

# Non Parametric Continuous Bayesian Belief Nets in Reliability of Bridges under Traffic Load.

O. Morales Nápoles & R.D.J.M. Steenbergen  
*TNO Built Environment and Geosciences*

R.M. Cooke

*Resources for the Future & TU Delft, Faculty of EEMCS.*

**ABSTRACT:** Probabilistic design of infrastructure is based on estimates of design values. These are mostly done using classical statistical models. The statistical procedures used in structural reliability are restricted due to lack of data. The lack of data is most evident when the estimation is conditional on infrequent situations. Non Parametric Continuous Bayesian Belief Nets (NPCBBN) in combination with structured expert judgment could help approximate situations that are difficult to observe in the data. The advantage of using NPCBBNs is that data can be generated artificially once a model has been quantified. Additionally, the use of NPCBBN provides the user with the possibility to do fast updating on the joint distribution once evidence becomes available. This could be of great advantage for decision makers. In this paper first NPCBBN are described. As an example, with data from highway RW16 in the Netherlands for April 2008 a BBN for bridge design load is quantified. Variables included are axle load, number of axles per vehicle, velocity, total vehicular weight, and vehicular length. Possibilities to extend the model with structured expert judgment are discussed.

## 1 INTRODUCTION

One of the main concerns in structural reliability is the estimation of design values. Usually these correspond to return periods of hundreds or thousands of years. Data is often available for limited periods of time to estimate design values. Often, parametric probability distributions are fitted to these observations and then extreme quintiles are determined. However the associated return period is large compared with the length of the period of observations. For example while in the Netherlands a return period of 76,500(NEN 2002) years is required for bridge design loads, observations used to compute the design load might include only one month. A similar situation occurs for flood discharges where a return period of 1,250 years in river dikes is of interest while flood discharges are only available for a period of approximately a hundred years (van Noortwijk, Kalk, and Chhab 2003).

The statistical procedures used in structural reliability are restricted due to lack of data. For example, if design loads for bridges were to be computed with the conditional distribution of axle load given vehicles with 11 axles and total weight equal to 1,000kN the data available would be severely restricted. Non Para-

metric Continuous Bayesian Belief Nets (NPCBBN) in combination with structured expert judgment could help approximate situations that are difficult to observe in the data such as the one previously described. Data can be generated artificially for such situations. Additionally, the use of NPCBBN provides the user with the possibility to do fast updating on the joint distribution once evidence becomes available. This could be of great advantage when rapid exploration of a data base is required.

In this paper first NPCBBN are described. Then, with data from highway R16 in the Netherlands for April 2008 a BBN for bridge design load is quantified. Variables included are axle load, number of axles per vehicle, velocity, total vehicular weight, and vehicular length. Possibilities to extend the model with structured expert judgment are discussed.

## 2 CONCEPTS & DEFINITIONS.

Copulae are part of the building blocks of the graphical models to be used in this paper and for that reason basic concepts and definitions regarding them are introduced. The book by (Nelsen 1998) presents an introduction to the subject.

Bivariate copulae will be of special interest for us.

In this paper by copula (or copulae) we mean a bivariate copula (or bivariate copulae) unless otherwise specified. The *bivariate copula* or simply the *copula* of two random variables  $X$  and  $Y$  is the function  $C$  such that their joint distribution can be written as:

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)).$$

Copulae are functions that allow naturally the investigation of association between random variables. Measures of association such as the rank correlation or Kendall's tau may be expressed in terms of copulae (Nelsen 1998). The measures of association to be used in this paper are described in the appendix.

Of special interest in this paper will be the normal copula. Denote by  $\Phi_\rho$  the bivariate standard normal cumulative distribution function with correlation  $\rho$  and  $\Phi^{-1}$  the inverse of the univariate standard normal distribution function then

$$C_\rho(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)); (u,v) \in [0,1]^2$$

is called the *normal copula*.

Notice that  $\rho$  is a parameter of the normal copula. In the case of a conditional bivariate copula the parameter  $\rho_{1,2;3,\dots,n}$  is used instead.

## 2.1 Bayesian belief networks.

*Bayesian Belief Networks (BBNs)* are directed acyclic graphs whose nodes represent univariate random variables and whose arcs represent direct influences between adjacent nodes. These influences may be probabilistic or deterministic<sup>1</sup>. The graph of a BBN induces a non unique ordering of variables and stipulates that each variable is conditionally independent of its non-descendants given its parents. The parent set of variable  $X_i$  will be denoted as  $Pa(i)$ . Hence, to specify a joint distribution through a BBN the graph must be specified together with conditional probability functions of each variable given its parents (equation 1).

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i | x_{Pa(i)}) \quad (1)$$

If  $Pa(i) = \emptyset$  then  $f(x_i | x_{Pa(i)}) = f(x_i)$ . A BBN is then a concise and complete representation of the joint distribution. In the case that all nodes in the BBN are discrete then the functions to be specified are conditional probability tables (CPT) of each node given its parents. When variables are continuous, one

<sup>1</sup>When an influence is deterministic, nodes will be called *functional*. The discussion presented next refers to probabilistic influences unless otherwise specified.

possibility is to discretize them into a large enough number of states and use discrete BBNs. This approach might however turn out to be infeasible even for a modest sized model mainly because of the number of parameters to be specified. In general, the number of probabilities to be assessed  $K$  for a discrete BBN on  $n$  nodes with  $k_i$  states for each  $X_i$  for  $i = 1, \dots, n$  is:

$$K = \sum_{j \in S} k_j - |S| + \sum_{l \in C} (k_l - 1) \prod_{m \in Pa(l)} k_m \quad (2)$$

where  $S = \{X_j | Pa(j) = \emptyset\}$  and  $C = \{X_l | Pa(l) \neq \emptyset\}$  and  $|S| + |C| = n$ . It is clear from equation (2) that  $K$  grows rather quickly as the number of states of each  $X_i$  grow. This is one of the main drawbacks of discrete BBNs. Some of the drawbacks of discrete BBNs were discussed in (Hanea, Kurowicka, and Cooke 2006) and (Cowell, Dawid, Lauritzen, and D.J. 1999). We list a summary of them next:

1.  $K$  imposes an assessment burden that might lead to informal and indefensible quantification or a drastic discretization or reduction of the model.
2. Marginal distributions are often available from data. Marginal distributions for children nodes is calculated from probability tables and this could impose severe restrictions in a quantification process.
3. Discrete BBNs are flexible with respect to recalculation and updating however they are not flexible with respect to modelling changes. If a parent node is added then the child nodes must be completely re-quantified.

Continuous-discrete non-parametric BBNs (Kurowicka and Cooke 2005), (Hanea, Kurowicka, and Cooke 2006) have been developed to cope with some of the drawbacks that discrete (and discrete-normal) models impose. These will be discussed next.

## 2.2 Non-Parametric Continuous BBNs.

Non-parametric Continuous BBNs and their relationship to other graphical models were presented in (Kurowicka and Cooke 2005) and extended in (Hanea, Kurowicka, and Cooke 2006). A *non-parametric continuous (or continuous-discrete) BBN (NPCDBBN)* is a directed acyclic graph whose nodes represent continuous univariate random variables and whose arcs are associated with parent-child (un)conditional rank correlations. For each variable  $X_i$  with parents  $X_j, \dots, X_{|Pa(i)|}$  associate the arc

$X_{Pa(i)-k} \rightarrow X_i$  with the conditional rank correlation:

$$\begin{cases} r_{i,Pa(i)}, & k = 0 \\ r_{i,Pa(i)-k|Pa(i),\dots,Pa(i)-k+1}, & 1 \leq k \leq Pa(i) - 1 \end{cases} \quad (3)$$

The assignment is vacuous if  $\{X_j, \dots, X_{Pa(i)}\} = \emptyset$ . These assignments together with a copula family indexed by correlation and with conditional independence statements embedded in the graph structure of a BBN are sufficient to construct a unique joint distribution. Moreover, the conditional rank correlations in 3 are algebraically independent, hence any number in  $(-1,1)$  can be attached to the arcs of a NPCDBBN (Hanea, Kurowicka, and Cooke 2006).

Any copula with an invertible conditional cumulative distribution function may be used as long as the chosen copula possesses the zero independence property<sup>2</sup>. Choosing the normal copula presents advantages with respect to other copulae for building the joint distribution. Observe that for the normal copula relation (9) holds and since conditional correlations are equal to partial correlations then formula (8) may be used to compute the correlation matrix corresponding to the graph. Moreover since for the joint normal distribution, conditional distributions are also normal (Tong 1990, p.33), then analytical updating is possible by this choice (Hanea, Kurowicka, and Cooke 2006, p.724).

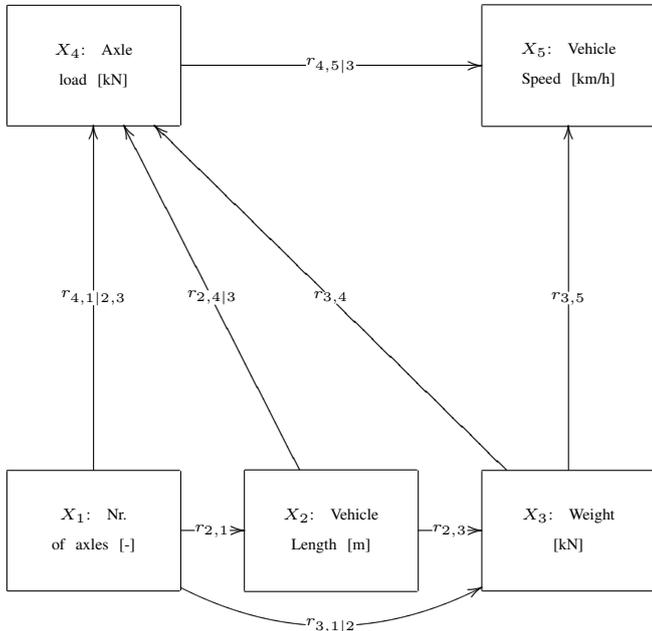


Figure 1: BBN for Axle Loads in Highway R16 (Left Lane) in the Netherlands.

Figure 1 shows an example of a NPCDBBN on 5

<sup>2</sup>A copula with an analytic form for the conditional and inverse conditional cumulative distribution function accelerates the sampling procedure. One example of such a copula is Frank's copula. See (Nelsen 1998)

nodes. To illustrate the assignment of rank and conditional rank correlations consider variable  $X_4$  with parents  $Pa(4) = \{X_1, X_2, X_3\}$ . According to the assignment in (3), the unconditional rank correlations  $r_{4,3}$  is used first. Then the conditional rank correlation  $r_{4,3-1|3} = r_{4,2|3}$  is assigned. Finally the conditional rank correlation  $r_{4,3-2|3,2} = r_{4,1|3,2}$  is assigned. Other (un)conditional rank correlations have been assigned in a similar way.

The indices in the assignment (3) refer to the index of the variable in the ordering in the parent set and hence it does not necessarily need to coincide with the variable's original index. However once the correlations have been assigned the variables' original indices are used.

The structure of the BBN gives information about the dependence structure in the joint distribution. For example variables  $X_1$  and  $X_2$  are correlated.  $X_1$  influences  $X_3$  directly but also through  $X_2$ . This is described in terms of the conditional rank correlation  $r_{1,3|2}$ . In general every variable in the graph is conditionally independent of its *ancestors* given its *parents*. For example  $X_5$  is conditionally independent of  $\{X_1, X_2\}$  given  $\{X_3, X_4\}$ . For a more complete description of the semantics of the graph of a BBN see (Pearl 1988, ch. 3).

### 3 NON-PARAMETRIC CONTINUOUS BBNs FOR DESIGN LOADS IN BRIDGES.

NPCBBNs have been used before in risk and uncertainty analysis. In (Ale, Bellamy, Boom, Cooper, Cooke, Goossens, Hale, Kurowicka, Morales, Roelen, and Spouge 2007) and (Morales Nápoles 2009) a large scale BBN for aviation safety<sup>3</sup> is discussed. In the public health field, in (Jesionek and Cooke 2009) a BBN for benefits and risks associated with food consumption is presented. An environmental application is presented in (Hanea, Kurowicka, Cooke, and Ababei 2010) where a NPCBBN is quantified from data of polluting gases in the USA. With respect to risks in civil infrastructure in (Morales Nápoles and Delgado Hernández 2009) a BBN for earth dam safety and its quantification through expert judgment is presented.

The BBN from Figure 1 is presented in this paper as an example for the use of these kind of models for bridge design loads calculations. The BBN in Figure 1 could have a number of advantages for users:

- A rapid visualization of conditional distributions could be available for analysts once the model is quantified.
- Artificial data can be generated based on the

<sup>3</sup>Approximately 1,500 nodes and 5,000 arcs.

model to reduce the uncertainty related to lack of data in estimating extreme quantiles.

- The model could be used not only for investigating extreme axle loads but any other univariate marginal not conditionalized in the model.

The quantification of the model was done from data partially as in (Hanea, Kurowicka, Cooke, and Ababei 2010). The quantification of the model is described next.

### 3.1 Model quantification

We use data from the Weigh in Motion (WIM) system in the Netherlands. The data records the variables presented in figure 1 for different locations. We use highway RW16 in the left lane for the month of April 2008. In total 239,994 vehicles are contained in the data. If joint data were available then a full quantification of the model would be possible following (Hanea, Kurowicka, Cooke, and Ababei 2010). This is however not the case.

Observe that a single measurement per vehicle is observed for variables  $X_1$ ,  $X_2$ ,  $X_3$  and  $X_5$ . The marginal distribution for  $X_4$  refers to axle load. Each vehicle can have from 2 to 13 observations for  $X_4$ . Hence its marginal distribution is available. However, the required rank correlations cannot be assessed directly from data because we have more observations for  $X_4$  than for all other variables.

Instead, a smaller model with variables  $\{X_1, X_2, X_3, X_5\}$  is quantified as in (Hanea, Kurowicka, Cooke, and Ababei 2010). The quantification consists roughly in computing the rank correlations of interest in figure 1 from the data. Then the full correlation matrix of the BBN using the normal copula ( $\Sigma_{BBN_S}$ ) may be computed. To validate the model 1) the determinant of the rank correlation matrix obtained by transforming the univariate distributions to standard normals, and then transforming the product moment correlations to rank correlations using Pearson's transformation<sup>4</sup> ( $\Sigma_{NOR_S}$ ) is compared with the distribution of the determinant of the BBN; and 2) the determinant of the empirical rank correlation matrix ( $\Sigma_{EMP_S}$ ) is compared with the sample distribution of the determinant of the rank correlation matrix obtained by transforming the univariate distributions to standard normals, and then transforming the product moment correlations to rank correlations using Pearson's transformation.

The motivation to use the determinants of correlation matrices as measures for model validation is described in (Hanea, Kurowicka, Cooke, and Ababei 2010). The determinant of the correlation matrix is

a measure of linear dependence in a joint distribution. If all variables are independent, the determinant is 1, and if there is linear dependence between the variables, the determinant is 0. Intermediate values reflect intermediate dependence. The determinant of a NPCBBN can be factorized in terms of the partial correlations (after transforming conditional rank correlations with equation 9) attached to its arcs.

Finally after the smaller model containing variables  $\{X_1, X_2, X_3, X_5\}$  has been quantified, variable  $X_4$  is introduced and the remaining rank correlations estimated. The rank correlations of interest for the model with  $\{X_1, X_2, X_3, X_5\}$  computed from data are:  $r_{2,1} = 0.5197$ ,  $r_{2,3} = 0.452431$ ,  $r_{3,1|2} = 0.5284$  and  $r_{5,3} = -0.0954$ . The rank correlation matrices of interest for the same model are shown in 4 to 6.

$$\Sigma_{BBN_S} = \begin{pmatrix} 1 & 0.52 & 0.641 & -0.0629 \\ 0.52 & 1 & 0.452 & -0.0448 \\ 0.641 & 0.452 & 1 & -0.0954 \\ -0.0629 & -0.0448 & -0.0954 & 1 \end{pmatrix} \quad (4)$$

$$\Sigma_{EMP_S} = \begin{pmatrix} 1 & 0.495 & 0.611 & -0.0518 \\ 0.495 & 1 & 0.44 & 0.0333 \\ 0.611 & 0.44 & 1 & -0.0737 \\ -0.0518 & 0.0333 & -0.0737 & 1 \end{pmatrix} \quad (5)$$

$$\Sigma_{NOR_S} = \begin{pmatrix} 1 & 0.52 & 0.641 & -0.0811 \\ 0.52 & 1 & 0.452 & -0.0045 \\ 0.641 & 0.452 & 1 & -0.0954 \\ -0.0811 & -0.0045 & -0.0954 & 1 \end{pmatrix} \quad (6)$$

Observe that  $\Sigma_{NOR_S}$  and  $\Sigma_{BBN_S}$  differ only with respect to  $r_{5,1}$  and  $r_{5,2}$ . This is because the subgraph consisting of  $\{X_1, X_2, X_3\}$  is complete. When introducing  $X_5$  the assumption that it is independent of  $X_1$  and  $X_2$  given  $X_3$  is used to compute  $r_{5,1}$  and  $r_{5,2}$  in  $\Sigma_{BBN_S}$ .

In our case  $\det(\Sigma_{BBN_S}) = 0.411637$ ,  $\det(\Sigma_{EMP_S}) = 0.448537$  and  $\det(\Sigma_{NOR_S}) = 0.410093$ . In order to give an indication of whether our model is a good representation of the normal data we obtain from a 100 simulations the sample distribution of  $\det(\Sigma_{BBN_S})$ . The 5<sup>th</sup> and 95<sup>th</sup> percentiles of this distribution are 0.40893 and 0.41523 respectively.  $\det(\Sigma_{NOR_S}) \in (\det(\widehat{\Sigma_{BBN_S}})_{5^{th}}, \det(\widehat{\Sigma_{BBN_S}})_{95^{th}}) = (0.40893, 0.41523)$  and hence we could have some confidence that the model represents adequately the normal data.

To have an idea of how the normal data represents the original data we obtain the sampling distribution of  $\Sigma_{NOR_S}$  and compare  $\det(\Sigma_{EMP_S})$  with it. In our case  $\det(\Sigma_{EMP_S}) \notin (\det(\widehat{\Sigma_{NOR_S}})_{5^{th}}, \det(\widehat{\Sigma_{NOR_S}})_{95^{th}}) =$

<sup>4</sup>Equation 9

(0.40765, 0.41307). This could be an indication that the normal data could not be the best representation of the empirical data. The different sign that we obtain between  $\Sigma_{EMPS}$  and the remaining two matrices for  $r_{2,5}$  is the cause that the normal data seems to be not representative of the data. In this case however since we still have to estimate the remaining dependence structure outside the data we decide to leave our model as it is warning about this discrepancy between the original data and the normal representation.

Adding  $X_4$  to our model implies the estimation of  $r_{3,4}$ ,  $r_{2,4|3}$ ,  $r_{1,4|2,3}$  and  $r_{4,5|3}$ . These can be estimated through experts as in (Morales, Kurowicka, and Roelen 2008) and (Morales-Nápoles, Kurowicka, Cooke, and van Baren 2009). For our presentation we estimate them by computing  $r_{3,4_j}$ ,  $r_{2,4|3_j}$ ,  $r_{1,4|2,3_j}$  and  $r_{4,5|3_j}$  for  $j = 2, \dots, 13$  for each axle. Finally we average over  $j$  to obtain our final estimates. By doing so we obtain  $r_{3,4} = 0.8605$ ,  $r_{2,4|3} = -0.2081$ ,  $r_{1,4|2,3} = 0.3647$  and  $r_{4,5|3} = 0.3979$ . The correlation matrix of the final BBN consisting of variables  $\{X_1, X_2, X_3, X_4, X_5\}$  is shown in equation (7). The determinant of the rank correlation matrix of our model from figure 1 after the final quantification is  $\det(\Sigma_{BBN}) = 0.0766$ .

$$\Sigma_{BBN} = \begin{pmatrix} 1 & 0.52 & 0.641 & 0.4 & 0.0634 \\ 0.52 & 1 & 0.452 & 0.295 & 0.0365 \\ 0.641 & 0.452 & 1 & 0.860 & -0.0954 \\ 0.4 & 0.295 & 0.860 & 1 & -0.277 \\ 0.0634 & 0.0365 & -0.0954 & -0.277 & 1 \end{pmatrix} \quad (7)$$

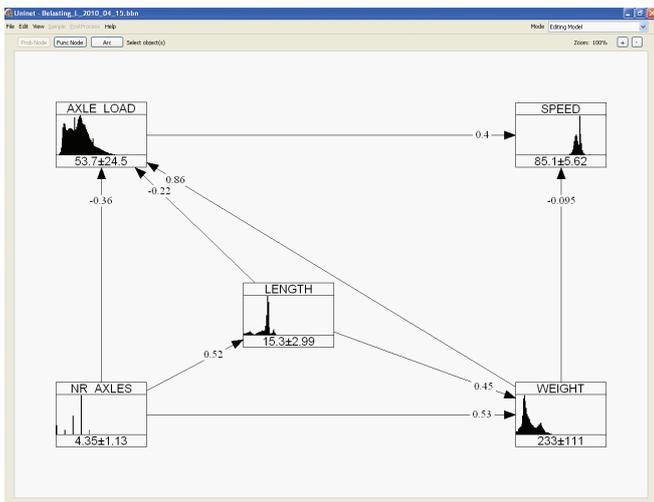


Figure 2: Quantified BBN in UniNet

Figure 2 presents the model in UniNet (Hanea 2008). Each node in figure 2 presents the univariate marginal distribution of  $\{X_1, \dots, X_5\}$ . In the bottom part of each node the mean and standard deviation (after the  $\pm$  sign) are shown. UniNet is a stand alone software that allows the manipulation of NPCBBNs.

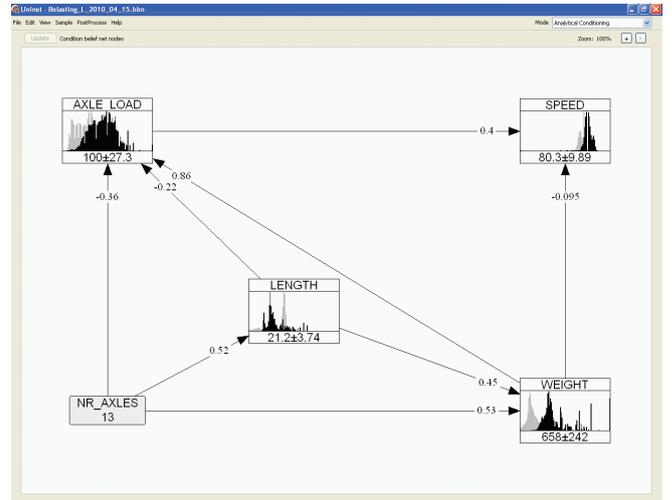


Figure 3: BBN in UniNet Conditional on  $X_1 = 13$

It has been used in the past in measuring health risks (Jesionek and Cooke 2009) and air transport safety (Ale, Bellamy, Cooke, Goossens, Hale, Roelen, and Smith 2006).

UniNet allows analytical conditioning over the BBN when evidence becomes available. this is shown in figure 3. Observe for example that conditional on 13 axles vehicles, the expectation of vehicular weight is about 2.8 times higher than the expectation of the unconditional distribution. In the case of the axle load distribution, the expectation grows approximately 1.8 times. The univariate margins of vehicular length and speed are updated similarly.

Interval conditioning is also available in UniNet through sample based conditioning. The conditional means and standard deviations for axle load and vehicular weight are computed from the 29 samples where  $X_1 \geq 9$  and  $X_5 \in [10, 60]$ . Situations as those described in figures 3 and 4 could be of interest for researchers. However, they are difficult to observe in the data alone. The use of NPCBBNs in such situa-

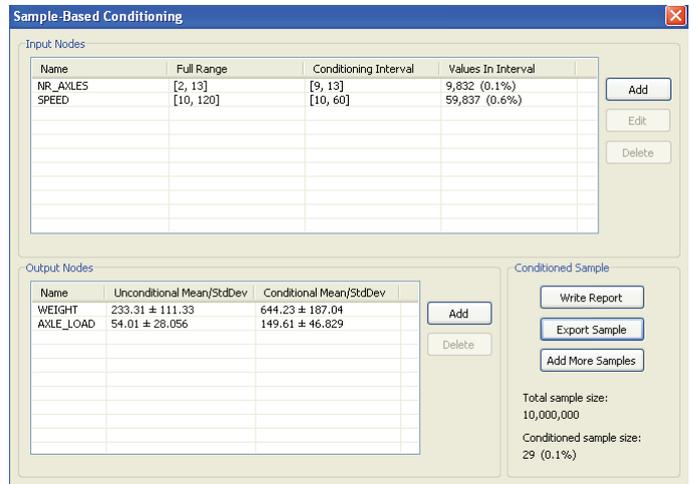


Figure 4: Sample based conditioning in UniNet

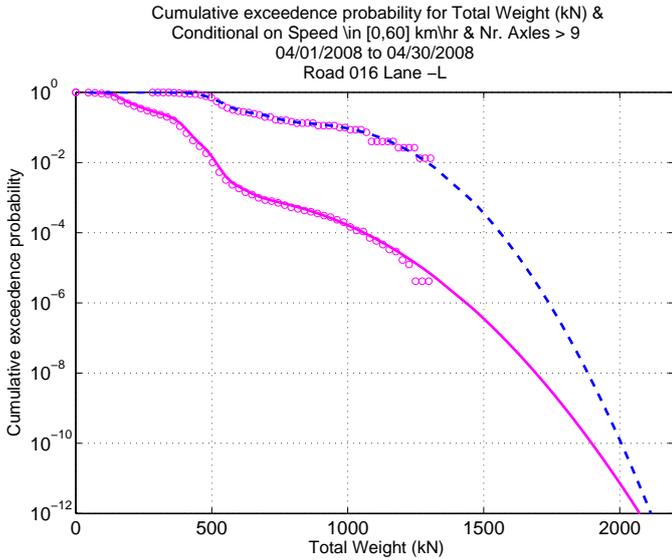


Figure 5: Distribution of  $X_2$  and  $X_2|x_1 \geq 9, x_5 \in [10, 60]$

tions is illustrated in the next section.

#### 4 EXTREME WEIGHTS FOR BRIDGES IN THE NETHERLANDS.

In (Steenbergen and Morales Nápoles 2010a) and (Steenbergen and Morales Nápoles 2010b) techniques for finding extreme values for axle loads and total vehicular weight for bridge design in the Netherlands are discussed. The data used in section 3.1 for the quantification of the BBN here presented is part of the data used in (Steenbergen and Morales Nápoles 2010b). These data together with the techniques described in (Steenbergen and Morales Nápoles 2010b) will be used further to exemplify the use of NPCBBNs in structural reliability. More specifically for the calculation of extreme values.

From our model in figure 1 we would like to see the effect of number of axles and vehicle speed on weight and axle load distributions. We suspect that vehicles with more than 9 axles and with speed smaller than  $60 \frac{\text{km}}{\text{h}}$  will be of more concern because they are heaviest. In our original database only two vehicles fulfill these characteristics.

In order to overcome this lack of data, we could use our BBN to generate data artificially. Once our model has been fully quantified we can generate data similarly to figure 4 and use it to estimate our conditional distribution. We generate 50 million samples of which 151 vehicles have more than 9 axles and speed smaller than  $60 \frac{\text{km}}{\text{h}}$ .

Figure 5 show the unconditional and conditional empirical exceedence probabilities. A mixture of normals is fitted to each empirical distribution accord-

ing to the procedures described in (Steenbergen and Morales Nápoles 2010b). The statistical toolbox of matlab 7.10.0(R2010a) is used for the fitting. For details see (Mclachlan and Peel 2000) and (Steenbergen and Morales Nápoles 2010b).

As expected the conditional distribution indicates smaller probabilities of exceedence for extreme values. For example observe that the probability of exceeding 2,000kN is about  $1 \times 10^{-11}$  in the unconditional distribution and one order of magnitude larger in the conditional distribution. Other extreme values for vehicle weight can easily be determined in this way for both the conditional and unconditional distributions.

#### 5 DISCUSSION.

In this paper we have presented the NPCBBNs in the context of structural reliability. A model for vehicular weight, axle load, vehicular speed, number of axles and vehicular length has been quantified for Moerdijk highway R16 in the left lane for the month of April 2008. The normal copula assumption for variables  $\{X_1, X_2, X_3, X_5\}$  needs to be further investigated. Because the model is hybrid in the sense that the relationship between  $\{X_1, X_2, X_3, X_5\}$  and  $X_4$  is investigated through expert judgment we decide to allow for the normal copula assumption.

The model in figure 1 quantified in this paper will evolve into a model whose base should be the BBN in figure 6. Observe that a load distribution is included for each axle in each vehicle type. In the same way, for each vehicle type a distribution for vehicular velocity and vehicular length is computed. The weight distribution is a function of the number of axles, and each axle load variable. For example if the model consisted only of two and three axles vehicles the the weight distribution would be computed as with the following formula: Weight = if(Number Axles = 2, axle 1 of 2 axle vehicles + axle 2 of 2 axle vehicles, if(Number Axles = 3, axle 1 of 3 axle vehicles + axle 2 of 3 axle vehicles + axle 3 of 3 axle vehicles)). The weight formula may be written similarly for vehicles with up to 13 axles. Whenever data is missing for the quantification of arcs, they may be computed through structured expert judgment as in (Morales, Kurowicka, and Roelen 2008).

Similar models as in figure 6 may be computed for other locations in the Netherlands. These include locations in 4 different highways for the left and right lanes. Thus a total of eight BBNs similar to figure 6 may be quantified. All of these can be combined into a single BBN and analyzed in real time using UNINET. The final model would look as in figure 7. The BBN in figure 7 consists of 697 nodes and 2,175 arcs. In the center of the model the *time to rush hour* variable connects the separate locations. This

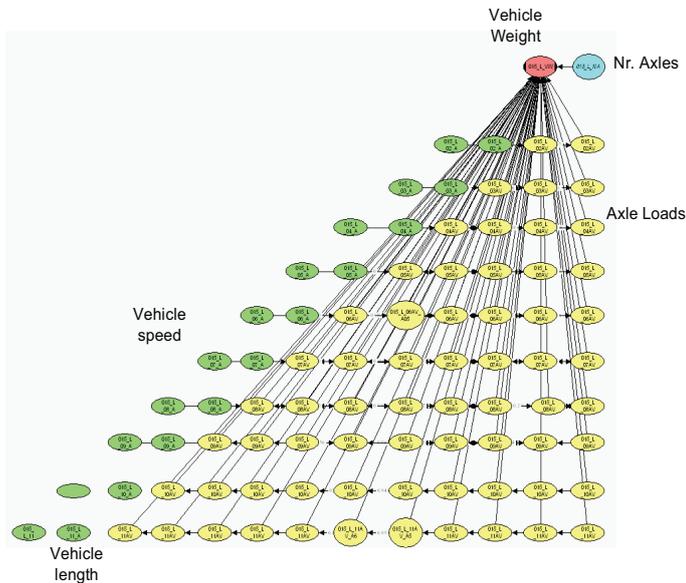


Figure 6: BBN for the Weight in Motion system in the Netherlands (one location).

model is at the moment of the writing being quantified. The reader may observe that the dependence structure of the model is still complex. However, the BBN helps clarifying and reducing the complexity of the dependence structure present in the data. With UNINET sampling such a BBN an conditionalizing is done in minutes.

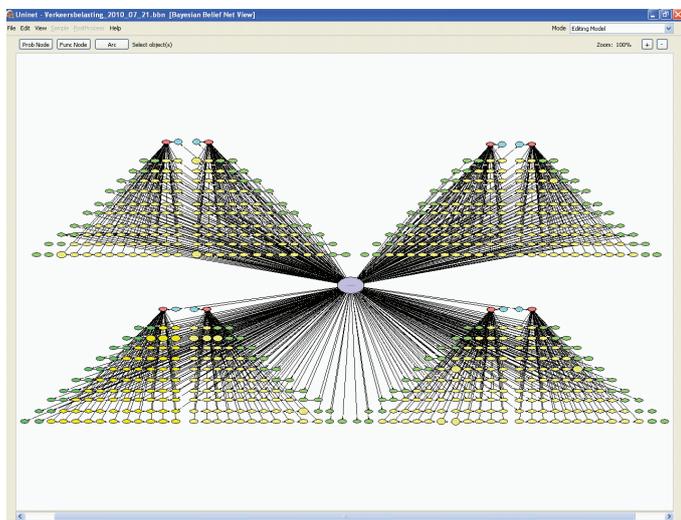


Figure 7: BBN for the Weight in Motion system in the Netherlands (8 locations).

Applications similar to the one described in this paper may profit from the use of NPCBBNs and structured expert judgment. Whenever the computation of design values is severely restricted due to lack of data the approach presented in this paper may be of advantage.

## REFERENCES

- Ale, B., L. Bellamy, R. d. Boom, J. Cooper, R. Cooke, L. Goossens, A. Hale, D. Kurowicka, O. Morales, A. Roelen, and J. Spouge (2007). Further developments of a causal model for air transport safety (cats); building the mathematical heart. In *ESREL*, pp. 1431–1439.
- Ale, B., L. Bellamy, R. Cooke, L. Goossens, A. Hale, A. Roelen, and E. Smith (2006). Towards a causal model for air transport safety an ongoing research project. *Safety Science* 44(8), 657–673.
- Cowell, R., A. Dawid, S. Lauritzen, and S. D.J. (1999). *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer.
- Hanea, A. (2008). *Algorithms for Non-parametric Bayesian belief nets*. Ph. D. thesis, TU Delft, Delft, the Netherlands.
- Hanea, A., D. Kurowicka, and R. Cooke (2006). Hybrid method for quantifying and analyzing bayesian belief nets. *Quality and reliability Engineering International* 22, 709–729.
- Hanea, A., D. Kurowicka, R. Cooke, and D. Ababei (2010, March). Mining and visualising ordinal data with non-parametric continuous bbns. *Computational Statistics & Data Analysis* 54(3), 668–687.
- Jesionek, P. and R. Cooke (2009). A bayesian belief network model for the benefit-risk analysis of fish consumption in the beneris project. Technical report, european union project, TU DELFT.
- Kurowicka, D. and R. Cooke (2005). Distribution-free continuous bayesian belief nets. In K.-M. S. Wilson A., Limnios N. and A. Y. (Eds.), *Modern Statistical and mathematical Methods in Reliability*, pp. 309–323.
- McLachlan, G. and D. Peel (2000, October). *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley-Interscience.
- Morales, O., D. Kurowicka, and A. Roelen (2008). Eliciting conditional and unconditional rank correlations from conditional probabilities. *Reliability Engineering & System Safety* 93(5), 699 – 710. Expert Judgement.
- Morales Nápoles, O. (2009). *Bayesian belief nets in aviation safety and other applications*. Ph. D. thesis, Delft University of Technology.
- Morales Nápoles, O. and D. Delgado Hernández (2009, November). Non-parametric continuous bayesian belief networks for earth dam safety. In P. V. Van Gelder (Ed.), *proceedings of the 7th International Probabilistic Workshop*, Delft, the Netherlands, pp. 295–312.
- Morales-Nápoles, O., D. Kurowicka, R. Cooke, and G. van Baren (2009). Expert elicitation methods of rank and conditional rank correlations: An example with human reliability models in the aviation industry. *Submitted to RE&SS*.
- Nelsen, R. B. (1998, October). *An Introduction to Copulas (Lecture Notes in Statistics)*. Springer.
- NEN (2002). *Eurocode EN 1990, Basis of Structural Design*. Nederlands Normalisatie-instituut.
- Pearl, J. (1988, September). *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann.
- Steenbergen, R. and O. Morales Nápoles (2010a, September). Determination of design axle loads for bridges from measurements. In *Proceedings of ESREL 2003 - European Safety and Reliability Conference 2010*, Rhodes, Greece.

Steenbergen, R. and O. Morales Nápoles (2010b). Statistical analysis wim registrations april 2008. report TNO report 2010-DWARS-M002, Netherlands Organization for applied Scientific Research.

Tong, Y. (1990). *The multivariate Normal Distribution*. Series in Statistics. Springer.

van Noortwijk, J., H. Kalk, and E. Chbab (2003, 15-18 June). Bayesian computation of design discharges. In T. Bedford and P. van Gelder (Eds.), *Safety and Reliability, Proceedings of ESREL 2003 - European Safety and Reliability Conference 2003*, Maastricht, The Netherlands, pp. 1179–1187. Lisse: Swets & Zeitlinger.

Yule, G. and M. Kendall (1965.). *An introduction to the theory of statistics* (14<sup>th</sup> ed.). Belmont, California.: Charles Griffin & Co.

The relationship between  $r$  (the rank correlation of the normal variables) and the parameter  $\rho$  or  $\rho_{1,2,3,\dots,n}$  is known and given by the following formula ((Kurowicka and Cooke 2005, p.55)):

$$\rho = 2 \sin\left(\frac{\pi}{6}r\right). \quad (9)$$

## Appendix

The *product moment correlation* of random variables  $X$  and  $Y$  with finite expectations  $E(X)$ ,  $E(Y)$  and finite variances  $var(X)$ ,  $var(Y)$  is:

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{var(X)var(Y)}}.$$

The *rank correlation* of random variables  $X$ ,  $Y$  with cumulative distribution functions  $F_X$  and  $F_Y$  is:

$$r_{X,Y} = \rho_{F_X(X), F_Y(Y)} = \frac{E(F_X(X)F_Y(Y)) - E(F_X(X))E(F_Y(Y))}{\sqrt{var(F_X(X))var(F_Y(Y))}}.$$

The rank correlation is the product moment correlation of the ranks of variables  $X$  and  $Y$ , and measures the strength of monotonic relationship between variables. The rank correlation always exists, is independent of marginal distributions and is invariant under non-linear strictly increasing transformations.

The conditional rank correlation of  $X$  and  $Y$  given  $Z$  is:

$$r_{X,Y|Z} = r_{\tilde{X}, \tilde{Y}}$$

where  $(\tilde{X}, \tilde{Y})$  has the distribution of  $(X, Y)$  given  $Z = z$ .

The (conditional) rank correlation is the dependence measure of interest because of its close relationship with conditional copulae used in non-parametric continuous BBNs (see section 2.2). One disadvantage of this measure however is that it fails to capture non-monotonic dependencies.

Rank correlations may be realized by copulae, hence the importance of these functions in dependence modeling. Partial correlations are also of interest in this paper.

Partial correlations can be computed recursively from correlations (see (Yule and Kendall 1965)):

$$\rho_{1,2,3,\dots,n} = \frac{\rho_{1,2,4,\dots,n} - \rho_{1,3,4,\dots,n} \cdot \rho_{2,3,4,\dots,n}}{\left((1 - \rho_{1,3,4,\dots,n}^2) \cdot (1 - \rho_{2,3,4,\dots,n}^2)\right)^{\frac{1}{2}}} \quad (8)$$