



How Tree Risk Assessment Methods Work: Sensitivity Analyses of Sixteen Methods Reveal the Value of Quantification and the Impact of Inputs on Risk Ratings

By M.B. Norris and G.M. Moore

Abstract. Sixteen tree risk assessment methods were subjected to sensitivity analysis to determine which factors most influenced the output of each method. The analyses indicate the relative influence that the input variables exert on the final risk value. Excel was used to create a simple $\pm 25\%$ or ± 1 rank change (depending on the method) for each criterion, with the change to the output recorded as a percentage. Palisade's @Risk software was used to undertake a Monte Carlo (with Latin Hypercube sampling) simulation of 5000 iterations based on the input variables and output formula. From the simulation, multivariate stepwise regression was undertaken to determine the influence of each method's input variables in determining the output values. Results from the sensitivity analysis indicate some clear and strong differences amongst the 16 methods, reflecting that the underlying mathematics, input categories, ranges, and scaling influence the way that different methods process and express risk. It is not surprising that methods perform differently in different circumstances and express risk level differently. The analyses demonstrated that most methods placed too great an emphasis on limited aspects of risk assessment. Most methods strongly focused on the hazard or defect aspects of assessment and the likelihood of failure rather than the consequence aspect of an assessment. While methods were uniquely different, they could be placed into 3 broad groups: Group 1 methods produced a normal distribution with most values around the mean; Group 2 methods produced outputs at the lower end of the risk scale; and Group 3 methods produced outputs evenly if not continuously across the risk scale. Users of tree risk assessment should understand the strengths and weaknesses of any method used, as it could be relatively simple to challenge the results of a risk assessment based on limitations inherent in the underlying methodology.

Keywords. Risk Assessment; Risk Consequence; Risk Likelihood; Sensitivity Analysis; Tree Risk.

INTRODUCTION

Prior to the 1980s, with the exception of work by the USDA Forestry Service (USDAFS), few tree risk assessment methods existed (Paine 1971; Johnson 1981; Robbins 1986). It is now commonplace for arborists to undertake tree risk assessments, which are significant elements of urban tree management, particularly for community-owned trees growing in public open spaces, such as parks, road reserves, and around public buildings (Koeser et al. 2016; Smiley et al. 2017). Based on the definitions used in Australian Standard AS/NZS 4360:2004 and adopted for this research, risk management is defined as: *The systematic application of management policies, procedures, and practices to the tasks of communicating,*

establishing the context, identifying, analysing, estimating, evaluating, treating, monitoring, and reviewing risk. Confusion can arise over the use of *hazard* and *risk*, but this paper defines hazard as being associated with the source of danger, such as a rotten branch, and risk being associated with elements of uncertainty, such as the likelihood of failure, the consequences of failure, and the time of exposure to the hazard.

In some parts of the world, local government agencies may be protected from legal action by an extension of sovereign immunity, but even in the USA, this may not completely exclude actions for damages caused by the negligence of an agency or its employees (Cornell Law School 2020). In Australia, tree risk management occurs largely because insurers of local

governments insist that members have a formalised risk management system (Hewett et al. 2003; State of Victoria 2017; VAGO 2018). Local government to a degree transfers its responsibility for risk by purchasing insurance. The insuring bodies require a risk management plan to minimise their exposure and therefore transfer some of the expense back to local government. It is assumed that if tree risk assessments are to be undertaken, then the context of the possible interaction of the tree, people, and property has been set, and only the level of risk an organisation is willing to accept needs to be defined. This means that risk assessment is done on a specified tree with potential elements of risk identified. A ranking system based on an accepted risk level will allow for variation of risk acceptability (Koeser et al. 2016).

Some widely acknowledged tree biology underpins tree risk assessment and the identification of hazards (Table 1). While the risks to property and persons posed by urban trees are small (Helliwell 1990; Lonsdale 1999, 2007; Ball and Watt 2013; Hartley and Chalk 2019), the issues and liability arising from foreseeability and the resultant duty of care are significant. The risks associated with trees are so low that an exaggerated perception of these risks is likely to be the greater concern (Slovic 2000), but the insurers of local governments in Australia dedicate significant resources to the liability mitigation aspects of public tree management (Hewett et al. 2003; VAGO 2018).

A good tree risk assessment methodology should be complete, robust, valid, repeatable, available, usable,

and credible (Lowrance 1976). Any person using a risk assessment process should understand the limitations of the decision theory underpinning the particular methodology, industry heuristics, and their own cognitive processes and biases. Since risk = likelihood \times consequences (Standards Australia 2004a), risk assessment must be undertaken in a climate of uncertainty, as both components are estimates and predictive. An assessor is in the business of uncertainty, and tree risk assessment systems utilise professional experience and judgement (Coder 1996, 2000). Risk assessment is a logical-step-driven decision-making process (Figure 1) that assists with processing data into intelligence that provides the information required to achieve competent decisions (Haimes 1998; Tomao et al. 2015). A primary rationale in using a method is to improve the validity and repeatability of assessments by reducing the potential for errors and biases by using defined parameters, terms, and variables.

In general risk assessment, once the consequence of a risk has been established, credibility will be affected by the technical community's best estimate of magnitude, how good the technical community believes the estimate to be, and how far the technical community's judgement can be trusted (Fischhoff 1994). In dealing with tree risk assessment, the technical community is made up of arborists who have the level of professional competence to undertake risk assessments. The establishment of credibility becomes the responsibility of the arboriculture profession. While tree risk assessment potentially aligns

Table 1. Elements of tree biology considered in tree risk assessment.

| Aspect of tree biology | Considerations in risk assessment |
|--|--|
| Every tree will eventually fail. | Unless the tree is removed first. |
| Trees are biological entities. | They are variable genetically and in responding to their environment. |
| Trees compartmentalise. | They seal off and grow over damaged tissues. |
| Most trees have some defects. | The extent of the defects varies. |
| Trees are slow to react. | Tree growth is "adaptive," therefore "visualisation" within the context of a longer tree time is required to understand tree actions and reactions. |
| Trees are regenerating systems. | Trees do not heal and must develop new material to continue growing. |
| Trees are shedding/biological organisms. | They create litter (leaves, bark, twigs, branches). |
| Healthy trees can drop limbs or break. | They may do so naturally or during certain weather conditions. |
| Many defects are not visible. | Technology may assist, but judgement is important. |
| Limited knowledge of tree biology and physiology may be a limitation to risk assessment. | Experience and judgement may be important in risk assessment, particularly in predicting failure or development of hazards with a greater degree of certainty. |

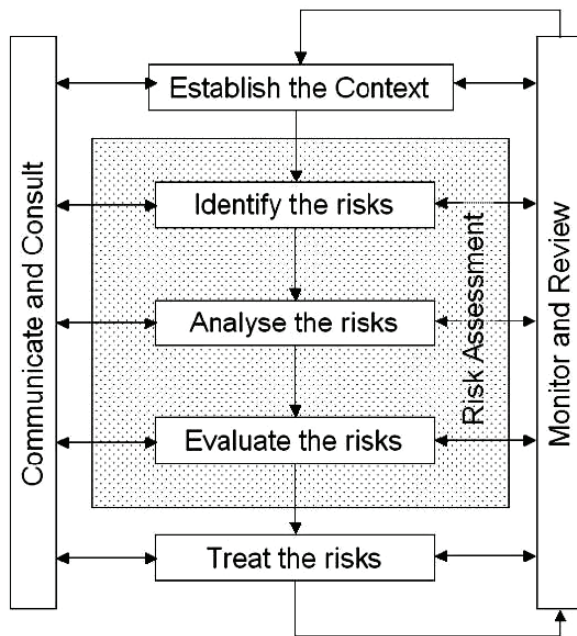


Figure 1. Risk Assessment process derived from the AS/NZS 4360:2004. (Note that Establishing Context and Treating Risk are considered risk management functions that sit outside the assessment process.)

well with established general risk frameworks such as ANSI or Standards Australia standards, there is still a lack of agreed upon arboriculture industry standards, as the diversity of tree risk assessment methods used in different countries and by different organisations within countries attests.

Despite the wider use of methods such as Tree Risk Assessment Qualification (TRAQ) (Dunster et al. 2017) or Quantitative Tree Risk Assessment (QTRA) (Ellison 2005a, 2005b), which provide an opportunity for the standardisation of terminology and processes, many other and often older methods are still in wide use. These newer methods may be used by a sector of the arboriculture industry, particularly those who are members of professional and industry associations, but they can be expensive of both time and money and require users to register and to commit to ongoing professional development. As a result, they are used by a minority of those undertaking tree risk assessment. There is concern about the credibility and validity of existing methods, limited use of hard data, and highly judgemental and subjective inputs. In many industries, it is rare to find disclosure of the degree of judgement used in defining a risk due to the belief that disclosure would result in the public

perceiving the assessment process as dishonest. However, it is important that any method is transparent and legally defensible (Fischhoff 1994).

To create an outcome, tree risk assessment methods need to answer a series of questions, such as: “What can go wrong?” “What is the likelihood that it will go wrong?” “What are the consequences if it does go wrong?” “How does this risk compare?” “Is the risk acceptable or unacceptable?” “What treatment options are available?” and “What is the residual risk?” (Haimes 1998; Standards Australia 2004a, 2004b; Koeser et al. 2016; Smiley et al. 2017). From an arboricultural perspective, these questions can be grouped. “What can go wrong?” is an assessment of identifiable structural defects and their severity, other non-tree hazards created by the tree (raised pavements, low branches, poisonous fruits), and an assessment of the existing and potential targets (Figure 2).

A second, more difficult question is composed of two elements: the likelihood of failure and the likelihood of the target being affected (a target being present when failure occurs). The latter can be analysed and a probability derived (possibly predefined in a method); however, the former is essentially judgement-based with a high degree of uncertainty. Arborists are rarely trained to quantify consequences of failure, which involves identifying the target, determining how often the target is in the vicinity of the tree, and what level of injury or damage might be done to the target if failure occurs. Most tree risk assessment methods use size of part to define consequence, which often fails to reflect the likely consequences well. The question of risk evaluation including acceptability and treatments is dependent on the earlier questions. Without broadly accepted industry guidelines, what is considered acceptable risk will continue to remain in the domain of individual assessors.

Trees are currently assessed for risk under three types of systems:

Qualitative: where the person undertaking the assessment visually inspects the tree for risk and provides a written (oral) report and recommendations. Subjective terms tend to be used to describe the risk, such as high, medium, or low.

Quantitative: quantitative risk assessment (QRA) uses true numerical values and is common in many industries but rare for tree risk assessment. For tree risk assessment, each and every likelihood and consequence element of the method would be assigned a numerical value.

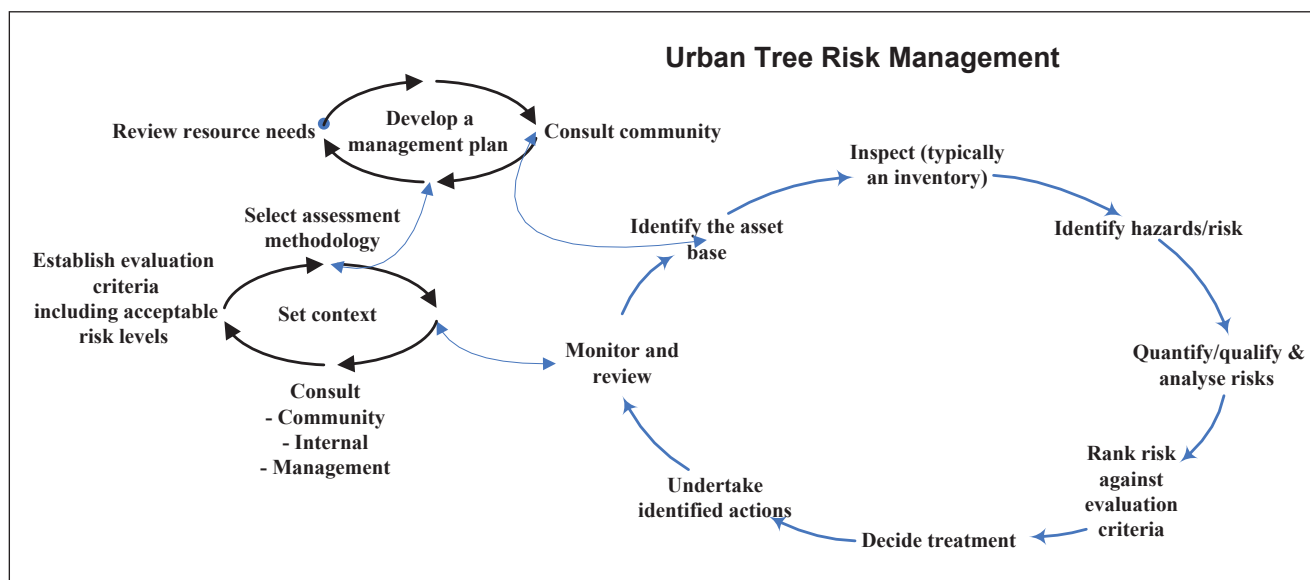


Figure 2. Summarised public urban tree risk management process (based on the Australian Risk Management Standard AS/NZS 4360:2004).

Semi-quantitative/qualitative: an inspection is undertaken, documented, and the elements that compose the likelihood and consequence components are assigned numerical or alphanumerical coding (typically ordinal numbers that are summed to produce a risk value). The output values are ordinal only (a risk score of 6 is greater than 3, but there is no numerical relationship, and it is not twice the score of 3). These methods suit large tree populations, where process speed and the advantages of coding are of high value.

A list of assessment methodology criteria that meets the needs of public authorities was provided by Hickman et al. (1989) using “hazard” for what today is called risk. Methods should be predictive, rapid, allow prioritisation and reduce liability concerns, address the legal issues of foreseeability, requirements for inspection, and record keeping, and be defensible while being pragmatic, cost effective, and easy to administer. Most modern tree risk assessment methods cover these criteria (Table 2). Tree inspections are a quick visit to a tree to determine if a more detailed assessment is required, and the majority of street- and park-tree inspections are free of issues and, to be cost effective, cannot exceed 3 to 5 minutes. Trees identified with more complex issues should be assessed separately. Ellison (2005a) promotes a targeted approach where only trees in areas that have use above certain levels are assessed, based on the premise that in low use areas, the risk is acceptable regardless

of the significance of the hazard. Several methods have recognised the requirement for quick assessments: Forbes-Laird’s (2003) Tree Hazard: Risk Evaluation and Treatment System (Threats) method can be used in an extensive mode suitable for full risk assessments and in an abridged form for rapid inspections, and Matheny and Clark (1994) had a shortened “quick survey” format. The Colorado Tree Coalition (2004) method is 2-tiered, where a more intensive assessment is only triggered when certain thresholds are reached, and QTRA (Ellison 2005a) has a walk-over survey category that pre-empts more extensive inspection.

Acceptable risk is a fundamental tenet of risk management (nothing is risk free). The likelihood of a non-workplace tree-related death is extremely low (Hartley and Chalk 2019). Trees provide value to the community, and a level of acceptable risk is implicit in the community’s desire for trees. However, the discussion of acceptable risk is largely absent from arboriculture (Sreetheran et al. 2011). The degree of risk relating to trees is rarely qualified, and obtaining data is difficult (Helliwell 1990, 1991; Ellison 2005b). The majority of existing tree risk assessment methods do not discuss or quantify acceptable risk levels (Matheny and Clark 1994; Pokorny et al. 2003; Bloniarz 2004; USDA Forest Service 2007), which is a significant deficiency given their widespread use. Broad, acceptable risk guidelines exist in many areas

Table 2. Examples of tree risk assessment data collection (Hickman et al. 1989; Matheny and Clark 1994; Forbes-Laird 2003; Hewett et al. 2003; Pokorny et al. 2003; Ellison 2005a; van Wassenae and Richardson 2009; State of Victoria 2017).

| Category | Typical data collected |
|-------------------------------|---|
| Assessment details | Date, name of assessor, scope of works, authority to undertake work |
| Location data | Identify site and tree within site; could include tree inventory (what species of tree and its location), tree record number, owner |
| Site data | Area use type, soil type and issues, site restrictions (utilities, paving, buildings), site history, exposure to weather |
| Tree data | Identification, history, age, physical data canopy and trunk (height, spread, DBH), physical data root system (root crown exposed/damaged roots) |
| Tree health data | Health, vigour (or vitality), pests, or diseases |
| Defects | Defects (decay, cracks, lean, poor architecture, deadwood, mechanical damage), defect symptoms (dieback, included bark, fungal fruiting bodies), site environment defects (low branches, raised path), severity and location of defects |
| Target damage (consequence) | Degree of damage possible, injury severity, “value” of target (children) |
| Target occupancy | Amount of time the area is occupied and sometimes “value” factors, influence of weather |
| Risk rating/ranking/value | Rating for various likelihood and consequence factors. Methods tend to rate size of part, likelihood of failure, and target ratings, and combine these to form a risk score |
| Treatment/controls/management | Options to mitigate risk, including accepting risk, often includes tree work recommendations |
| Residual risk level | The level of risk remaining after the assessment’s recommendations have been undertaken (this includes the “take no action” assessment) |

of public risk, but not in relation to trees, and owners of tree risk should develop defensible approaches to risk in line with national risk standards.

The level of acceptable risk depends upon site management objectives and management’s perceptions and expectations of tree performance (Corder 1996). The Australian risk standard AS/NSZ 4360:2004 states that it is the function of management to set the context and evaluation criteria against which risk assessments are to be evaluated. Helliwell (1990) suggested that a 1:10,000 (0.0001) risk per annum would be a suitable figure for trees based on everyday risk fatality rates. Lonsdale (1999) used a United Kingdom Health and Safety Executive report of the Interdepartmental Liaison Group on Risk Assessment (HSE 1996) to suggest that the 1:10,000 was perhaps appropriate, and Ellison (2005a) supported 1:10,000 based on the papers and reports cited by Helliwell (1990) and Lonsdale (1999). These reports explain that whilst 1:10,000 is the point where risk is totally unacceptable, and that an individual risk of death of 1:1 million is a very low level of risk and should be considered as broadly acceptable, the area between is the tolerability region and subject to context and other considerations.

The QTRA method definitively sets acceptable/unacceptable criteria (1:10,000) but tempers this by stating that risk owners might adopt the 1:10,000 limit of acceptable risk or choose to operate to a higher or lower level (Ellison 2005a). Matheny and Clark (1994) state that ratings have a relative meaning, but that the greater the hazard rating, the greater the risk associated with a tree. Many methods use qualitative terms to describe the risk, such as Low, Slight, and Critical. Table 3 summarises a range of terms from different methods. The Australian fatality risk level for trees is less than 1:5 million per annum (Hartley and Chalk 2019).

To defend due diligence, an authority needs to inspect trees, but at what frequency and at what standard? Most tree assessment methods suggest a frequency of between 0.5 to 3 years, with 9 of the 16 methods reviewed suggesting annual general inspections (Harris 1983; Grey and Deneke 1986; Robbins 1986; Rushforth 1987; Matheny and Clark 1994; Mattheck and Breloer 1994; Albers et al. 1996; Fraedrich 1999; Lonsdale 2000; Dockter 2001; Hayes 2001; Kane et al. 2001; Dunster 2003; Forbes-Laird 2003; Wildlife Tree Committee 2003; Ellison 2005b). Such annual inspections are not a full risk assessment, but may lead to such an assessment for a subset of selected

Table 3. Examples of terms used in tree risk assessment methods to describe levels of risk.

| Method | Acceptable risk category |
|--|---|
| Matheny and Clark (1994) | None; authors suggest use “Hazard Rating” for ranking only, but provide examples. |
| Threats (Forbes-Laird 2006) | Qualitative method that uses 7 broad categories from insignificant to extreme. Action is taken above the “slight” rating. |
| Bartlett (Smiley et al. 2002) | Qualitative method that is not specific, but uses terms such as Low, Moderate, High, and Critical Risk. |
| United States Department of Agriculture Forestry Service (Pokorny et al. 2003) | Several methods developed by the USDAFS are qualitative and don’t define acceptable risk levels. Most use a numerical score similar to Matheny and Clark (1994). |
| QTRA (Ellison 2005a; 2005b) | A quantitative probabilistic method that develops a probability-based output value. Probabilities < 1/10,000 <i>per annum</i> are unacceptable. |
| Colorado Tree Hazard Tree System (Colorado Tree Coalition 2004) | A 2-tiered qualitative system. Once the initial risk score exceeds 36/60, or certain categories reach a threshold, a second level inspection is triggered which prioritises. Fundamentally output values < 36 are deemed acceptable risk. |

trees. In inspecting large numbers of trees in public areas, time becomes a significant cost and resource limitation. In Australia, current street-tree inspections are being conducted at a rate of about 400 trees per day (50 trees per hour) with an average inspection time of 72 seconds, including travel time between trees. Longer inspection times increase costs significantly, and few public authorities have such resources. Frequency of inspections is typically based on the need to identify defects, but the frequency of use by a risk target is probably a better approach.

Many Victorian councils manage about 50,000 (range 15,500 to 106,000) street trees (Beer et al. 2001). At an inspection rate of 400 trees per day, an annual inspection would require 125 labour days per annum, with an inspection time of around 1 minute per tree. Several authors discuss the use of zoning districts, generally designated by usage levels, to prioritise areas for inspection (Lonsdale 2000, 2007; Pokorny et al. 2003; Ellison 2005a). It would also be possible to designate zones by tree factors, such as age, species, or size. Ellison (2005a) suggests that in low-use sites, the risk could be so low that inspections would be unnecessary. Zoning fits well with the variable fatality risk criteria that have higher requirements to provide a safe environment for particular land use types (Department of Planning NSW 1992) and with the “target” rating used in risk assessment methods such as Threats (Forbes-Laird 2006) and QTRA (Ellison 2005b). If zoning is to be used, it may change the weighting on any assessment system that uses target rating. For example, the Matheny and Clark (1994) method allows a target rating of 1 to 4 depending on

frequency of use. If zoning was used, it probably should pre-set the target value rating for the assessment or be removed from the assessment model.

The purpose of this research was to investigate a range of tree risk assessment methods using sensitivity analyses to determine the underlying model factors and their influence on the output value of risk. It was hypothesised that methods would differ in their treatment of likelihood and consequences, their outputs, the availability of scores in their output ranges, and their tendency to rate the level of risk. Sensitivity analyses provide an indication for each method of the relative influence that the input variables exert on the final risk value, which makes users of a method aware of its strengths and weaknesses. The analyses can also assist in the choice of method and in determining which method might be most appropriate to a particular situation. Fourteen of the sixteen methods of tree risk assessment analysed in this research existed before the study commenced, and two that conformed to the Australian Standard on Risk Assessment AS/NZS 4360:2004 were developed as part of the research (Standards Australia 2004a, 2004b). While a number of tree risk assessment methods exist, many are modifications of a few established methods.

MATERIALS AND METHODS

Twenty-six tree risk methods were gathered from sources such as the literature, methods created for in-house use by local governments, and from various arboricultural consultants in Australia. Sixteen methods were chosen for further analysis and comparison (Table 4) using the selection criteria that each had to

be sufficiently different from other methods to offer valid comparisons and be sufficiently well-documented for thorough analysis. One method (QTRA) was assessed twice because an infield assessment tool (the wheel) had the potential to provide significantly different results when compared to the full QTRA method. Two methods—Tree Risk Evaluation Quantitative (TRE QT) and Tree Risk Evaluation Qualitative (TRE QL)—that were consistent with the risk matrix methodology of the Australian Standard on Risk Assessment AS/NZS 4360:2004 (Standards Australia 2004a, 2004b) were developed for this research. The quantitative method was designed to provide a high level of risk quantification, and the qualitative method was designed for ease and speed of use in the field. Both allowed for ground truth testing of the 2 methods of sensitivity analyses and informed the interpretation of outcomes of the analyses for all of the methods.

The methods not selected for further analysis tended to be minor modifications of Matheny and Clark (1994) or were too incomplete to allow full analysis. Six methods originated from sources in the United States of America, three came from the United Kingdom, and seven originated from Australia (including the two methods developed for this research). Table 4 lists and summarises the methods used. Several methods are titled “Private” because they were created by consulting arborists who asked that the source not be disclosed. Hence, whilst the methodology is detailed, no reference is provided, but a copy of each method is available by request from the corresponding author of this study.

The 16 tree risk assessment methods were subject to 2 methods of sensitivity analysis with a goal of determining which assessment factors most influenced the output of each method. The sensitivity analyses are inherently linked to uncertainty and can identify control points and critical data and assist in validating a model (Frey and Patil 2002; Frey et al. 2004). Standards Australia (2004a) suggests that due to the imprecision of some risk estimates, sensitivity analysis should be used to determine the effect of uncertainty on assumptions and data and potentially to test the appropriateness and effectiveness of potential controls and risk treatment options.

The more sensitive an input parameter is, arguably the greater the precision required to accurately estimate the input's value. The basic process of sensitivity analysis is to change one variable at a time whilst holding the others constant and to measure the effect on

the output rating. While 2-way sensitivity analysis is relatively common, typically Monte Carlo simulation methods are used (Thompson 2002). Sensitivity analysis is widespread in disciplines such as engineering and risk assessment and management (Vose 1996; Haines 1998; Saltelli et al. 2000; Saltelli 2002; Patil and Frey 2004) but does not appear to have been widely used or discussed for tree risk assessment methods.

Excel was used to create a simple $\pm 25\%$ or ± 1 rank change (depending on risk method) to the mean (rounded to an integer with ordinal data methods) for each assessment criterion; the change to the output score was recorded as a percentage change. Palisade's @Risk software (Palisade 2002) was used to undertake a Monte Carlo (multivariate stepwise regression with Latin Hypercube sampling) simulation of 5000 iterations based on the input variables and output formula, assuming a uniform distribution for each input. The advantage of this approach was that the entirety of possible input ranges was utilised, and all possible outputs were analysed, which identified variations not readily found with simpler approaches. The regression data provided a quantifiable determination of the variation provided by each input element and produced a theoretical output distribution for the method. From the simulation, multivariate stepwise regression was undertaken to determine the influence of each method's input variables in determining the size of the output values. When the corresponding regression R^2 value was below 0.60, indicating a limited linear relationship, rank order correlation analysis was used to determine variance. While some valid probability distribution data possibly exist for the various tree risk assessment method inputs (defect frequency and severity), none were identified. Hence, the distribution frequency used for this analysis was based on a uniform probability of each possible input variable occurring.

Due to differences in input variables, mathematics, scaling, and other factors, it was not possible to directly compare methods, and so descriptive statistics were used that summarise the range of values created by each method. The data were based on ordinal inputs, no assumption of normality was made, and non-parametric approaches were used as applicable. To permit inter-method comparisons, three approaches were used: output data were transformed quantitatively via rank ordering, standardised scores (z-scores) were created, and a scale range from 1 to 10 was assigned qualitatively based on guidelines or thresholds provided by each method. A scale range from 1 to 10 was assigned

Table 4. Brief descriptions of the 16 tree risk assessment methods analysed (full explanations are available by request from the corresponding author).

| Name of method (Reference) (Country) (Coding) | Brief description and derivation of method |
|---|---|
| Bartlett Tree Experts (Smiley et al. 2002) (USA) (Bartlett) | Bartlett was a 2-assessment–category summation model (“Failure Potential/Defect Severity” and “Consequence of Failure”) using ordinal ranked values. The assessment criteria varied in range and number of values. The risk rating scores ranged from 2 to 15. Reasonably strong descriptors for each criterion were provided that relied on the assessor being suitably qualified and experienced. The final risk rating score was quantified with scores being aligned to 4 categories: < 7 = Low Risk, 7 to 9 = Moderate Risk, 10 to 12 = High Risk, and 13 to 15 = Critical Risk. Bartlett withdrew the method after Tree Risk Assessment Qualification (TRAQ) was endorsed by the International Society of Arboriculture (ISA). |
| Colorado Tree Coalition (Colorado Tree Coalition 2004) (USA) (CTC) | A qualitative method from the USA with 2 tiers of assessment and several base criteria that can trigger a second tier inspection. For this research, only the initial risk assessment was analysed. Three fields, each with different ranges (“Tree Species,” “Potential Target,” and “Defects Present”), were multiplied to create a risk score. The output scores ranged from 1 to 60. A unique element was the “Tree Species Index” category, where species were allocated to one of 5 groups ranging from Very Low (1) to Very High (5) Hazard. The authors suggested such groupings be developed from local consensus. The total rating (output score) was grouped into 3 predefined ratings: Low (1 to 14), Medium (15 to 35), and High (36 to 60). |
| Hume City Council (Hume City Council, nd) (Australia) (HCC) | A version of the Matheny and Clark (1994) method, highly modified by a Melbourne (Australia) local government authority and 2 consulting arborists. It was a complex model with 5 assessment criteria and a final risk score created by a combination of addition and multiplication. The assessment criteria were “Failure Potential,” “Failure Size,” “Target Usage,” “Target Value,” and “Damage Probability.” The first 4 were evenly scored (5 values from 0.5 to 4), whilst the “Damage Probability” criterion was scored in 5 values from 0.2 to 1. Risk output values ranged from 3 to 64. No definitions or quantification of the output score values were provided. |
| Matheny and Clark (Matheny and Clark 1994) (USA) (MandC) | This 3-category summation model is one of the older published urban tree risk methods. “Failure Potential,” “Size of Part,” and “Target Rating” were each scored from 1 to 4 and summed to derive a “Hazard Rating.” The method was superseded by the TRAQ training program endorsed by the ISA (Matheny and Clark 2007, 2009; Dunster et al. 2017). However, it is probably the most widely used method in Australasia and in North America by those not affiliated with the ISA who have not undertaken the fee-for-service TRAQ program. The output scores ranged from 3 to 12. No definition or quantification of the output score values was provided. |
| Kenyon 1993 (Kenyon 1993) (Australia) (Kenyon) | An early quantitative method that attempted to quantify risk rather than the more common approach of ordinal ranked values: at the time a novel tree risk assessment. It multiplied 4 fields, the “Probability of Failure,” “Target Value” (in dollars), “Target Risk Time,” and a “Damage Factor” to give a dollar value. Two of the fields with predefined ranges and outputs were expressed financially ranging from \$0 to more than AUD \$4.5 million. No definition or quantification of the output score values was provided. |
| Private 1 (no reference) (Australia) (Private 1) | A qualitative method developed by an arboricultural business in Melbourne, Australia. The method was provided confidentially and used 3 assessment criteria. The final risk score was created by (Likelihood of Failure × Likelihood of Impact)/2 × Consequences. The assessment criteria were discontinuous ordinal ranked numbers, and the Likelihood of Failure criterion had a different scale (1, 2, 4, 6, 8, and 10) to the others (1, 4, 6, 8, and 10). The final risk output values ranged from 1 to 500. The levels of risk were defined: 1 to 125 points = Very Low Risk, 125 to 250 points = Low Risk, 250 to 375 points = Medium Risk, 375 to 500 points = High Risk. |
| Private 2 (no reference) (Australia) (Private 2) | A modification of the Matheny and Clark method. The method was provided confidentially and had 6 assessment fields; each field was summed to create a “Hazard Rating” of 6 to 24. This method’s novelty was the inclusion of 3 categories: “Wind Alignment,” “Defect Height,” and “Remediation.” The “Hazard Rating” scores are quantified: 5 to 10 = Low Risk, 10 to 16 = Moderate Risk, 16 to 20 = High Risk, 20 to 24 = Extreme Risk. |

(Table 4 continued on next page)

Table 4. continued

| Name of method (Reference) (Country) (Coding) | Brief description and derivation of method |
|--|---|
| Private 3 (no reference) (Australia) (Private 3) | This method based on the Australian Risk Standard matrix (AS/NZS 4360:2004) was provided confidentially as a simple risk matrix composed of 5 values each for the “Likelihood” and “Consequence” categories. Descriptors provided guidelines for each category. The 2 scores were multiplied to produce a risk score. The risk score ranged from 1 to 25. The meaning of these scores was quantified: 1 to 6 = Low Risk, 7 to 10 = Medium Risk, 11 to 12 = Significant Risk, 15 to 20 = High Risk, > 21 = Immediate Risk. |
| QTRA – Full (Ellison 2005a, 2005b) (United Kingdom) (QTRA) | Quantified Tree Risk Assessment was a probabilistic method which multiplied 3 assessment criteria to derive an output risk score. The assessment criteria were “Probability of Failure,” “Size of Part” (Impact Potential), and “Target.” The assessor was required to quantify category data inputs, such as probability of the defect failing within the inspection period, estimated number of people passing the tree, number of vehicles per day, and size of part based on fractionising the estimated dry weight of a 600-mm-diameter tree using data derived from an allometric equation. With this method, the risk score was usually reported as the inverse of the decimal (a ratio) and termed Risk of Harm (RoH). The risk output values are expressed as a probability from 0 to 1, expressed as a ratio ranging from 1:1 to > 1:1000 billion (infinite). Acceptable risk was a RoH above 1:10,000. |
| QTRA – Wheel Version (Ellison 2005a, 2005b) (United Kingdom) (QTRA W) | A field tool for calculating QTRA risk levels, but values used in the various categories were not equal, and the RoH values differed from the full QTRA method. The changes effectively lowered the “Probability of Failure” input by one order of magnitude. Wheels were produced in 2005 and 2007, and this analysis was of the 2007 version, which had changes to the probability of failure range and to the vehicle use ranges within the “Target” category. |
| Threats (Forbes-Laird 2006) (United Kingdom) (Threats) | A qualitative United Kingdom method, Threats used 3 assessment categories (“Likelihood of Failure,” “Target Score,” and “Impact Score”) that were multiplied to create a risk score. This method could create a zero score. The “Hazard Ratings” ranged from 0 to 20,000, and are quantified: 0 to 49 = Insignificant, 50 to 159 = Minimal, 160 to 349 = Slight, 350 to 999 = Moderate, 1000 to 2000 = Significant, 2001 to 3999 = Serious, and 4000+ = Extreme. |
| Tree Risk Evaluation (TRE) Quantitative (Standards Australia 2004a, 2004b) (Australia) (TRE QT) | A probabilistically based quantitative method developed for this research aligning with the Australian Standard of Risk Management (AS/NZS 4360:2004). Three categories, “Probability of Failure” (Pf), “Probability of Impact” (Pi), and “Consequences” (Co), were assigned a probability (or ratio). “Consequences” was fractionised based on the maximum value of a statistical life developed from data used in Australia and indexed by consumer price index. These categories were multiplied to create a risk score expressed as a probability, ratio, or a financial value. At scores below 1:10,000 risk was unacceptable; for risk scores from 1:10,000 to 1:100,000 the risk should be reduced if practicable; risk scores > 1:100,000 were considered acceptable. |
| TRE Qualitative (Standards Australia 2004a, 2004b) (Australia) (TRE QL) | A method designed for quick field application, which was developed for this research, based on the same principles as TRE QT. A user interface was created with assessment categories of 5 to 7 values; the assessment values were summable by converting them to logarithmic scales. Risk scores from 3 to 300 were possible and could be converted to probabilities. At scores of 200 to 300 (> 1:10,000), risk was unacceptable; for risk scores 170 to 199 (1:10,000 to 1:100,000), risk should be reduced if practicable; scores below 170 (< 1:100,000) were considered acceptable. |
| US Department of Agriculture Forestry Service Method 1 (Pokorny et al. 2003) (USA) (USDAFS 1) | One of several qualitative methods developed by the USDAFS. This method summed the values from 4 assessment criteria: “Probability of Failure,” “Size of Part,” “Probability of Impact,” and “Other Risk Factors.” These categories had differing ranges, and the “Other Risk Factors” category could be used “if professional judgement suggested the need to increase the risk rating.” The output risk rating ranges from 3 to 12. No definition or quantification of the output score values was provided. |

Table 4. continued

| Name of method (Reference) (Country) (Coding) | Brief description and derivation of method |
|--|---|
| USDAFS Method 2 (Pokorny et al. 2003) (USA) (USDAFS 2) | The second USDAFS used 2 categories, “Targets” and “Defects” (each covers a different range), which were multiplied to create a “Hazard Rating.” This method can produce a zero score, and the “Hazard Rating” ranged from 0 to 6. No definition or quantification of the output score values was provided. |
| Tree Risk Assessment Qualification (TRAQ) (Dunster et al. 2017) (USA) (TRAQ) | This method endorsed by the ISA is wholly verbal or ordinal in the description of risk and has 3 assessment criteria. To allow analysis, scores were allocated for the “Likelihood” components (“Impact” and “Failure”), ranging from 1 to 4, which were multiplied. The product was then re-ranked, with scores of 2 to 6 being unlikely and re-ranked as 1, while scores > 12 were very likely and re-ranked as 4. These were then multiplied by the “Consequences” factor, again scored from 1 to 4, giving a Risk Rating (RR) range from 1 to 16. In both multiplications, it is impossible to obtain scores of 5, 7, 10, 11, 13, 14, or 15. The final risk rating scores were quantified and aligned to 4 categories: 1 to 4 = Low, 6 to 8 = Moderate, 9 to 12 = High, and > 12 = Extreme. |

to the different methods to allow comparison of their risk ratings (Table 5). If clear guidelines on the meaning of the output were not provided, the rating was based on the width of the method’s risk scaling. For example, a Matheny and Clark score of 4/12 was designated a risk scale rank of 3 or lower (because it was only 1 above the minimum possible score), while a score of 7/12 was designated a risk scale rank of 6 (or medium/moderate). For wholly verbal or ordinal methods such as TRAQ, points were assigned to the verbal descriptions provided to allow analysis. The results were directly compared using descriptive statistics and correlation analysis.

The “standardised score” is a method permitting comparisons amongst different methods’ outputs. The conversion of raw scores to units of standard deviation is often termed “z-scores” (Urdan 2005). Minitab™ (Minitab 2006) was used to generate the standardised scores using the default “subtract mean from raw value and divide by standard deviation” method:

$$\left(z = \frac{X - \mu}{\sigma}\right)$$

where X is the raw score, μ is the mean, and σ the standard deviation.

Positive values are those above the mean, and negative values those below the mean. The standardised score represents the number of standard deviations. Hence, a standardised score of -1 is 1 standard deviation

below the mean. Whilst all of the qualitative methods passed this test for normality ($p > 0.05$), the 4 probabilistic methods failed ($p < 0.05$) due to their inherent logarithmic nature. The raw output values from these methods were log-transformed and then passed the test for normality.

RESULTS

The results are presented in two parts. The first presents data on the general outcomes of the 2 sensitivity analyses for all methods, while the second part presents the data on the analysis of each method separately.

Comparison of Sensitivity Analyses for All Methods

The 1 rank or 25% change produced a range of outcomes that readily identified inputs that had a significant influence on outcomes. This approach can be extended by undertaking a 2 rank or 50% change to each input. A limitation occurs with methods that do not use an even range across the input scales (e.g., 1, 2, 3, and 4 versus 1, 3, 7, and 10). The second sensitivity approach (@Risk multivariate stepwise regression) more successfully identifies the range of variations found with methods. Matched input scales, such as MandC or probabilistic methods using true probabilities such as QTRA, rather than scaled probabilities, such as Kenyon and QTRA W, will produce even changes.

Table 5. Qualitative risk score scaling assigned to tree risk assessment methods to allow comparison of the different outputs and risk ratings for use in Minitab™.

| Risk scale rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------|---------------|----------|----------|--------------|----------|-----------------|----------|---------------|------|---------|
| Risk descriptor | Insignificant | Very low | Low | Minor/slight | Median | Medium/moderate | Moderate | Somewhat high | High | Extreme |
| Bartlett | 2 | 3 | 5 | 6 | - | 7 | 9 | 10 | 13 | 15 |
| CTC | 1 | - | - | 14 | 15 | - | 35 | 36 | 60 | N/A |
| HCC | 3 | - | - | - | - | - | - | - | 64 | N/A |
| MandC | - | - | 4 | - | - | 7 | - | - | 12 | N/A |
| Kenyon | < \$100 | - | - | - | \$10,000 | - | - | - | - | > \$1m |
| Private 1 | 1 | - | 125 | 250 | - | 375 | - | - | 375 | 500 |
| Private 2 | - | 5 | 10 | - | 10 | - | 16 | - | 20 | 24 |
| Private 3 | - | 1 | 6 | - | 7 | - | 10 | 12 | 15 | 21 |
| QTRA | > 1:1m | - | - | < 1:10k | - | - | - | - | - | N/A |
| QTRA W | > 1:1m | - | - | < 1:10k | - | - | - | - | - | N/A |
| Threats | > 50 | 159 | 160 | 349 | 350 | - | 999 | 1000 | 2000 | 4000 |
| TRE QT | - | - | > 1:100k | - | - | - | - | < 1:10k | - | N/A |
| TRE QL | - | - | > 1:100k | - | - | - | - | < 1:10k | - | N/A |
| USDAFS 1 | - | - | 3 | - | - | 7 | - | - | 12 | N/A |
| USDAFS 2 | 0 | 1 | 2 | 3 | - | 4 | - | - | 6 | N/A |
| TRAQ | 1 | 2 | 3 | 4 | 6 | 8 | - | 9 | 12 | 16 |

Note. For some methods, the highest risk rating was extreme and for others high. For clarity and simplicity, this table only shows method scores that directly match the 10 point scale and/or the verbal descriptor of risk.

Consequently, individual input analysis for methods such as MandC, Private 2, QTRA, and TRE QT identified that each input produced the same influence on the output value. Probabilistic methods recorded a percentage change that reflected the input change, where a 25% change to an input modified the output value by 25%. For qualitative methods using ordinal ranks such as MandC, Private 1, and Private 2, the percentage change to the output values varied significantly. For MandC, a 1 rank change from the mean modified the output value by 17%; for Private 2, the change was 8%. Private 3 recorded a 33% change to the output value, and for TRAQ, a 1 rank change to impact or failure moved the risk rating 1 step, but a change to consequences did not alter the risk rating.

Of the methods reviewed, 8 were designed such that each input category exhibited a different influence on the their respective output value, and due to category scaling differences, a 1 rank change in different directions could create a large change in the output values. For example, QTRA W, due to its logarithmic nature, exhibited extreme intra-category variation (a + 1 rank change increased the output by 900%, whilst a 1 rank decrease reduced the output value by 81%). Methods such as USDAFS 2, Threats,

and Kenyon exhibit very large movements in both inter- and intra-category values for simple 1 rank changes (Table 6).

The 1 rank or 25% change sensitivity analysis clearly and simply demonstrated that different risk assessment methods approached the input weighting and scaling differently and combined them mathematically differently. Hence, a change to 1 input value could create a wide range of changes to the output values and hence risk rating. Only the QTRA and the 2 TRE methods provided an explanation for the range and scaling of inputs, weighting of categories, or combination mathematics, which makes working with other methods more difficult and less certain.

The Monte Carlo simulation for the 16 methods aligned with the simpler rank-change analysis (Table 6). Simple linear models that sum the inputs, such as MandC, TRE QL, USDAFS 1, and Private 2, produced distributions that approximated normal distribution curves, with the majority of outputs towards the centre of the distribution. The 2 QTRA approaches and TRE QT are mathematically probabilistic and hence produce logarithmic distributions. These methods produced output scores over a wide and sometimes extreme range, where, for example, the lowest value

Table 6. Summarised results of sensitivity analysis and combined consequence (Co) and likelihood (Li) weightings and ratios for 16 tree assessment methods.

| Method | Change to output of 1 rank or 25% change | Change to output of Monte Carlo simulation stepwise regression | Fitted distribution curve and projected mean | Combined consequence and likelihood weighting and ratio |
|------------------|--|---|---|---|
| Bartlett | 43% change to Failure/Defect 29% change to Consequence | Failure/Defect 80% variance 19% Consequence | Even spread of scores. Projected mean 8.5. Some output values cannot be produced. | Co = 19% Li = 80% Co:Li = 0.24 |
| CTC | 13% each category Changed with 2 rank variation | Species 38% variance Defects 33% Target 26% | Strong tendency to generate lower end scores. Projected mean 15.1. | Co = 23% Li = 58% Co:Li = 0.40 |
| HCC | 25% change with 4 categories Damage 33% | Damage 31% variance All others 13% | Tends to produce lower end score. Projected mean 10.5. | Co = 57% Li = 26% Co:Li = 2.19 |
| MandC | 17% each category | 33% variance all categories (linear model) | Generated scores towards the centre. Projected mean 7.5. | Co = 33% Li = 66% Co:Li = 0.5 |
| Kenyon | Target Time and Target Value 25% Damage -60% + 60% Failure -57% + 14% | Damage 46% of the variance Failure 13% Other categories 10% | Tendency to generate lower end scores—logarithmic distribution. Projected mean \$276,000. | Co = 56% Li = 21% Co:Li = 2.66 |
| Private 1 | 33% each category except Failure where a + 1 rank change results in a 67% increase | Failure 30% variance Remaining fields for 23% of the variance | Strong tendency to generate lower end scores. Projected mean 87. | Co = 28% Li = 59% Co:Li = 0.47 |
| Private 2 | 8% each category | 16% all categories | Generates scores towards the centre. Projected mean 15. | Co = 34% Li = 49% Co:Li = 0.69 |
| Private 3 | 33% each category | 45% each category | Generates scores towards lower to median. Projected mean 9. | Co = 0% Li = 92% Co:Li = 0.00 |
| QTRA | 25% each category | 25% each category | Generates lower end scores—logarithmic distribution. Projected mean 0.125 equal to a 1:8 ratio. | Co = 25% Li = 49% Co:Li = 0.51 |
| QTRA W | Size -81% + 900% Failure -77% + 852% Target -72% + 900% | Target 45% Size 21% Failure 31% | See QTRA W section for discussion. | Co = 4% Li = 58% Co:Li = 0.07 |
| Threats | Impact -33% + 67% Target \pm 25% Failure -75% + 525% | Failure 51% Target 20% Impact 3% | Generates scores towards the lower end of the range. Projected mean 1100. | Co = 3% Li = 72% Co:Li = 0.04 |
| TRE QL | 17% each category | 32% each category | Generates scores towards the centre. Projected mean 150. | Co = 32% Li = 64% Co:Li = 0.5 |

(Table 6 continued on next page)

Table 6. continued

| Method | Change to output of 1 rank or 25% change | Change to output of Monte Carlo simulation stepwise regression | Fitted distribution curve and projected mean | Combined consequence and likelihood weighting and ratio |
|-----------------|--|--|--|---|
| TRE QT | 25% each category | 25% each category | Generates lower end scores—logarithmic distribution. Projected mean 0.125 equal to a 1:8 ratio. | Co = 25% Li = 49% Co:Li = 0.51 |
| USDAFS 1 | 14% each category | Failure 39% 20% all others | Generates centre range scores skewed slightly towards high. Projected mean 7.5. | Co = 21% Li = 59% Co:Li = 0.36 |
| USDAFS 2 | Defect \pm 50% Target -0% + 100% | Defect 77% of variation Target 15% | Tends to generate risk score outputs of 0 and 2. Projected mean 2.6. | Co = 0% Li = 92% Co:Li = 0.00 |
| TRAQ | One rank change to failure or impact moved risk rating one step. One rank increase to Co did not change RR from Moderate. One rank decrease to Co changed RR from Moderate to Low. | Failure or Impact increase 80% variance, decrease 60% variance Consequence increase 80% variance, decrease 24% variance | Strong tendency to generate lower end scores. Projected mean 4.06. Some output values cannot be produced. | Co = 40% Li = 60% Co:Li = 0.66 |

created on the QTRA W is 1:3,000 billion (3×10^{-12}). Hence a distribution chart would be difficult to interpret. The distribution charts for other methods illustrated a wide range of results: Kenyon and Threats appeared logarithmic, while TRAQ, HCC, and Private 1 appeared to generate the majority of their output scores towards the lower end of their respective potential ranges, as would be expected from methods that multiply inputs. The CTC method generated a large number of output ratings towards the lower end of its scaling and had a significant number of gaps over its range.

A consequence:likelihood ratio (Co:Li) was created by dividing the respective weightings of consequence and likelihood (Table 6), with values greater than 0 representing methods that favour consequence, and values less than 0 representing methods that favour likelihood; the closer to 1 the ratio, the more balanced the method. Some methods had a Co:Li = 0.00; only consequence or likelihood was measured by those methods (Figure 3; Table 6).

The correlations (Figure 4) between the ranked values and standardised scores were very high (mean 0.89). Only the Threats method produced a significantly

lower correlation due to the zero-risk values that influenced the standardised values.

The outputs from MandC did not exhibit strong correlations with methods that other authors claimed were derived from its methodology, such as HCC (0.21) and Private 2 (0.43). The 3 probabilistic methods, Kenyon, QTRA, and TRE, have strong relationships with each other; given that the QTRA and the QTRA W methods and the TRE QT and TRE QL methods are respectively direct analogues, this high correlation was to be expected.

An Abridged Summary of Sensitivity Analysis of Each Method

The full data set for the individual analysis of all methods is very lengthy, and for each method, the summary included figures for a sensitivity 1 rank (sometimes also a 2) change, regression sensitivity, combined consequences and likelihood regression sensitivity, and @Risk Monte Carlo distribution. There are nearly 60 figures in total, and so only 2 for each method are included here, but all of them with relevant text are available by request from the corresponding author of this study.

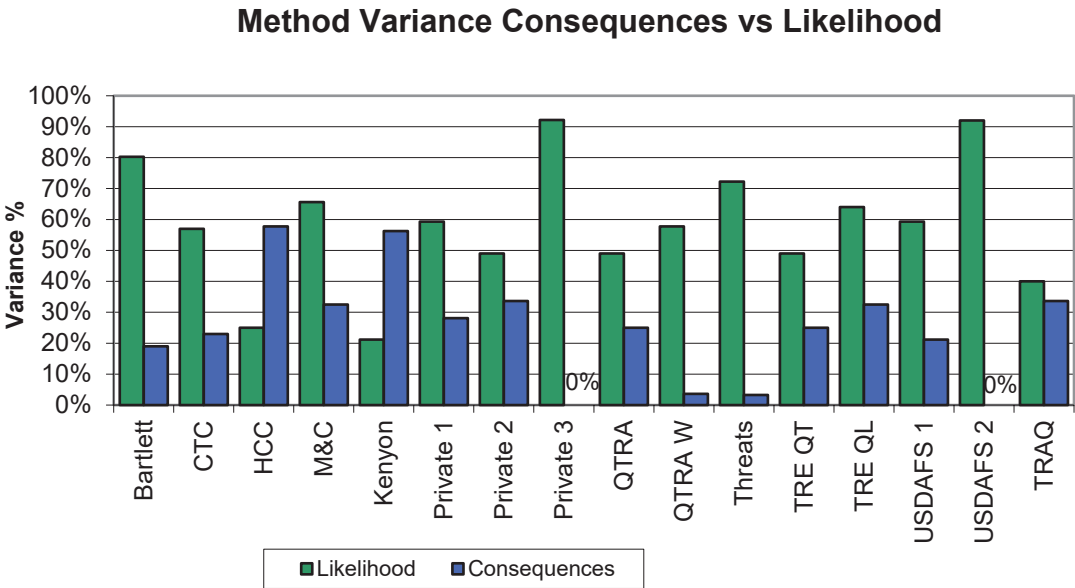


Figure 3. Comparison of weighting given by each method to likelihood (Li) and consequence (Co) elements of the risk assessment.

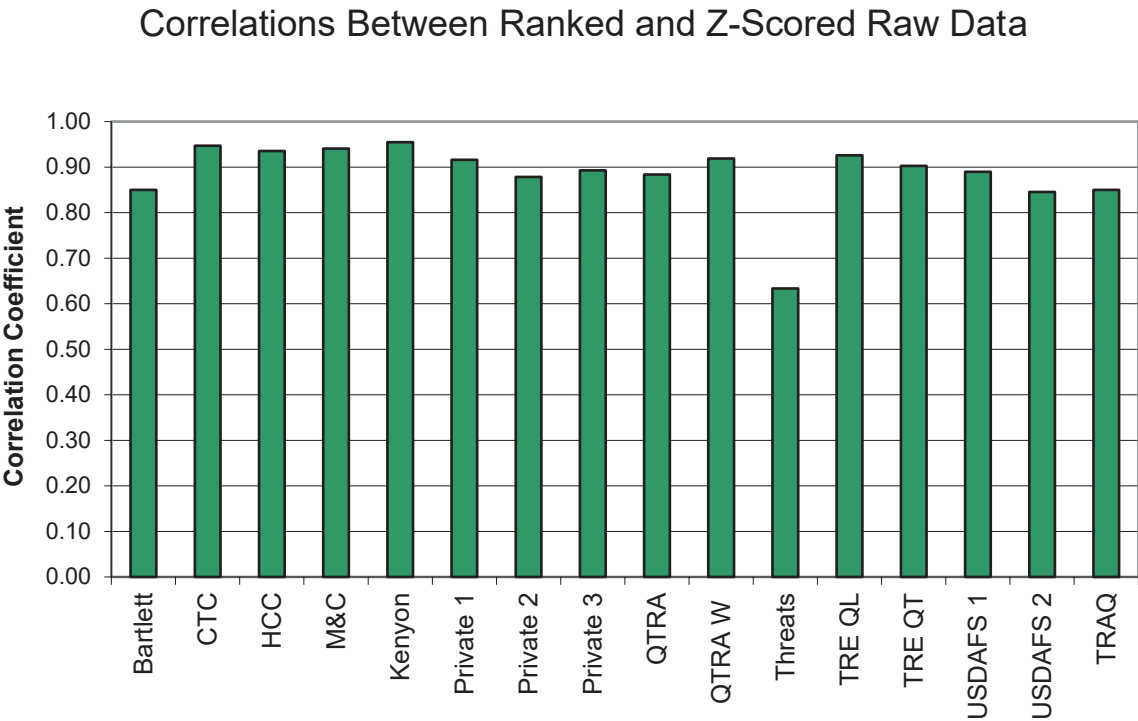


Figure 4. Correlations between ranked and standardised scored raw data for the 16 methods of tree assessment.

Bartlett Tree Experts (Smiley et al. 2002)

The assessment criteria varied in range and number of values, and risk rating scores ranged from 2 to 15. As a 2-assessment–category summation model (“Failure Potential/Defect Severity” and “Consequence of Failure”) using ordinal ranked values and uneven input scaling, a simple 1 rank change (\pm) to the “Failure Potential/Defect Severity” input resulted in a 43% change to the output value, whilst the same change to the “Consequence of Failure” inputs resulted in a 29% change to the output value.

@Risk’s stepwise regression also identified this trait (Figure 5), with the “Failure Potential/Defect Severity” input accounting for some 80% (0.896) of the variation in the outputs, whilst the “Consequence of Failure” input only accounted for 19% (0.436). The method used 2 inputs that fit within the $\text{Risk} = \text{Li} \times \text{Co}$ definition. The likelihood elements of an assessment rated more strongly than consequences elements based on regression analysis ($R^2 = 0.99$), as consequence accounted for 19% and likelihood 80% of the variation. As illustrated in the distribution chart (Figure 6), the Bartlett method produces a uniform distribution, except for 2 values (3 and 14) that cannot be

produced. Bartlett produces 12 output values from its range of 2 to 15.

Colorado Tree Coalition (CTC)(2004)

Three inputs, “Tree Species,” “Potential Target,” and “Defects Present,” were multiplied to create a risk output score. Each of the 3 assessment categories had different ranges (scaling). The “Total Rating” (output score) was grouped into 3 predefined ratings: Low (1 to 14), Medium (15 to 35), and High (36 to 60).

The 1 rank change confirmed that each input variable had an equal effect on the output (a 13% change), but a 2 rank change showed this was not constant throughout the input ranges. The regression sensitivity (Figure 7) reflected this variation within the model created by each element at “Tree Species” (38%), “Defects Present” (33%), and “Potential Target” (26%). Combining the various inputs into 2 categories (consequences and likelihood) resulted in CTC significantly favouring the inputs that affect likelihood over consequences, with consequences accounting for 23% (0.48) of the variation in the method and likelihood 58% (0.76). The Monte Carlo distribution (5000 iterations) suggested that this method would produce far more values at the lower end of the risk scale

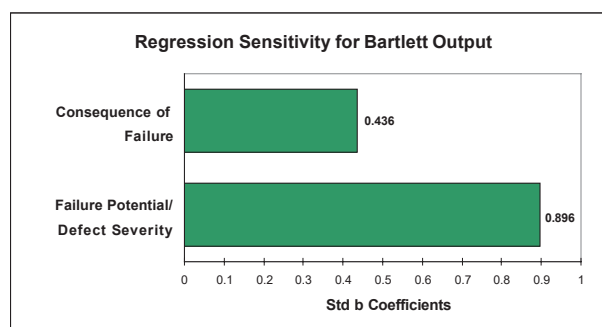


Figure 5. Bartlett regression sensitivity.

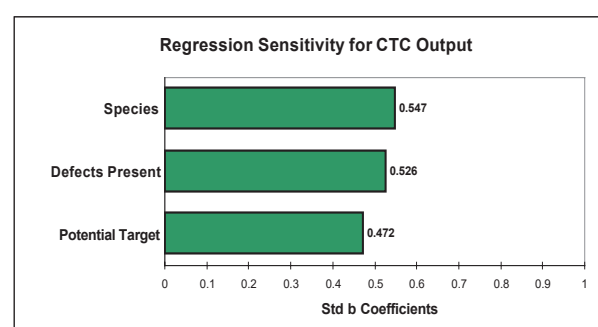


Figure 7. CTC regression sensitivity.

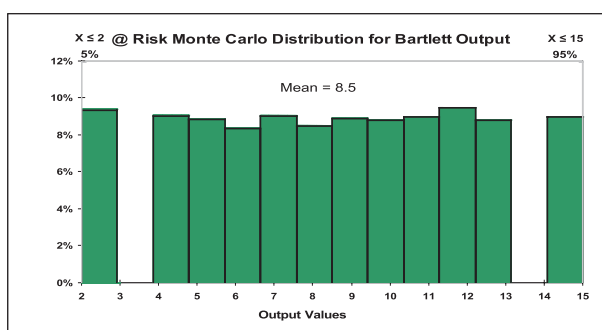


Figure 6. Bartlett @Risk Monte Carlo distribution.

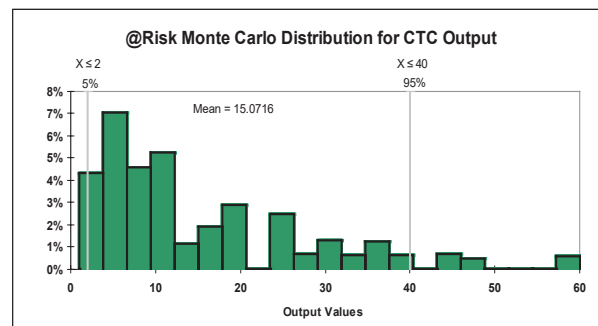


Figure 8. CTC @Risk Monte Carlo distribution.

(Figure 8), with a mean value of 15 (lowest end of the method's definition of "medium risk"). The CTC approach is more complex than modelled in the analyses, as it requires the species to have predefined ratings set, and if either the "Potential Target" or "Defects Present" rating has a rating of "3" or the overall output is ≥ 36 , then the second level of the assessment is triggered. This analysis was limited by these additional factors.

Hume City Council (HCC)(nd)

This method was a modified version of Matheny and Clark (1994) that had 5 assessment criteria and a final risk score created by a combination of addition and multiplication. The assessment criteria were "Failure Potential," "Failure Size," "Target Usage," "Target Value," and "Damage Probability." The first 4 were evenly scored (5 values from 0.5 to 4), whilst the "Damage Probability" criterion had 5 values from 0.2 to 1. No definitions or quantification of the resultant output ratings were provided by the Hume Method.

Four of the five inputs were scaled and mathematically combined identically, as illustrated by both the 1 rank change and the regression sensitivity (Figure 9). Any change to these inputs had the same influence

on the output. The "Damage Probability" input is a multiplier applied at the end of the method and had the strongest influence on the output values. The regression sensitivity identified that each of the 4 same scale inputs accounted for 13% of the output variability, whilst the "Damage Probability" element accounted for some 31% of the output.

Combining the various inputs into 2 categories (consequences and likelihood) resulted in HCC significantly favouring the inputs that affected consequences over likelihood, with consequences accounting for some 57% (0.76) of the variation in the method and likelihood accounting for 26% (0.51). This consequence weighting was different to most other methods. The Monte Carlo distribution (5000 iterations) suggested that this method would typically produce far more values at the lower end of the risk scale, with a mean value of 10.5 (Figure 10).

Matheny and Clark (MandC)(1994)

A photographic guide to the evaluation of hazard trees in urban areas (Matheny and Clark 1994), MandC is a 3-category summation method. "Failure Potential," "Size of Part," and "Target Rating" were each scored from 1 to 4 and summed to derive a "Hazard Rating." Output scores range from 3 to 12, with no definitions or quantifications of the output values provided. MandC stated the output as a ranking and provided only a relative meaning.

All of the inputs were scaled and mathematically combined the same. As illustrated by both the 1 rank change and the regression sensitivity (Figure 11), outputs changed by the same amount for each changed input. In the regression sensitivity, each of the 3 same scaled inputs accounted for 33% of the output variability. Combining the various inputs into 2 categories (consequences and likelihood) resulted in MandC significantly favouring the inputs that affect likelihood over consequences. Consequence factors accounted for some 33% (0.57) of the variation in the method, and likelihood accounted for 66% (0.81). Given that "Size of Part" is the only factor that accounts for consequences, this outcome was logical. The Monte Carlo distribution (5000 iterations) suggested that this method will typically produce a "normal" distribution with most values towards the centre (Figure 12).

Kenyon (1993)

Kenyon was a probabilistic method that multiplied 4 fields, the "Probability of Failure," "Target Value" (in dollars), "Target Risk Time," and a "Damage Factor,"

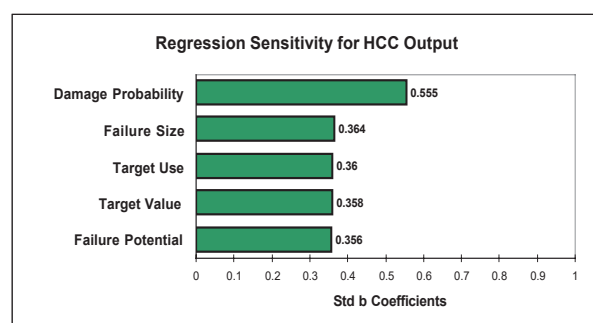


Figure 9. HCC regression sensitivity.

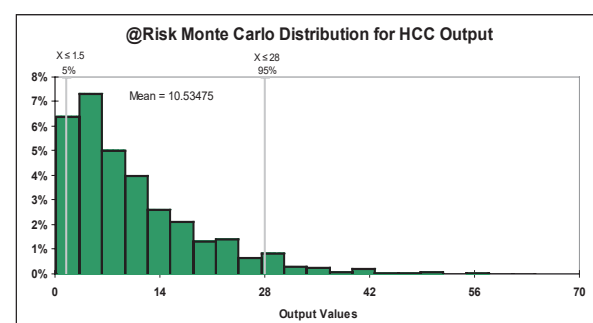


Figure 10. HCC @Risk Monte Carlo distribution.

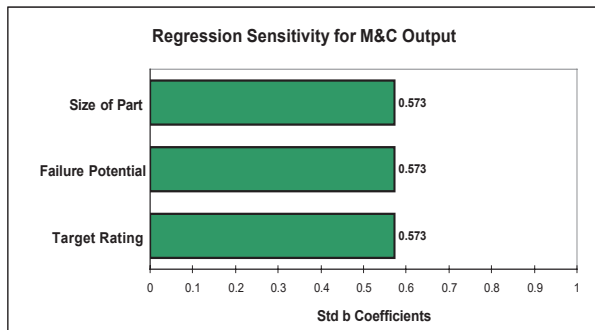


Figure 11. MandC @Risk regression sensitivity.

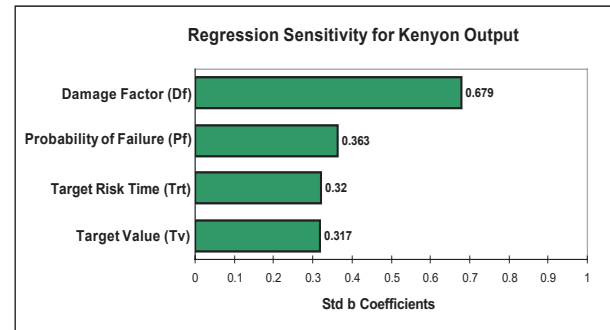


Figure 13. Kenyon @Risk regression sensitivity.

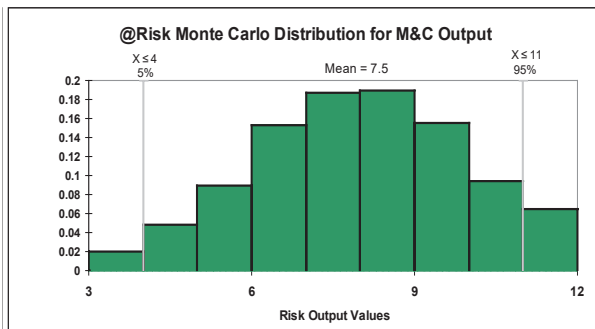


Figure 12. MandC @Risk Monte Carlo distribution.

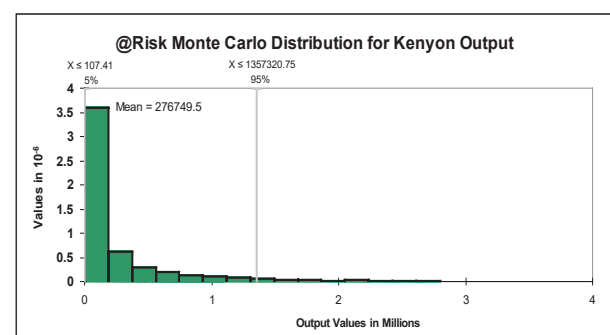


Figure 14. Kenyon @Risk Monte Carlo distribution.

to give risk quantification as financial values. No definitions or quantification of the output score values were provided. Due to the mixed scaling and probabilistic nature of this method, both the ± 1 rank and $\pm 25\%$ were used. This method was more complex than most: the “Probability of Failure” and “Damage Factor” inputs were based on scaled ranges from 0.05 to 1 and 0.001 to 1, respectively; “Target Risk Time” was based on the number of minutes of exposure per week, and hence was scaled from 0.0059 to 1; whilst “Target Value” was the value of damage from \$0 to \$4.5 million (AUD). Therefore, the 1 rank change and the correlation sensitivity (Figure 13) showed that different inputs had differing influences on the output.

The $\pm 25\%$ or 1 rank changes produced predictable outcomes for the “Target Risk Time” and “Target Value” elements, because these were probabilities. The “Damage Factor” and “Probability of Failure” categories were scaled; a 1 rank change to the “Damage Factor” resulted in a $\pm 60\%$ change to the output value, whilst the “Probability of Failure” element produced an uneven result due to the very uneven scaling of the input. Because the regression R^2 was below 0.6 (0.55), correlation coefficients were used

to express the Monte Carlo modelling. Similar to the ± 1 rank, changing the “Damage Factor” had a very significant influence on this method’s output values (46%). Combining the various inputs into consequences and likelihood resulted in Kenyon significantly favouring the inputs that affect consequences over likelihood, with consequences accounting for some 56% (0.75) of the variation in the method and likelihood 21% (0.46). The Monte Carlo distribution (5000 iterations) illustrated the logarithmic nature of any probabilistic approach (Figure 14).

Private 1

Private 1 used 3 assessment criteria and the risk score was created by (Likelihood of Failure \times Likelihood of Impact)/2 \times Consequences. The assessment criteria were ordinal ranked numbers, but not continuous, and the Likelihood of Failure criterion was a different scale (1, 2, 4, 6, 8, and 10) to the other two (1, 4, 6, 8, 10). The final risk score could range from 1 to 500. The levels of risk these scores represent were defined: 1 to 125 points = Very Low Risk Tree; 125 to 250 points = Low Risk Tree; 250 to 375 points = Medium Risk Tree; 375 to 500 points = High Risk Tree. A 1 rank change illustrated the difference that the different

scaling in the Likelihood of Failure category created. A single change to any input created significant changes to the output and a large change in the Likelihood of Failure category. The effect of the failure criterion was more objectively seen in the Monte Carlo regression ($R^2 = 0.74$), where the Likelihood of Failure category accounted for 29.7% of the method's variation, whilst the remaining categories accounted for some 22% each (Figure 15).

Combining the various inputs into 2 categories (consequences and likelihood) resulted in Private 1 significantly favouring the inputs that affect likelihood over consequences, with consequences accounting for some 28% (0.53) of the variation in the method and likelihood 59% (0.77). The Monte Carlo distribution illustrated that this method would produce the majority of its outputs at the lower end of the scale (mean 86.7), with 88% below the 250 level termed Low Risk Tree and 72% below 125 defined as a Very Low Risk Tree (Figure 16).

Private 2

Private 2 was a significantly modified version of the MandC method that had 6 assessment inputs, each summed to create a "Hazard Rating" range from 6 to 24.

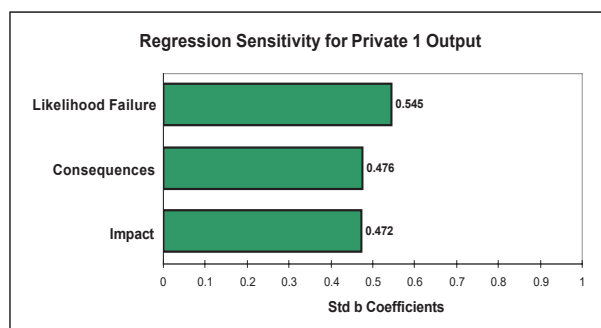


Figure 15. Private 1 @Risk regression sensitivity.

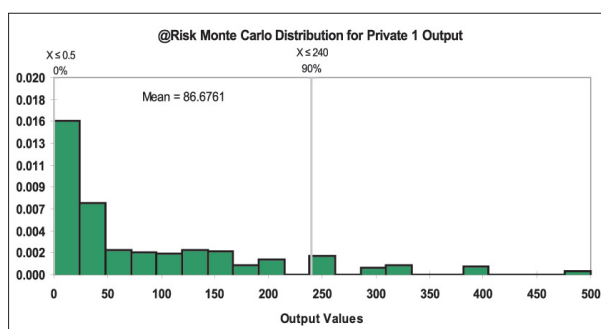


Figure 16. Private 1 @Risk Monte Carlo distribution.

The "Hazard Rating" scores were quantified: 6 to 10 = Low Risk, 10 to 16 = Moderate, 16 to 20 = High, 20 to 24 = Extreme Risk. This method used the same approach as MandC, with 3 additional input categories, and had the same sensitivities. Because each category was scaled the same, each provided the same weighting to the output. The difference to MandC is that the additional 3 categories effectively halved the influence any single input had on the method. A 1 rank change produced only an 8% change to the subsequent output value. The regression sensitivity ($R^2 = 1.0$) showed that each input affected the variability in the model by 16% (Figure 17). This represented 8% in each direction.

Combining the various inputs into 2 categories (consequences and likelihood) was not possible with this method, because 1 assessment input did not align with these categories ("Remediation"), but the method favoured likelihood (0.70, 49% of the variation) over consequences (0.58, 34% of the variation). The Monte Carlo distribution illustrated that this method would tend to produce the vast majority of its outputs in the form of a normal distribution curve (Figure 18) with a mean of 15 (method range 4 to 24).

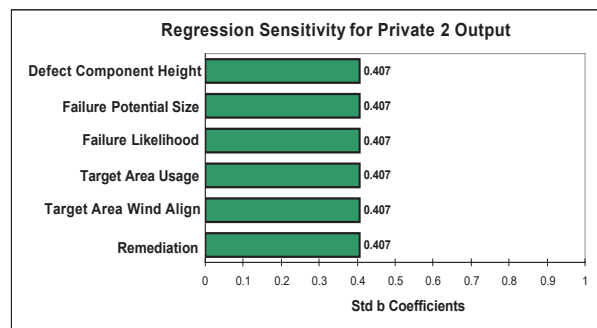


Figure 17. Private 2 @Risk regression sensitivity.

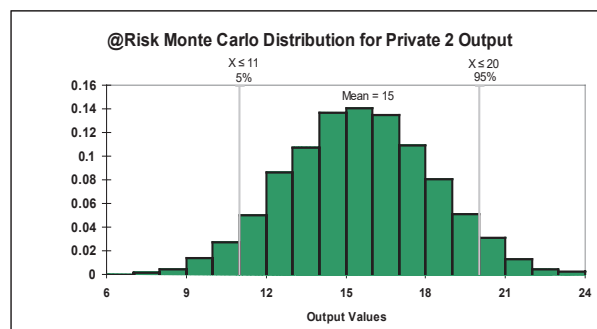


Figure 18. Private 2 @Risk Monte Carlo distribution.

Private 3

Private 3 was a method based on the Australian Risk Standard risk matrix (AS/NZS 4360:2004) composed of 5 values in each of the likelihood and consequence categories. A range of descriptors provided guidelines for each category. The 2 scores were multiplied to derive a risk score. The risk score could range from 1 to 25. The meaning of these scores was quantified: 1 to 6 = Low Risk, 7 to 10 = Medium Risk, 11 to 14 = Significant Risk, 15 to 20 = High Risk, > 21 = Immediate Risk. Given the 5×5 matrix and equal scaling of the input values, this method exhibited identical changes for each input, shown by both the 1 rank change method and the Monte Carlo regression analysis. A 1 rank change to any of the risk inputs created a 33% change to the output. The Monte Carlo regression ($R^2 = 0.90$) identified that each category accounted for 45% of the method's variation (Figure 19).

This method used 2 input categories termed "Likelihood" and "Consequence." However, this method's matrix inputs only considered the likelihood of a failure (the "Likelihood" input) and the likelihood of it impacting a risk target (termed "Consequence" by the authors). It did not attempt to quantify the degree of

harm or potential damage. In effect, this method did not measure consequence. The Monte Carlo distribution illustrated that the method tended to produce most values towards the lower end of the scale because the method cannot mathematically produce 9 of the integers between 10 and 25 (Figure 20).

QTRA (Ellison 2005a, 2005b)

QTRA was a probabilistic method, which multiplied inputs from the 3 assessment categories to derive an output risk score. The input categories were "Probability of Failure," "Size of Part" (Impact Potential), and "Target." With this method, the risk score was usually reported as the inverse of the decimal (a ratio) termed Risk of Harm (RoH). The range of output ratings was infinite (depending on the inputs), and a range of 0 to > 1 billion was feasible. QTRA stated acceptable risk was below a RoH of 1:10,000 (e.g., 1:20,000). Being a probabilistic method with all inputs ranging between 0 and 1, both the $\pm 25\%$ and regression sensitivity confirmed that each input influenced the output equally (Figure 21).

Combining the various inputs into 2 categories (consequences and likelihood) was interesting with this method, because the input categories changed

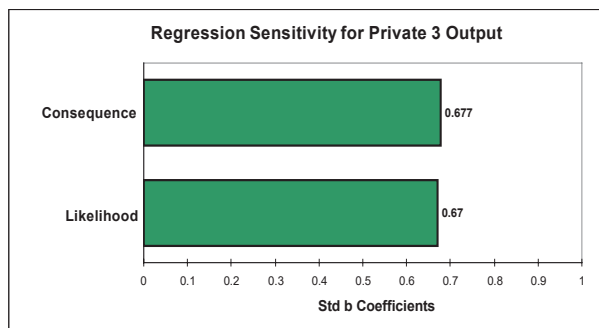


Figure 19. Private 3 @Risk regression sensitivity.

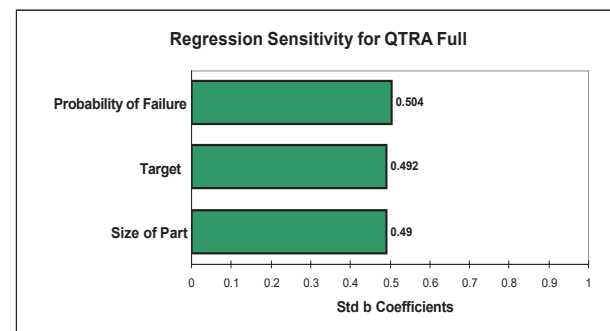


Figure 21. QTRA @Risk regression sensitivity.

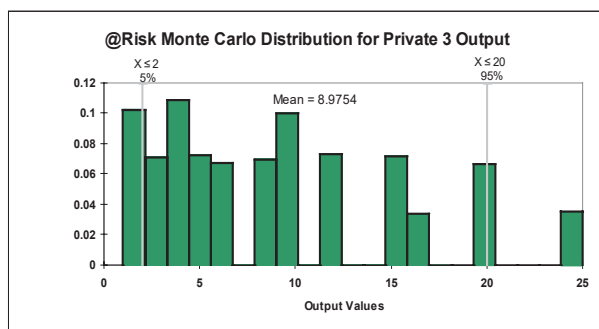


Figure 20. Private 3 @Risk Monte Carlo distribution.

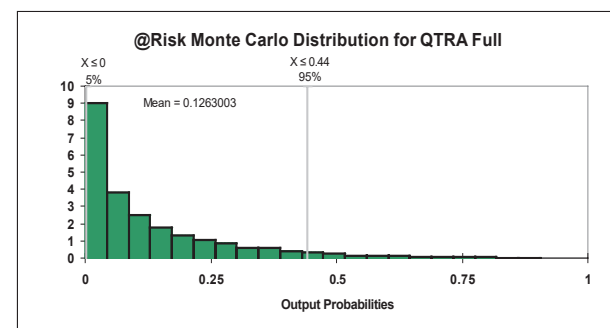


Figure 22. QTRA @Risk Monte Carlo distribution.

depending on the risk target. For people and vehicles, the method more strongly weighted likelihood (49%) over consequences (25%), but when assessing structures it reversed the weightings. The distribution curve generated from the Monte Carlo simulation showed the logarithmic curve expected from the mathematics (Figure 22); the mean is 0.126, which equated to a risk value of 1:8, which was very high given the level of acceptable risk is deemed to be 1:10,000.

QTRA W Version (Ellison 2005a, 2005b)

QTRA W was a simple tool for calculating quick in-field QTRA risk levels that could generate different RoH values to the full QTRA assessment method. The most striking variation was that an increase of 1 rank changed the output by 852% to 900%, whereas a reduction in 1 rank only changed the output by 72% to 81%. The QTRA W produced a low R^2 value when regression was applied ($R^2 = 0.14$), hence correlation coefficients were used; as shown in Figure 23, the various inputs differed in their influence on the output values, with “Target” influencing 45% of variation, “Probability of Failure” influencing 31%, and “Size of Part” influencing 21%.

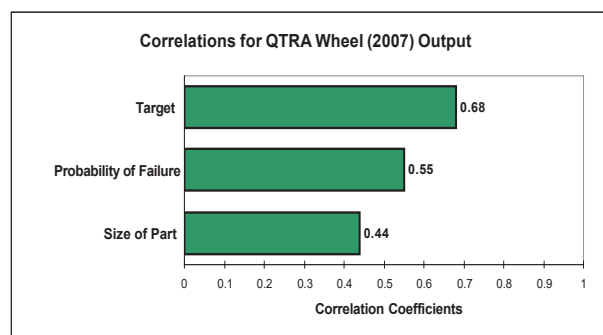


Figure 23. QTRA W @Risk ranked correlation sensitivity.

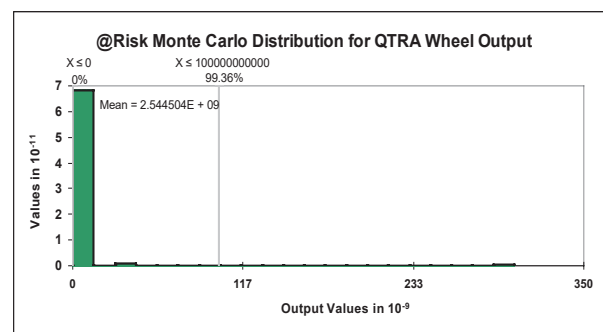


Figure 24. QTRA W @Risk Monte Carlo distribution.

As with the full QTRA version, the likelihood and consequences categories changed depending on the risk target. For people and vehicles, the method more strongly weighted likelihood (49%) over consequences (25%); whereas when assessing for structures, the method reversed the weightings. The distribution curve generated from the Monte Carlo simulation showed an extreme logarithmic curve (Figure 24); with output values based in the 10^{-9} scale (billions). The Monte Carlo mean value was 2.5×10^{-9} . This extreme value occurred because the wheel generated values from 1:1 to 3×10^{-12} using two 5-scaled ranges and one 6-scaled range; and these were expressed over a 150-point range. Due to scaling and limited categories, the QTRA W produced different outputs to the full version of QTRA, and users should be aware of the differences between the two approaches.

Threats (Forbes-Laird 2006)

Threats was a qualitative method with 3 assessment categories, “Likelihood of Failure,” “Target Score,” and “Impact Score.” Threats used differing ranks and scale ranges for each input, hence the significantly differing effects of the various inputs. The 1 rank change showed the large variation within each input category and between the 3 inputs, with a 1 rank increase in the “Likelihood of Failure” category modifying the output value by 525%. The @Risk Monte Carlo simulation and subsequent regression produced an R^2 of 0.53, hence correlations coefficients are reported. Figure 25 confirmed that the “Likelihood of Failure” input category had the greatest influence on the outputs but was the input with the greatest uncertainty. The “Likelihood of Failure” category accounted for some 51% of the variation, whilst “Target Score” and “Impact Score” accounted for 20% and 3%, respectively.

Combining the various inputs into consequences and likelihood indicated a significant favouring of inputs that affect likelihood over consequences, due to the strong influence of the “Likelihood of Failure” category, with likelihood accounting for some 72% of the variation in the method and consequences a mere 3%. The Monte Carlo distribution showed that the method produced many more values at the lower end of the risk scale (Figure 26). However, whilst the Threats risk output scale ranges from 0 to 20,000, the maximum risk level was reached at an output of 4000. Based on this simulation, Threats would produce

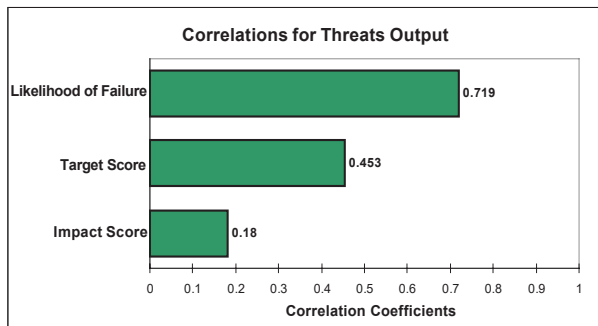


Figure 25. Threats @Risk correlation sensitivity.

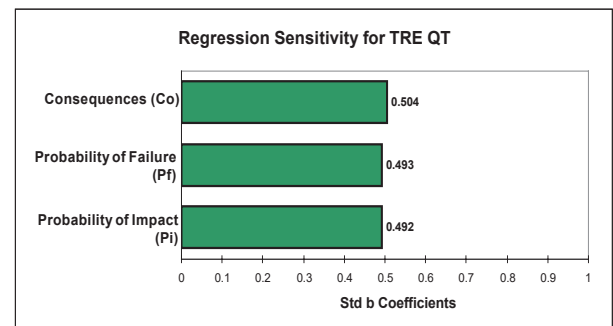


Figure 27. TRE QT @Risk regression sensitivity.

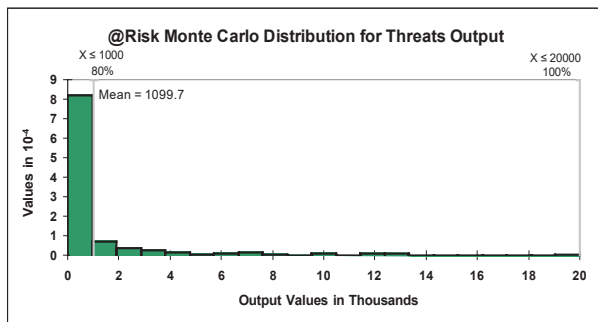


Figure 26. Threats @Risk Monte Carlo distribution.

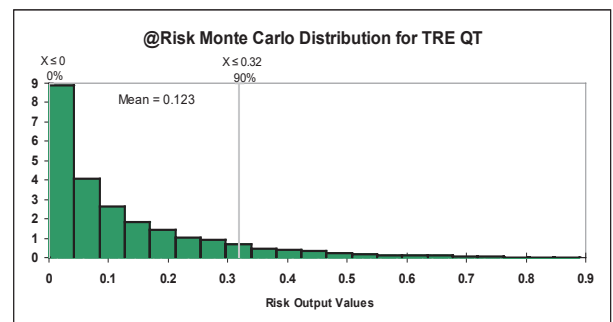


Figure 28. TRE QT @Risk Monte Carlo distribution.

69% of its output values below the defined risk level of “slight” (350) and 80% below the level of “moderate” (1000); as shown in Figure 26, there were few values above 1000. The mean value was 1100, categorised as “significant” (range 1000 to 2000).

TRE Quantitative (TRE QT)

TRE Quantitative was a quantitative method aligned with the Australian Standard of Risk Management definition $\text{Risk} = \text{Likelihood} \times \text{Consequences}$ (AS/NZS 4360:2004) and had 3 categories: “Probability of Failure” (Pf), “Probability of Impact” (Pi), and “Consequences” (Co). These categories were multiplied to create a risk score expressed as a probability, ratio, or financial value. Acceptable risk was divided into 2 categories: at risk scores above 1:10,000, risk was unacceptable, and for risk scores from 1:10,000 to 1:100,000, risk should be reduced if practicable; risk scores $> 1:100,000$ were considered broadly acceptable. A probabilistic method with all inputs between 0 to 1 and multiplied together, TRE QT had a logical consistency, as a 25% change to any input changed the output by 25%, as confirmed by the regression sensitivity (Figure 27) where each input influenced the output by the same amount (25%).

Combining the inputs into 2 categories (consequences and likelihood) resulted in TRE QT favouring likelihood over consequences, with consequences accounting for 25% of the method’s variation and likelihood 49%. The Monte Carlo simulation distribution was a logarithmic curve as expected (Figure 28). The mean was 0.123, giving a risk value of 1:8, which is very high. The distribution was identical to that of QTRA.

TRE Qualitative (TRE QL)

TRE Qualitative was a simplified quantitative method designed for quick field application based on the same principles as TRE QT. A “user interface” was developed using the 3 categories of TRE QT, with five 7-value ranges, and making the assessment values summable by converting them to logarithmic scales. Risk scores from 3 to 300 were possible. These could be converted to probabilities. At risk scores of 200 to 300, risk was considered unacceptable; for risk scores 170 to 199, risk should be reduced if practicable; and risk scores below 170 were considered acceptable.

As TRE QL was based on log-converted values from the TRE QT method, both the 1 rank and regression

sensitivities (Figure 29) illustrated that each input variable influenced the output identically. The combined consequences and likelihood values were similar to the TRE QT values; however, due to the linear rather than logarithmic nature of the method, the regression relationship was stronger, with the likelihood elements providing 64% of the variation and consequences 32%. Figure 30 illustrated the expected distribution curve, which was typical of the shape expected for a strong linear model. The mean was 154 (as expected from a normal distribution curve with a range of 0 to 300), indicating that given a full range of input values, this method generated values towards the mean.

USDAFS 1 (Pokorny et al. 2003)

USDAFS 1 sums the values from 4 assessment criteria, “Probability of Failure,” “Size of Part,” “Probability of Impact,” and “Other Risk Factors.” The output “Risk Rating” ranged from 3 to 12. No definitions or quantification of the output score values were provided. The 1 rank change suggested a linear model, with 14% to the output from any 1 rank change; however, the regression chart (Figure 31) reflected the wider range and higher possible value of available

input values in the “Probability of Failure” category, and hence the greater influence on the outputs. Each input influenced 20% of the output, except the “Probability of Failure” category that influenced some 39% of the variability in the method.

Combining the various inputs into 2 categories (consequences and likelihood) was not possible with this method because the “Other Risk Factors” category allows an assessor to add up to 2 extra points for any reason. Hence, 3 categories are shown, as with the majority of methods; the likelihood elements had the strongest influence on this method, accounting for some 59% of the variation, whilst consequences accounted for 21%. The “Other Risk Factors” category had a significant potential effect on the risk output values. The distribution curve (Figure 32) was typical of a linear method; the mean of 7.5 was higher than a method using matched scaling. Nevertheless, as expected, the majority of this method’s output values were mid-range.

USDAFS 2 (Pokorny et al. 2003)

USDAFS 2 used 2 categories, “Targets” and “Defects,” to create a “Hazard Rating.” The influences of the differing input ranges in this method were obvious in

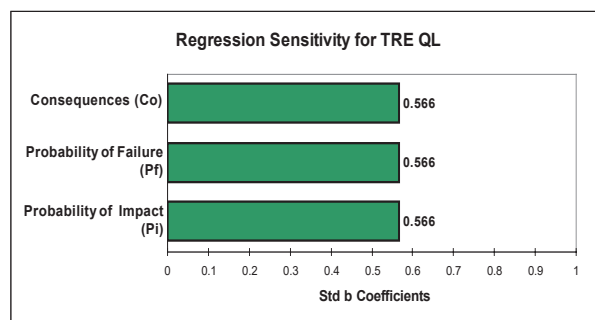


Figure 29. TRE QL @Risk regression sensitivity.

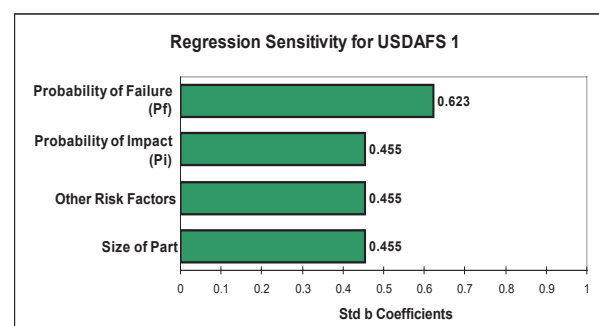


Figure 31. USDAFS 1 @Risk regression sensitivity.

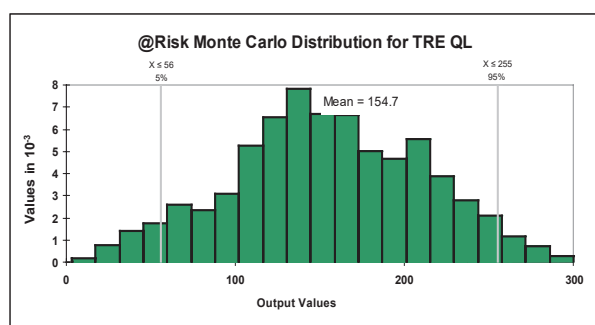


Figure 30. TRE QL @Risk Monte Carlo distribution.

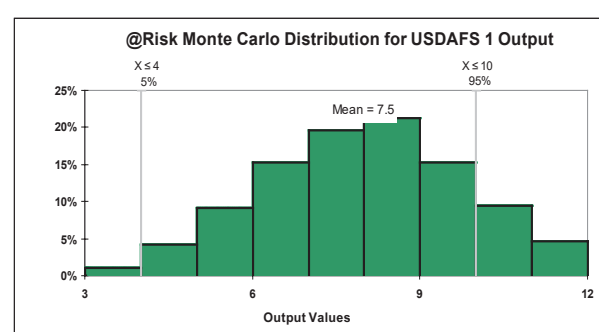


Figure 32. USDAFS 1 @Risk Monte Carlo distribution.

both the 1 rank change and the regression (Figure 33). The limit of only 2 input categories and choice of only the values 1 or 2 for the “Targets” rating versus a range from 0 to 4 for the “Defects” category, plus the multiplication method employed to calculate the output values, produced large variability. A single rank change to the target value could make zero difference to the output or increase it by 100%. The regression indicated that “Defects” account for some 77% of the variation in the method, whilst “Targets” account for 15%.

The limitations of the method became apparent when an attempt was made to combine the inputs into the 2 categories of consequences and likelihood. This method does not account for consequences in the inputs. The “Targets” category merely considered likelihood. Hence, the regression suggested that likelihood factors accounted for some 92% of the variation and consequences 0%. The distribution curve (Figure 34) reflected the somewhat different nature of this method, with the outputs 0 and 2 dominating (50% of the available frequencies), whilst it was not possible to create an output value of 5. The mean predicted value was 2.26.

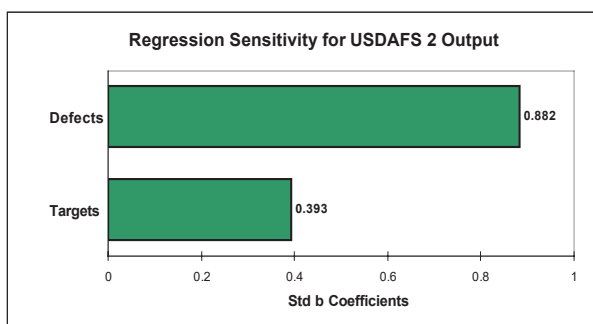


Figure 33. USDAFS 2 @Risk regression sensitivity.

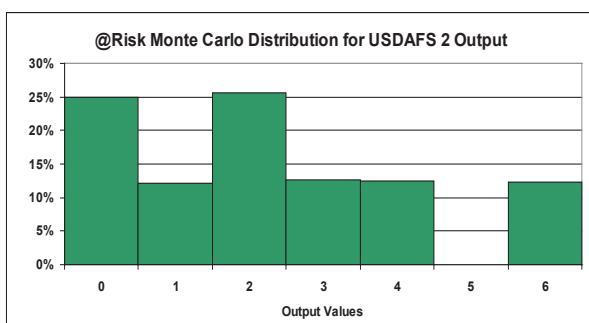


Figure 34. USDAFS 2 @Risk Monte Carlo distribution.

TRAQ (Dunster et al. 2017)

The TRAQ method used wholly verbal, ordinal assessment criteria, and to allow analyses, scores were assigned to “Likelihood of Impact,” “Likelihood of Failure,” and “Consequences” (ranges 1 to 4). The assigned scores, however, work exactly as the words do in assigning a risk rating. A 1 rank change for “Likelihood of Impact” and “Likelihood of Failure” change the risk rating equally by 50%, but the same change for “Consequences” only changed the rating by 33% (Figure 35). Combining the various inputs into 2 categories (consequences and likelihood) resulted in TRAQ favouring the inputs that affect likelihood over consequences. Given that size of part was the only factor accounting for consequences, this outcome was logical.

The Monte Carlo distribution suggested that this method would produce far more values at the lower end of the risk scale, with a mean value of 4.06

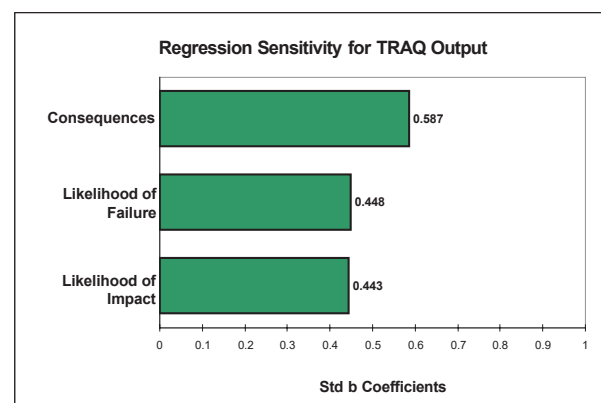


Figure 35. TRAQ @Risk regression sensitivity.

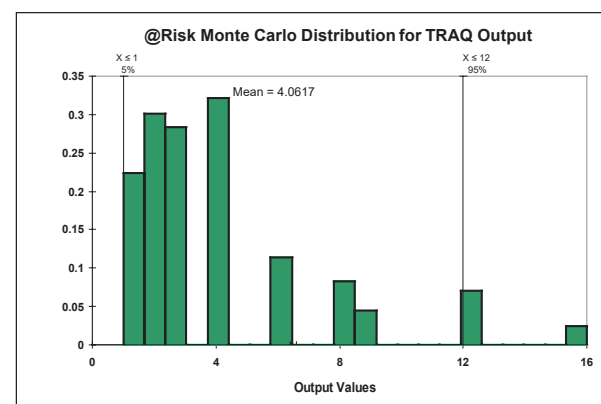


Figure 36. TRAQ @Risk Monte Carlo distribution.

(Figure 36). The distribution clearly shows the impact of the impossibility of obtaining risk ratings of 5, 7, 10, 11, 13, 14, or 15 in the method. The Best Management Guide that accompanies the TRAQ program presented and explained the various risk inputs and the meaning of the outputs well.

DISCUSSION

Sensitivity analyses demonstrated that most methods placed too great an emphasis on limited aspects of a risk assessment, and in most instances, tree risk assessment methods strongly focus on the likelihood of failure or defect aspect of a risk assessment. This is not surprising, given that much of the literature focuses strongly on identification of tree defects, in many cases downplaying the importance or relevance of target usage and particularly consequences in the assessment of tree risk. Both methods of sensitivity analysis identified significant differences between the methods trialled. Failure to understand the influence of inputs on the subsequent risk rating can make it easy to unfairly question the validity of a method. Some methods exhibited very large output changes with little movement of an input, and in many cases this was different for each of a method's input categories. This was due to different scaling and/or ranges used for the inputs and the combination mathematics.

It was surprising to find that 2 methods did not measure consequences, therefore failing to measure risk as it is commonly defined. Most methods placed more emphasis on the likelihood inputs than those influencing consequences, having in most instances 2 likelihood inputs but only 1 measure of consequences. What the ratio should be has not been defined, but in 13 of the methods, likelihood outweighed consequences by a factor of at least 2. If equal likelihood and consequences factor ratios were used, then both components would be reflected in the risk score, which would more accurately represent the real level of risk as it is defined. Having a better balance between likelihood and consequences in tree risk assessment would result in lower risk ratings in most cases, which would be more in line with the very low actual tree-related injury and death rates. The emphasis on likelihood and the associated higher risk rating is likely to result in unnecessary tree removals. Most tree risk assessment methods focus more on the tree than on the target and the real risk that the tree might pose to the target. Such weightings could result in

different methods, creating quite different risk ratings for the same scenario.

The output distributions created from the Monte Carlo simulation highlighted that significant differences existed between the tree risk assessments and indicated that different methods created dissimilar risk values due to the differing input ranges, scales, and methods of mathematically combining the inputs. Modelled distributions were used to indicate how a method was likely to operate in the real world and where various strengths and weaknesses existed. For simple, linear methods with constant scaling and ranges for each of the input categories, such as MandC, Private 2, and TRE QL, the regression model proportionally identified the same changes to outputs and influence of variables. The full probabilistic methods, such as QTRA and TRE QT, displayed identical changes and percentage variation. Methods with differing inter- or intra-category ranges or scales yielded differing percentage variations from that of the 1 rank or 25% sensitivity analysis. This was generally due to the multivariate stepwise regression using the full range of available input values, which more accurately represented the overall performance of each method.

The Monte Carlo simulation produced theoretical or predicted distribution profiles and tended to be more conservative, because a uniform distribution was used for the modelling (all inputs were equally likely), whereas for most managed tree populations, the number of high-risk trees would be lower than low-risk trees. The Monte Carlo simulation identified that some methods tended to produce larger frequencies of particular output values; few methods produced flat or even output distributions; and several methods, including TRAQ, could not produce a full range of the assigned output values (representing word combinations) due to the scales, ranges, and mathematics used.

Ten methods had input categories with differing input values, ranges, or scaling. With these methods, the designers had apportioned differential weightings to each input category, so that each influenced the output differently. Some methods appeared to apply a disproportionate weighting to a single input variable. In Bartlett, the "Failure Potential/Defect Severity" category accounted for 80% of the method's variance, whilst the USDAFS 2 "Defects" category accounted for 77% of the variance. In contrast, the Threats input

category “Impact Score” appeared to only influence the output by 3%, which raised questions as to the value of this input. The analysis showed that the QTRA W method applied different weightings to its inputs to the full QTRA method, with “Target” accounting for 45%, “Probability of Failure” accounting for 31%, and “Size of Part” accounting for 21% of the variance, indicating that in many instances, different risk ratings under the same circumstances would be produced, depending on the method (full or wheel) used.

Whilst most tree risk assessment methods used similar input categories, the multivariate stepwise regression of the Monte Carlo simulation demonstrated that significant differences existed between methods because of scaling, range of input categories, and the mathematics. With the exception of the full version of QTRA, TRE, and TRAQ, no method provided an explanation of its weighting, scaling, or mathematics. Risk assessors often assume that outputs of methods provide a full range of possibilities and that all output scores are equally likely, but the analyses show this is not the case. Few of the method authors clearly discussed the reasons behind the range, weighting, and mathematics of their methods, but TRAQ and QTRA do it better than most. It seemed that few, if any, authors of the assessment methods had analysed the effects of the inputs on the subsequent measurement of risk. Fair and reasonable explanations may exist, but in failing to provide them, many methods lack transparency, which could leave them open to legal criticism or challenge.

The Monte Carlo simulation permitted the creation of probability distribution profiles. The purpose of the distribution profiles was to determine the range of values generated by the input ranges, scaling, and underlying mathematics, and the probable frequency of these output values, in order to represent these as a probability distribution curve. The distribution output charts presented in each method’s summary indicated that some methods tend to produce larger frequencies of particular output values, that few methods produced flat or even output distributions, and that in some instances, methods could not produce a full range of output values. For instance, USDAFS 2 had a theoretical output range of 0 to 6; however, because of the mathematics used, this method cannot produce the output value 5, and generated the numbers 2 and 0 at 2.5 times the frequency of other values. TRAQ could not generate 7 of the outputs in its assigned

range from 1 to 16. Two methods had the possibility of producing a zero risk output, which raised the question as to whether it is possible to rate any tree-related risk as zero. Given the size, age, and context of trees that are subjected to risk assessment, it would seem probable that there is always some element of risk, no matter how small that risk might be, especially for trees growing in public spaces.

In analysing the sensitivity data, while each method was unique and different from the others, they could be placed into 3 broad groups (Table 7):

- Group 1 methods produced a normal distribution with most values around the mean.
- Group 2 methods produced outputs at the lower end of the risk scale.
- Group 3 methods produced outputs evenly, if not continuously, across the risk scale (with Private 3 producing outputs at the lower end of the risk scale as in Group 2).

This grouping shows that the method chosen to risk assess a tree can impact the score derived. For example, Group 2 methods will usually provide a lower risk rating, while Group 1 methods will tend to provide a score closer to the mean for the range of possible scores, and Group 3 methods provide scores evenly across their range. The variations identified between methods in both the ranked order and standardised score approaches strongly aligned, suggesting that both approaches identified similar aspects of each method.

This research defined risk as a variant of $R = Li \times Co$, and methods were assessed in relation to the formula. Using the data generated from the multivariate regression analysis, the various inputs for each method were placed into the consequences or likelihood categories. There is no reason that the weightings for likelihood and consequences should be equal; and this did not apply in any instance (Figure 3). The closest to balanced methods were QTRA (Co:Li = 0.51), TRE QT (Co:Li = 0.51), and TRAQ (Co:Li = 0.66). Private 2 (Co:Li = 0.69) was mathematically the closest to balanced, but its “Other” category weakened the effect of other inputs. USDAFS 2 (Co:Li = 0.00), Private 3 (Co:Li = 0.00), Threats (Co:Li = 0.04), and QTRA W (Co:Li = 0.07) were very strongly weighted to the likelihood inputs. For 14 of the 16 methods, the weighting favoured likelihood over consequence inputs, and only HCC (Co:Li = 2.19) and Kenyon (Co:Li = 2.66) favoured consequences, with both using a “damage factor” as a final modifier. Some

Table 7. Grouping of methods with similar characteristics. Group 1 methods produce a normal distribution with most values around the mean; group 2 methods produce outputs at the lower end of the risk scale; and group 3 methods produce outputs evenly across the risk scale.

| Risk assessment methods | | | | | | | | |
|-------------------------|-------------|-----------|----------|--------|--------|-----|-----|------|
| Group 1 | MandC | Private 2 | USDAFS 1 | Tre QL | | | | |
| Group 2 | QTRA/QTRA W | Private 1 | Threats | Tre QT | Kenyon | CTC | HCC | TRAQ |
| Group 3 | Bartlett | Private 3 | USDAFS 2 | | | | | |

inputs provided a significantly stronger influence on the output than other inputs. This is demonstrated by comparing the two versions of QTRA. QTRA equally weighted each input category (a probability between 0 and 1), and so each input provided the same influence on the output value; but QTRA W weighted the 3 input categories differently (Target = 45%, Size of Part = 21%, and Failure = 31%).

The tendency to more strongly favour likelihood inputs can be partly explained by the mathematics of most methods. Most methods use 3 inputs, 2 of which were typically a likelihood of failure and a probability of impact, which are likelihood factors, but used only a single measure of consequence (typically size of part). Therefore, for simple, equally scaled summation methods, this resulted in a 2:1 ratio favouring likelihood over consequence. While multiplication complicated the process, the skewed weighting remains. Methods that had the reverse weighting typically have a “damage factor” modifier in the final equation. For QTRA, the weighting changed depending on what was being assessed.

CONCLUSION

Most of the tree risk assessment methods analysed would be suitable for managing large urban tree populations, provided the user understands their strengths and weaknesses. The better methods have a balanced set of inputs that consider both likelihood and consequence and produce a full and even range of output values. Whether they meet the criteria of completeness, credibility, reliability, repeatability, robustness, and validity was not tested by the sensitivity analyses and so remains in question. However, methods that explain the meaning of inputs and outputs and which train users in their procedures are more likely to be reliable and repeatable. Furthermore, in arboriculture, widely recognised, acceptable risk levels do not exist, and so an industry-based approach should form the

foundation standard of what is considered to be acceptable tree-related risk. It is important that users of tree risk assessment methods understand the relationship between consequence and likelihood, and the influence that range, weighting, scaling, and number of input variables have on distribution curves and output values.

In general it can be concluded:

Risk methods varied in how they measured risk and will give different results. The choice of tree risk assessment method will influence assessed risk levels, so assessors should be aware of the influence that the method they use may have on risk ratings.

If it is accepted that risk is defined as $R = Li \times Co$, then methods that better balance the 2 components, such as TRAQ and QTRA, better express the risk than methods which do not, such as USDAFS 2, QTRA W, or Threats.

Methods that do not balance likelihood and consequence tend toward higher tree risk outputs, which could lead to unnecessary tree removals.

Methods that utilise a full range of risk ratings will be superior to those with gaps in the range of outputs, such as TRAQ, Bartlett, Private 1, Private 3, Threats, and USDAFS 2. Gaps are easily seen in the outputs of numeric methods, but they can occur in methods using ordinal outputs, as the assigning of numeric values for analysis revealed.

Methods where there is a clear, balanced, and logical relationship between input and output values will be more defensible than those methods with inconsistent inputs, ranges, and mathematics. MandC, QTRA, TRAQ, TRE QT, and TRE QL, with balanced relationships between input values, will be more defensible than other methods. Private 2 is similar to MandC upon which it is based. Private 3 has balanced inputs for likelihood, but does not consider consequences, and CTC, HCC, Private 1, and USDAFS 1 have most but not all of their inputs balanced.

An over-emphasis on an input in a method can raise the risk rating considerably, which again could lead to the unnecessary removal of trees which actually pose a very low risk.

Users of methods should be aware of how they work and the extent of their outputs. For example, for methods which do not allow a full range of outputs, ensure that you do not report an output that cannot be obtained, either due to typographical error, or in the mistaken belief that an intermediate value better represents the perceived risk.

For wholly qualitative (word based, ordinal) methods, such as MandC, CTC, or TRAQ, the changes made in 1 rank sensitivity analyses matter less to the actual change in risk rating than for fully quantitative methods. For example, a 1 rank increase in TRAQ's "Likelihood of Failure" or "Likelihood of Impact" changes the risk rating from moderate to high (1 category), and a decrease changes the risk rating from moderate to low (2 categories); but a 1 rank increase in "Consequences" has no effect on the risk rating, while a decrease changes the risk rating from moderate to low (2 categories). This raises questions as to the balance of the method.

The inputs of most methods were highly subjective due to a lack of data related to the inputs. Often methods do not include well-considered descriptors, which created uncertainty and variability in the interpretation of the meaning of an output, which may affect the robustness, credibility, repeatability, or validity of the methods.

Methods such as TRAQ and QTRA, which have a strong training program, are more likely to provide consistent assessments of risk.

Risk assessments are often assumed to be a matter of "common sense." Such an assumption is incorrect. Even well-trained and experienced assessors may have widely differing opinions as to levels of risk.

Users of tree risk methods should understand the relative strength and weaknesses of any method that they use. It could be relatively simple to challenge the results of a tree risk assessment in a court of law based on the limitations inherent in the underlying methodology, particularly if the user was not aware of them and had failed to take them into account in interpreting the outcomes. The sensitivity analyses indicated clear and strong differences between the methods, reflecting that the underlying mathematics, input categories, ranges, and scaling influence the way that different tree risk assessment methods

process and express risk. Hence, it is not surprising that methods will perform differently in different circumstances and will express risk levels differently.

LITERATURE CITED

- Albers J, Eiber T, Hayes E. 1996. How to recognize hazardous defects in trees. St. Paul (MN, USA): Minnesota Department of Natural Resources and USDA Forest Service. NA-FR-01-96. 20 p.
- Ball DJ, Watt J. 2013. The risk to the public of tree fall. *Journal of Risk Research*. 16:261-269.
- Beer R, Frank S, Waters G. 2001. Overview of street tree populations in Melbourne—turn of the 21st century. *Journal of Arboriculture*. 32:155-63.
- Bloniarz DV. 2004. A community tree risk calculator for pocket PCs. Amherst (MA, USA): USDA Forestry Service, Northeast Center for Urban and Community Forestry. [Accessed 12 April 2014]. <http://www.umass.edu/urbantree/hazard/pda.shtml>
- Coder KD. 1996. Tree risk management and hazard assessment: a general overview. Athens (GA, USA): Warnell School of Forestry and Natural Resources, University of Georgia. Extension Publication FOR 96-033. 10 p.
- Coder KD. 2000. Tree risk management. *City Trees*. 36(6):26-29.
- Colorado Tree Coalition. 2004. Tree Hazard Assessment System 2004. Broomfield (CO, USA): Colorado Tree Coalition. 93 p.
- Cornell Law School. 2020. Sovereign immunity. Ithaca (NY, USA): Wex, Legal Information Institute (LII). [Accessed 20 January 2020]. https://www.law.cornell.edu/wex/sovereign_immunity
- Department of Planning NSW. 1992. Risk criteria for land use safety planning. Parramatta (NSW, Australia): Department of Planning, New South Wales Government. Hazardous Industry Planning Advisory Paper No. 4.
- Dockter D. 2001. *Tree technical manual*. 1st Ed. Palo Alto (CA, USA): The City of Palo Alto Department of Planning and Community Environment. 161 p.
- Dunster JA. 2003. An overview of legal considerations. In: *Wildlife/danger tree assessor's course workbook—parks and recreation sites course module*. Victoria (BC, Canada): Wildlife Tree Committee of British Columbia.
- Dunster JA, Smiley ET, Matheny NP, Lilly SJ. 2017. *Tree risk assessment manual*. 2nd Ed. Champaign (IL, USA): International Society of Arboriculture. 194 p.
- Ellison M. 2005a. *Quantified tree risk assessment—user manual*. Macclesfield (UK): Quantified Tree Risk Assessment Ltd.
- Ellison M. 2005b. Quantified tree risk assessment used in the management of amenity trees. *Journal of Arboriculture*. 31:57-65.
- Fischhoff B. 1994. The psychology of risk characterization. In: Sahlin NE, Brehmer B, editors. *Future risks and risk management*. Dordrecht (The Netherlands): Springer. 256 p.
- Forbes-Laird J. 2003. *Threats—tree hazard: risk evaluation and treatment system—a method for recording, identifying, and managing hazard trees*. Bedford (UK): Forbes-Laird Arboricultural Consultancy Ltd.
- Forbes-Laird J. 2006. *Threats—tree hazard: risk evaluation and treatment system—a method for identifying, recording, and managing hazards from trees*. Bedford (UK): Forbes-Laird Arboricultural Consultancy Ltd.

- Fraedrich BR. 1999. Tree structure evaluation. Charlotte (NC, USA): Bartlett Tree Research Laboratories.
- Frey HC, Mokhtari A, Zheng J. 2004. Recommended practice regarding selection, application and interpretation of sensitivity analysis methods applied to food safety process risk models. Raleigh (NC, USA): Computational Laboratory for Energy, Air, and Risk, Department of Civil, Construction, and Environmental Engineering, North Carolina State University. Prepared for the Office of Risk Assessment and Cost-Benefit Analysis, USDA, Washington, DC, USA. 148 p.
- Frey HC, Patil SR. 2002. Identification and review of sensitivity analysis methods. *Risk Analysis*. 22:553-578.
- Grey GW, Deneke FJ. 1986. *Urban forestry*. New York (NY, USA): John Wiley and Sons. 279 p.
- Haimes YY. 1998. *Risk modelling, assessment, and management*. New York (NY, USA): John Wiley and Sons. 726 p.
- Harris RW. 1983. *Arboriculture: care of trees, shrubs, and vines in the landscape*. Englewood Cliffs (NJ, USA): Prentice-Hall. 688 p.
- Hartley MA, Chalk JJ. 2019. A review of deaths in Australia from accidental tree failure. Marlestone (Australia): Arboriculture Australia. 18 p.
- Hayes E. 2001. *Evaluating tree defects: a field guide*. Rochester (MN, USA): Safetrees. 30 p.
- Helliwell D. 1990. Acceptable levels of risk associated with trees. *Arboricultural Journal*. 14:159-162.
- Helliwell D. 1991. Acceptable levels of risk—letter to the editor. *Arboricultural Journal*. 15:179-181.
- Hewett P, Fakes J, Barnes R, Davidson C. 2003. *Best practice manual—trees and tree root management*. Version 2. Sydney (Australia): Statewide Mutual Insurances. 40 p.
- Hickman GW, Caprile J, Perry E. 1989. Oak tree hazard evaluation. *Journal of Arboriculture*. 15:177-184.
- HSE. 1996. Use of risk assessment within government departments. Report prepared by the Interdepartmental Liaison Group on Risk Assessment. Sudbury (Suffolk, UK): Health and Safety Executive Books. 48 p.
- Hume City Council. No date. Tree Risk Management. Melbourne (Australia): City of Hume.
- Johnson D. 1981. Tree hazards: recognition and reduction in recreation sites. Golden (CO, USA): USDA Forest Service, Rocky Mountain Region. Technical Report R2-1. 17 p.
- Kane B, Ryan D, Bloniarz D. 2001. Prioritizing risk trees in a community. *Tree Care Industry*. 11(7):45-49.
- Kenyon P. 1993. Tree maintenance and hazard assessment. In: Yau DP, editor. *Tree maintenance: proceedings of a seminar conducted by the Royal Australian Institute of Parks and Recreation (Victorian Region), Victorian College of Agriculture and Horticulture (Burnley)*. Dickson (Australia): Royal Australian Institute of Parks and Recreation.
- Koeser AK, Hauer RJ, Peterson W. 2016. Municipal tree risk assessment in the United States: findings from a comprehensive survey of urban forest management. *Arboricultural Journal*. 38:218-229.
- Lonsdale D. 1999. *Principles of tree hazard assessment and management*. London (UK): The Stationery Office. 388 p.
- Lonsdale D. 2000. *Hazards from trees: a general guide*. Edinburgh (UK): Forestry Commission. 35 p.
- Lonsdale D. 2007. Current issues in arboricultural risk assessment and management. *Arboricultural Journal*. 30:163-174.
- Lowrance W. 1976. *Of acceptable risk*. Los Altos (CA, USA): William Kaufmann. 180 p.
- Matheny NP, Clark JR. 1994. *A photographic guide to the evaluation of hazard trees in urban areas*. Savoy (IL, USA): International Society of Arboriculture. 85 p.
- Matheny NP, Clark JR. 2007. Managing risk in the urban forest—part two: tree risk assessment. *Arborist News*. 16(3):9-16.
- Matheny N, Clark JR. 2009. Tree risk assessment: what we know (and what we don't know). *Arborist News*. 18(1):28-33.
- Mattheck C, Breloer H. 1994. Field guide for visual tree assessment (VTA). *Arboricultural Journal*. 18:1-23.
- Minitab. 2006. Minitab v. 15.1. State College (PA, USA): Minitab LLC. <https://www.minitab.com/en-us>
- Paine LA. 1971. Accident hazard evaluation and control decisions on forested recreation sites. Berkeley (CA, USA): USDA Forest Service, Pacific Southwest Forest and Range Experiment Station. Research Paper PSW-68. 10 p.
- Palisade. 2002. @RISK v. 4.5. Ithaca (NY, USA): Palisade. <https://www.palisade.com/risk/default.asp>
- Patil SR, Frey HC. 2004. Comparison of sensitivity analysis methods based on applications to a food safety risk assessment model. *Risk Analysis*. 24:573-585.
- Pokorny JD, O'Brien J, Hauer R, Johnson G, Albers J, Bedker P, Mielke M. 2003. Urban tree risk management—a community guide to program design and implementation. St. Paul (MN, USA): USDA Forest Service, Northeastern Area State and Private Forestry. NA-TP-03-03. 204 p.
- Robbins K. 1986. HOW to recognize and reduce tree hazards in recreation sites. St. Paul (MN, USA): USDA Forest Service, Northern Area State and Private Forestry. NA-FR-31. 28 p.
- Rushforth K. 1987. Trees and the law and damage by trees. In: *The Hillier book of tree planting and management*. Newton Abbot (UK): David and Charles. 224 p.
- Saltelli A. 2002. Sensitivity analysis for importance assessment. *Risk Analysis*. 22:579-590.
- Saltelli A, Chan K, Scott E, editors. 2000. *Sensitivity analysis: gauging the worth of scientific models*. Chichester (UK): John Wiley and Sons. 504 p.
- Slovic P. 2000. *The perception of risk*. London (UK): Earthscan Publishing. 511 p.
- Smiley ET, Fraedrich BR, Hendrickson N. 2002. Tree risk management. Charlotte (NC, USA): Bartlett Tree Research Laboratories.
- Smiley ET, Matheny NP, Lilly SJ. 2017. Tree risk assessment. In Ferrini F, Konijnendijk van den Bosch CC, Fini A, editors. *Routledge handbook of urban forestry*. London (UK): Taylor and Francis. 574 p.
- Sreetheran M, Adnan M, Khairil Azuar AK. 2011. Street tree inventory and tree risk assessment of selected major roads in Kuala Lumpur, Malaysia. *Arboriculture & Urban Forestry*. 37:226-235.
- Standards Australia. 2004a. AS/NZS 4360:2004: risk management. Strathfield (Australia): Standards Association of Australia.
- Standards Australia. 2004b. HB 436:2004: risk management guidelines—companion to AS/NZS 4360:2004. Strathfield (Australia): Standards Association of Australia.

- State of Victoria, Department of Environment, Land, Water and Planning. 2017. Tree hazard pictorial guide: look up - look down - look around. Melbourne (Australia): State of Victoria Department of Environment, Land, Water and Planning. [Accessed 22 July 2018]. <https://files-em.em.vic.gov.au/public/Safety/TreeHazardPictorialGuide-2017.pdf>
- Thompson K. 2002. Variability and uncertainty meet risk management and risk communication. *Risk Analysis*. 22:647-654.
- Tomao A, Secondi L, Corona P, Giuliarelli D, Quatrini V, Agrimi M. 2015. Can composite indices explain multidimensionality of tree risk assessment? A case study in an historical monumental complex. *Urban Forestry & Urban Greening*. 14:456-465.
- Urdan TC. 2005. *Statistics in plain English*. East Sussex (UK): Psychology Press, a Taylor & Francis Group. 200 p.
- USDA Forest Service. 2007. Rocky Mountain Region: guide to hazard tree management. USDA Forest Service. [Accessed 15 October 2007]. <http://www.fs.fed.us/r2/fhm>
- VAGO. 2018. Local government insurance risks: July 2018. Independent assurance report to parliament 2018–19:1. Melbourne (Australia): Victorian Auditor-Generals Office. PP no 422, Session 2014–18. 117 p. <https://www.audit.vic.gov.au/sites/default/files/2018-07/20180725-Insurance-Risks.pdf>
- van Wassenae P, Richardson M. 2009. A review of tree risk assessment using minimally invasive technologies and two case studies. *Arboricultural Journal*. 32:275-292.
- Vose D. 1996. *Quantitative risk analysis: a guide to Monte Carlo simulation modelling*. New York (NY, USA): John Wiley and Sons. 340 p.
- Wildlife Tree Committee of British Columbia. 2003. Wildlife/danger tree assessor's course workbook—parks and recreation sites course module. Victoria (BC, Canada): Wildlife Tree Committee of British Columbia.

ACKNOWLEDGMENTS

M.B. Norris was a member of the expert panel consulted in the development of TRAQ.

M.B. Norris
University of Melbourne
Burnley College
School of Ecosystem and Forest Sciences
500 Yarra Boulevard
Richmond, Australia

G.M. Moore (corresponding author)
University of Melbourne
Burnley College
School of Ecosystem and Forest Sciences
500 Yarra Boulevard
Richmond, Australia
gmmoore@unimelb.edu.au

Conflicts of Interest:

The authors reported no conflicts of interest.

Résumé. Seize méthodes d'évaluation des risques liés aux arbres ont été soumises à une analyse de sensibilité afin de déterminer les facteurs ayant le plus influencé le résultat de chaque méthode. Les analyses ont établi l'influence relative que les variables d'entrée exercent sur la valeur finale du risque. Un tableau Excel a été utilisé pour créer une simple variation de rang de $\pm 25\%$ ou ± 1 (selon la méthode) pour chaque critère, la variation de la sortie étant enregistrée en pourcentage. Le logiciel Palisade's @Risk fut utilisé afin d'effectuer une simulation de type Monte Carlo (avec échantillonnage Hypercube latin) de 5000 itérations basées sur les variables d'entrée et la formule de sortie. À partir de la simulation, une régression graduelle multivariée fut entreprise afin de déterminer l'influence des variables d'entrée de chacune des méthodes dans l'établissement des valeurs de sortie. Les résultats de l'analyse de sensibilité montrèrent des différences nettes et marquées parmi les 16 méthodes, reflétant le fait que les mathématiques sous-jacentes, les catégories d'entrée, les écarts et l'échelle influencent la manière dont les différentes méthodes traitent et expriment le risque. Il n'est pas étonnant que les méthodes fonctionnent différemment selon les circonstances et expriment ainsi différemment le niveau de risque. Les analyses ont démontré que la plupart des méthodes mettaient trop l'accent sur des aspects limités de l'évaluation des risques. La plupart des méthodes sont fortement axées sur les aspects de l'évaluation liés aux dangers ou aux défauts et sur la probabilité de bris plutôt que sur l'aspect des conséquences d'une évaluation. Bien que les méthodes soient bien distinctes, il était possible de les regrouper en trois grandes catégories : les méthodes du Groupe 1 généraient une distribution normale avec la plupart des valeurs avoisinant la moyenne; les méthodes du Groupe 2 ont produit des résultats situés dans la partie basse de l'échelle des risques; et les méthodes du Groupe 3 ont généré des résultats répartis de manière uniforme, voire continue, sur l'échelle des risques. Les utilisateurs de l'évaluation des risques liés aux arbres doivent comprendre les forces et les faiblesses de toute méthode utilisée, car il pourrait être relativement facile de contester les résultats d'une évaluation des risques en se fondant sur les limitations inhérentes à la méthodologie sous-jacente.

Zusammenfassung. Sechzehn Methoden zur Risikobewertung von Bäumen wurden einer Sensitivitätsanalyse unterzogen, um festzustellen, welche Faktoren die Ergebnisse der einzelnen Methoden am stärksten beeinflussten. Die Analysen zeigen den relativen Einfluss, den die Eingangsvariablen auf den endgültigen Risikowert ausüben. Mit Hilfe von Excel wurde eine einfache Rangänderung von $\pm 25\%$ oder ± 1 (je nach Methode) für jedes Kriterium erstellt, wobei die Änderung des Ausgangs als Prozentsatz erfasst wurde. Die @Risk-Software von Palisade wurde verwendet, um eine Monte-Carlo-Simulation (mit Latin-Hypercube-Stichprobenziehung, LHS) von 5000 Iterationen auf der Grundlage der Eingabevariablen und der Ausgabeformel durchzuführen. Anhand der Simulation wurde eine multivariate schrittweise Regression durchgeführt, um den Einfluss der Eingabevariablen jeder Methode auf die Bestimmung der Ausgabewerte zu bestimmen. Die Ergebnisse der Sensitivitätsanalyse weisen auf einige klare und starke Unterschiede zwischen den 16 Methoden hin, was darauf hinweist, dass die zugrunde liegende Mathematik, die Eingangskategorien, die Messbereiche und die Skalierung die Art und Weise beeinflussen, wie die verschiedenen Methoden das Risiko verarbeiten und ausdrücken. Es ist

nicht überraschend, dass die Methoden unter verschiedenen Umständen unterschiedlich arbeiten und das Risikoniveau unterschiedlich ausdrücken. Die Analysen zeigten, dass die meisten Methoden einen zu großen Schwerpunkt auf begrenzte Aspekte der Risikobewertung legen. Die meisten Methoden konzentrierten sich stark auf die Gefahren- oder Fehleraspekte der Bewertung und die Wahrscheinlichkeit eines Fehlschlags und weniger auf den Aspekt der Folgen einer Bewertung. Die Methoden unterschieden sich zwar in einzigartiger Weise, konnten aber in drei große Gruppen eingeteilt werden: Methoden der Gruppe 1 ergaben eine Normalverteilung mit den meisten Werten um den Mittelwert; Methoden der Gruppe 2 erbrachten Ergebnisse am unteren Ende der Risikoskala; und Methoden der Gruppe 3 erbrachten Ergebnisse, die gleichmäßig, wenn auch nicht kontinuierlich über die gesamte Risikoskala verteilt waren. Benutzer von Baum-Risikobewertungen sollten die Stärken und Schwächen jeder verwendeten Methode verstehen, da es relativ einfach sein könnte, die Ergebnisse einer Risikobewertung aufgrund der zugrundeliegenden Methode eigenen Einschränkungen in Frage zu stellen.

Resumen. Se sometieron dieciséis métodos de evaluación del riesgo de árboles a análisis de sensibilidad para determinar cuáles factores influyen más en cada método. Los análisis indican la influencia relativa que las variables de entrada ejercen sobre el valor final de riesgo. Se utilizó Excel para crear un cambio de rango simple de 25% o 1 (dependiendo del método) para cada criterio; con el cambio en la salida registrado como un porcentaje. El software @Risk de Palisade se utilizó para llevar a cabo una simulación de Monte Carlo (con muestreo de Latin Hypercube) de 5000 iteraciones basadas en las variables de entrada y la fórmula de salida. A partir de la simulación, se llevó a cabo una regresión escalonada multivariante para determinar la influencia de las variables de entrada de cada método en la determinación de los valores de salida. Los resultados del análisis de sensibilidad indican algunas diferencias claras y fuertes entre los 16 métodos, lo que refleja que las matemáticas subyacentes, las categorías de entrada, los rangos y la escala influyen en la forma en que los diferentes métodos procesan y expresan el riesgo. No es de extrañar que los métodos funcionen de manera diferente en diferentes circunstancias y expresen un nivel de riesgo diferente. Los análisis demostraron que la mayoría de los métodos ponen demasiado énfasis en aspectos limitados de la evaluación del riesgo. La mayoría de los métodos se centraron fuertemente en los aspectos de daños o defectos de la evaluación y la probabilidad de falla en lugar del aspecto de consecuencia de una evaluación. Si bien los métodos eran únicos, podían colocarse en 3 grupos amplios: los métodos del Grupo 1 producían una distribución normal con la mayoría de los valores alrededor de la media; los métodos del Grupo 2 dieron resultados en el extremo inferior de la escala de riesgo; y los métodos del Grupo 3 produjeron salidas uniformes si no continuamente a través de la escala de riesgo. Los usuarios de la evaluación del riesgo de los árboles deben comprender las fortalezas y debilidades de cualquier método utilizado, ya que podría ser relativamente sencillo cuestionar los resultados de una evaluación del riesgo basada en limitaciones inherentes a la metodología subyacente.