



**HAL**  
open science

## Expert judgment in life-cycle degradation and maintenance modelling for steel bridges

A Kosgodagan, O Morales-Nápoles, J Maljaars, W Courage

### ► To cite this version:

A Kosgodagan, O Morales-Nápoles, J Maljaars, W Courage. Expert judgment in life-cycle degradation and maintenance modelling for steel bridges. Fifth International Symposium on Life-Cycle Civil Engineering, Oct 2016, Delft, Netherlands. hal-01517169

**HAL Id: hal-01517169**

**<https://hal.science/hal-01517169v1>**

Submitted on 2 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Expert judgment in life-cycle degradation and maintenance modelling for steel bridges

A. Kosgodagan

*IRCCyN, École des Mines de Nantes, France*

*Structural reliability, Netherlands Organization for Applied Scientific Research, The Netherlands*

O. Morales-Nápoles

*Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands*

J. Maljaars

*Structural reliability, Netherlands Organization for Applied Scientific Research, The Netherlands*

*Faculty of Built Environment, Eindhoven University of Technology, The Netherlands*

W. Courage

*Structural reliability, Netherlands Organization for Applied Scientific Research, The Netherlands*

**ABSTRACT:** Markov-based models for predicting deterioration for civil infrastructures are widely recognized as suitable tools addressing this mechanism. The objective of this paper is to provide insights regarding a network of orthotropic steel bridges in terms of degradation. Consequently, a model combining a dynamic Bayesian network and a Markov chain is first introduced that builds up the network in a concise way. In an attempt to represent a network composed of two general classes of orthotropic steel bridges, the classical method of structured expert judgment is carried out as a quantification procedure. The first objective is to elicit indirectly transition probabilities for a Markov chain that describes how each bridge type deteriorates in time. Second, experts are asked to provide estimates on required conditional probabilities related to the Bayesian network. An in-depth analysis of the results is presented so that remarks and observations are subsequently pointed out and, finally conclusions are drawn.

## 1 INTRODUCTION

Ensuring a satisfactory level of safety and driving comfort are generally the primary objectives for motorway bridge managers. Throughout a bridge service life, numerous maintenance type of interventions need to be performed to keep the structure above such levels. If a newly constructed bridge is considered to be in a perfect condition and the degradation phenomenon assumes a monotonic decreasing-shape function, a bridge's condition can then be described as a function in time bouncing up and down between these two phases. A schematic illustration of these cycles is proposed in Fig. 1 where two different maintenance plans are implemented. One strategy typically proposes a corrective-and-rehabilitation option for maintaining the bridge (solid line) while the other one's purpose is to extend its service lifetime by coupling preventive and corrective maintenance deci-

sions postponing a full renovation to the latest (dashed line). Substantial financial investments are initiated in order to perform these repairs and costs are typically non-linear especially when considering a full rehabilitation compared to preventive or corrective actions. These are generally considered the three principal maintenance categories available to decision makers. In Fig. 1 the areas separated by the dotted line labelled *Preventive maintenance level* divides preventive (area above) and corrective (below) maintenance options. When the bridge degradation function hits the solid line *Minimum acceptable level* it necessarily entails a repair. Hence a well-timed maintenance strategy aims to save money without jeopardizing safety and functionality.

Degradation modelling is of utter importance in such a context as future maintenance plans are determined based upon the shape (slope and monotonicity) of the degradation curve. Both determinis-

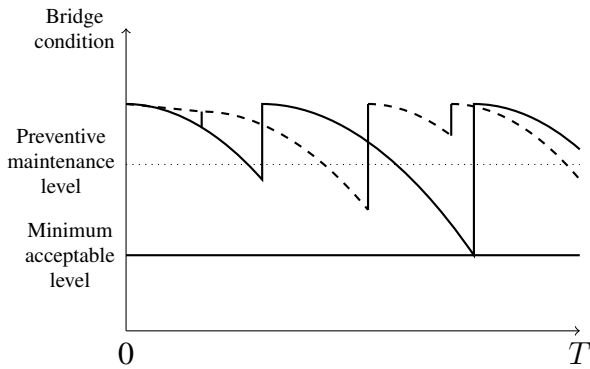


Figure 1: Schematic representation of bridge degradation and maintenance cycles

tic and stochastic models have been widely surveyed to assess deterioration mechanism in the bridge reliability field (Morcoux 2011). In practice, a significant number of countries have integrated a so-called bridge management system (BMS) that opts for a discrete-time stochastic Markov process (or chain) as standard support tool to describe the degradation behaviour in time (Mirzaei, Adey, Thompson, & Klatter 2014). The goal of this system is also to bring forward knowledge at a network scale. Indeed optimizing locally at the single bridge scale may not comply with the network-scale optimization requirements, for instance if personnel and equipment available are limited. However information per bridge does not necessarily facilitate the choice for decision makers because dealing with sometimes hundreds of elements makes it difficult to prioritize. Hence a full probabilistic degradation model is sought encompassing both the Markov framework and the network level case.

In this paper the deterioration phase is governed by a combination of a Markov chain embedded in a Bayesian network that provides in a compact way probabilistic information to a bridge inventory. We draw much attention in the way both of these tools are quantified. In fact, the objective is to construct a network of bridges whose structure resembles that of the Dutch bridge network. In particular, motorway orthotropic steel deck bridges are of central attention. To properly quantify our model we use the classical, or Cooke’s, method for structured expert judgement (Cooke 1991). It is frequently used when field data is missing, difficult to obtain or of poor quality. In this case, variables that are needed to be assessed refer to degradation inputs for moveable and fixed types of steel bridges through transition durations between consecutive deterioration states.

## 2 DEGRADATION MODELLING FOR ORTHOTROPIC STEEL BRIDGES

As we want to represent a network of steel bridges whose purpose is to resemble as accurately as possible that of the Dutch motorway steel bridges network, two classes of steel bridge are considered: fixed and moveable. They do not refer to specific existing

bridges but describe more conventionally each type of fixed and moveable steel bridges through various characteristics (key geometry aspects, type and thickness of overlay, deck plate thickness, and so on). Fatigue cracking is generally considered as the main phenomenon driving degradation for orthotropic steel bridges. It results from fluctuating stresses caused by the crossing of heavy vehicles. Typically, loading and traffic characteristics are key quantities when studying fatigue mechanism in this context. The nature of these two variables is reasonably assumed to be random (Morales-Nápoles & Steenbergen 2014). Specifically we are looking at cracks located in the deck plate and in ‘trough to deck plate’ parts as suggested in Fig. 2. Their number together with their size are crucial parameters to monitor. The condition of a bridge is then broken down into several states featuring characteristics on various degrees of severity on crack size, location and number. These states subsequently stand for the state space  $S$  of a Markov chain  $\{M_t, t \geq 0\}$ . The latter describes probabilistically the evolution of a bridge’s condition in time. It is assumed that a bridge can either stay in the same state or move to its next worst state at the next time step given its current condition state, thus  $p_{i,i}, p_{i,i+1} > 0$  where  $p_{i,j} = P(M_{t+1} = j | M_t = i)$  with  $i, j \in S$ . In this paper, one of the goals is to quantify the  $p_{i,j}$ ’s through expert elicitation as detailed in section 3. To then address the network-scale maintenance problem the Markov chain  $\{M_t\}$  acts as time sequenced nodes in a dynamic Bayesian network (DBN).

A Bayesian network (BN) is a directed acyclic graph (DAG) whose nodes represent random variables and whose arcs designate probabilistic dependencies between nodes. Most of the applications use discrete BNs where marginal distributions are specified for the nodes with no parents, and conditional probability tables for child nodes. A BN encodes in a compact way the probability density or mass function on a set of variables by specifying a set of conditional independence statements in the directed acyclic graphs associated with a set of conditional probability functions. More specifically, a BN consists of a qualitative part, the DAG structure, and a quantitative part, the set of conditional probability distributions. A full characterization of a BN lies entirely in these two parts. The graphical property called *directional separation* (abbreviated as *d-separation*) asserts conditional independence statements. This attribute covers three different possible layouts for which variables can be d-separated. The attractiveness of BNs comes thus partly from the ability to model high dimensional probability distributions in a relatively intuitive visual way. In addition, knowledge, on a state of a variable for instance, can be inserted and propagated throughout the graph. This way, the marginal distributions of other nodes for which evidence is not available are updated accordingly using algorithms developed for this purpose (Jordan 1999). This mechanism is called

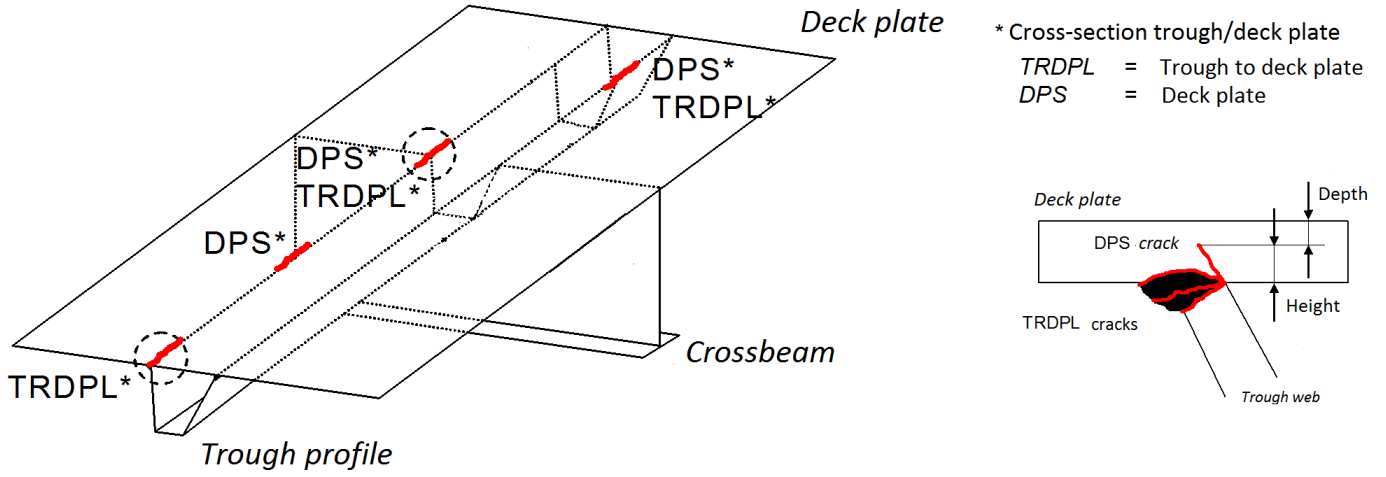


Figure 2: Three-dimensional view of the bridge considered cracks' location (left); longitudinal cross-section with 'trough to deck plate' (TRDPL) crack location (right);

*probabilistic inference*. Readers are referred to Pearl (1988) for a full mathematical treatment on BNs and foundations therein.

It is often sought in reliability modelling the need to describe dynamically, in the sense of time-indexed, the evolution of degradation as opposed to the static or stationary case. A special type of BN called dynamic BN (DBN) deals with domains containing recurring networks that evolve over time. This is particularly desirable when stochastic processes are involved (Straub 2009). The complete DBN model is presented in Fig. 3. Nodes  $T_t^{(k)}$  and  $L_t^{(k)}$  denote respectively traffic and loading variables where superscript ( $k$ ) refers to the bridge number. At each time slice, the structure suggests that load depends on traffic and the degradation process  $\{M_t\}$  depends on the load in turn. We assume that this sequential connexion is a reasonable way to first describe that explanatory variables  $T_t^{(k)}$  and  $L_t^{(k)}$  impact degradation in this manner. Second traffic quantities link consecutively every traffic node proper to each bridge so that the network is set up.

Various methods have been tested to quantify Markov chain's transition probabilities using field data, however since we are constructing general classes of bridges we do not possess such material at hand. Additionally, BN's conditional probabilities have to be assessed as well. In practice, again, collected data generally provides the sufficient quantification material to feed the BN with. In the absence of it, expert judgment is applied to fill it out. The light blue arrows in Fig 3 correspond to the links for which missing conditional probabilities are quantified by expert opinions. For the remainder of the conditional distributions, field measurements are used to quantify  $T_t^{(k)}$  and  $L_t^{(k)}$  where each can have three condition states, *High*, *Medium* and *Low*, and *Heavy*, *Normal*

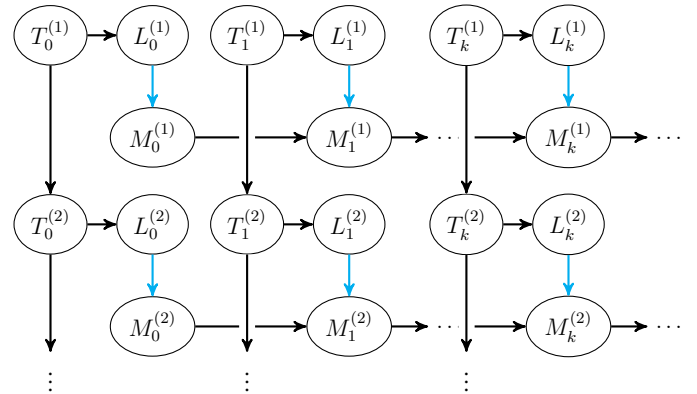


Figure 3: The DBN structure for the network of bridges

and *Light*, respectively.

Since the distance between degradation condition state in state space  $S$  is not necessarily constant and, in addition, assumption is made on the distance pattern (whether it is linear or not), we narrow down the number of states to four,  $S = \{1, 2, 3, 4\}$ . Indeed, the number of probabilities of transition to elicit for the Markov chain as well as the conditional probabilities for the DBN is a direct consequence of the size of  $S$ ; the larger it gets the more tedious it is for experts. On this basis, experts answered a total of 24 questions of interest detailed in Table 1. We mention that items for Question 2 (V13 to V24) were not directly elicited in this way. Rather, out of a sample of size  $N$ , experts are asked to give a proportion of it.

### 3 STRUCTURED EXPERT JUDGMENT

Eliciting data from expert's opinion using Cooke's method is a growing popular way tested and applied in numerous fields (Cooke & Goossens 2008). The goal of applying structured expert judgment fosters rational consensus as opposed to political consensus. Opinions are combined via different possible

| Variable ID | Description  |
|-------------|--|
| Question 1  | Expected duration (in years) to transition between the following condition states <ul style="list-style-type: none"> <li>• under a normal load for <ul style="list-style-type: none"> <li>a moveable bridge <ul style="list-style-type: none"> <li>V1 1 → 2</li> <li>V2 2 → 3</li> <li>V3 3 → 4</li> </ul> </li> <li>a fixed bridge <ul style="list-style-type: none"> <li>V4 1 → 2</li> <li>V5 2 → 3</li> <li>V6 3 → 4</li> </ul> </li> </ul> </li> <li>• under a heavy load for <ul style="list-style-type: none"> <li>a moveable bridge <ul style="list-style-type: none"> <li>V7 1 → 2</li> <li>V8 2 → 3</li> <li>V9 3 → 4</li> </ul> </li> <li>a fixed bridge <ul style="list-style-type: none"> <li>V10 1 → 2</li> <li>V11 2 → 3</li> <li>V12 3 → 4</li> </ul> </li> </ul> </li> </ul>   |
| Question 2  | Probability that bridges transitioning to their next worse state conditional on a given load and state at previous time step for <ul style="list-style-type: none"> <li>• a moveable bridge <ul style="list-style-type: none"> <li>V13 <math>P(M_t = 2   M_{t-1} = 1, L_t = Normal)</math></li> <li>V14 <math>P(M_t = 3   M_{t-1} = 2, L_t = Normal)</math></li> <li>V15 <math>P(M_t = 4   M_{t-1} = 3, L_t = Normal)</math></li> <li>V16 <math>P(M_t = 2   M_{t-1} = 1, L_t = Heavy)</math></li> <li>V17 <math>P(M_t = 3   M_{t-1} = 2, L_t = Heavy)</math></li> <li>V18 <math>P(M_t = 4   M_{t-1} = 3, L_t = Heavy)</math></li> </ul> </li> <li>• a fixed bridge <ul style="list-style-type: none"> <li>V19 <math>P(M_t = 2   M_{t-1} = 1, L_t = Normal)</math></li> <li>V20 <math>P(M_t = 3   M_{t-1} = 2, L_t = Normal)</math></li> <li>V21 <math>P(M_t = 4   M_{t-1} = 3, L_t = Normal)</math></li> <li>V22 <math>P(M_t = 2   M_{t-1} = 1, L_t = Heavy)</math></li> <li>V23 <math>P(M_t = 3   M_{t-1} = 2, L_t = Heavy)</math></li> <li>V24 <math>P(M_t = 4   M_{t-1} = 3, L_t = Heavy)</math></li> </ul> </li> </ul> |

Table 1: Variable of interest elicited as part of the expert opinion workshop aiming to quantify probabilistic inputs for the degradation of motorway orthotropic steel bridges.

weighted averaging schemes, where the weights are based on performance measures. The classical model is extensively formalized in Cooke (1991). The main procedure and objectives are introduced below.

A group of experts are asked to assess their uncertainty of continuous quantities for which the realizations are known post hoc. These variables are chosen to resemble the quantities of interest, and/or to draw on the sort of expertise which is required for the assessment of the variables of interest. They are called *calibration* or *seed* variables. Experts then provide their uncertainty estimates through pre-chosen quantiles (usually the 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup>). Note that variables of interest are assessed in a similar way. Concisely, *calibration* measures the degree to which experts are statistically accurate with respect to estimates provided for the seed questions. In turn, *information* measures the degree to which experts uncertainty estimates are concentrated relative to a background measure (uniform or log-uniform generally). "Good expertise" corresponds to good calibration (typically greater than 0.05) and high information.

More precisely, assume from expert  $e = 1, \dots, E$ , each provide their uncertainty estimates through the 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> quantiles on items (or calibration variables)  $i = 1, \dots, N$ . For each item, experts divide their belief range into four inter-quantile intervals, for which the corresponding probabilities of occurrence are:  $p_1 = 0.05$  for a realization value less or equal than the 5<sup>th</sup>,  $p_2 = 0.45$  for a realization value in the inter-quantile range  $(5^{th}, 50^{th}]$ ,  $p_3 = 0.45$  for a realization value in the inter-quantile range  $(50^{th}, 95^{th}]$  and  $p_4 = 0.05$  for a realization value strictly greater than the 95<sup>th</sup> percentile. Empirically we thus get for each expert  $e = 1, \dots, E$  the probability of the relative frequency that realizations fall in the inter-quantile bins  $(0.05, 0.45, 0.45, 0.05)$  denoted by the vector  $\mathbf{s}(e) = (s_1(e), \dots, s_4(e))$ . The calibration score is given by

$$C(e) = 1 - \chi_n^2(2NI(\mathbf{s}(e), \mathbf{p})) \quad (1)$$

where  $I(\mathbf{s}(e), \mathbf{p}) = \sum_{i=1}^4 s_i(e) \ln \left( \frac{s_i(e)}{p_i} \right)$  and  $\chi_n^2$  is the Chi-square distribution with  $n$  degrees of freedom. On the other hand the information score is computed per expert as

$$I(e) = \sum_{i=1}^N f_{e,i} \ln \left( \frac{f_{e,i}}{g_i} \right) \quad (2)$$

where  $f_{e,i}$  and  $g_i$  are the expert  $e$ 's density and the background measure on item  $i$  respectively.

Subsequently, scores are combined to form weights. These weights are constructed to be a strictly proper scoring rule in an appropriate asymptotic sense, that is, experts receive their maximal expected long-run weight by stating their true belief. Important to mention that statistical accuracy dominates informativeness, in other words poor calibration cannot be compensated by high information. Calibration and information constitute the essential metrics to weight the experts in view to combine their opinions. The weighted combined uncertainty distribution is called the *decision maker* (DM) in the sense of linear pooling. The DM is thus a weighted linear pool of experts' individual weight. Consider the following weighting score for expert  $e$

$$w_\alpha(e) = \mathbf{1}_\alpha(C(e)) \times C(e) \times I(e) \quad (3)$$

where  $\mathbf{1}_\alpha(x) = 0$  if  $x < \alpha$  and  $\mathbf{1}_\alpha(x) = 1$  otherwise. This weighting score is referred to as *global* weighted score (GL) and complies with the above mentioned scoring rule criterion. Let  $DM_\alpha(i)$  be the result of linear pooling for seed item  $i$  with weights proportional to (3):

$$DM_\alpha(i) = \sum_{e=1, \dots, E} w_\alpha(e) f_{e,i} / \sum_{e=1, \dots, E} w_\alpha(e) \quad (4)$$

Moreover,  $\alpha$  can be chosen so as to maximize the DM combined score, we then speak of optimized DM. It

| Item ID | Measurement technique | Location of crack | Year 1 <sup>st</sup> measurement | Crack length 1 <sup>st</sup> (mm) | Year 2 <sup>nd</sup> measurement | Crack length 2 <sup>nd</sup> (mm) |
|---------|-----------------------|-------------------|----------------------------------|-----------------------------------|----------------------------------|-----------------------------------|
| S1      | Crack-PEC             | DPS               | 2008                             | 200                               | 2009                             | <b>360</b>                        |
| S2      | Crack-PEC             | DPS               | 2008                             | 250                               | 2009                             | <b>350</b>                        |
| S3      | Crack-PEC             | DPS               | 2006                             | 100                               | 2009                             | <b>1040</b>                       |
| S4      | Crack-PEC             | DPS               | 2006                             | 200                               | 2009                             | <b>500</b>                        |
| S5      | Crack-PEC             | DPS               | 2006                             | 300                               | 2009                             | <b>350</b>                        |
| S6      | UT                    | DPS               | 2009                             | 30                                | 2010                             | <b>50</b>                         |
| S7      | UT                    | DPS               | 2009                             | 80                                | 2010                             | <b>90</b>                         |
| S8      | UT                    | DPS               | 2009                             | 100                               | 2010                             | <b>100</b>                        |
| S9      | UT                    | DPS               | 2009                             | 550                               | 2010                             | <b>590</b>                        |
| S10     | VO                    | TRDPL             | 2008                             | 100                               | 2009                             | <b>250</b>                        |
| S11     | VO                    | TRDPL             | 2008                             | 100                               | 2010                             | <b>250</b>                        |
| S12     | Crack-PEC             | DPS               | 2010                             | 400                               | 2011                             | <b>500</b>                        |

Table 2: Seed variables elicited as part of the expert opinion workshop aiming to quantify probabilistic inputs for the degradation of motorway orthotropic steel bridges.

must be mentioned that other weighting scores are available to the analyst. For the *equal weight* (EQ) score every expert receives the same weight, it is the usual arithmetic weighted average. Then for the *item weight* score (IT), calibration and information are computed per item as opposed to the global weight score where it is used an average information scores. Note that the optimized DM only applies to GL and IT DMs. Recall that the goal of the proposed DM is to reach rational consensus.

### 3.1 Data on fatigue cracking

To come up with the seed questions, we exploited data coming from crack measurements performed at the Tacitus bridge. The latter is a steel box girder cable stayed bridge located in the Dutch province of Gelderland. These measurements were performed using three different techniques, namely Crack Pulsed Eddy Current, further denoted as Crack-PEC, Ultrasonic Testing (UT) and visual observation (VO). A detailed explanation of each technique can be found in Jong (2007). Next, the measurements were carried out at various spots on the bridge, essentially located at the deck plate (DPS) when performing Crack-PEC and UT techniques and at the *trough to deck plate* (TRDPL) spot for the VO measurements (see Fig. 2 for details). These inspections were done between 30 to 35 years after the bridge was in service. The questions then used combinations of the above variables so that experts were asked to assess crack lengths. The seed variables are listed in Table 2 where each row reads as follows:

"A crack was detected by the measurement technique to be crack length 1<sup>st</sup> (mm) in Year 1<sup>st</sup> measurement, what would be its length (mm) in Year 2<sup>nd</sup> measurement using the same measurement technique?"

The realization of each question refers to the last column *Crack length 2<sup>nd</sup>*. The expertise calls on experts' reasoning, experience and ability to quantify own uncertainty on how a crack develops between two crack length records. This way, a total number of 12 seed variables were obtained and elicited from the expert

panel. The 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> percentiles of estimates of each expert for these 12 seed questions are presented in Fig. 4 including the DMs assessments as well as the realization (vertical red line). Together with the variables of interest, we end up having 36 items that need to be assessed.

### 3.2 Results

For the elicitation, the pool of experts consists of  $E = \{1, 2, 3\}$  whose field of expertise is in the steel bridge management and reliability community, including various type of inspections and decision-making more generally.

After answering the 12 seed questions and the 24 variables of interest, the estimates are processed in the EXCALIBUR software (Cooke & Solomatine 1992). Calibration and relative information scores together with experts' weight according to the different DMs (GL, EQ and IT) are presented in Table 3. Among the three experts' calibration score, none of them exceeds the cut-off level (0.05) as the greatest calibration value is obtained by expert 3 (6E-4). Theoretically, a panel in which one or more experts' calibration score is greater than this threshold means that all the other experts are attributed a zero weight. Regarding the three different DMs, they all have the same score (0.446) which desirably proves to be significantly larger than individual calibrations. As for relative information, both sub-columns ('Total' and 'Realization') refer to information scores computed with respect to all the items and only the seed variables respectively. Interesting to notice that expert 1 was quite informative regarding the overall questionnaire (2.42) but much less when looking at only the seed variables (0.52). The same observation applies to expert 2 (1.79 and 1.21 respectively) with a lesser difference than for expert 1. Expert 3 shows consistently a very similar degree of information between all the variables (0.84) and the seed variables (0.91). For the DMs, information naturally decreases between 'Total' and 'Realization' while IT DM gets the highest score in both (1.093 and 0.49). Experts commented

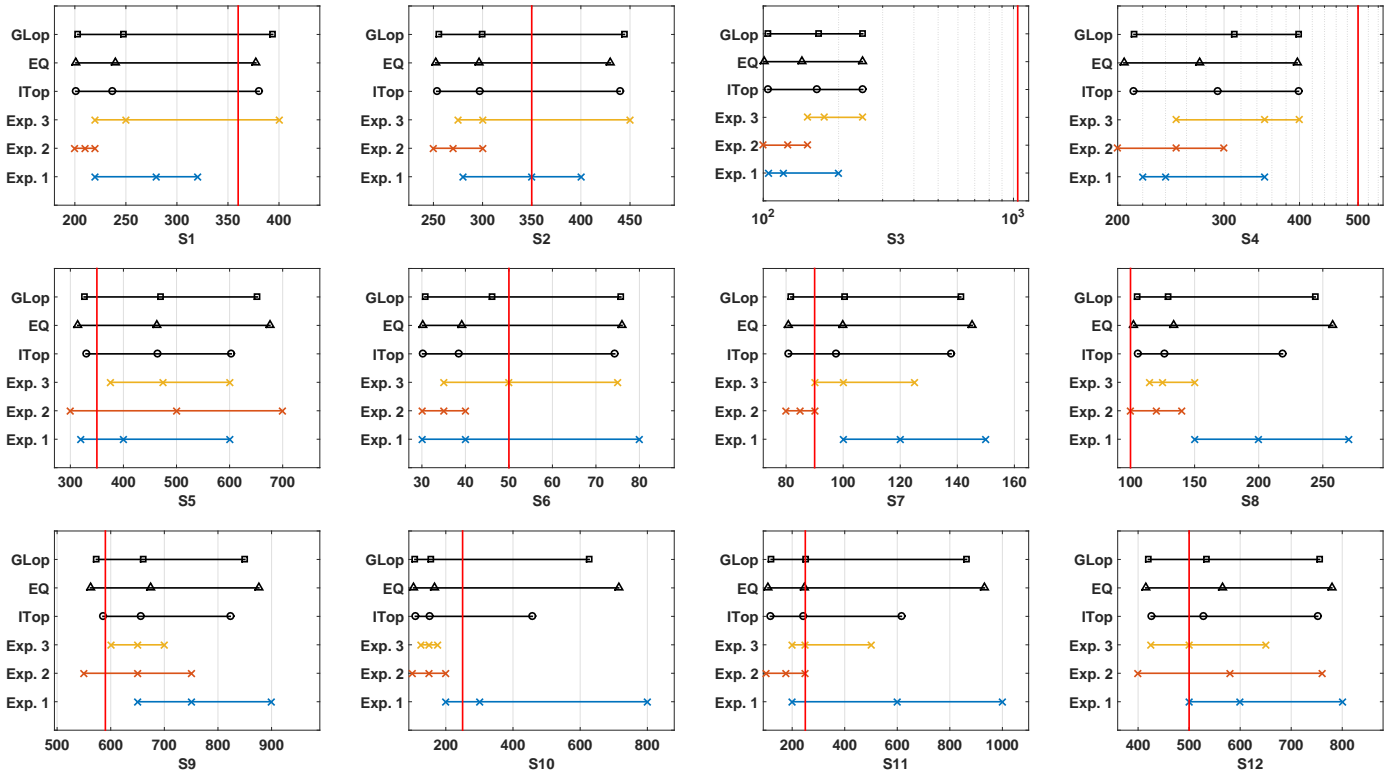


Figure 4: Distributions for the 12 seed variables as represented by their 5th, 50th and 95th percentiles for 3 experts and combined distributions derived from the item weight optimized DM (ITop), the equal weight DM (EQ) and the global weight optimized DM (GLop). The vertical red line in each plot shows the true value for the seed variable.

| Expert ID     | Calibration | Relative information |             | Normalized weight without DM |       | Normalized weight with DM |        |        |
|---------------|-------------|----------------------|-------------|------------------------------|-------|---------------------------|--------|--------|
|               |             | Total                | Realization | Global                       | Equal | Global                    | Equal  | Item   |
| Exp. 1        | 2.7E-4      | 2.42                 | 0.52        | 0.17                         | 1/3   | 7.9E-4                    | 8.5E-4 | 6.3E-4 |
| Exp. 2        | 9.8E-5      | 1.79                 | 1.21        | 0.15                         | 1/3   | 6.8E-4                    | 7.3E-4 | 5.4E-4 |
| Exp. 3        | 6E-4        | 0.84                 | 0.91        | 0.68                         | 1/3   | 3.1E-3                    | 3.4E-3 | 2.5E-3 |
| Equal weight  | 0.446       | 0.445                | 0.36        |                              |       |                           | 0.995  |        |
| Global weight | 0.446       | 0.23                 | 0.39        |                              |       | 0.995                     |        |        |
| Item weight   | 0.446       | 1.093                | 0.49        |                              |       |                           |        | 0.996  |

Table 3: Results of the performance assessment for 3 experts and three different decision makers (DMs) were compared: the equal weight DM, the global weight DM, and item weight DM.

unanimously that were more comfortable in eliciting seed question compared to the variables of interest. Though it is interesting to observe that informativeness is greater when looking at the overall score than when focusing only the seed variables. In terms of weight attribution, the columns 'Normalized Weights' (with and without DM) are used in determining the DM. For 'Normalized Weights without DM' only GL and EQ DMs are computed since the weights used for the IT DM vary from item to item. Expectedly, expert 3 gets the biggest weight (0.68) for the GL DM while expert 1 (0.17) and 2 (0.15) contributions are low. When accounting for the DM, for all three schemes the DM gets almost the whole weight (0.99) whereas all three experts contribute marginally ( $< 0.003$ ).

### 3.3 Robustness tests

Part of the post hoc analysis of the results includes robustness tests to estimate how stable the combined DMs outcomes are to (sets of) experts or calibration items. For instance item-wise, one calibration question is removed at a time and the DMs scores are re-computed. The similar procedure can be done expert-wise. Typically in our case, all three experts missed to capture within their  $[5^{th}, 95^{th}]$  quantile range the realization for S3 and S4 (see Table 2) as they all underestimated it. This is illustrated in Fig. 4 where the chosen abscissa scale is logarithmic due the fact that the realization is located too far away on the right from each of the experts' distribution. In other words, the latter fell in their upper inter quantile range, i.e. above the  $95^{th}$  percentile. By removing only S3, the DMs' calibration score improves substantially by a factor almost as large as 2 having again all three the same

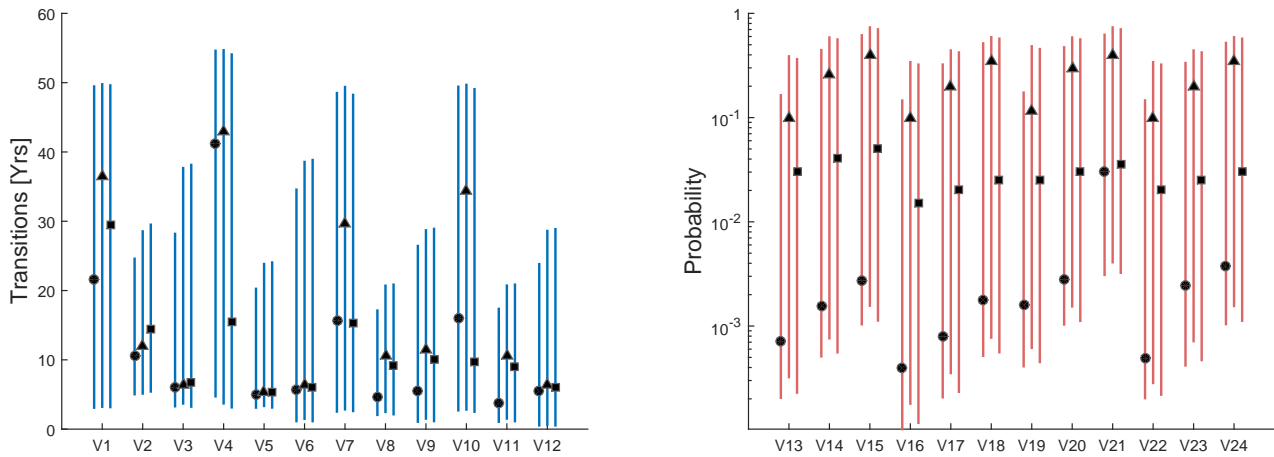


Figure 5: The decision makers distribution estimate of question 1 (left) and question 2 (right) from table 1, expressed by the 5<sup>th</sup> and 95<sup>th</sup> percentiles through the segments lower and upper tips respectively, and the 50<sup>th</sup> by the related symbol for the item weight (○), the global weight (△) and the equal weight (□).

value (0.852). Similar to the general case, IT DM outperforms the other decision makers having the highest information score (1.021) by a factor greater than 2 compared to EQ DM (0.41) and by 5 to GL DM (0.19). We mention that robustness test on experts was performed too but did not lead to any improvement. This is likely due to the small size of the panel (3 experts).

The combined distributions for the variables of interest taking into account the outcome on the robustness test are given in Fig. 5. The uncertainty intervals are narrower for the item weight DM, than for the other DMs. In spite of this, rather large uncertainties are expressed especially for variable V1, V4, V6, V7 and V10 for question 1 and for V14, V15, V18, V20, V21, V24. Specifically for V1, it reads that there is 0.9 probability that under a normal solicited load a moveable bridge would take between 3.09 and 49.45 years to transition between states 1 and 2, with a median equal to 21.62 years. We also observe that items regarding transition from state 1 to 2 (V1, V4, V7 and V10) show a great uncertainty interval compared to the other transitions asked to experts no matter the type of bridge nor its loading configuration. Similarly, V15 and V21 possess a larger uncertainty interval and have in common to address the exact same question that only differs in the type of bridge considered.

### 3.4 Discussion

Remarks coming from experts were partly related to the usage of the method as well as the degradation modelling approach in this context. Narrowing down fatigue cracking only to the deck plate and the trough-to-deck-plate locations was indeed addressed by the pool of experts.

A successful implementation of Cooke's method lies on a large extent on finding suitable seed variables. As mentioned, those should in principle resemble as much as possible variables of interest. Indeed experts' performance on the seed variables should

be judged indicative for their performance on the variables of interest. In our case, the link refers to cracking condition and development for the seed variables. In terms of the variables of interest, this type of knowledge was integrated to bridge condition as quantitative thresholds separating the different states (Question 1) and further extended to conditional probabilistic assessments (Question 2). Undoubtedly, the latter turned out to be challenging as many experts argued. However, the way conditional probabilities were assessed through proportions out of a sample mitigated the risk of getting zeros or ones in the estimates.

It is worth mentioning that the expert pool number here limits to three which claims to be rather small compared to surveys using Cooke's method (Cooke & Goossens 2008) where the number of experts usually ranges from 4 to 45. A larger panel of experts should likely enrich current results by bringing together additional experts' knowledge to the current combined DMs. Concretely, it could also entail having one or more experts whose calibration score is greater than the cut-off level (0.05).

The combined distributions for the variables of interest obtained under the item weight DM can readily be used to provide the input parameters for the degradation model, since this DM obtained the highest performance before and after performing robustness tests.

## 4 CONCLUSION & PERSPECTIVES

The research presented in this paper proposed a structured expert judgment method to quantify a degradation model composed of a combination of a Bayesian network and a Markov chain. The use of the classical method to combine opinion was elaborated to fulfill two objectives. First to explore the usefulness of applying the well-established classical method of expert judgment elicitation to the field of steel bridge



reliability and maintenance. In fact, the ambition of this study is to provide insights in this particular domain via uncertainty assessments. In that sense, this can possibly highlight the limited knowledge as well as attempting to give another viewpoint that current practice has. Furthermore, although substantial material is available in various fields including in the domain of infrastructure reliability using the classical method, no records were found for this particular class of structures. Second, in either a little- or no-data scenario, the probabilistic framework provided by Cooke's method complies with first objective. Though in this regard, addressing the quantification problem, especially when using discrete BNs whose requirements through probabilistic assessments can be very demanding, demonstrates a rather great uncertainty interval proving how challenging this task still is.

The use of the expert judgment outcome in this model will be presented in another study. As a perspective, a more extended model could address the possibility of jumping by more than one state when deteriorating, hence allowing for transitions probabilities  $p_{1,3}$ ,  $p_{1,4}$ , and so on, or even considering maintenance actions entailing for instance  $p_{i,j}$  with  $i > j$ , to be non-null. An undesirable consequence though would be a larger number of items to add to the current questionnaire.

## REFERENCES

- Cooke, R. & D. Solomatine (1992). *EXCALIBUR Integrated System for Processing Expert Judgements version 3.0*. Delft University of Technology and SoLogic Delft, Delft.
- Cooke, R. M. (1991). *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press.
- Cooke, R. M. & L. L. Goossens (2008, may). TU delft expert judgment data base. *Reliability Engineering & System Safety* 93(5), 657–674.
- Jong, F. d. (2007). *Renovation techniques for fatigue cracked orthotropic steel bridge decks*. Ph. D. thesis, Delft University of Technology.
- Jordan, M. I. (1999). *Learning in Graphical Models*. MIT PR.
- Mirzaei, Z., B. T. Adey, P. Thompson, & L. Klatter (2014). The IABMAS bridge management committee overview of existing bridge management systems. Technical report.
- Morales-Nápoles, O. & R. D. J. M. Steenbergen (2014). Analysis of axle and vehicle load properties through bayesian networks based on weigh-in-motion data. *Reliability Engineering and System Safety* 125, 153–164.
- Morcous, G. (2011). Developing deterioration models for nebraska bridges. Technical report.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kauffman Publishers.
- Straub, D. (2009, October). Stochastic modeling of deterioration processes through dynamic bayesian networks. *Journal of Engineering Mechanics* 135, 1089–1099.