality and Reliability Engineering International*, *22*(6), 709–729. Retrieved 2023-05-01, from https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.808 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.808) doi: 10.1002/qre.808

Hanea, A. M., Morales-Nápoles, O., & Ababei, D. (2015, December). Non-parametric Bayesian networks: Improving theory and reviewing applications. *Reliability Engineering & System Safety*, *144*, 265–284. Retrieved 2023-02-15, from https://linkinghub.elsevier.com/retrieve/pii/S0951832015002331 doi: 10.1016/j.ress.2015.07.027

Hanea, D., Jagtman, H., & Ale, B. (2012, April). Analysis of the Schiphol Cell Complex fire using a Bayesian belief net based model. *Reliability Engineering & System Safety*, *100*, 115–124. Retrieved 2023-07-12, from https://linkinghub.elsevier.com/retrieve/pii/S095183201200004X doi: 10.1016/j.ress.2012.01.002

Hart, B. T., & Pollino, C. A. (2009). Bayesian modelling for risk-based environmental water allocation.

Hassanain, M. A., Fatayer, F., & Al-Hammad, A.-M. (2016, April). Design-Phase Maintenance Checklist for Electrical Systems. *Journal of Performance of Constructed Facilities*, *30*(2), 06015003. Retrieved 2023-03-20, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000774 (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)CF.1943-5509.0000774

Hauashdh, A., Jailani, J., Abdul Rahman, I., & Al-Fadhali, N. (2022, April). Factors affecting the number of building defects and the approaches to reduce their negative impacts in Malaysian public universities' buildings. *Journal of Facilities Management*, *20*(2), 145–171. Retrieved 2023-03-08, from https://www.emerald.com/insight/content/doi/10.1108/JFM-11-2020-0079/full/html doi: 10.1108/JFM-11-2020-0079

Hoffmann, T. (2018, November). *Why Quality Installation Matters in the HVAC Industry.* Retrieved 2023-06-14, from https://tomhoffmannairconditioning.com/why-quality-installation-matters-in-the-hvac-industry/

Hosamo, H. H., Nielsen, H. K., Kraniotis, D., Svennevig, P. R., & Svidt, K. (2023, February). Digital Twin framework for automated fault source detection and prediction for comfort performance evaluation of existing non-residential Norwegian buildings. *Energy and Buildings*, *281*, 112732. Retrieved 2023-01-27, from https://linkinghub.elsevier.com/retrieve/pii/S0378778822009033 doi: 10.1016/j.enbuild.2022.112732

Hosamo, H. H., Svennevig, P. R., Svidt, K., Han, D., & Nielsen, H. K. (2022, April). A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics. *Energy and Buildings*, *261*, 111988. Retrieved 2023-04-17, from https://linkinghub.elsevier.com/retrieve/pii/S0378778822001591 doi: 10.1016/j.enbuild.2022.111988

Hu, J., Xiong, B., Zhang, Z., & Wang, J. (2022, December). A continuous Bayesian network regression model for estimating seismic liquefaction-induced settlement of the free-field ground. *Earthquake Engineering & Structural Dynamics*, eqe.3804. Retrieved 2023-03-01, from https://onlinelibrary.wiley.com/doi/10.1002/eqe.3804 doi: 10.1002/eqe.3804

Hua, Y., Göçer, ., & Göçer, K. (2014, September). Spatial mapping of occupant satisfaction and indoor environment quality in a LEED platinum campus building. *Building and Environment*, *79*, 124–137. Retrieved 2023-03-08, from https://linkinghub.elsevier.com/retrieve/pii/S0360132314001334 doi: 10.1016/j.buildenv.2014.04.029

Hänninen, M. (2014, December). Bayesian networks for maritime traffic accident prevention: Benefits and challenges. *Accident Analysis & Prevention*, *73*, 305–312. Retrieved 2023-02-23, from https://linkinghub.elsevier.com/retrieve/pii/S0001457514002735 doi: 10.1016/j.aap.2014.09.017

Ilozor, B., Okoroh, M., Egbu, C., & Archicentre. (2004, March). Understanding residential house defects in Australia from the State of Victoria. *Building and Environment*, *39*(3), 327–337. Retrieved 2023-03-27, from https://linkinghub.elsevier.com/retrieve/pii/S0360132303001616 doi: 10.1016/j.buildenv.2003.07.002

Ishak, S. N. H., Chohan, A. H., & Ramly, A. (2007, August). Implications of design deficiency on building maintenance at post-occupational stage. *Journal of Building Appraisal*, *3*(2), 115–124. Retrieved 2023-04-01, from https://doi.org/10.1057/palgrave.jba.2950061 doi: 10.1057/palgrave.jba.2950061

Islam, R., Nazifa, T. H., & Mohamed, S. F. (2019, June). Factors Influencing Facilities Management Cost Performance in Building Projects. *Journal of Performance of Constructed Facilities*, *33*(3), 04019036. Retrieved 2023-03-12, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0001284 doi: 10.1061/(ASCE)CF.1943-5509.0001284

Islam, R., Nazifa, T. H., Mohammed, S. F., Zishan, M. A., Yusof, Z. M., & Mong, S. G. (2021, July). Impacts of design deficiencies on maintenance cost of high-rise residential buildings and mitigation measures. *Journal of Building Engineering*, *39*, 102215. Retrieved 2023-03-08, from https://linkinghub.elsevier.com/retrieve/pii/S2352710221000711 doi: 10.1016/j.jobe.2021.102215

ISO. (2012). *ISO IDT 15686-2:2012 - Buildings and constructed assets - Service-life planning - Part 2: Service life prediction procedures.*

ISO. (2014). *ISO 55000:2014 - Asset management - Overview, principles and terminology.*

Josephson, P. E., & Hammarlund, Y. (1999, August). The causes and costs of defects in construction: A study of seven building projects. *Automation in Construction*, *8*(6), 681–687. Retrieved 2023-03-27, from https://www.sciencedirect.com/science/article/pii/S0926580598001149 doi: 10.1016/S0926-5805(98)00114-9

Kabir, S., & Papadopoulos, Y. (2019, June). Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: A review. *Safety Science*, *115*, 154–175. Retrieved 2023-02-23, from https://linkinghub.elsevier.com/retrieve/pii/S0925753518305435 doi: 10.1016/j.ssci.2019.02.009

Kaikkonen, L., Parviainen, T., Rahikainen, M., Uusitalo, L., & Lehikoinen, A. (2021). Bayesian Networks in Environmental Risk Assessment: A Review. *Integrated Environmental Assessment and Management*, *17*(1), 62–78. Retrieved 2023-02-27, from https://onlinelibrary.wiley.com/doi/abs/10.1002/ieam.4332 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ieam.4332) doi: 10.1002/ieam.4332

Kian, P. S. (2001). A REVIEW OF FACTORS AFFECTING BUILDING DEFECTS IN SINGAPORE. *Civil Engineering Dimension*, *3*(2).

Kitson, N. K., Constantinou, A. C., Guo, Z., Liu, Y., & Chobtham, K. (2023, January). A survey of Bayesian Network structure learning. *Artificial Intelligence Review*. Retrieved 2023-03-06, from https://doi.org/10.1007/s10462-022-10351-w doi: 10.1007/s10462-022-10351-w

Koot, P., Mendoza-Lugo, M. A., Paprotny, D., Morales-Nápoles, O., Ragno, E., & Worm, D. T. (2023, February). PyBanshee version (1.0): A Python implementation of the MATLAB toolbox BANSHEE for Non-Parametric Bayesian Networks with updated features. *SoftwareX*, *21*, 101279. Retrieved 2023-07-24, from https://linkinghub.elsevier.com/retrieve/pii/S2352711022001972 doi: 10.1016/j.softx.2022.101279

Kurowicka, D., & Cooke, R. M. (2005). Distribution-free continuous Bayesian belief. In *Modern statistical and mathematical methods in reliability.*

Kurowicka, D., & Cooke, R. M. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling.* John Wiley & Sons. (Google-Books-ID: dRVGNYU7RskC)

Kusiak, A., & Li, M. (2010, March). Cooling output optimization of an air handling unit. *Applied Energy*, *87*(3), 901–909. Retrieved 2023-04-26, from https://linkinghub.elsevier.com/retrieve/pii/S0306261909002499 doi: 10.1016/j.apenergy.2009.06.010

Kyrimi, E., Dube, K., Fenton, N., Fahmi, A., Neves, M. R., Marsh, W., & McLachlan, S. (2021, June). Bayesian networks in healthcare: What is preventing their adoption? *Artificial Intelligence in Medicine*, *116*, 102079. Retrieved 2023-02-23, from https://www.sciencedirect.com/science/article/pii/S0933365721000725 doi: 10.1016/j.artmed.2021.102079

Kyrimi, E., McLachlan, S., Dube, K., Neves, M. R., Fahmi, A., & Fenton, N. (2021, July). A comprehensive scoping review of Bayesian networks in healthcare: Past, present and future. *Artificial Intelligence in Medicine*, *117*, 102108. Retrieved 2023-03-01, from https://linkinghub.elsevier.com/retrieve/pii/S0933365721001019 doi: 10.1016/j.artmed.2021.102108

Lai, C. (1993, June). Approaching fault-free electrical installation in building. In *1993 International Conference on Electrical Installation Engineering in Europe* (pp. 110–113).

Lee, H., & Cha, J. H. (2016, November). New stochastic models for preventive maintenance and maintenance optimization. *European Journal of Operational Research*, *255*(1), 80–90. Retrieved 2022-12-30, from https://linkinghub.elsevier.com/retrieve/pii/S0377221716302399 doi: 10.1016/j.ejor.2016.04.020

Li, C., & Mahadevan, S. (2018, March). Sensitivity Analysis of a Bayesian Network. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, *4*(1), 011003. Retrieved 2023-08-06, from https://asmedigitalcollection.asme.org/risk/article/doi/10.1115/1.4037454/369982/Sensitivity-Analysis-of-a-Bayesian-Network doi: 10.1115/1.4037454

Lind, H., & Muyingo, H. (2012, February). Building maintenance strategies: planning under uncertainty. *Property Management*, *30*(1), 14–28. Retrieved 2022-12-18, from https://www.emerald.com/insight/content/doi/10.1108/02637471211198152/full/html doi: 10.1108/02637471211198152

Lupășteanu, V., Lupășteanu, R., & Chingălată, C. (2022, April). Condition assessment of buildings in Romania: A proposed method and case study. *Journal of Building Engineering*, *47*, 103814. Retrieved 2023-02-27, from https://linkinghub.elsevier.com/retrieve/pii/S2352710221016727 doi: 10.1016/j.jobe.2021.103814

Mendoza-Lugo, M. A., & Morales-Nápoles, O. (2023, July). Version 1.3-BANSHEE—A MATLAB toolbox for Non-Parametric Bayesian Networks. *SoftwareX*, *23*. Retrieved 2023-09-11, from https://www.softxjournal.com/article/S2352-7110(23)00175-9/fulltext (Publisher: Elsevier) doi: 10.1016/j.softx.2023.101479

Mendoza-Lugo, M. A., Morales-Nápoles, O., & Delgado-Hernández, D. J. (2022, March). A Non-parametric Bayesian Network for multivariate probabilistic modelling of Weigh-in-Motion System Data. *Transportation Research Interdisciplinary Perspectives*, *13*, 100552. Retrieved 2023-08-11, from https://linkinghub.elsevier.com/retrieve/pii/S259019822200015X doi: 10.1016/j.trip.2022.100552

Mirnaghi, M. S., & Haghighat, F. (2020, December). Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review. *Energy and Buildings*, *229*, 110492. Retrieved 2023-04-12, from https://www.sciencedirect.com/science/article/pii/S037877882031327X doi: 10.1016/j.enbuild.2020.110492

Morales-Nápoles, O., Delgado-Hernández, D. J., De-León-Escobedo, D., & Arteaga-Arcos, J. C. (2014, May). A continuous Bayesian network for earth dams' risk assessment: methodology and quantification. *Structure and Infrastructure Engineering*, *10*(5), 589–603. Retrieved 2023-02-15, from http://www.tandfonline.com/doi/abs/10.1080/15732479.2012.757789 doi: 10.1080/15732479.2012.757789

Morales-Nápoles, O., Hanea, A. M., & Worm, D. T. (2014). Experimental results about the assessments of conditional rank correlations by experts: Example with air pollution estimates. In *Safety, reliability and risk analysis: beyond the horizon: proceedings of the European Safety and Reliability Conference, ESREL 2013, Amsterdam, The Netherlands, 29 September-2 October 2013/.*

Morales-Nápoles, O., Kurowicka, D., & Roelen, A. (2008, May). Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety*, *93*(5), 699–710. Retrieved 2023-02-15, from https://linkinghub.elsevier.com/retrieve/pii/S095183200700097X doi: 10.1016/j.ress.2007.03.020

Morales-Nápoles, O., & Steenbergen, R. D. J. M. (2015, January). Large-Scale Hybrid Bayesian Network for Traffic Load Modeling from Weigh-in-Motion System Data. *Journal of Bridge Engineering*, *20*(1), 04014059. Retrieved 2023-08-15, from https://ascelibrary.org/doi/10.1061/%28ASCE%29BE.1943-5592.0000636 doi: 10.1061/(ASCE)BE.1943-5592.0000636

Morato, P., Papakonstantinou, K., Andriotis, C., Nielsen, J., & Rigo, P. (2022, January). Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes. *Structural Safety*, *94*, 102140. Retrieved 2023-03-02, from https://linkinghub.elsevier.com/retrieve/pii/S0167473021000631 doi: 10.1016/j.strusafe.2021.102140

Nakagawa, T. (2005). *Maintenance theory of reliability.* London: Springer.

NEN. (2006). *NEN 2767 Conditiemeting van bouw- en installatiedelen – Deel 1: Methodiek (Condition Assessment of Building and Installation Components – Part 1: Methodology).* NEN, Delft.

Nogal, M., Morales Nápoles, O., & O'Connor, A. (2019, September). Structured expert judgement to understand the intrinsic vulnerability of traffic networks. *Transportation Research Part A: Policy and Practice*, *127*, 136–152. Retrieved 2023-07-20, from https://linkinghub.elsevier.com/retrieve/pii/S0965856418301794 doi: 10.1016/j.tra.2019.07.006

Nojavan A., F., Qian, S. S., & Stow, C. A. (2017, January). Comparative analysis of discretization methods in Bayesian networks. *Environmental Modelling & Software*, *87*, 64–71. Retrieved 2023-07-18, from https://www.sciencedirect.com/science/article/pii/S1364815216308672 doi: 10.1016/j.envsoft.2016.10.007

Olanrewaju, A. L., Khamidi, M. F., & Idrus, A. (2010, May). Quantitative analysis of defects in Malaysian university buildings: Providers' perspective. *Journal of Retail & Leisure Property*, *9*(2), 137–149. Retrieved 2023-03-20, from http://link.springer.com/10.1057/rlp.2010.2 doi: 10.1057/rlp.2010.2

Olanrewaju, A. L., Tan, Y. Y., & Soh, S. N. (2021, January). Defect characterisations in the Malaysian affordable housing. *International Journal of Building Pathology and Adaptation*, *40*(4), 539–568. Retrieved 2023-03-12, from https://doi.org/10.1108/IJBPA-11-2018-0095 (Publisher: Emerald Publishing Limited) doi: 10.1108/IJBPA-11-2018-0095

Olubodun, F., & Mole, T. (1999, September). Evaluation of defect influencing factors in public housing in the UK. *Structural Survey*, *17*(3), 170–178. Retrieved 2022-12-26, from https://www.emerald.com/insight/content/doi/10.1108/02630809910291352/full/html doi: 10.1108/02630809910291352

Pan, W., & Thomas, R. (2015, August). Defects and Their Influencing Factors of Posthandover New-Build Homes. *Journal of Performance of Constructed Facilities*, *29*(4), 04014119. Retrieved 2023-03-27, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0000618 doi: 10.1061/(ASCE)CF.1943-5509.0000618

Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018, July). Machine Learning approach for Predictive Maintenance in Industry 4.0. In *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)* (pp. 1–6). Oulu: IEEE. Retrieved 2023-01-27, from https://ieeexplore.ieee.org/document/8449150/ doi: 10.1109/MESA.2018.8449150

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Wagenaar, D., Castellarin, A., Carisi, F., … Schröter, K. (2021, February). A probabilistic approach to estimating residential losses from different flood types. *Natural Hazards*, *105*(3), 2569–2601. Retrieved 2023-09-11, from https://doi.org/10.1007/s11069-020-04413-x doi: 10.1007/s11069-020-04413-x

Paton-Cole, V. P., & Aibinu, A. A. (2021, February). Construction Defects and Disputes in Low-Rise Residential Buildings. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *13*(1), 05020016. Retrieved 2023-03-27, from https://ascelibrary.org/doi/10.1061/%28ASCE%29LA.1943-4170.0000433 (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)LA.1943-4170.0000433

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Elsevier. Retrieved 2023-08-11, from https://linkinghub.elsevier.com/retrieve/pii/C20090276094 doi: 10.1016/C2009-0-27609-4

Pearl, J., & Russell, S. (2000). Bayesian Networks. *Citeseer*.

Phan, T. D., Smart, J. C. R., Capon, S. J., Hadwen, W. L., & Sahin, O. (2016, November). Applications of Bayesian belief networks in water resource management: A systematic review. *Environmental Modelling & Software*, *85*, 98–111. Retrieved 2023-02-27, from https://www.sciencedirect.com/science/article/pii/S1364815216304698 doi: 10.1016/j.envsoft.2016.08.006

Pitchforth, J., & Mengersen, K. (2013, January). A proposed validation framework for expert elicited Bayesian Networks. *Expert Systems with Applications*, *40*(1), 162–167. Retrieved 2023-08-06, from https://www.sciencedirect.com/science/article/pii/S0957417412008810 doi: 10.1016/j.eswa.2012.07.026

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., & Hart, B. T. (2007, August). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software*, *22*(8), 1140–1152. Retrieved 2023-08-06, from https://linkinghub.elsevier.com/retrieve/pii/S1364815206000788 doi: 10.1016/j.envsoft.2006.03.006

PwC. (2017). *Predictive maintenance 4.0 - Predict the unpredictable.*

Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. *Water Resources Research*, *51*(5), 3070–3092. Retrieved 2023-08-06, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016527 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR016527) doi: 10.1002/2014WR016527

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., … Maier, H. R. (2021, March). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, *137*, 104954. Retrieved 2023-08-16, from https://linkinghub.elsevier.com/retrieve/pii/S1364815220310112 doi: 10.1016/j.envsoft.2020.104954

Renooij, S. (2001, September). Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, *16*(3), 255–269. Retrieved 2023-06-05, from https://www.cambridge.org/core/journals/knowledge-engineering-review/article/probability-elicitation-for-belief-networks-issues-to-consider/9FF39B517D1C4FEF4C962B1276421BE4 (Publisher: Cambridge University Press) doi: 10.1017/S0269888901000145

Rongen, G., Morales-Nápoles, O., & Kok, M. (2022, August). Expert judgment-based reliability analysis of the Dutch flood defense system. *Reliability Engineering & System Safety*, *224*, 108535. Retrieved 2023-02-15, from https://linkinghub.elsevier.com/retrieve/pii/S0951832022001880 doi: 10.1016/j.ress.2022.108535

Rongen, G., Morales-Nápoles, O., & Kok, M. (2023). *Structured expert elicitation of dependence between river tributaries using Non-Parametric Bayesian Networks.*

Rosone, M. C. (2023, March). *Ductwork Sizing & Other Mistakes That Hinder Your HVAC Comfort.* Retrieved 2023-06-14, from https://aristair.com/ductwork-sizing-other-mistakes-that-hinder-your-hvac-comfort-2/

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., … Wu, Q. (2019, April). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. Retrieved 2023-08-16, from https://linkinghub.elsevier.com/retrieve/pii/S1364815218302822 doi: 10.1016/j.envsoft.2019.01.012

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010, February). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. Retrieved 2023-08-06, from https://linkinghub.elsevier.com/retrieve/pii/S0010465509003087 doi: 10.1016/j.cpc.2009.09.018

Santiago, A. R., Antunes, M., Barraca, J. P., Gomes, D., & Aguiar, R. L. (2019, April). Predictive Maintenance System for Efficiency Improvement of Heating Equipment. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 93–98). Newark, CA, USA: IEEE. Retrieved 2023-03-08, from https://ieeexplore.ieee.org/document/8848208/ doi: 10.1109/BigDataService.2019.00019

Scutari, M., Graafland, C. E., & Gutiérrez, J. M. (2019, December). Who learns better Bayesian network structures: Accuracy and speed of structure learning algorithms. *International Journal of Approximate Reasoning*, *115*, 235–253. Retrieved 2023-03-02, from https://linkinghub.elsevier.com/retrieve/pii/S0888613X19301434 doi: 10.1016/j.ijar.2019.10.003

Shiguihara, P., Lopes, A. D. A., & Mauricio, D. (2021). Dynamic Bayesian Network Modeling, Learning, and Inference: A Survey. *IEEE Access*, *9*, 117639–117648. Retrieved 2023-03-01, from https://ieeexplore.ieee.org/document/9514854/ doi: 10.1109/ACCESS.2021.3105520

Silva, A., & de Brito, J. (2019). Do we need a buildings' inspection, diagnosis and service life prediction software? *Journal of Building Engineering*, *22*, 335–348. doi: 10.1016/j.jobe.2018.12.019

Sklar, M. (1959). *Fonctions de répartition à N dimensions et leurs marges.*

Sommerville, J., & McCosh, J. (2006, January). Defects in new homes: an analysis of data on 1,696 new UK houses. *Structural Survey*, *24*(1), 6–21. Retrieved 2023-03-14, from https://doi.org/10.1108/02630800610654397 (Publisher: Emerald Group Publishing Limited) doi: 10.1108/02630800610654397

Straub, A. (2002, October). Using a condition-dependent approach to maintenance to control costs and performances. *Journal of Facilities Management*, *1*(4), 380–395. Retrieved 2022-12-16, from https://www.emerald.com/insight/content/doi/10.1108/14725960310808079/full/html doi: 10.1108/14725960310808079

Straub, A. (2012). Maintenance and Repair. In *International Encyclopedia of Housing and Home* (pp. 186–194). Elsevier. Retrieved 2022-12-18, from https://linkinghub.elsevier.com/retrieve/pii/B9780080471631005233 doi: 10.1016/B978-0-08-047163-1.00523-3

Sui Pheng, L., & Wee, D. (2001, January). Improving maintenance and reducing building defects through ISO 9000. *Journal of Quality in Maintenance Engineering*, *7*(1), 6–24. Retrieved 2023-03-28, from https://doi.org/10.1108/13552510110386865 (Publisher: MCB UP Ltd) doi: 10.1108/13552510110386865

Sullivan, G., Pugh, R., Melendez, A. P., & Hunt, W. D. (2010, August). *Operations & Maintenance Best Practices - A Guide to Achieving Operational Efficiency (Release 3)* (Tech. Rep. Nos. PNNL-19634, 1034595). Retrieved 2023-01-03, from http://www.osti.gov/servlets/purl/1034595/ doi: 10.2172/1034595

Taal, A., & Itard, L. (2020a, May). Fault detection and diagnosis for indoor air quality in DCV systems: Application of 4S3F method and effects of DBN probabilities. *Building and Environment*, *174*, 106632. Retrieved 2023-03-29, from https://linkinghub.elsevier.com/retrieve/pii/S0360132319308455 doi: 10.1016/j.buildenv.2019.106632

Taal, A., & Itard, L. (2020b, October). P&ID-based automated fault identification for energy performance diagnosis in HVAC systems: 4S3F method, development of DBN models and application to an ATES system. *Energy and Buildings*, *224*, 110289. Retrieved 2023-03-30, from https://linkinghub.elsevier.com/retrieve/pii/S0378778819337843 doi: 10.1016/j.enbuild.2020.110289

Taal, A., & Itard, L. (2022, May). Automated energy performance diagnosis of HVAC systems by the 4S3F method. *CLIMA 2022 conference*, 2022: CLIMA 2022 The 14th REHVA HVAC World Congress. Retrieved 2023-03-29, from https://proceedings.open.tudelft.nl/clima2022/article/view/183 (Publisher: CLIMA 2022 conference) doi: 10.34641/CLIMA.2022.183

Uusitalo, L. (2007, May). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, *203*(3-4), 312–318. Retrieved 2023-02-27, from https://linkinghub.elsevier.com/retrieve/pii/S0304380006006089 doi: 10.1016/j.ecolmodel.2006.11.033

Uusitalo, L., Tomczak, M. T., Müller-Karulis, B., Putnis, I., Trifonova, N., & Tucker, A. (2018, May). Hidden variables in a Dynamic Bayesian Network identify ecosystem level change. *Ecological Informatics*, *45*, 9–15. Retrieved 2023-03-01, from https://linkinghub.elsevier.com/retrieve/pii/S1574954117301747 doi: 10.1016/j.ecoinf.2018.03.003

Waddicor, D. A., Fuentes, E., Sisó, L., Salom, J., Favre, B., Jiménez, C., & Azar, M. (2016, June). Climate change and building ageing impact on building energy performance and mitigation measures application: A case study in Turin, northern Italy. *Building and Environment*, *102*, 13–25. Retrieved 2023-03-14, from https://linkinghub.elsevier.com/retrieve/pii/S0360132316300737 doi: 10.1016/j.buildenv.2016.03.003

Wang, F., Li, H., Dong, C., & Ding, L. (2019, November). Knowledge representation using non-parametric Bayesian networks for tunneling risk analysis. *Reliability Engineering & System Safety*, *191*, 106529. Retrieved 2023-09-17, from https://www.sciencedirect.com/science/article/pii/S0951832018315710 doi: 10.1016/j.ress.2019.106529

Watt, D. (2007). *Building pathology: principles and practice* (2nd ed ed.). Oxford ; Malden, MA: Blackwell Publishing. (OCLC: ocm85814127)

Weeks, D. J., & Leite, F. (2021, April). Facility Defect and Cost Reduction by Incorporating Maintainability Knowledge Transfer Using Maintenance Management System Data. *Journal of Performance of Constructed Facilities*, *35*(2), 04021004. Retrieved 2023-03-12, from https://ascelibrary.org/doi/10.1061/%28ASCE%29CF.1943-5509.0001569 doi: 10.1061/(ASCE)CF.1943-5509.0001569

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., & Morales-Nápoles, O. (2017, May). Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions. *European Journal of Operational Research*, *258*(3), 801–819. Retrieved 2023-07-07, from https://www.sciencedirect.com/science/article/abs/pii/S0377221716308517 (Publisher: North-Holland) doi: 10.1016/j.ejor.2016.10.018

Yacob, S., Ali, A. S., & Au-Yong, C. P. (2022). *Managing Building Deterioration: Prediction Model for Public Schools in Developing Countries*. Singapore: Springer Singapore. Retrieved 2023-03-14, from https://link.springer.com/10.1007/978-981-16-5860-0 doi: 10.1007/978-981-16-5860-0

Yan, Y., Cai, J., Li, T., Zhang, W., & Sun, L. (2021, June). Fault prognosis of HVAC air handling unit and its components using hidden-semi Markov model and statistical process control. *Energy and Buildings*, *240*, 110875. Retrieved 2023-04-17, from https://linkinghub.elsevier.com/retrieve/pii/S0378778821001596 doi: 10.1016/j.enbuild.2021.110875

Yan, Y., Luh, P. B., & Pattipati, K. R. (2020, October). Fault Prognosis of Key Components in HVAC Air-Handling Systems at Component and System Levels. *IEEE Transactions on Automation Science and Engineering*, *17*(4), 2145–2153. (Conference Name: IEEE Transactions on Automation Science and Engineering) doi: 10.1109/TASE.2020.2979166

Yu, Y., Woradechjumroen, D., & Yu, D. (2014, October). A review of fault detection and diagnosis methodologies on air-handling units. *Energy and Buildings*, *82*, 550–562. Retrieved 2023-04-19, from https://linkinghub.elsevier.com/retrieve/pii/S0378778814005246 doi: 10.1016/j.enbuild.2014.06.042

Zalejska, J. A., & Hungria, G. R. (2019, January). Defects in newly constructed residential buildings: owners' perspective. *International Journal of Building Pathology and Adaptation*, *37*(2), 163–185. Retrieved 2023-03-14, from https://doi.org/10.1108/IJBPA-09-2018-0077 (Publisher: Emerald Publishing Limited) doi: 10.1108/IJBPA-09-2018-0077

Zerrouki, H., Estrada-Lugo, H. D., Smadi, H., & Patelli, E. (2019). Applications of Bayesian Networks in Chemical and Process Industries: A Review. In *Proceedings of the 29th European Safety and Reliability Conference (ESREL)* (pp. 3122–3129). Research Publishing Services. Retrieved 2023-03-01, from http://rpsonline.com.sg/proceedings/9789811127243/html/0914.xml doi: 10.3850/978-981-11-2724-3_0914-cd

Zhang, G., & Thai, V. V. (2016, August). Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review. *Safety Science*, *87*, 53–62. Retrieved 2023-02-27, from https://linkinghub.elsevier.com/retrieve/pii/S092575351630025X doi: 10.1016/j.ssci.2016.03.019

Zhang, N. L., & Poole, D. (1996, December). Exploiting Causal Independence in Bayesian Network Inference. *Journal of Artificial Intelligence Research*, *5*, 301–328. Retrieved 2023-03-22, from https://jair.org/index.php/jair/article/view/10178 doi: 10.1613/jair.305

Zhang, X.-Y., Trame, M., Lesko, L., & Schmidt, S. (2015). Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT: Pharmacometrics & Systems Pharmacology*, *4*(2), 69–79. Retrieved 2023-08-06, from https://onlinelibrary.wiley.com/doi/abs/10.1002/psp4.6 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/psp4.6) doi: 10.1002/psp4.6

Zhao, H.-Z., & Magoulès, F. (2012, August). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592. Retrieved 2023-06-07, from https://

www.sciencedirect.com/science/article/pii/S1364032112001438 (Publisher: Pergamon) doi: 10.1016/j.rser.2012.02.049

Appendix A

NEN 2767-1

This appendix is based on NEN (2006). Any change in recent versions is thus not integrated.

Created in the early 2000s, the NEN 2767 standard has been implemented in the Netherlands in an attempt to limit the subjectivity of inspections in the context of building maintenance. The method provided in 2767-1 allows asset owners to obtain an objective diagnostic of the condition of their assets through a score ranging from 1 to 6.

First, buildings are decomposed in a set of components: floor, walls, roof, stairs ... Three different levels of decomposition are prescribed in NEN 2767-2 and are the basis for the standard. Then, for each component, the observed defects are reported and assessed on three aspects: significance, extent and intensity.

Significance	**Definition**	**Example of defects**
Severe defect	Adversely affects the function of the building or installation component	Wood rot, cracks in a central heating boiler's system
Serious defect	Causes degradation of the building or installation component without directly affecting its functionality	Weathering, erosion, a defect that leads to installations leaking
Minor defect	Does not adversely affect the functionality of the building or installation component	Discolouration due to aging, improper attachment of sub-components

Table A.1: Subdivision significance.

Extent score	**Percentage**	**Description**
Extent 1	$< 2\%$	The defect occurs occasionally
Extent 2	2% to 10%	The defect occurs locally
Extent 3	10% to 30%	The defect occurs regularly
Extent 4	30% to 70%	The defect occurs considerably
Extent 5	$\geq 70\%$	The defect occurs generally

Table A.2: Subdivision extent.

Intensity score	**Designation**	**Explanation**
Intensity 1	Initial stage	The defect is hardly observable
Intensity 2	Advanced stage	The defect is clearly observed on the surface
Intensity 3	Final stage	The defect is easily observed, the process of degradation is irreversible and can hardly develop

Table A.3: Subdivision intensity.

Afterwards, the condition score of the individual component is obtained using the criteria evaluated above using a matrix provided in the standard. An example is given in the case of a *minor defect* (Table A.4):

Intensity/extent	Extent 1	Extent 2	Extent 3	Extent 4	Extent 5
Intensity 1	1	1	1	2	
Intensity 2	1	1	1	2	3
Intensity 3	1	1	2	3	4

Table A.4: Matrix of resulting condition class - minor defects.

Similar matrix are defined for serious and severe defects but not presented here in a concern of space. A condition score is now obtained for individual building components. Note that cases exist where a component is affected by multiple defects for which the standard provides an alternative method not presented here. Then, each defect/component is attributed a weight factor according to Table A.5 use in the following way: the factor is multiplied by the extent of the defect (in %), and this new measure is summed on all the components inspected. The result is finally divided by the sum of the extents, and a final outcome is obtained.

Condition class	**Weight factor**
1	1
2	1.02
3	1.1
4	1.3
5	1.7
6	2

Table A.5: Weight factor as a function of component's condition score.

For instance, five items defined by a couple (weight factor, extent) $= (w_k, e_k), k \in [1, 5]$ would lead to a final outcome x of : $x = \frac{1}{e_1+e_2+e_3+e_4+e_5} \sum_{k=1}^{5} w_k * e_k$

Finally, the outcome x is used to get the condition score of the whole building through Table A.6.

Outcome	**Condition class**
Outcome < 1.01	1
$1.01 <$ Outcome ≤ 1.04	2
$1.04 <$ Outcome ≤ 1.15	3
$1.15 <$ Outcome ≤ 1.4	4
$1.4 <$ Outcome ≤ 1.78	5
$1.78 <$ Outcome	6

Table A.6: Building's condition score as a function of the outcome x.

Appendix B

Probability of concordance, probability of exceedance and rank correlation

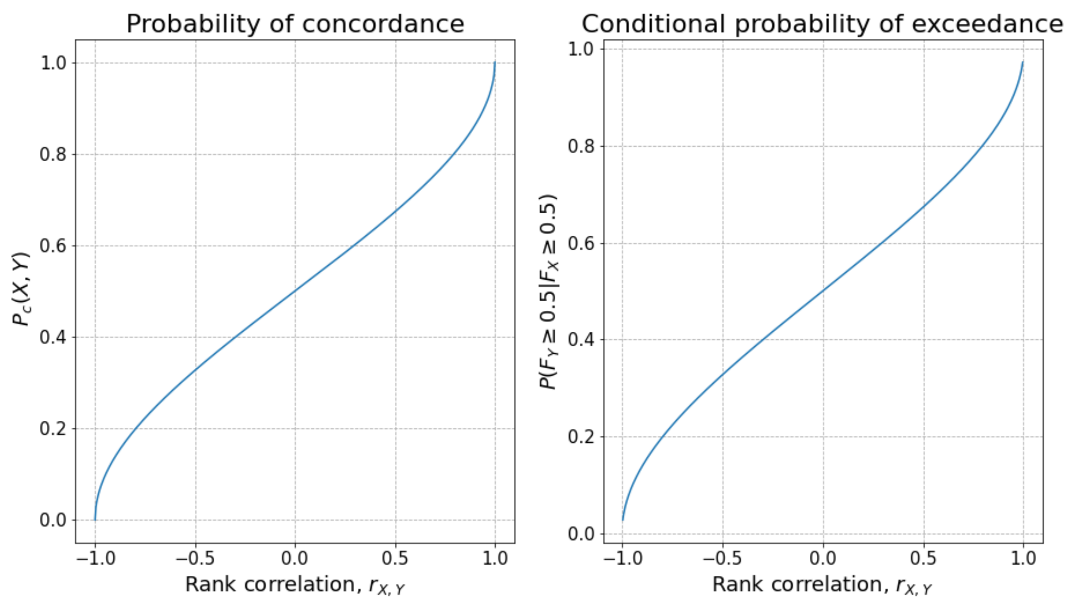


Figure B.1: Probability of concordance (left) and conditional probability of exceedance (right) as a function of rank correlations.

Appendix C

Questionnaire for experts' judgments

Date
June 2023

From:
Benjamin Ramousse
b.ramousse@student.tudelft.nl

Questionnaire

Information

By participating in this interview, you accept the terms laid in the additional consent terms presented at the end of this document.

Name:
Company and Role:
Years of experience:

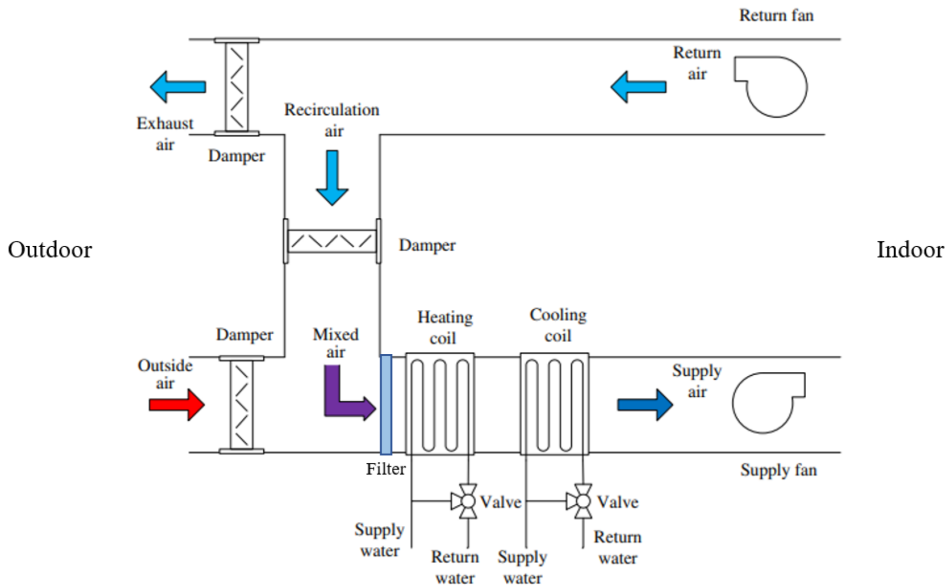
Introduction

This questionnaire aims to quantify statistical correlations[1] between several building components and other exogenous variables. The following paragraphs introduce the system of interest as well as the context of the research.

Consider two buildings taken randomly from all non-residential buildings in the Netherlands. The building's physical characteristics (materials, methods of construction) and occupancy are similar. These buildings are both equipped with a central Air Handling Unit (Luchtbehandelingkast) with the configuration shown in the figure below.

Figure 1: Schematic diagram of an AHU

[1] Statistical correlations measure the extent to which two variables are (linearly) related, and is in the range [-1,1]. A correlation of -1 entails that two variables move exactly in opposite directions (as one increases, the other decreases), 0 that these variables are independent and 1 that these variables move exactly in the same direction.



This questionnaire focuses on the **gradual** deterioration of three elements: the (supply/exhaust) **fans**, (heating/cooling) **coils** and (supply/exhaust) **filters**. Their deterioration is affected by several factors, including exogenous variables (e.g age) and the condition of other building components. The network below attempts to model these dependencies, which are represented by arrows.

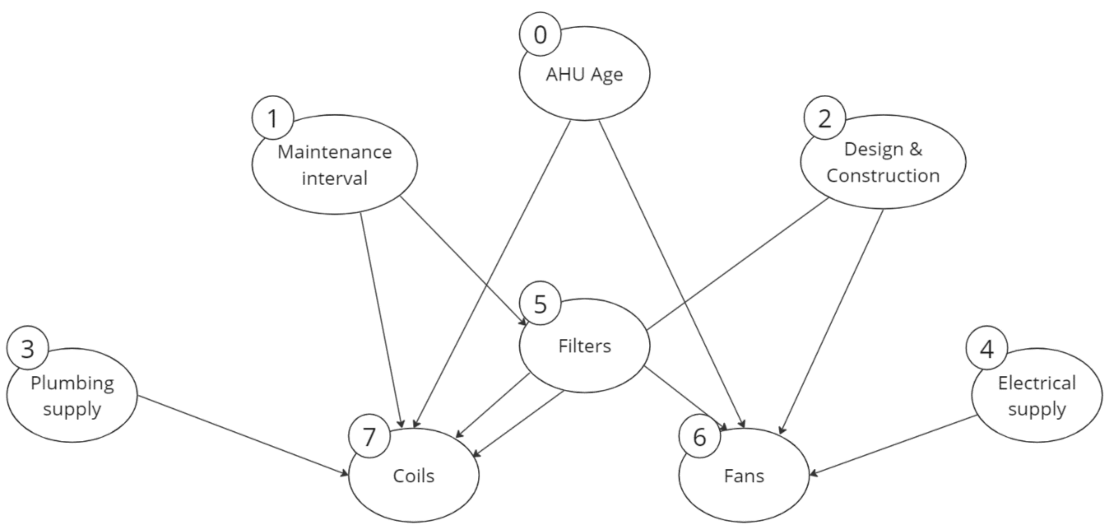


Figure 2: Structure of the network of an AHU.

Each variable is defined as follows:

V0 = Age of the Air Handling Unit (LBK)

V1 = Maintenance interval (in years) of the Air Handling Unit

V2 = Design and Construction quality (1-5: very bad to very good)

V3 = Plumbing supply: boilers, chillers, pumps.

V4 = Electrical supply: transformers, emergency power generators.
V5 = Filters
V6 = Fans
V7 = Coils

Variables 3 to 7 represent the **condition score** of different building components as defined in NEN 2767 (1-6), where 1 represents an excellent condition aand 6 a very bad one.

In the questions below, $P(x_1 \leq x_2 | y_1 \leq y_2)$ refers to the probability that a particular value of the variable X_1 is inferior to a particular value of variable X_2 **given that** a particular value of variable $\boldsymbol{Y_1}$ is inferior to a particular value Y_2.

Questions

1. You observe that the AHU in building 1 is maintained more regularly than in building 2 (i.e $v_{1,1} \leq v_{1,2}$), what is the probability that the filters are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{5,1} \leq v_{5,2}$)
$$P\left(v_{5,1} \leq v_{5,2} \middle| v_{1,1} \leq v_{1,2}\right) =$$

2. You observe that the AHU in building 1 is more recent than in building 2 (i.e $v_{0,1} \leq v_{0,2}$), what is the probability that the fans are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{6,1} \leq v_{6,2}$)
$$P\left(v_{6,1} \leq v_{6,2} \middle| v_{0,1} \leq v_{0,2}\right) =$$

3. You observe that the design and the installation of the AHU are of lesser quality in building 1 than in building 2 (i.e $v_{2,1} \leq v_{2,2}$), what is the probability that the fans are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2? $v_{6,1} \leq v_{6,2}$)
$$P\left(v_{6,1} \leq v_{6,2} \middle| v_{2,1} \leq v_{2,2}\right) =$$

4. You observe that the electrical supply system is in better condition in building 1 than building 2 (i.e $v_{4,1} \leq v_{4,2}$), what is the probability that the fans are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{6,1} \leq v_{6,2}$)
$$P\left(v_{6,1} \leq v_{6,2} \middle| v_{4,1} \leq v_{4,2}\right) =$$

5. You observe that the filters are in better condition in building 1 than building 2 (i.e $v_{5,1} \leq v_{5,2}$), what is the probability that the fans are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{6,1} \leq v_{6,2}$)
$$P\left(v_{6,1} \leq v_{6,2} \middle| v_{5,1} \leq v_{5,2}\right) =$$

6. You observe that the AHU in building 1 is more recent than in building 2 (i.e $v_{0,1} \leq v_{0,2}$), what is the probability that the coils are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{7,1} \leq v_{7,2}$)
$$P\left(v_{7,1} \leq v_{7,2} \middle| v_{0,1} \leq v_{0,2}\right) =$$

7. You observe that the AHU in building 1 is maintained more regularly than in building 2 (i.e $v_{1,1} \leq v_{1,2}$), what is the probability that the coils are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{7,1} \leq v_{7,2}$)
$$P\left(v_{7,1} \leq v_{7,2} \middle| v_{1,1} \leq v_{1,2}\right) =$$

8. You observe that the design and the installation of the AHU are of lesser quality in building 1 than in building 2 (i.e $v_{2,1} \leq v_{2,2}$), what is the probability that the coils are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{7,1} \leq v_{7,2}$)
$$P\left(v_{7,1} \leq v_{7,2} \middle| v_{2,1} \leq v_{2,2}\right) =$$

9. You observe that the plumbing supply system is in better condition in building 1 than building 2 (i.e $v_{3,1} \leq v_{3,2}$), what is the probability that the coils are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{7,1} \leq v_{7,2}$)
$$P\left(v_{7,1} \leq v_{7,2} \middle| v_{3,1} \leq v_{3,2}\right) =$$

10. You observe that the filters are in better condition in building 1 than building 2 (i.e $V_{5,1} \leq V_{5,2}$), what is the probability that the coils are in better condition in building 1 than in building 2? (i.e the condition score is higher in building 2, $v_{7,1} \leq v_{7,2}$)
$$P\left(v_{7,1} \leq v_{7,2} \middle| v_{5,1} \leq v_{5,2}\right) =$$

Seed questions

Introduction

We consider in this second part the hourly precipitations (uursom van de neerslag) measured in three weather stations: Rotterdam, Gilze-Rijen (nearby Breda) and Eindhoven between the 1st January 2023 and the 18th June 2023.

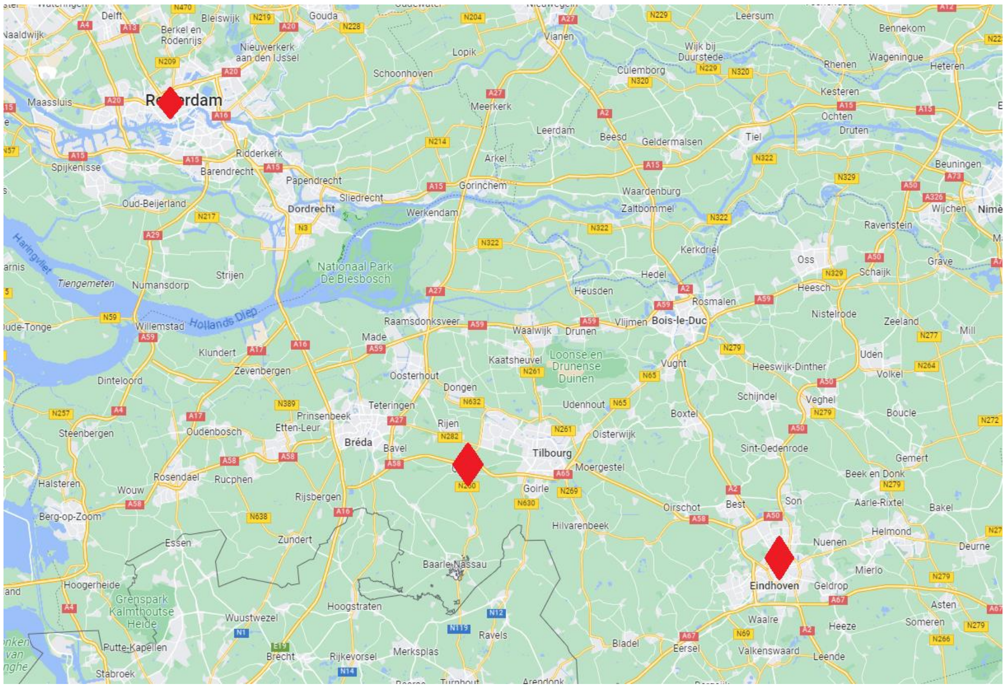
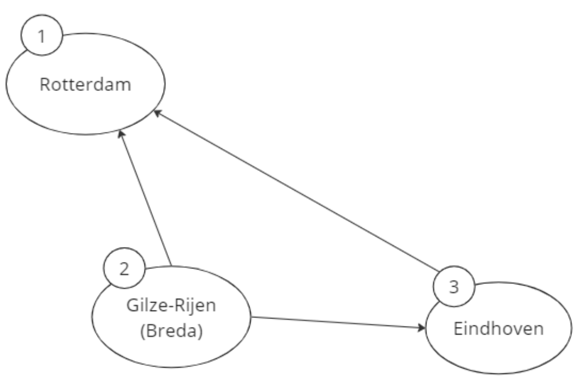


Figure 3: Location of the weather stations

Similarly to the first part, the graph below represents (assumed) statistical correlation between the variables, here the precipitation. Let us consider two moments defined by the hour (H1 and H2) taken randomly between the 01/01/2023 and 18/06/2023 (which could be different days/months/years).



1. You observe that the hourly precipitation in Gilze-Rijen is higher at H2 than H1 (i.e $v_{2,1} \leq v_{2,2}$), what is the probability that the hourly precipitation is higher in Rotterdam at H2 than H1?

$$P(v_{1,1} \leq v_{1,2} | v_{2,1} \leq v_{2,2}) =$$

2. You observe that the hourly precipitation in Eindhoven is higher at H2 than H1 (i.e $v_{3,1} \leq v_{3,2}$), what is the probability that the hourly precipitation is higher in Rotterdam at H2 than H1?

$$P(v_{1,1} \leq v_{1,2} | v_{3,1} \leq v_{3,2}) =$$

3. You observe that the hourly precipitation in Gilze-Rijen is higher at H2 than H1 (i.e $v_{2,1} \leq v_{2,2}$), what is the probability that the hourly precipitation is higher in Eindhoven at H2 than H1?

$$P(v_{3,1} \leq v_{3,2} | v_{2,1} \leq v_{2,2}) =$$

Additional questions:

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
1. I felt comfortable assessing probabilities.					
2. I consider my knowledge more practical than theoretical.					
3. The integration of such model to the condition assessment is relevant.					

- Should have environmental conditions been accounted for? Why?

Consent terms

You are being invited to participate in a research study titled "*An application of Bayesian Networks to MEP systems condition assessment in the Netherlands"*. This study is being done by Benjamin Ramousse from the TU Delft.

The purpose of this research study is to determine correlations between AHU components and other variables and will take you approximately 60 minutes to complete. The data will be used for publication in a master thesis. We will be asking you to assess some probability related to AHUs.

Your participation in this study is entirely voluntary and you can withdraw at any time. Your personal information (name, position, years of experience) will appear as an Appendix of the final deliverable but will not be associated with your answers.

Benjamin Ramousse – 0683831944
b.ramousse@student.tudelft.nl

Appendix D

List of questionnaire respondents

Name	Role	Organization	Experience (years)
Boris Hadzisejdic	Maintenance specialist	TU Delft	1.5
Marcel Klok	Maintenance engineer	TU Delft	43
Frans Strik	Installations advisor	Van Dorp	25
Arie Taal	Lecturer (indoor climate, energy transition)	De Haagse Hogeschool	40
Ziao Wang	PhD candidate	TU Delft	3

Table D.1: List of respondents to the questionnaire and details.

Appendix E

Correlation matrices

This appendix contains the correlation matrices obtained from experts' judgments for both the *main* and *seed* questions.

E.1 Expert A

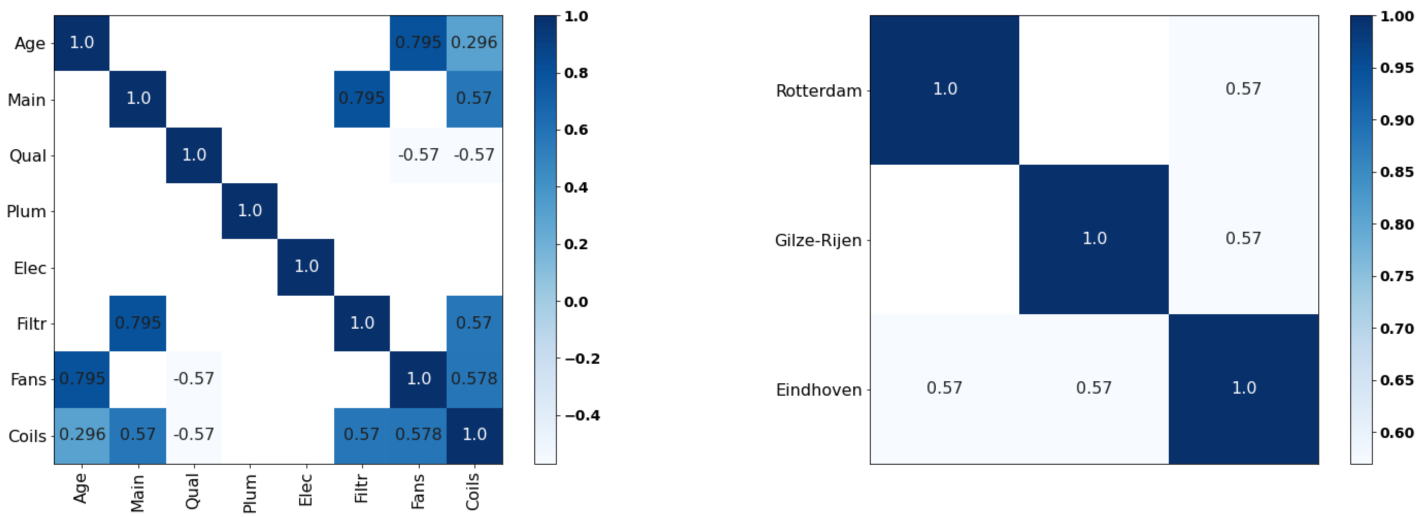


Figure E.1: Correlation matrices retrieved from expert A.

E.2 Expert B

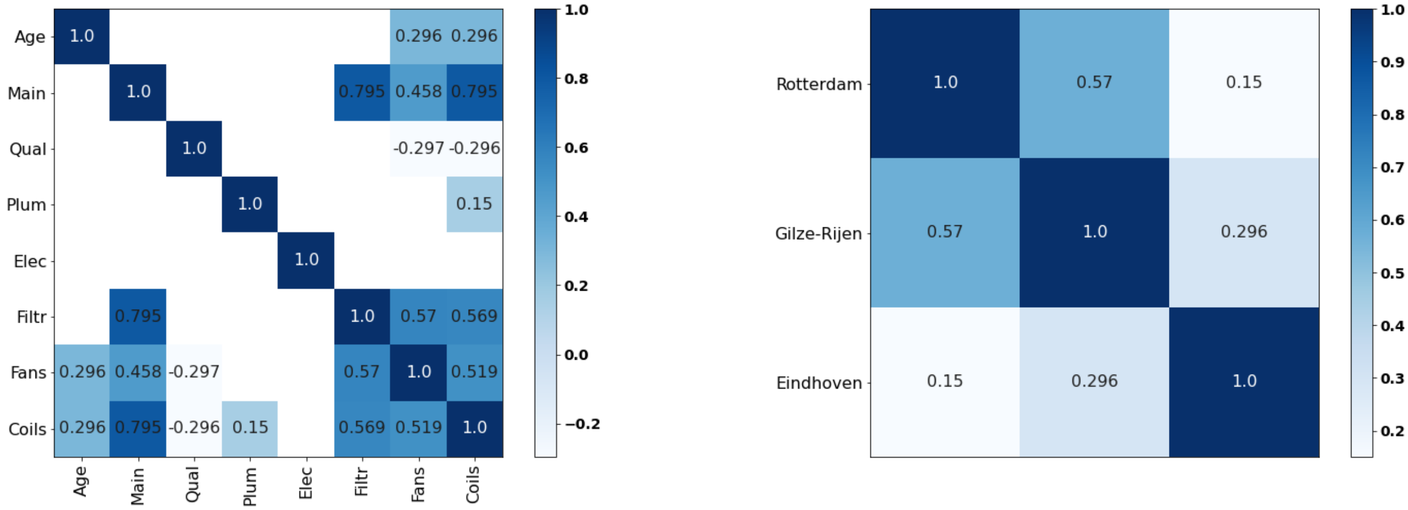


Figure E.2: Correlation matrices retrieved from expert B.

E.3 Expert C

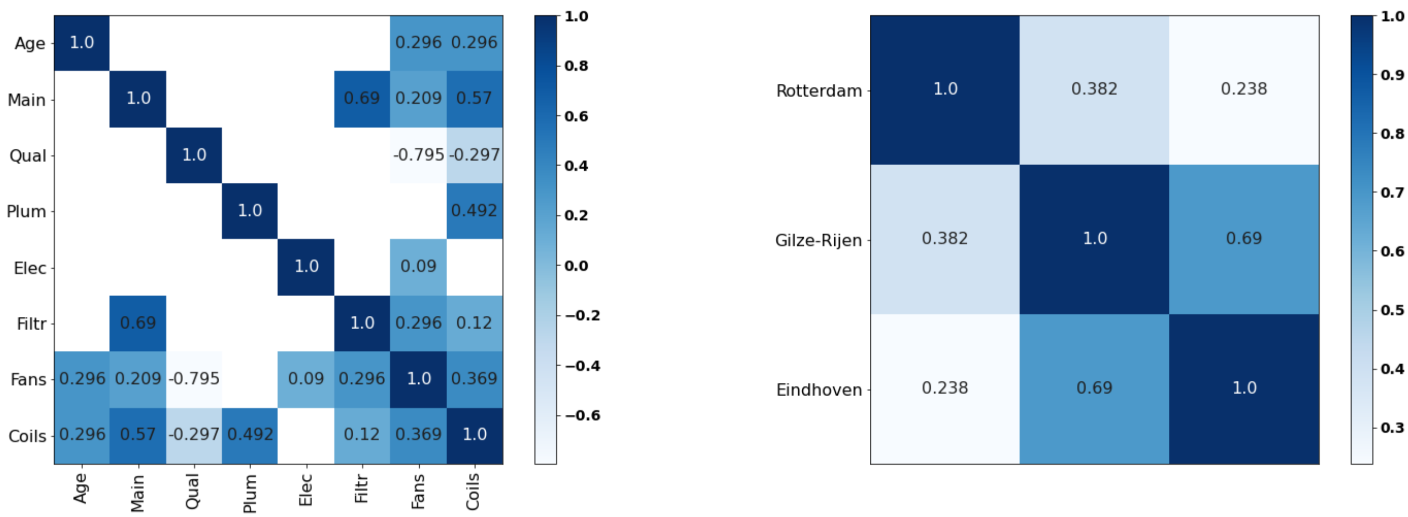


Figure E.3: Correlation matrices retrieved from expert C.

E.4 Expert D

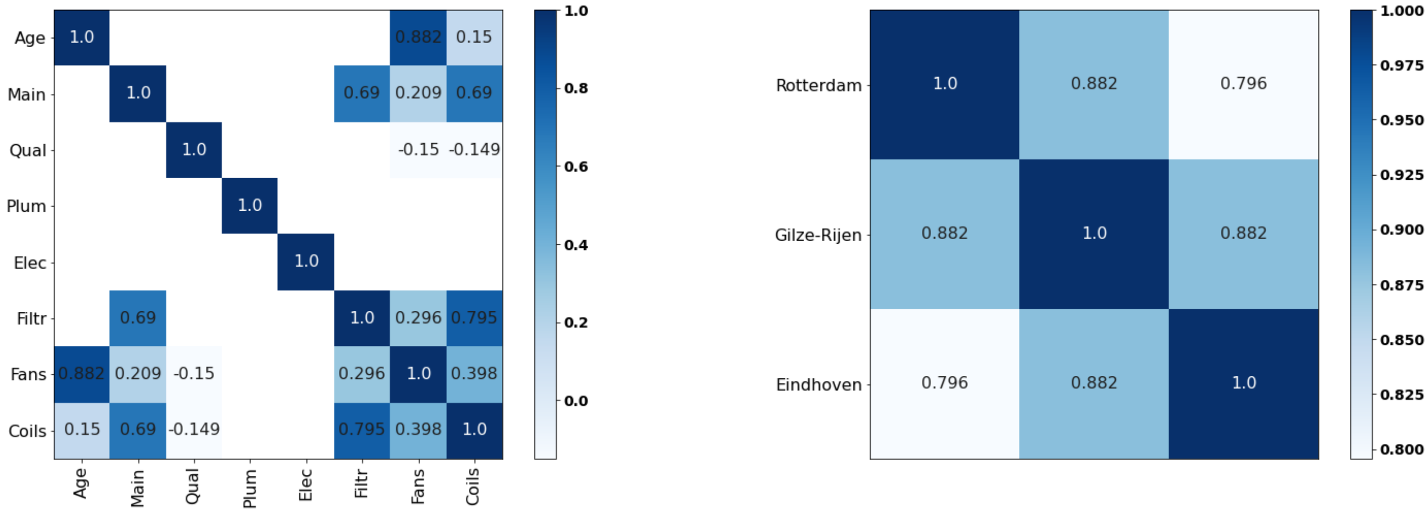


Figure E.4: Correlation matrices retrieved from expert D.

E.5 Expert E

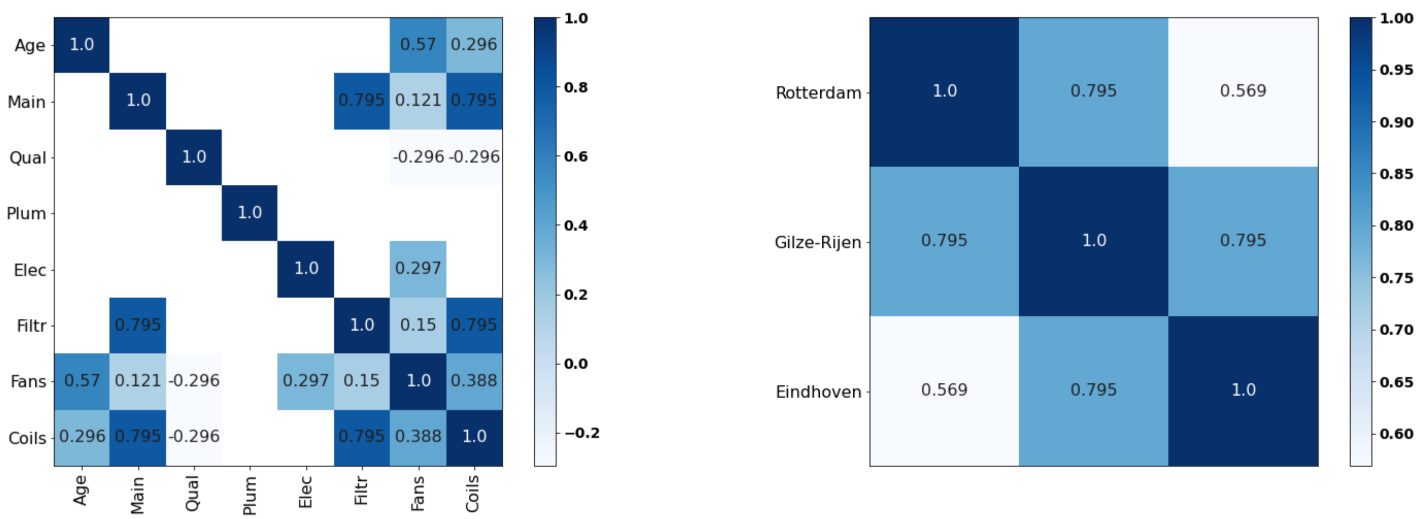


Figure E.5: Correlation matrices retrieved from expert E.

Appendix F

Implementation of Sobol's first- and total-order indices in Python

Note: the scripts below require the definition of the BN in pyBanshee for the use of the inference method (Koot et al., 2023). Modifications may be required to adjust to other modules.

```
def fo_sobol(input_var, output_var, dist:list, sample_size=100):
    ```
    Algorithm for the computation of Sobol's first-order sensitivity indices.

    Arguments:
    input_var: contains the node id of the input variables.
    output_var: contains the nodes Ids of the output variables.
    dist: list of the network's marginal distributions.
    sample_size: number of samples generated in the outer loop.

    Returns:
    to_indices: np.ndarray of Sobol's first-order sensitivity indices of dimension (len(output_var), len(input_var))
    '''

    start_time = time.time()
    n = len(output_var)
    k = len(input_var)
    fo_indices = np.zeros((n, k))
    for i in range(n):
        variance = np.var(data.iloc[:,output_var[i]])
        for j in range(k):
            means = []
            u = np.random.rand(sample_size)
            sample = dist[input_var[j]].ppf(u)
            for x in sample:
                F = inference([input_var[j]],
                          [x],
                          RBN,
                          data,
                          Output='mean',
                          SampleSize=1000)[0]
                means.append(F[output_var[i]-1])
            fo_indices[i,j] = np.var(means)/variance
    print("--- The algorithm ran in %s seconds ---" % (time.time() - start_time))
    return fo_indices

```

Listing F.1: Algorithm for Sobol's first-order indices.

```
def to_sobol(input_var:list, output_var:list, dist:list, sample_size=100):
    ```
    Algorithm for the computation of Sobol's total-order sensitivity indices.
```

```

    Arguments:
    input_var: contains the node id of the input variables.
    output_var: contains the nodes Ids of the output variables.
    dist: list of the network's marginal distributions.
    sample_size: number of samples generated in the outer loop.

    Returns:
    to_indices: np.ndarray of Sobol's total-order sensitivity indices of dimension (len(output_var), len(input_var))
    '''

    start_time = time.time()
    n = len(output_var)
    k = len(input_var)
    to_indices = np.zeros((n, k))
    for i in range(n):
        variance = np.var(data.iloc[:,output_var[i]])
        for j in range(k):
            list_var = []
            sample = np.random.rand(sample_size,6)
            #variables = input_var[:j] + input_var[j+1:]
            dist_temp = dist[:input_var[j]] + dist[input_var[j]+1:output_var[i]] + dist[output_var[i]+1:]
            variables = list(range(8))
            variables.pop(input_var[j])
            variables.pop(output_var[i]-1)
            for u in sample:
                x = [dist_temp[z].ppf(u[z]) for z in range(len(dist_temp))]
                F = inference(variables,
                          x,
                          RBN,
                          data,
                          Output='full',
                          SampleSize=1000)[0]
                list_var.append(np.var(F[-1]))
                print(list_var[-1])
            to_indices[i,j] = np.mean(list_var)/variance
    print("--- The algorithm ran in  %s seconds ---" % (time.time() - start_time))
    return to_indices
```

Listing F.2: Algorithm for Sobol's total-order indices.

Appendix G

Demonstration of formulas to transform conditional concordance probabilities in rank correlations

Let us first recall the set of equations used to retrieve an unconditional rank correlation (r) from an unconditional probability of concordance P_c:

$$\tau = 2P_c - 1 \tag{G.1}$$

$$\rho = sin\left(\frac{\pi\tau}{2}\right) \tag{G.2}$$

$$r = \frac{6}{\pi}arcsin\left(\frac{\rho}{2}\right). \tag{G.3}$$

This section aims to demonstrate the applicability of the formulas above in the event a *conditional* concordance probability is elicited. Let X, Y be two random variables with undetermined marginal distributions, (x_1, x_2) and (y_1, y_2) two random realizations of X and Y, $\mathbf{Z}$ a vector of covariates and $\mathbf{z} \in \mathbb{R}^p$ any realization of $\mathbf{Z}$. Then, the conditional concordance probability $P_{c|\mathbf{z}}$ is defined as follows:

$$P_{c|\mathbf{z}} = P(x_1 \leq x_2 | y_1 \leq y_2, \mathbf{Z} = \mathbf{z}).$$

For the remainder of the demonstration, the normal copula for NPBN is applied, i.e. all (conditional) bivariate copulas are considered normal. $P_{c|\mathbf{z}}$ is associated to the conditional Kendall's tau by (Derumigny & Fermanian, 2019):

$$\tau(X, Y|\mathbf{Z} = \mathbf{z}) = 2P_{c|\mathbf{z}} - 1.$$

Then, we know by Theorem 3.1 in Fang et al. (2002) that Equation G.2 is true for all pairs of random variables with a meta-elliptical distribution. Let $\tilde{X} = (X|\mathbf{Z} = \mathbf{z})$ and $\tilde{Y} = (Y|\mathbf{Z} = \mathbf{z})$. Resulting from the normal copula assumption, the copula $C(F_X, F_Y|F_{\mathbf{Z}}(\mathbf{z})) = C(F_{\tilde{X}}, F_{\tilde{Y}})$ is also normal and $(\tilde{X}, \tilde{Y})$ follows a meta-elliptical distribution. That implies:

$$\rho(\tilde{X}, \tilde{Y}) = sin\left(\frac{\pi\tau(\tilde{X}, \tilde{Y})}{2}\right).$$

$$\Leftrightarrow \rho(X, Y|\mathbf{Z} = \mathbf{z}) = sin\left(\frac{\pi\tau(X, Y|\mathbf{Z} = \mathbf{z})}{2}\right).$$

Lastly, Equation G.3 still applies under the normal copula assumption as the copula of $\tilde{X}$ and $\tilde{Y}$ is normal. Therefore (Kurowicka & Cooke, 2006):

$$r(X, Y|\mathbf{Z} = \mathbf{z}) = \frac{6}{\pi} arcsin\left(\frac{\rho(X, Y|\mathbf{Z} = \mathbf{z})}{2}\right).$$

For more information on meta-elliptical distributions, the reader is referred to Fang et al. (2002); for more information on conditional concordance probabilities and conditional Kendall's tau, see Derumigny and Fermanian (2019).