Using Bayesian belief networks to predict the reliability of military vehicles

by Martin Neil, Norman Fenton, Simon Forey and Roger Harris

Predicting the reliability of military vehicles has traditionally concentrated on estimation using failure data gathered during trials or use. However, it is increasingly recognised that predicting reliability earlier in the life cycle, using design and process capability evidence, is one way of improving predictions and positively influencing reliability. This article presents the use of Bayesian belief networks (BBNs) as a decision support tool to achieve these twin goals. The BBN models presented are built into the TRACS software tool, which is in daily use within DERA Land Systems.

On behalf of the UK Ministry of Defence (MOD), DERA assesses procurement bids from various companies seeking to build new vehicle systems against MOD specifications. Typically the Data Analysis and Reliability Section (DARS) at DERA Land Systems compares several bids on the basis of reliability data for individual systems.

In terms of predictive accuracy, DERA's previous generation of reliability evaluation methods were found to be optimistic because they were based on data collected from mature designs and so did not take explicit account of design faults at the vehicle level. Much of classical reliability theory has focused on using trials to evaluate reliability. Once an unreliable vehicle has been built it is very expensive to re-design and re-test to improve reliability. The challenge for reliability engineers is to predict reliability *before* vehicles are tested or in some cases even built. Having historical reliability data, gathered in use or from testing, can provide a helpful baseline here but ultimately prediction also depends on non-statistical sources of information.

Recent initiatives within DERA and the MOD have led to a number of new standards and innovations to encourage a more holistic and wide-ranging view of reliability. A prime example is the MOD's 00-42 guidelines that present the idea of an R&M (reliability and maintainability) case whereby the manufacturer is expected to argue for R&M properties using all available relevant evidence.[1]

This need to take account of all relevant evidence for reliability assessment is what drives the work described in this article. DERA engineers have accrued considerable knowledge about vehicle failures and their causes over the last 30 years and have collected extensive statistical data. However, most of this information is 'hard' (for example, failure counts, modes and exposure periods) in the sense that it only covers the areas required by traditional reliability evaluation methods. In contrast, much of the root-cause data is 'soft' knowledge and is contained only in the heads of engineers. Prime examples of soft knowledge included things like manufacturer reputation, design staff experience, and whether the design looks 'solid'. Despite the tacit acceptance that soft factors did affect reliability the fact that the accepted 'objective' models of reliability evaluation did not, and could not, factor-in subjective information meant that such information was ignored or played down.

Fortunately, Bayesian probability allows the expression of uncertain knowledge, be it statistical or subjective, within a rational and unifying framework.

Hence, in 1997 DERA contracted a feasibility study with the Centre for Software Reliability (CSR) at City University to investigate whether BBNs could help improve on DERA's approach to reliability prediction and bid assessment. Following the positive results of this study DERA contracted further work to develop and build a system, TRACS (Transport Reliability, Assessment and Calculation System), to predict the probability that non-combat land vehicles will fulfil the requirements of their specified mission. TRACS uses soft and hard data in a single decision support model.

In the following sections, we describe the requirements of TRACS and the underlying technology (BBNs) and we explain the basic Bayesian assumptions underlying TRACS. It turns out that the BBN model for TRACS was built using a number of subnets. There are subnets corresponding to each subsystem of the vehicle under assessment (this is described in the section on predicting subsystem reliability) and there are further subnets corresponding to the design, production, and test processes described later. We also describe how we built the models, including how we elicited the relevant probabilities, and the validation of TRACS in practice.

TRACS requirements

DERA's high-level technical goals for TRACS were to predict failure rate, reliability and MDBF (mean distance between failures) for vehicles and subsystems by taking account of evidence about a manufacturer's design and production process capability when assessing reliability. These predictions would be made under different scenarios and would need to be updated through the life-
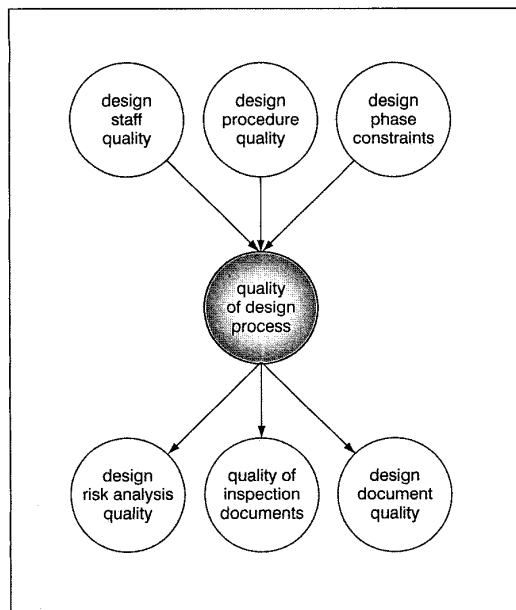

Fig. 1 Example BBN (this is a fragment of the TRACS BBN model

cycle from concept stage to tender stage, through production and into use. At the vehicle level DERA needs to:

- Assess vehicles at the tender stage using design and production process information and combine this with historical baselines for subsystem reliabilities.
- Use real data gathered from vehicle trials and combine this with data from subsystem or design level analysis.
- Assess production-line vehicle reliability and reconcile this with design and trial test information.
- Use informal information based on inspections of the vehicle as a whole by DERA staff (kicking the tyres!).

In addition to an extensive range of reliability computations and graphical outputs, TRACS provides all the supporting facilities necessary for an audited reliability assessment. This includes a detailed questionnaire/explorer view front-end that shields the user from the underlying BBN models, tools for working with reusable libraries of subsystems and components, and a range of tools for documenting assumptions and observations.

What are Bayesian belief networks?

Almost all realistic decision or prediction problems involve reasoning with uncertainty. Bayesian belief networks are an increasingly popular formalism for solving such problems. A BBN is a directed graph, like the one in Fig. 1, which is a fragment of the TRACS BBN. The nodes represent uncertain variables and the edges are the causal or influential links between the variables. For example, the value of the uncertain variable 'design staff quality' (which can be: 'very low', 'low', 'average', 'high', 'very high') will influence the 'quality of design process', which in turn will influence the 'design document quality'. Associated with each node is a set of conditional probability values that model the uncertain relationship between the node and its parents. For example, associated with the node 'design staff quality' is a set of probabilities for each value (such as 'very low') given the combinations of values of the parent nodes.

The underlying theory of BBNs combines Bayesian probability theory and the notion of conditional independence. For introductory tutorial material on BBNs see References 2 and 3. Although Bayesian probability theory has been around for a long time (Bayes developed his theorem in the 18th century) it is only in the last few years that efficient algorithms (and tools to implement them) taking advantage of conditional independence have been developed. The recent explosion of interest in BBNs is due to these developments, which mean that realistic size problems can now be solved. These recent developments, in our view, make BBNs the best method for reasoning about uncertainty.

Once a BBN is built it can be executed using an appropriate propagation algorithm, such as the Hugin algorithm. This involves calculating the joint probability
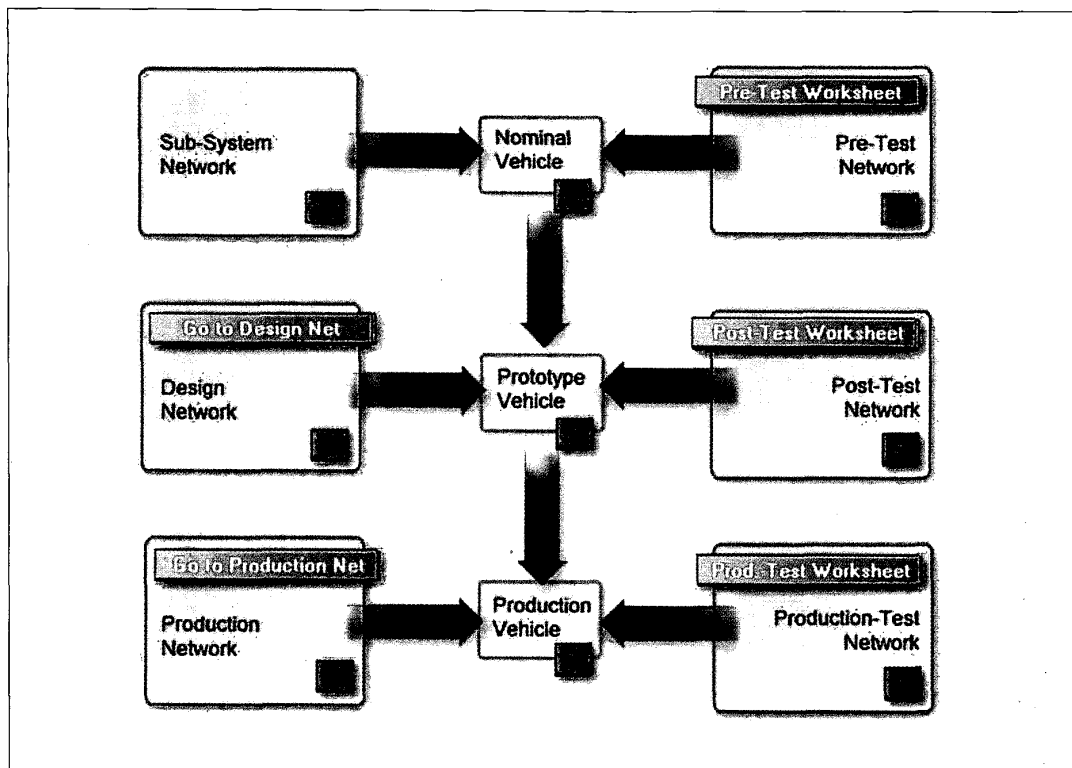
Fig. 2 Schematic of the TRACS BBN and prediction process

table for the model (probability of all combined states for all nodes) by exploiting the BBN's conditional probability structure to reduce the computational space (see Reference 4). Once a BBN has been compiled it can be executed and exhibits the following two key features:

• The effects of observations entered into one or more nodes can be propagated throughout the net, in any direction, and the marginal distributions of all nodes updated.
• Only *relevant* inferences can be made in the BBN. The BBN uses the conditional dependency structure and the current knowledge base to determine which inferences are valid.

To date BBNs have proven useful in applications such as medical diagnosis and diagnosis of mechanical failures. Their most celebrated recent use has been by Microsoft where BBNs are used in the intelligent help assistants in Microsoft Office.[5] Our interest in applying BBNs stems from the problem of predicting reliability of complex systems. Our objective was to improve predictions about these systems by incorporating diverse evidence, such as subjective judgments about the quality of the design process, along with objective data such as the test results themselves. Since 1993 we have been involved in many collaborative R&D projects in which we have built BBNs for real.

Because of our extensive practical use of BBNs we are well aware of their benefits in modelling uncertain domains. However, we are also aware of the problems. Practitioners wishing to use BBNs to solve *large-scale* problems have faced two significant barriers that have dramatically restricted exploitation. The first barrier is that of producing the 'right' graph topology, one that it is a sensible model of the types of reasoning being applied. The second barrier occurs when eliciting the conditional probability values, from a domain expert. For a graph containing many combinations of nodes, where each may have a large number of discrete or continuous values, this is infeasible. Although there has been extensive theoretical research on BBNs there is little guidance in the literature on how to tackle these two problems of scale. In our research projects we arrived at what we feel are very good partial solutions to both problems.[6-9] As a result we have been able to build what we believe are some of the largest BBN models ever constructed. In particular, the BBN models that underlie TRACS are unsurpassed in terms of scale and it would have been impossible to build them without our recent research breakthroughs.

Fig. 2 shows schematically the TRACS BBN structure and prediction process. Each component of the Figure

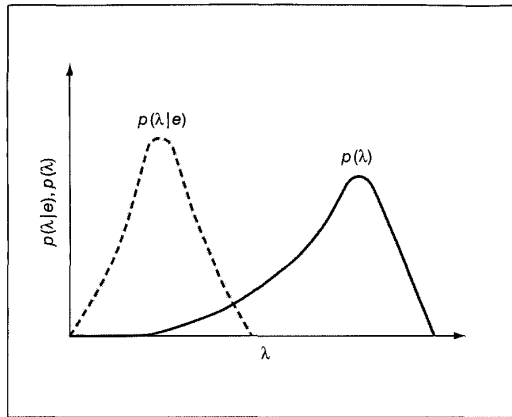Fig. 3 Prior (bold line) and posterior (dotted line) failure rate distributions

variables is shown in eqn. 1. Here our prior belief about the failure rate, $p(\lambda)$, is updated by the likelihood of the evidence, $p(e|\lambda)$, to give a posterior failure rate distribution, $p(\lambda|e)$:

$$p(\lambda|e) = \frac{p(e|\lambda)p(\lambda)}{p(e)} \qquad (1)$$

Fig. 3 shows how new evidence, e, changes the prior failure rate distribution into the posterior. The degree of change in the posterior failure rate depends on how evidence supports different values of λ and how certain we are in our prior belief about λ.

Predicting subsystem reliability using a BBN

Each vehicle is assumed to consist of a number of subsystems that have not yet been reliability-tested or used. Corresponding to each subsystem is a BBN. The goal of the subsystem BBN is to model the effects of trials evidence, gathered from previous 'similar' subsystems, on our belief about the failure rate of the subsystem under evaluation.

The idea of learning a failure rate distribution from samples of similar subsystems is a very flexible one and suits our objective to define a product family on a case-by-case basis using historical data. For example, in the
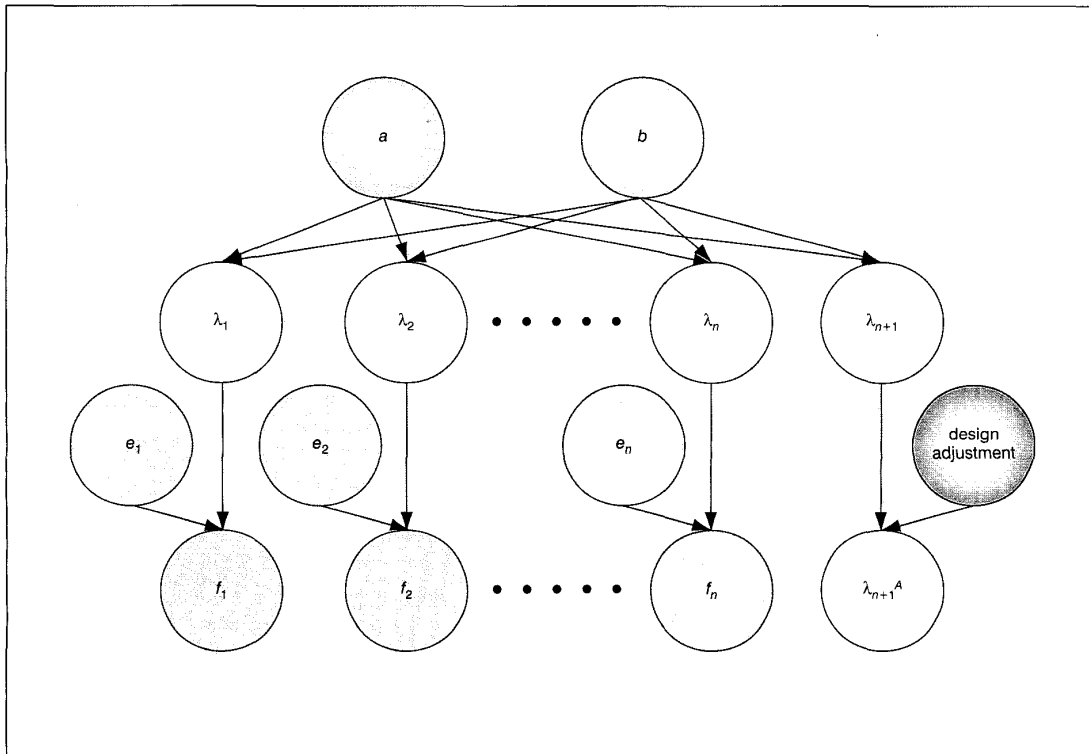
represents a subnet. In this article we will concentrate on the subsystem network and the design network, as building these parts of the model presented the greatest research challenges.

We use evidence about historical subsystems and the capability of design and production processes to update our prior knowledge about the failure rate probability distribution of the vehicle. Bayes' theorem with these

Fig. 4 Subsystem Bayesian network for predicting failure rate from population parameters
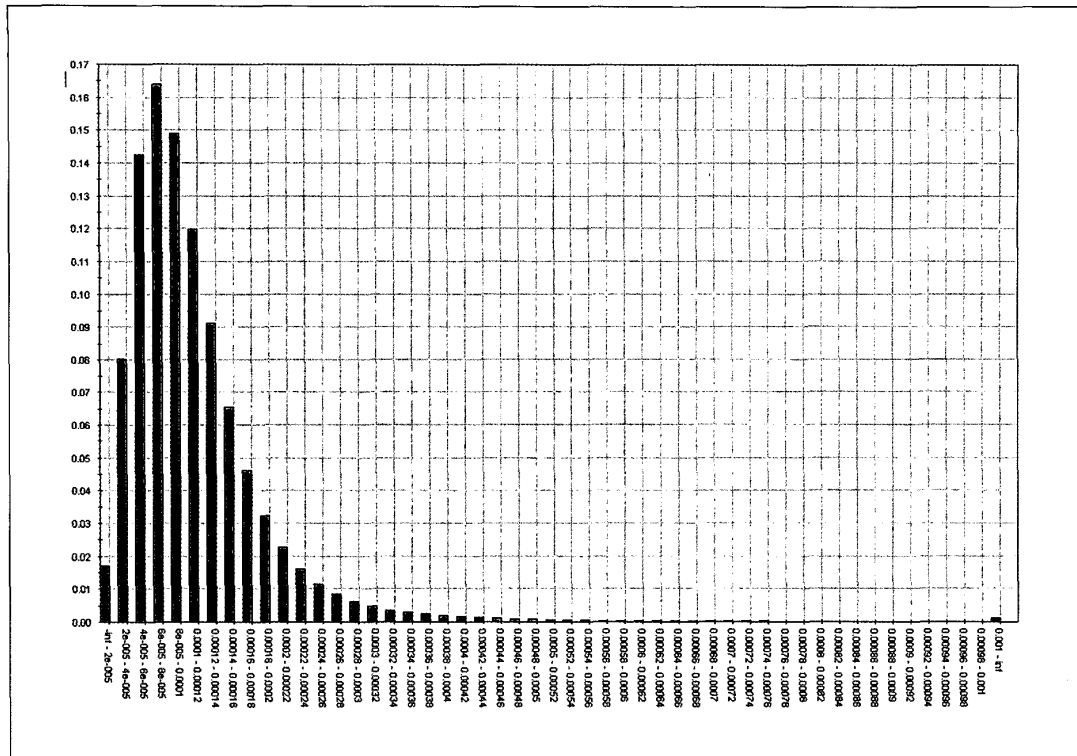
Fig. 5 Posterior failure rate distribution for $\lambda_{n+1}{}^A$ from two data sources (weight = 1 each). Mean = $1 \cdot 12 \times 10^{-4}$, median = $9 \cdot 3 \times 10^{-5}$, variance = $1 \cdot 17 \times 10^{-8}$

case of an engine, likely scenarios for this might be where the following is known about the new vehicle being evaluated:

● It will be a petrol engine of a particular size. The product family is then defined as all engines in the database with these characteristics from different manufacturers.
● It will be a diesel engine. The product family is therefore all diesel engines.
● It might be either petrol or diesel. The product family is drawn from all petrol and diesel engines.

A family of products known to have a very high variation between subsystem failure rates will be described in the failure rate distribution of λ. For a product chosen at random from this distribution the high variation of its failure rate distribution will reflect this, even if similar subsystems all had their own individually tight, but different, failure rate estimates. The subsystem BBN is shown in Fig. 4.

Each subsystem, from $i = 1$ to n, used to characterise the product family is represented by the nodes, f_i (failure count), e_i (exposure distance), and λ_i (failure rate). The failure rate distribution of the subsystem being predicted is represented by λ_{n+1}. The hyper-parameters repre-

senting the product family are shown as nodes a, b. The implemented BBN allows up to 50 subsystems to be used for inference.

Data from historical subsystems, such as contained in the TRACS data library, are entered into the BBN to define the product family of interest. These are represented by the failure count and exposure distances (f_i, e_i) for each subsystem in the BBN. This new evidence updates the distributions for each λ_i, which in turn propagate to update the product family hyper-parameter distributions (a, b). In turn this is propagated by the BBN to update all λ_i and finally λ_{n+1}. It is worthwhile noting that each subsystem failure rate changes as it is updated by its own actual failure data *and* the propagated effects from other subsystem rates, via the product family parameters

In TRACS we can use weights to model bias towards one or more subsystem data sources over others in situations where we might believe that these are more likely to be representative of the unobserved subsystem being predicted. For example, suppose we wish to predict the reliability of a subsystem based on two data sources. One is a subsystem tested for 100 000km that experienced five failures, while the other experienced 20 failures in 100 000km. If we are indifferent to the choice between these two subsystem data sources then the
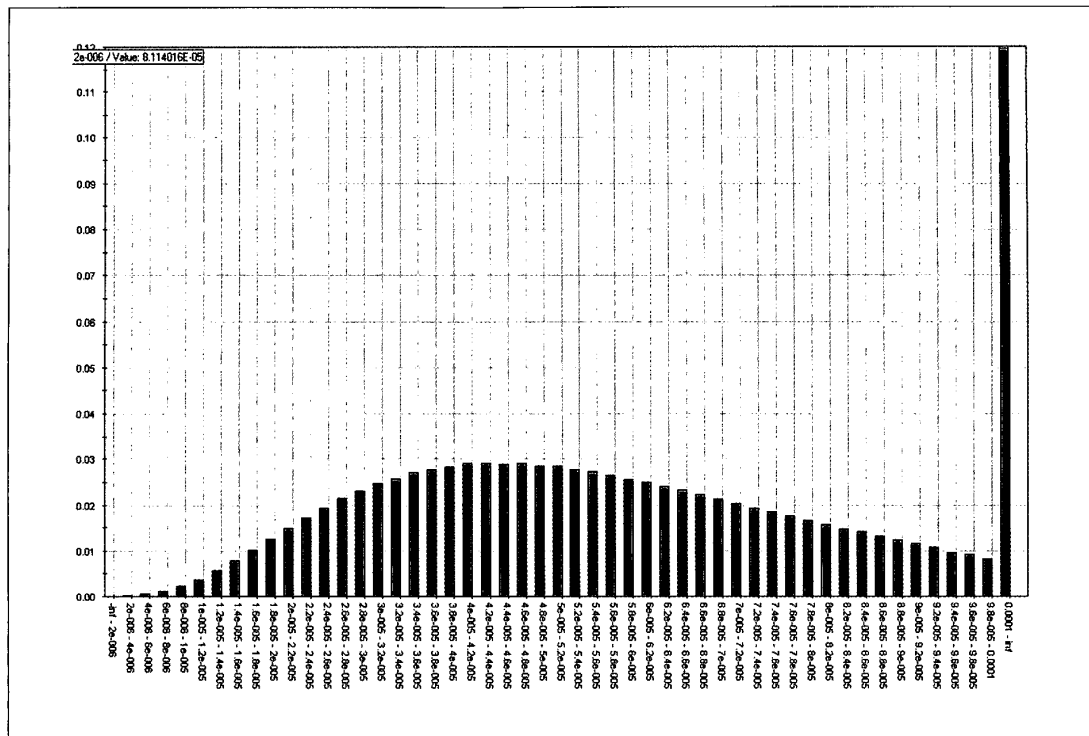
Fig. 6 Posterior failure rate distribution for $\lambda_{n+1}{}^A$ from two data sources (weight = 10 and 1, respectively).
Mean = $6\cdot1 \times 10^{-5}$, median = $5\cdot5 \times 10^{-5}$, variance = 112×10^{-9}

posterior failure rate for the unobserved subsystem is shown in Fig. 5. However, if we believe that the first subsystem is much more representative of the specific subsystem to be used by the manufacturer for the new vehicle being considered then we might give it more weight. The result of giving it a weight 10 is shown in Fig. 6. Notice that the mean has decreased considerably (in fact the scale here is in the 10^{-6} interval range rather than 10^{-5} in Fig. 5).

Design, production and test BBNs
BBN structure

The design, production and test BBNs are presented to the user as a modularised structure comprising five separate BBNs as an aid to understanding and navigation. However, in fact, they are represented within TRACS as a single BBN and so evidence entered in one module affects the others.

The structure of the design, production and test BBNs is shown in Figs. 7 and 8 (the dashed edge in Fig. 8 represents the joins between the 'BBN modules'). In summary the functions of the BBNs are:

● *Design and production BBNs:* Here the pre-vehicle failure rate, λ_{pre}, is adjusted to produce a prediction of the post-vehicle failure rate, according to the quality of

the design process applied and any direct observations made of the post-vehicle failure rate, λ_{post}. We adjust the pre-vehicle reliability distribution according to our beliefs about the differences, introduced over a period of time, between the historical (pre) vehicle and the vehicle of interest (post). This BBN is shown in Fig. 7. The production BBN has a similar structure.

● *Test BBNs:* If actual failure count and exposure data are available on the vehicle under assessment they can be used to update the failure rate distribution. Here exposure period and failure count data on the vehicle are used to update the failure rate distribution for the vehicle failure rate node, λ_{pre}. This BBN is shown in Fig. 8 by the three sets of node triads: representing the failure count, exposure period and failure rate respectively. For the pre-vehicle these are f_{pre}, e_{pre}, λ_{pre}, and similar conventions are used to denote the post-vehicle and production vehicle testing nodes.

Vehicle design BBN

The vehicle design BBN is shown in Fig. 7. TRACS users assess the design process in terms of the historic capability, the current capability of the organisation and the quality of the project processes in place. Likewise, direct assessment of the vehicle is done by observing intermediate products, such as design documents and

risk analyses, and the final product, the vehicle itself.

To give a flavour of the types of factors considered in the design BBN we list a subset below. These are organised according to the type of assessment being made (note the timeline of causality that permeates the design process):

(i) Assessment of reputation (how well did the organisation do in the past?)
● Historical design record: this is the track record of the organisation in terms of its staff and procedure quality.
● Relevance of historical design record: given that organisations change we need to assess how relevant the

historical record is in terms of what it can tell us about the current organisation.

(ii) Assessment of capability (what should the current team be capable of?)
● Design procedure quality: this is a measure of the quality of the written procedures, guidelines, standards and documentation in use to guide the design process.
● Complexity of design requirements: this represents the complexity of the MOD's requirements for the vehicle. Tougher requirements may be more difficult to satisfy.
● Schedule compression: schedules that are unreasonably tight can have a devastating effect on the reliability of
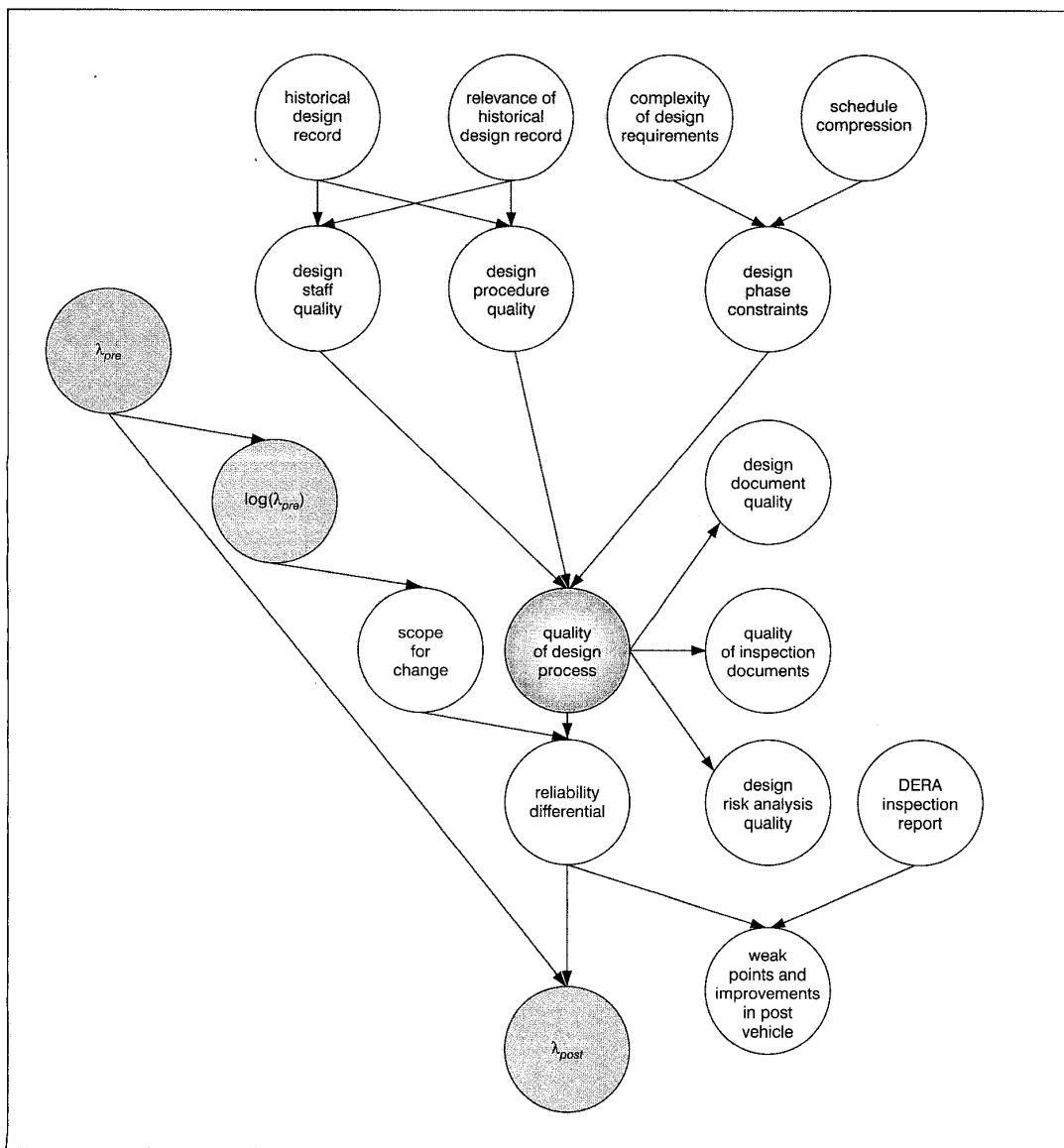

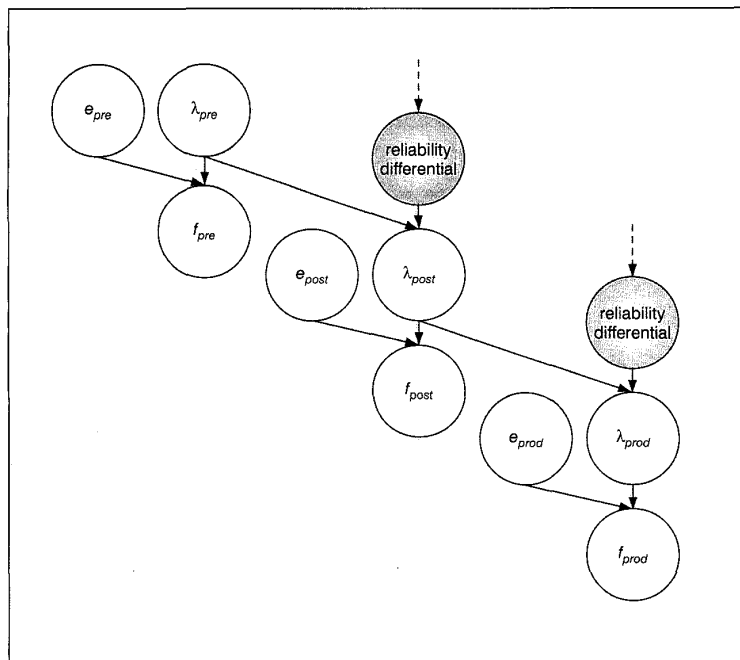**Fig. 7 The design BBN to predict** λ_{post}

Fig. 8 The pre, post and production vehicle test BBN fragments (the dotted arcs represent the rest of the design, production and test BBN)

the resulting vehicle.

(iii) Assessment of the actual design process (how good was the actual design process applied?)

- Quality of inspection documents: this is a measure of the extent to which an R&M plan was applied and whether interdisciplinary and QA procedures were applied.
- Design risk analysis quality: this is a measure of FTA, FMECA and other types of risk analysis that may have or may not have been carried out on the vehicle.

(iv) Assessment of the vehicle reliability (how good was the final product?)

- Reliability differential: this measures the expected difference in failure rates between the pre and post-vehicles and is influenced by our assessment of the quality of the design process.
- Scope for change: it will be more difficult to improve very or ultra reliable vehicles with correspondingly low failure rates regardless of the quality of the design process applied. This node acts as a brake on improvement to reliability.

As an example, we will show how TRACS uses negative process evidence to update post-vehicle prediction. Assume from an audit we had discovered the following:

- schedule = 'very tight'
- complexity of design requirements = 'high'

- design staff quality = 'low'.

Fig. 9 shows a TRACS screen dump with this evidence and the corresponding pre- and post-vehicle failure rate distributions shown alongside. The 'blue' histogram bars indicate that facts have been entered from those nodes in the design BBN. Red coloured distributions are shown for uncertain nodes of interest.

The resulting reliability differential is slightly skewed to negative values for this evidence, resulting in a lower median failure rate for the post-vehicle compared to the pre-vehicle. Note that the predictions for the intermediate products of the design process, such as risk analyses and design inspections, tend towards the negative also.

The post-vehicle failure rate distribution graph is shown in Fig. 10. The vehicle's median failure rate has increased from 1.9×10^{-4} (pre-failure rate) to 4.3×10^{-4} (post-failure rate) and demonstrates the effects of negative process evidence on our beliefs about reliability.

Probability elicitation and validation of predictions

There were two elements of the design and production BBNs to develop for TRACS. Firstly, the BBN graph topology had to be agreed and then the node probability tables (NPTs) were required for all probability tables in the graph.

The graph topologies for the design and production BBNs were developed by Agena and DERA senior staff together. This process was iterative and involved some negotiation to ensure that the factors that DERA identified and their relationships could be placed correctly according to sensible causal rules and in a way consistent with the idioms employed by Agena to build BBNs.

The design and production BBN NPTs were elicited, typically, in sessions that included from three to ten engineers with differing levels of experience and seniority. During these elicitation sessions anecdotes and past cases were used to illustrate the likely shape of probability distributions and this process in a way served to cement a core of reliability knowledge between the more senior and junior members of the team. Clearly the probabilities elicited here are not personal beliefs but are products of a consensus of belief.

Complete validation of prediction systems the size of

TRACS is difficult. The timescale of vehicle selection, deployment and use militates against collection of consistent useful data. Also, because of the historical focus on traditional 'objective' reliability evaluation methods little effort has gone into collecting any data on design and production processes. Therefore evaluation of the 'soft' elements of the BBN is difficult to do in the short term. But now that a data collection system exists, in the form of TRACS, we can use past studies to determine the extent to which the factors identified by DERA really do cause vehicle unreliability.

Retrospective studies of vehicle assessments using TRACS already confirmed that TRACS generally fixed the problem of over-optimistic predictions encountered with previous methods. Under a recent 12 month validation study by DERA (1999-2000) TRACS has been used on four further vehicle assessments, ranging from battlefield ambulances to motorcycles. Although the reliability predictions are confidential the results are extremely promising. DERA felt the predictions made appeared to be sensible. While it is therefore too early to make firm statements about predictive accuracy, we can claim that TRACS enhances repeatability of assessments

and provides a decision support framework of considerable explanatory power.

Conclusions

The TRACS system serves as a good exemplar for the practical use and benefits of BBN technology. By producing TRACS we have shown that very large BBNs can be constructed in a modular fashion in a way that allows the overall BBN model to grow arbitrarily large. Likewise we have developed a very large 'learning' BBN, the subsystem BBN that learns the probability distribution of a subsystem failure rate from a population of similar subsystems.

We have elicited very large subjective probability distributions from DERA relating to design and production process maturity using novel techniques. This experience has taught us that probability elicitation is neither as difficult as we expected nor as difficult as the literature would have us believe. The use of causal assumptions in building the graph topologies and our automatic methods for generating NPTs has significantly improved the efficiency and effectiveness of BBN construction.
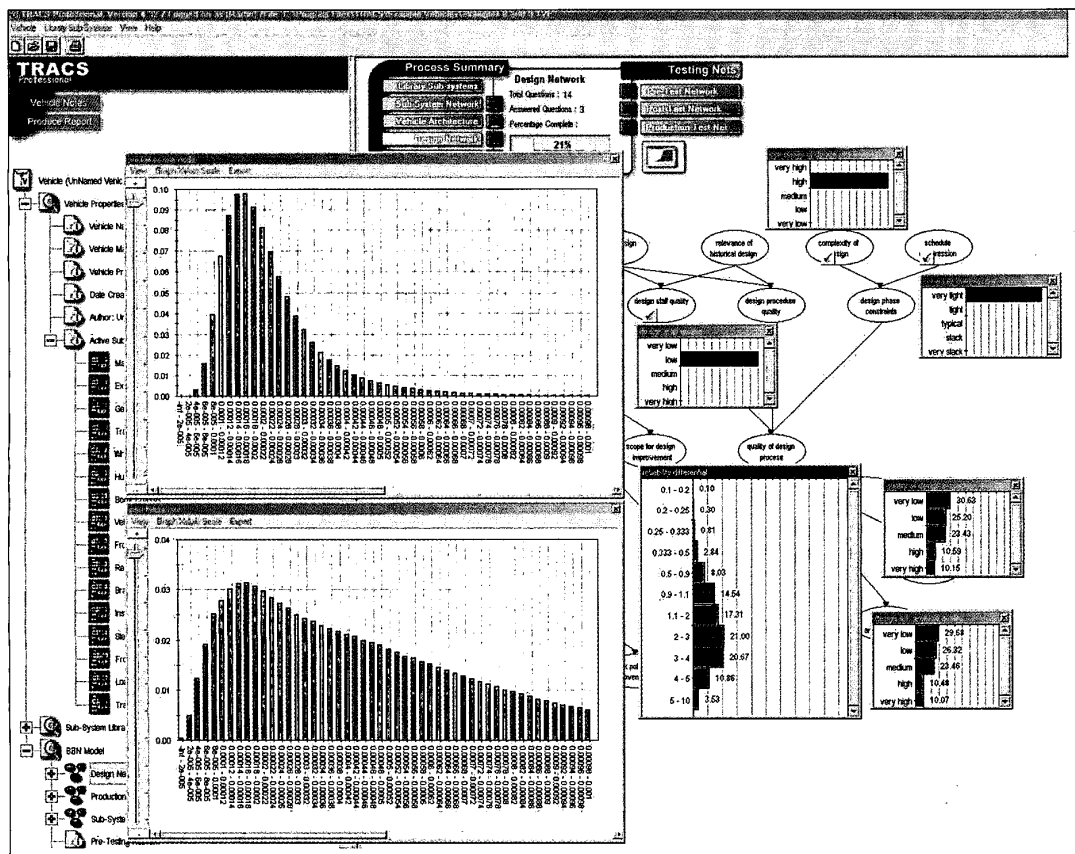

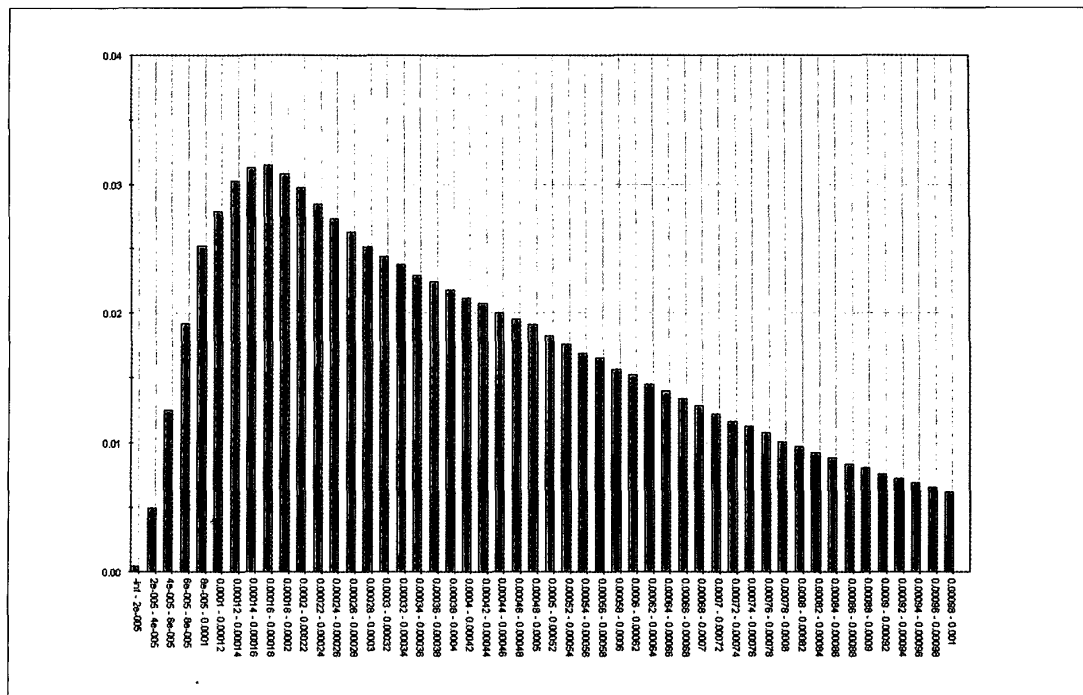Fig. 9 TRACS screen dump showing positive process evidence and its effects

Fig. 10 Updated post-vehicle failure rate distribution, λ_{post}, when negative soft evidence entered.
Mean = $5 \cdot 87 \times 10^{-4}$, median = $4 \cdot 3 \times 10^{-4}$, variance = $3 \cdot 59 \times 10^{-7}$

TRACS has been very well received within DERA and its customers both within industry and the MOD. Where predictions have been made, in validation trials, the results have been excellent. However, it is important not to assess a decision support system on the basis of its predictive accuracy alone. The real benefit from the approach is being able to help DERA assessors think through the reasons that lie behind a prediction and also, perhaps more importantly, identify to the manufacturer areas for potential improvement.

TRACS was designed to meet DERA's specific requirements and should not currently be considered as a general R&M package. However, some of the principles used in its development are general and can be easily tailored to other types of reliability analyses in other domains where reliability concerns are high priority.

In the near future DERA plans to expand TRACS to cope with reliability growth trials, maintainability and availability assessment and the evaluation of the effectiveness of fixes made to vehicles during trials.

Acknowledgments

References

1 UK Ministry of Defence, Defence Standard 00-42 (Part 3)/Issue 1. Reliability and Maintainability (R&M) Assurance Guidance, Part 3: R&M Case, 22nd October 1999.
2 'Bayesian belief nets', Agena Ltd., Cambridge, UK. Article at http://www.agena.co.uk, 2000
3 Hugin Expert A/S, Aalborg, Denmark, On-line brochure at http://www.hugin.dk
4 JENSEN, F. V.: 'An introduction to Bayesian networks' (UCL Press, London, 1996)
5 HECKERMAN, D., and HORVITZ, E.: 'A Bayesian approach to inferring a user's needs from free-text queries for assistance', in COOPER, G. F., and MORAL, S. (Eds.): Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (Morgan Kaufmann Publishers, Inc., San Francisco, 1998;) see also the Microsoft Research decision theory group at http://www.research.microsoft.com/research/dtg/
6 FENTON, N., LITTLEWOOD, B., NEIL, M., STRIGINI, L., SUTCLIFFE, A., and WRIGHT, D.: 'Assessing dependability of safety critical systems using diverse evidence', IEE Proceedings on Software Engineering, February 1998, 145, (1)
7 FENTON, N. E., and NEIL M.: 'Bayesian belief nets: a causal model for predicting defect rates and resource requirements', Software Testing and Quality Engineering, 2000, 2, (1), pp. 48-53
8 NEIL, M., FENTON, N. E., and NIELSEN, L.: 'Building large-scale Bayesian networks', Knowledge Engineering Review (to be published)
9 The SERENE method', ESPRIT 22187 project, http://www.dcs.qmw.ac.uk/~norman/serene.htm

Martin Neil, Norman Fenton and Simon Forey are with Queen Mary, University of London and Agena Ltd., UK. Roger Harris is with DERA Land Systems, Chertsey, UK.