A Risk-Based Approach to Assessing the 'Fitness for Use' of Spatial Data

Aggrey Agumya and Gary J. Hunter

Abstract: *In this paper the authors propose an alternative method for assessing the 'fitness for use' of spatial data. While the traditional standards-based approach to this problem has long been used (due primarily to the map production roots of spatial data handling), the authors believe this method has several limitations. As an alternative they propose a technique based on risk management practices, in which a study is made of the effect that uncertainty in the data has upon the ultimate decision to be made with it. In turn the adverse consequences of making a poor decision are quantified, and it is this information which enables a user to determine whether a data set is fit for use or not. The paper presents a six-step risk management process which includes a variety of options for reducing the risk that an agency may face when using spatial data for critical decision making. Finally, an example is provided which illustrates the entire process.*

Corresponding Author: *Dr Gary J. Hunter*

1. Introduction

This paper proposes an alternative approach to the traditional method of assessing the 'fitness for use' of spatial data. It begins by examining the shortcomings of the traditional standards-based approach and explains how an alternative option that focuses on assessing the impact or consequences of uncertainty may be more appropriate. The suggested method not only overcomes several limitations of the standards-based technique, but also offers mechanisms for managing the impact of uncertainty in data that is eventually accepted for use. It is argued that this latter feature is an important characteristic of the proposed approach because the problem of uncertainty neither ends with nor is limited to simply assessing fitness for use. Hence it is desirable that any fitness assessment provides significant inputs to the subsequent uncertainty management steps.

Aggrey Agumya *has an MSc from ITC in The Netherlands and is a PhD student with the Department of Geomatics at the University of Melbourne, Australia. His research specialty is 'fitness for use' assessment of spatial data and he has presented three papers on this subject at past URISA conferences.*

Dr Gary J. Hunter *is a Senior Lecturer and Deputy-Head of the Department of Geomatics, and his research specialties are modeling and communicating uncertainty in spatial data and decision-making under conditions of uncertainty. He has presented papers on this and other topics at URISA conferences for the past 11 years and has had his work previously published in the URISA Journal. In conjunction with Professor Michael Goodchild, he has twice been awarded URISA's Horwood Critique Prize.*

Furthermore, in those cases where conditions, traditions or the circumstances of users and their access to data preclude assessment of its fitness, a knowledge of how data uncertainty affects their decisions is essential for managing the possible consequences of uncertainty. The suggested method uses risk as a metaphor for quantifying the consequences that uncertainty in data may have upon the decisions for which it is used. The paper discusses limitations of the standards-based approach, then reviews the concept of risk, its subjective bias, and the factors that influence its perception. Finally, the paper outlines the risk management process for assessing fitness for use, presents an example of its application, and discusses the limitations of the approach and future research issues that will need to be resolved.

2. Limitations of The Standards-based Approach

The primary concern that end-users have about uncertainty in data is its potential impact upon their decisions. As such, the assessment of fitness for use is intended to avoid using data whose uncertainty is associated with consequences that are deemed unacceptable. Therefore fitness for use is essentially about, and should be ultimately determined by, whether the impact of uncertainty is either acceptable or unacceptable. The method traditionally employed to assess fitness for use — the standards-based method — compares data uncertainty with a set of standards that reflect acceptable levels of uncertainty in the data (Frank 1998). With this technique, fitness for use is assessed by directly comparing the quality elements of information against a set of standards that represent the corresponding acceptable quality components. To facilitate direct comparison, the standards are defined using the same elements as those used for describing data quality. These may include: scale (of the source document); Root Mean Square Error (RMSE); resolution; Percentage of Correctly Classified pixels (PCC); currency; and percentage completeness.

Its use as the method of choice for assessing fitness can be traced back to the map production roots of spatial data handling, where positional and attribute accuracy standards have long been used for quality control and assurance. Since uncertainty in the data is considered to be acceptable if its consequences are acceptable, then the standards approach inherently reflects a certain threshold of acceptability. However, the standards-based method does not provide for estimation of the consequences of uncertainty and it is argued that this is one of its key limitations.

While uncertainty in spatial data is composed of several well-known elements (Guptill and Morrison 1995), the obvious measurable ones are positional and attribute accuracy, logical consistency, completeness and currency. Unfortunately, measures of these elements cannot as yet be combined into a single meaningful composite index (Veregin and Hargitai 1995), which means that assessment of fitness for use entails inspecting each element separately. In turn, this requires specification of a separate standard for each element, and the necessary standards are derived by inverting the acceptable consequences of uncertainty into a set of the various elements of spatial data uncertainty (Frank 1998).

This inversion is perhaps the most vexed problem of the standards-based method, since it involves estimating several unknowns from a single value — a problem without a unique solution (since the solution is an infinite number of possible combinations of the unknowns). This makes the use of simplifying assumptions imperative — for example the standard may be based on a single uncertainty element to which all consequences will be attributed. The uncertainty element typically assigned this role is the one to which the decision is most sensitive, but clearly such assumptions compromise the validity of the estimated standards and ultimately the fitness for use assessment.

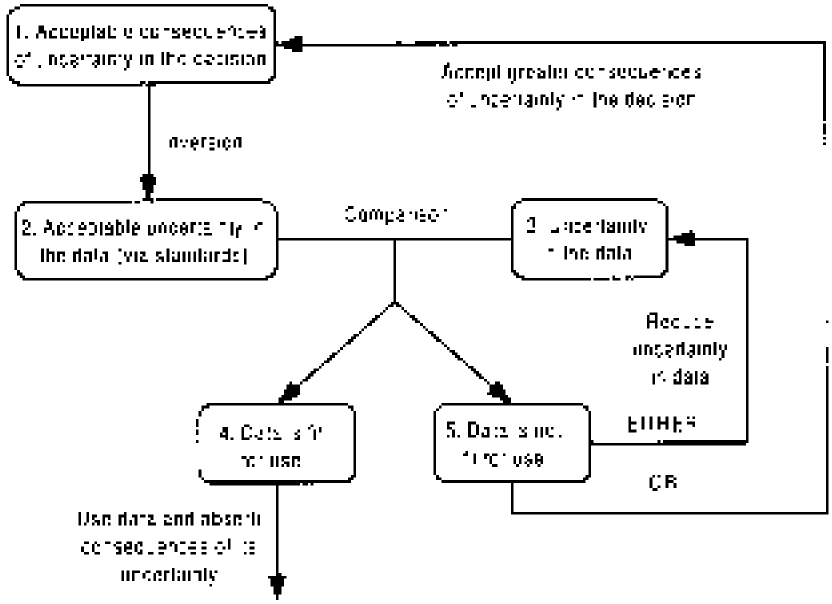
Assessment of the various uncertainty elements separately, according to a limited combination of standards, fails to account for possible trade-offs and compensations that occur when the elements are combined. For example, uncertainty in slope gradient is primarily influenced by the positional and attribute accuracy of elevation models. Using the standards-based method, a dataset is declared to be unsuitable if either its positional or attribute accuracy does not meet a specified standard, but this assessment does not take into account the fact that when considered jointly the two accuracies may well produce a result that is acceptable. In another example, it was rumored in the mid-1990s that an Australian emergency service agency was to take out a $20 million insurance policy to protect itself against liability claims arising from its use of a GIS-based dispatch system. While the street centerline and address databases were produced according to contract specifications (standards), the authors suggest there was no means of separating out the individual error effects of poor positional accuracy (of the street segments), attribute accuracy (street numbers and names), logical consistency (street network topology), completeness (missing street segments and addresses) or currency (out-of-date information), and as such the $20 million figure may well have been simply a calculated guess. Instead, with a risk-based approach it would have been possible to determine which of these factors contributed most to the potential liability associated with the system and to formulate a strategy for reducing this liability in the most cost-effective manner.

Furthermore, the reluctance by many users to formally assess the fitness of their data can be traced back to the difficulty of the task (Beard 1997), and the failure of the method of assessment to rationally justify the undertaking primarily because it does not warn users of the consequences of uncertainty. As Goodchild (1998) observes, politicians and decision makers often ignore uncertainty issues because uncertainty is not portrayed as a number to which they can easily relate. The authors agree with this contention and suggest that a more appropriate form which users would relate to more easily and effectively, is one that they associate with the consequences of uncertainty such as liability, pecuniary losses and mortality.

At the same time, there are instances when users are constrained to use particular datasets regardless of the uncertainty they contain because alternative datasets either do not exist, are not accessible or not feasible to use. For example, it is simply not feasible for a person or organization other than a national statistical agency to undertake a national census, and so users are constrained to use the data that is officially available. For users faced with these conditions the main concern regarding data uncertainty should be managing its impact on their decisions. This is also a concern shared by those who already assess fitness, because they too must eventually absorb any residual uncertainty and therefore need suitable mechanisms for managing its impact. Because the standards-based approach does not involve estimating the size of the impact of uncertainty, it offers limited value for managing uncertainty. In addition, the inability of the method to enable comparison between competing datasets that are all judged suitable, diminishes its usefulness when the problem extends to choosing the best among several suitable datasets.

From Figure 1, when a fitness assessment establishes that a dataset is unsuitable the initial reaction is to reduce uncertainty in the data. However other options exist such as accepting higher consequences of uncertainty and thus tolerating greater uncertainty, or reducing the vulnerability of the decision to uncertainty in the data by diminishing the influence that it exerts upon the decision. If acceptance

Figure 1: The standards-based approach to assessing fitness for use.



of higher consequences is chosen, then again the key issue is quantifying the extra burden of the consequences. The third option of reducing the vulnerability of the decision involves establishing an association between reduction of the influence of data on a decision and the resultant reduction in the consequences of its uncertainty. Again, the standards-based approach is not well suited to supporting this method. Quantifying the consequences of uncertainty is also necessary for stimulating interest amongst those that ignore the issue because they remain doubtful as to its impact upon their decisions. Furthermore, quantification can provide a compelling rationale to system managers for supporting further system development and implementation, by helping to determine the ratio between the cost of avoided misuse achieved by using suitable data, and the required funding.

Thus, the fundamental shortcomings of the standards-based method revolve around the fact that it does not involve quantitative estimation of the consequences of uncertainty. Knowledge of these consequences is essential for improving the utility of spatial data and motivating reluctant users who may be exposed to potential losses to undertake fitness assessments. Such knowledge is also necessary for managing the consequences, and accordingly the authors believe an alternative approach to the current standards-based method is required.

3. An Overview of The Risk-based Approach

Risk analysis has already been suggested as a plausible basis for characterizing and estimating consequences of uncertainty in spatial data (Goodchild 1992). Using risk to represent the consequences of uncertainty has the added benefit of being amenable to an established framework (risk management) for estimating and managing the consequences of uncertainty. In recognition of the limitations of relying on standards and of the potential benefits of a risk-based approach, the transition from standards has already been proposed by organizations such as the water resources planning and management division of the American Society of Civil Engineers (Haimes and Stakhiv 1986), the Australian National Commission On Large Dams (ANCOLD) (McDonald 1995), and the US Environmental Protection Agency (The Conservation Foundation 1985). According to the risk-based approach (Figure 2), assessment of fitness for use essentially involves establishing whether the adverse consequences of uncertainty in the data, expressed in terms of risk, are acceptable or not. The assessment process therefore elicits answers to two fundamental questions, viz.:

- What are the adverse consequences associated with the decision, in terms of risk, attributable to uncertainty in the data? and
- What are the acceptable consequences of uncertainty in terms of risk?

Answering the first question entails propagating data uncertainty, in its various elements, into risk. The propagation demands an understanding of how uncertainty in data interacts with the decision environment to adversely affect decisions, and the extent to which a particular dataset influences decisions — that is, the degree of utilization of the data (Zwart, 1991). If the utilization of data in a decision is minimal, such as when it is only referred to, then it is reasonable to expect uncertainty in that data to have a lesser impact in terms of risk than if the utilization was more significant, such as when the data has the power to change decisions. Indeed, when uncertainty in data is very high it is common for users to minimize their exposure to its impact by limiting the utilization of the data (Laws et al. 1989)

The second question entails establishing a threshold for the risk that is considered acceptable. Here, the word 'acceptable' prompts the crucial questions: 'acceptable to who, in whose view and in what terms?' (Lowrance 1976). These questions point to important characteristics of the risk threshold, and consequently of fitness for use, namely that the threshold is subjective and influenced by continually changing values (economic or other means), the constraints of parties to the risk, and the circumstances under which the risk is determined (Fischhoff *et al.* 1981). Each of these factors depends on the context of the decision and are subject to change over time. In turn, users may be categorized into two groups according to how they respond to uncertainty in data, viz.:

- those who establish the fitness of data before using it; and
- those who use data regardless of its fitness, either because they are compelled to use it or else they choose to ignore its uncertainty.

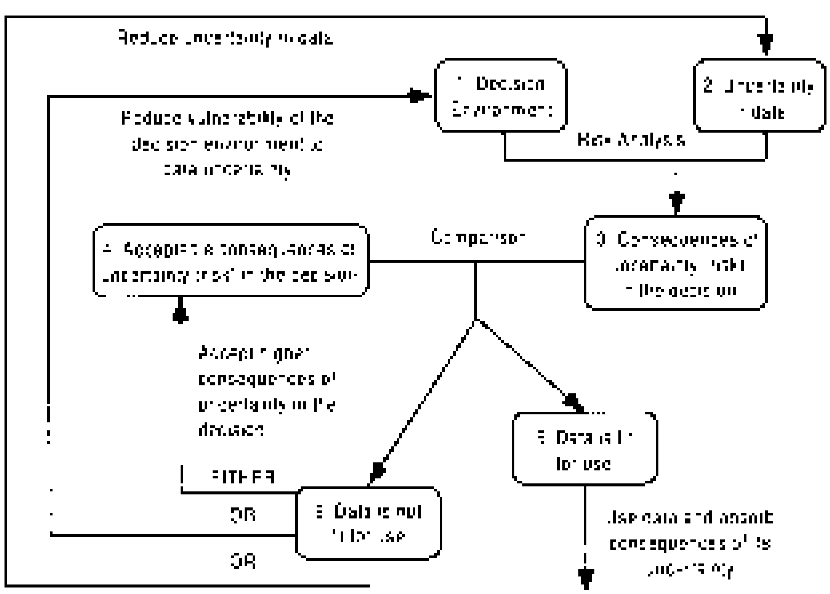
Because the risk-based approach provides for estimating consequences of uncertainty, its relevance extends even to those users who choose not to test data for fitness. For data that are judged to be unsuitable, the risk-based approach offers several possible actions (as in Figure 2) all of which are quite practicable. They include reducing the uncertainty, accepting more risk, making the decision environment less vulnerable to uncertainty in data, or any combination of the above. This range of options is a feature of the risk-based method that makes it more favorable than its standards-based counterpart.

Smith *et al.* (1991) and Burrough *et al.* (1996) show that the quality of information is a function of both the quality of the models (or algorithms) and the data used to derive the information. For unstructured decision problems the models are less well known, rendering the propagation of data uncertainty into risk more difficult. Furthermore, in such decisions the contribution of data uncertainty to the overall decision risk is likely to be less significant compared to the risk attributable to uncertainty in the models, which is largely unknown. Finally, the lack of decision structure implies that a single dataset is not likely to determine the decision, so in general the degree of utilization for individual datasets is low. For these reasons, the risk-based approach is not well-suited to unstructured problems.

4. The Concept of Risk

Risk is an abstract parameter for representing the impact of adverse events or actions about which there is no certainty (Thomson 1987). It is an important factor in decision making since it is one of the few quantitative variables that can determine the utility of decision alternatives. It consequently aids in providing a rational basis for comparing not only those alternatives, but also for allocating

Figure 2: *The risk-based approach to assessing fitness for use of spatial data.*



resources towards risk reduction and improving safety and public health (Starr 1987). Analysis or estimation of risk entails identifying the potential damage, injury or loss associated with events or undertakings, and estimating its likelihood. This knowledge is then used to determine the level of exposure and to judge whether the risk is tolerable or unacceptable (Fischhoff *et al.* 1981, Bohnenblust and Schneider 1987). Understanding the process and degree of exposure is essential for developing mechanisms, safeguards and responses for limiting the likelihood and/or magnitude of damage as well as the cost of the damage (Rasmussen 1981, Starr 1987). Indeed, as Reid (1992) points out, quite often simply understanding the process of exposure can be more useful and informative than deriving the ultimate quantitative values of risk.

4.1 Definition of Risk

'Risk' is a very overloaded term. It cuts across many disciplines and although it has been extensively discussed, it still lacks a standard definition. Fischhoff *et al.* (1984) suggest that the definition of risk is inherently controversial because of the important role that risk plays in policy issues. They point out that the choice of definition can depend on how well it suits particular policy positions and as such is an inherently political issue. However, there is general agreement that the notion of risk involves uncertainty about occurrences or outcomes of events, which are associated with some kind of loss or damage (Kaplan and Garrick 1981, Gratt 1987 and Williams 1995). Sage (1995) on the other hand, argues that risk is not limited just to adverse events but includes positive outcomes as long as their occurrence is uncertain. This interpretation is common in insurance and economics applications and is referred to as 'speculative risk'. Nevertheless, the sense in which the term risk is traditionally used is the one that connotes the possibility of loss or damage rather than gain. This does not necessarily mean that any undesirable event or outcome of a decision constitutes a risk, but rather that it only becomes a risk when there is uncertainty about its occurrence. In addition, there are some terms that are often erroneously used as synonyms for risk, the commonest examples being uncertainty and hazard.

4.2 Risk Versus Uncertainty

'Risk' and 'uncertainty' are so inextricably related that they are often used interchangeably especially in the disciplines of economics, insurance and heath science. For example, Knight (1921) defines risk as 'measurable uncertainty' while others variously describe it as 'the uncertainty of loss' or 'the objectified uncertainty regarding the occurrence of an undesirable event'. Nevertheless, decisions are said to be made under conditions of risk when the likelihood of their outcomes is known, whereas they are made under conditions of uncertainty when the likelihood of their outcomes is not known (Davis and Olson 1985). The predominantly held view, sometimes referred to as the engineering view, defines risk as a function of uncertainty or likelihood about the occurrence of an event that occasions loss, and the loss suffered if the event occurs. Some authors decline to specify the structure of the relationship between the two parameters (likelihood and loss), however many such as Gratt (1987), Raftery (1994) and Williams (1995) characterize risk as likelihood multiplied by loss. The 'likelihood times loss' characterization of risk depicts it as a single valued measure, and while this is desirable for comparing risks it can mask valuable information about the joint variability of likelihood and loss. Knowledge of this variability is important for understanding how the contribution to risk is distributed, which in turn aids in establishing where to best target risk reduction and other risk response efforts.

4.3 Risk versus Hazard

Another term that is closely associated with and sometimes confused with risk is 'hazard'. Kaplan and Garrick (1981) define a hazard as the source of adverse consequences, while from a social science viewpoint Pidgeon *et al.* (1992:89) define it as '… threats to people and the things they value'. From these perspectives, it can be argued that exposure to risk results from the presence of a hazard. It can hence be asserted that uncertainty in spatial information may constitute a hazard, especially if it is ignored. In seismology a hazard is the likelihood of an area being affected by potentially destructive seismic activity within a given period of time. Seismic risk is a function of (seismic) hazard, value and vulnerability, where value includes the number of people or amount of property exposed to the activity, while vulnerability is a measure of the proportion of value likely to be lost or damaged by the activity (Dobran 1995). In this paper the word 'hazard' will be used in an engineering sense, that is, it will refer to causes of adverse consequences.

4.4 Real versus Perceived Risk

The notion of real risk is controversial. Elms (1992) describes it as the risk that would be calculated if all the relevant information about the likelihood and consequences of an adverse event were known. This suggests that the difference between perceived and real risk lies in the completeness and certainty of knowledge about the likelihood and consequences. Kaplan and Garrick (1981) assert that the notion of real risk or absolute risk always ends up being somebody else's perceived risk. Hence the conflict between real and perceived risk can be thought of as conflict between risk perceptions of experts and those of the public. The expert's concept of risk is based on a narrow definition of risk which is limited to the likelihood and undesirable consequences of an event, whereas the public's conception of risk is much broader and includes a host of attitudes towards the event (Kasperson *et al.* 1988). The perception of risk involves the beliefs, judgements, feelings and values that people adopt towards adverse events (Pidgeon *et al.* 1992). These psychological, social and cultural preferences influence attitude to risk as well as acceptance and the acceptable level of risk. Hence they should be considered when making decisions that affect the public. The various characteristics upon which risk perception depends have been widely discussed, (Lowrance 1976, Rowe 1977 and Elms 1992), and include:

- Voluntariness of the risk: people are more averse to adverse events that they have no control over. In most cases, this is the characteristic that has greatest influence on perception of risk.
- Familiarity with the event or its consequences: familiarity tends to reduce the perceived risk.
- Extent of damage or loss: reaction to a single disaster that leads to monumental loss tends to be much stronger than that of a similar loss, which is scattered over several events. Hence risk due to the former is perceived to be greater.
- Cultural context: culture which shapes people's values, beliefs and their attitudes to loss in turn biases their perception of risk due to particular events;
- Personal context: the relative vulnerability of the public to consequences of an event amplifies its perceived risk, while its importance and anticipated benefits diminish its perceived risk.
- Nature of communication: when descriptive communication of adverse events emphasizes consequences at the expense of benefits, it increases their perceived risk and vice versa.
- Long-term versus short-term exposure: long term exposure is considered to be more serious than short-term exposure because the risk has to be lived with all the time. Hence risks associated with the former are perceived to be greater.
- Immediacy of consequences: there is greater aversion to events with consequences that are immediate than to those that manifest themselves at a later time.
- Availability of alternatives: risk associated with situations for which there are no practical alternatives is perceived more favorably than where alternatives exist.
- Reversibility of consequences: irreversibility of consequences tends to amplify the perceived risk and vice-versa.
- Whether exposure is essential: when exposure is necessary, such as in medical treatment, the risk is perceived more favorably than when the exposure is a luxury.
- Certainty with which risk is known: the natural aversion towards events with risks that are not well understood amplifies their perceived risk.

It is evident from the foregoing discussion that the questions: "Who is perceiving the risk?" and "Why they are perceiving it?" are very important. Therefore, analyses of risk due to uncertainty in spatial data should consider and specify the parties affected by the consequences of the uncertainty and accordingly account for their perception of risk. Such specification will determine the scope for use of the risk information, that is, whether it is relevant only to the user or to the broader community as well.

5. Quantitative Evaluation of Risk

Typically, an adverse event has multiple scenarios each with a corresponding likelihood and loss. Therefore a risk analysis entails asking the following three questions:

- What can happen or what can go wrong? (defining the scenario)
- How likely is it to happen? (estimating the likelihood)
- If it does happen, what are the consequences or losses? (estimating the impact)

Unfortunately, there are no neat mathematical relationships to assist risk analysis that can satisfactorily define the relationships between all possible scenarios, their likelihoods and their consequences. Instead, the typical approach is to arrange the different scenarios in order of increasing severity and then plotting the consequences against their cumulative likelihood. Smoothing the resulting staircase function produces a risk curve (Figure 3), which is the preferred portrayal for many types of risk such as those due to natural hazards, pollution and engineering structure failures (Starr 1965, McDonald 1995).

As noted earlier, the definition of risk as likelihood times consequences enables its quantification, which in turn aids comparison. However, the aggregation of likelihood and consequences into a single metric leads to loss of information and has been described by Rasmussen (1981) and Schneider (1987) as simplistic. The main complaint with this characterization of risk is that it equates low-likelihood and high-consequence scenarios with high-probability and low-consequence ones. To overcome this issue, alternative methods may be used which involve placing different weights on the consequences according to their unacceptability in the context of the decision to be made. Nonetheless, the simple multiplication of likelihood by consequences is effective in many cases, and this value in turn is multiplied by the degree of utility of the data used in the decision (with the utility value varying between zero and one).

Of course, the estimation of risk itself is subject to uncertainty for the following reasons:

- simplicity in the definition of risk;
- inevitability of subjective judgements in the analysis of risk;

- uncertainty in estimates of likelihood, consequences and any other variables (such as value of life) that may be used in estimating risk;
- poor identification of scenarios and their ranges; and
- the inability to exhaust all possible scenarios, including the inclination to focus on scenarios that are amenable to quantification at the expense of those that are not, which leads to underestimation of risk.

The main problem with uncertainty in risk estimates is not that it exists, but rather whether its magnitude is known and reported. Okrent (1982) warns that it is essential for risk analyses to provide statements of assumptions made in arriving at values used in the analysis and to point out any known uncertainties in the risk estimates. Evidently the same reasons used to argue for reporting uncertainty in spatial data under the concept of "truth in labeling", also apply to reporting uncertainty associated with risk estimates. It is clear that satisfactorily accounting for all sources and quantifying the uncertainty in risk is an elusive goal. In order to estimate the uncertainty associated with risk due to uncertainty in spatial data, meta-uncertainty of the data as well as the uncertainty associated with estimates of consequence scenarios are necessary. Unfortunately, this data is often not available.

Having identified the different adverse event scenarios, a risk analysis proceeds to estimate their likelihoods and consequences. Analysis of risk due to uncertainty in spatial data requires two sets of information:

- a report on uncertainty in the spatial data (data quality report); and
- the relationship between the magnitude of uncertainty and consequences, which is used to estimate the consequences of each scenario.

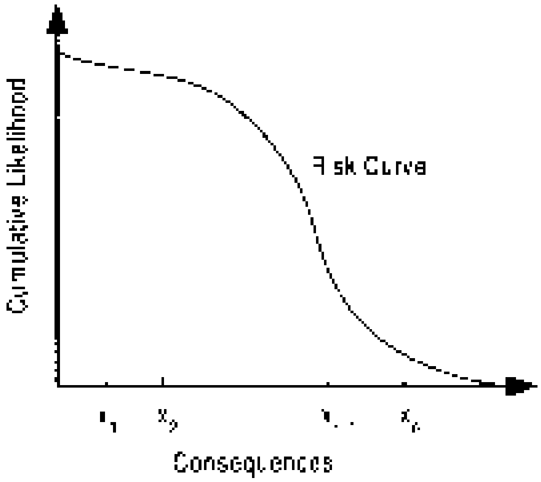
This shows that likelihood is the risk parameter directly associated with uncertainty in data, and that a typical data quality report provides only part of the information necessary for risk analysis. Probability, in the widest sense of the term, is invariably used as the measure of likelihood when estimating risk. Accordingly, the preferred representation of uncertainty is in probabilistic terms. However, not all types of spatial data uncertainty are modeled on probability theory. (Shi 1995) and Stoms (1987) categorize uncertainty in spatial data into three classes according to the theories suitable for modeling them. They include:

- uncertainties due to randomness or variability of error;
- incompleteness of evidence, such as when sampling has been applied or surrogate variables employed; and
- vagueness, which may result due to imprecision in taxonomic definitions.

Theories appropriate for modeling these uncertainties include: probability theory, Dempster-Shafer's theory of evidence and fuzzy set theory. Probability theory, which has a rigorous foundation, enjoys a long tradition dating as far back as the 17th century whereas Dempster-Shafer's theory and fuzzy set theory are relatively recent (introduced in 1967 and 1965 respectively). The Dempster-Shafer theory, or the mathematical theory of evidence, recognizes the existence of ambiguity or ignorance due to incomplete information. The absence of evidence to support a hypothesis is not assumed to constitute evidence against it. This means that what is known and what is not known are explicitly evaluated. The theory is based on complementary measures, namely *belief* and *plausibility*, and the interval between them¾*belief interval*. The scarcity of reports documenting application of the theory in GIS suggests that it has not yet attracted widespread use. However, some of the reported applications show that it has been applied in the classification of multi-spectral scanner data (Lee *et al.* 1987) and in a GIS application for finding optimal routes for military helicopters (Garvey 1987). In recognition of the importance of this theory in modeling uncertainty due to incomplete evidence (information), appropriate classifiers are now supported in recent versions of some software products such as IDRISI.

Fuzzy set theory was introduced by Zadeh (1965) to handle vague or imprecise concepts in a precise fashion that a computer can handle. It allows partial membership in a set and its basis lies in the assignment of a membership function which indicates the degree to which an observation, X, belongs to a set, A. The membership function may also be viewed as the measure of belief that X is an element of A, or as an index of the relative accuracy associated with assigning observation X to class A. Membership function values (membership grades) are real numbers ranging from zero to one, where values closer to one indicate full membership and those close to zero indicate non-membership. The vagueness modeled by fuzzy set theory pertains to definitions of classes (or sets) into which objects or phenomena are to be assigned. The boundaries of these classes are gradual such that membership grades of objects gradually transition from non-membership to full membership. On the other hand probability theory is based on classes that are crisp such that objects either wholly belong a class or they do not. Fuzzy sets are increasingly being used in remote sensing for image interpretation and classification, and the results have been shown to be useful in extracting more information than with conventional approaches based on crisp sets (Gopal and Woodcock 1994). They have also been applied in GIS to represent uncertainty and propagate it as data are transformed by various GIS functions (Veregin 1989).

Figure 3. The risk curve showing the cumulative likelihood for the consequences for each adverse event scenario as a result of data uncertainty.



6. The Risk Management Process

The overall risk management process involves a series of tasks (Figure 4), viz.: risk identification; risk analysis; risk appraisal; risk exposure; risk assessment and risk response. Risk identification is considered to be the most important step in the process on the grounds that "a risk identified is a risk controlled". It essentially involves determining what may go wrong and how it may happen. Risk analysis requires estimating the probabilities and expected consequences for the identified risks. The consequences of an adverse event will vary depending on the magnitude of the event and the vulnerability of the elements affected by the event. The outcome of risk analysis is risk exposure. Risk exposure is the total amount of risk exposed by the adverse event, and can be considered to be the summation of all the individual risks identified. Risk appraisal involves determining the magnitude of risk which is considered acceptable. It can be determined by analyzing and choosing the risk associated with the most favorable among the possible combinations of decision quality indicators, namely cost-benefits and risks. A complication of this step is the difficulty in quantifying the intangible benefits of a decision. Risk assessment requires comparing risk exposure with the results of risk appraisal. Depending on whether the risk exposure is acceptable or not, the decision maker must then consider taking an appropriate risk response. Risk response is the final stage in the process and also the ultimate objective of risk management—to help the decision maker make a prudent response in advance of a problem. The possible responses to risk exposure include: avoidance, retention, transfer, control, and insurance.

By way of example, we can illustrate how the risk management process might be used to assess the fitness for use of a particular geographic data set. Consider a manager responsible for emergency response in a region who needs to know what land parcels and residents would be threatened by a particular magnitude flood event. In order to prepare the emergency response plan, a DEM of the region is considered to be a key data set in the manager's decision-making processes, and may be assumed to have the highest utilization factor possible. The manager wishes to establish the fitness for use of the DEM before basing the emergency management plan upon it. The data quality component that the manager is particularly concerned about is the accuracy of elevations, and the data quality statement that accompanied the DEM indicates that for elevation values it has a Root Mean Square Error (RMSE) of 7 metres.

Step 1 – Risk Identification: The manager starts by identifying what could go wrong due to error in the data and how it may happen. The possible adverse event arising from uncertainty in the DEM is where locations declared clear of flood waters in fact become inundated due to elevations in the DEM being greater than their true values on the ground.

Step 2 – Risk Analysis: The manager then examines the likelihood or probability of the adverse event and its possible consequences. In this case, residents of properties that have been incorrectly designated as not subject to flooding will have been led to believe they are safe and may not take precautionary measures. When the area is flooded, the agency responsible for emergency response could find itself subject to compensation claims for negligence which may have caused damage to property, loss of income and personal injury or loss of life. The probability of the adverse event occurring (that is, the flooding of areas declared to be safe) is equivalent to the probability of the elevations in the DEM being greater than the nominated flood level — when in fact they are not. The consequences of the adverse event then need to be quantified in appropriate units, for instance dollars, injuries or lives lost. The magnitude of the consequences will depend on the vulnerability of properties and residents to the adverse event; the values of the properties; and the magnitude of the event (in this case, the depth of inundation of the areas that were erroneously declared free from flooding). All three parameters might vary in space such that geographical analysis may need to be applied to account for local variations when estimating the value of the consequences.

Step 3 - Risk Exposure: From the two profiles constructed at the end of the risk analysis stage (that is, probability vs magnitude and magnitude vs vulnerability), the user can compute estimates of the risk exposure for each magnitude of adverse event (that is, for different inundation depths). The exposure is the summation of identified risks and may be expressed as a combination of diverse units depending on the nature of the consequences — for example, dollars, injuries sustained and lives lost.

Step 4 - Risk Appraisal: The manager then establishes how much risk is tolerable. For example, this might be the maximum amount of money that can be set aside or for which the agency is insured to compensate victims for the its negligence. Tolerance limits may also be set for other consequences such as lives lost and harm to the reputation of the agency. The procedure for determining these limits is discussed further in Agumya and Hunter (1998).

Step 5 - Risk Assessment: This step compares outcomes from the risk exposure and risk appraisal stages. If the risk exposure is less than or equal to the tolerable risk appraisal, then the DEM is fit for use. However, comparing the risk exposure and tolerable risk can be complicated by the different units in which risk exposure and appraisal are expressed. Table 1 illustrates how this difficulty may manifest itself, and the manager may have trouble deciding whether the risk exposure is less than the tolerable risk. For example, use of the DEM may yield a risk exposure of a single loss of life and property damage worth a million dollars, compared to a tolerable risk appraisal of no deaths and up to two million dollars damage. However, criteria do exist for aggregating risk in various units into a single unit, such as dollars. For example, the value for a lost life might be based on an estimate of the amount of money that the deceased would have earned during an average life expectancy. At this point, there will clearly be many factors that will impact on what is and is not tolerable with respect to the degree of error associated with the declaration of flood affected land.

If the risk exposure exceeds the acceptable risk appraisal, then the manager is faced with three options regarding the fitness for use of the DEM:

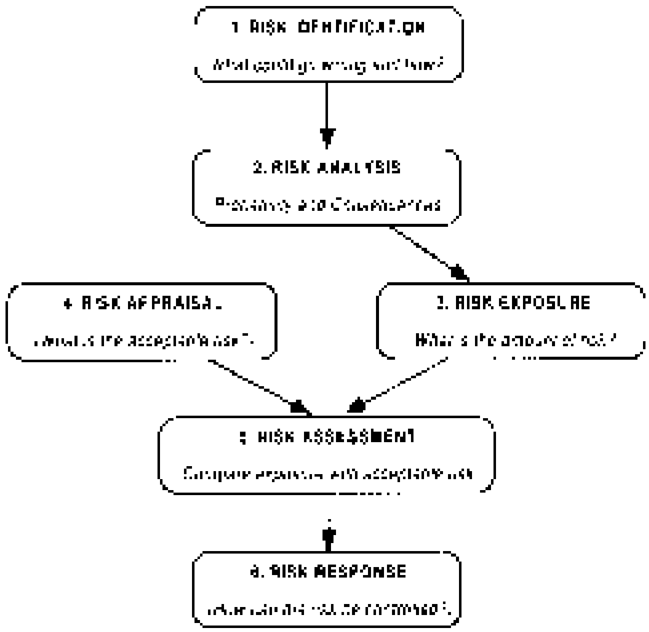
- reject it and secure another DEM of superior quality;
- retain the DEM but improve its quality; or
- use the DEM regardless of the consequences, since securing a superior data set or improving upon the quality of that available are not feasible options due to cost and/or time constraints.

At the end of the risk assessment stage the manager is aware of the uncertainty in terms of risk that the agency is being exposed to and will need to either reduce it or absorb (accept) it (Bedard, 1987). The manager may then use the knowledge gained about these risks to chose the best method of protection from their consequences.

Step 6 - Risk Response: Finally, a decision must be made regarding how to deal with the risk exposure. Again, there are several options available:

- **Risk Avoidance:** this is where the decision maker chooses to avoid the possibility of being exposed to risk. It is the 'do nothing' option, and in the DEM example above it is not available to the agency in question because risk avoidance would mean that the agency would have to cease carrying out its functions.
- **Risk Retention:** sometimes referred to as *risk assumption*. This is when the decision maker bears liability for either some or all of the consequences of an adverse event, either deliberately (active retention such as with a deductible in an insurance policy, or as in the case of some land title registries where the correctness of the information is guaranteed), or out of ignorance or indifference (passive retention). Retention is appropriate primarily for risks of low probability and relatively small consequences. In the example given, retention does not appear to be an appropriate choice because the consequences are likely to be relatively large.
- **Self Insurance:** another version of risk retention whereby parties faced with exposure to similar adverse events pool together their exposure in order to accurately predict the risks they will be exposed to. The difference between self insurance and other insurance operations is that the body underwriting the insurance is also part of the one it is insuring. The availability of this option depends on whether the agency has willing partners with whom to pool its exposure.
- **Risk transfer:** this is when the party exposed shifts the risk to another party, other than an insurance company. The transfer may be through a contract or a clause in an agreement or undertaking or through hedging (whereby the party exposed, the hedger, takes two simultaneous positions that offset each other so that no matter what the outcome of an adverse event, the hedger never loses or wins). If the agency in the example above were to respond to risk exposure by transfer, it could if possible secure legal liability protection, or transfer the risk to another government body. However, this may not be in the best interest of the reputation of the agency. Also the agency may be liable for such losses under statutory law.
- **Loss Control (prevention and reduction):** is concerned with reducing the probability of occurrence and consequences of an adverse before it occurs. In the example given, the probability of occurrence depends on the amount of uncertainty in the DEM. This uncertainty can be reduced where possible by purchasing a more accurate DEM or one with a higher resolution. The consequence is dependent on the vulnerability of the elements exposed, their value and the magnitude

Figure 4: *The risk management process.*



of the adverse event. The consequence can be reduced by taking precautionary measures in the most vulnerable areas such as where there is high population density, high property values and where property is in general vulnerable to damage by flooding. Loss prevention is only feasible as long as the benefits realized from the prevention are greater than the cost of the loss prevention program. Loss reduction is intended to diminish the consequences after the adverse event has occurred. Since for purposes of assessing fitness for use the user is particularly concerned with the risk posed by the adverse event before it occurs, this option is not relevant.

- **Insurance:** This represents a contractual transfer of risk from the party exposed, to an insurance company. Insurance is especially appropriate when the probability of occurrence of an adverse event is low and its consequences very high and is widely considered to be the most practical method of response to a major risk. Depending on the cost of the policy, it may be an attractive option for the user in this example (Epstein et al. 1998).

7. Limitations of The Approach and Further Research

Of course, the risk-based approach to quantifying and managing uncertainty in spatial data is not immune from problems and is itself subject to some degree of uncertainty. For instance, the assumptions employed may be open to challenge, and the perception and acceptance of the utility of risk management can be clouded when factors cannot be easily quantified in economic terms such as social issues, environmental damage and loss of life. Indeed, placing a value of the loss of life is one of the most vexing problems facing risk managers, and there are wide ranging views on how this should be performed. Furthermore, estimating the likelihood of events may be open to dispute, especially where it varies with location yet must be aggregated into simpler regional values. Risk analysis also requires special skills and can be an expensive to perform. Nevertheless, it is often noted that undertaking risk analysis can be extremely productive in itself and forces organizations to think deeply, often the first time, about the possible consequences of uncertainty in their data and its effect upon the decisions they make with it.

At this stage there are still several outstanding research issues that need to be resolved. The first is whether the proposed risk-based method is suitable for use with spatial data across a wide variety of applications. It may be that the approach is only viable for high-risk tasks such as in emergency response systems, in which case we need to identify (a) the threshold of its usefulness, (b) under what terms it becomes feasible to employ, and (c) whether it is suited for 'one-off' decisions that are rarely made. For other applications it should be determined whether the approach can be 'short circuited' in less important decision situations, yet still remain valid and resource effective.

8. Conclusion

In this paper an alternative approach to using standards for assessing the fitness for use of spatial data has been presented. The proposed risk-based method not only aids the assessment of fitness for use, but is also valuable to users for whom the primary concern about uncertainty in data is how to manage its impact. Risk management is widely used as a mechanism for dealing with consequences of uncertainty, however risk is a complex concept that must be radically simplified to make it quantifiable. Perception is an important characteristic of risk yet it is not easily measurable, and its subjectivity opens risk estimates to challenge. These factors combine to make risk estimates subjective, yet there is a tendency to believe that they are objective, especially considering that their determination often involves rigorous analysis. Nonetheless, risk has proven to be a very valuable tool for managing uncertainty in other disciplines and it is argued that further investigation of its potential for assessing the fitness for use of spatial data is now warranted.

Acknowledgements

The authors acknowledge funding support received under Australian Research Council Large Grant No. A49601183, "Modeling Uncertainty in Spatial Databases", and Austra-

Table 1. *The Problem of Comparing Risk Exposure and Appraisal*

Consequence	Exposure	Appraisal
Loss of lives	1	0
Injuries (severe)	0	5
Property damage	$1 million	$2 million

lian Research Council Small Grant No. S499692, "Development of Risk Management Techniques for Handling Uncertainty in Spatial Information".

References

Agumya, A. and Hunter, G.J., 1998, "Assessing Fitness for Use of Geographic Information: What Risk are we Prepared to Accept in our Decisions?". *Proceedings of the 3rd International Symposium on Spatial Data Accuracy Assessment in Natural Resources and Environmental Sciences*, Quebec City, Canada, 10 pp.

Beard, K., 1997, Representations of Data Quality. In *Geographic Information Research: Bridging the Atlantic*, Eds. Craglia, M. and Couclelis, H., (London: Taylor and Francis), pp. 280-294.

Bedard, Y., 1987, Uncertainties in Land Information Systems Databases. *Proceedings of the Auto-Carto 8 Conference*, Baltimore, Maryland, pp. 175-184.

Bohnenblust, H. and Schneider, T., 1987, Risk Appraisal: Can it be Improved by Formal Decision Models. In *Uncertainty in Risk Assessment, Risk Management and Decision Making*, Eds. Covello, V. T., Lave, L. B., Moghissi, A. and Uppuluri, V. R. R., (New York: Plenum Press), pp. 71-87.

Burrough, P. A., Rijn, R. v. and Rikken, M., 1996, Spatial Data Quality and Error Analysis. In *GIS and Environmental Modelling: Progress and Research Issues*, Eds. Goodchild, M. F., Stayaert, L. T. and Parks, B. O., (Fort Collins: GIS World), pp. 29-34.

Davis, G. B. and Olson, M. H., 1985, *Management Information Systems: Conceptual Foundations, Structure and Development*, (New York: McGrawHill).

Dobran, F., 1995, A Risk Assessment Methodology at Vesuvius based on the Global Volcanic Simulation. In *Natural Risk and Civil Protection*, Eds. Horlick-Jones, T., Amendola, A. and Casale, R., (London: E & FN Spon), pp. 131-136.

Elms, D. G., 1992, Risk Assessment. In *Engineering Safety*, Ed. Blockley, D. I., (Berkshire: McGraw-Hill), pp. 28-46.

Epstein, E., Hunter, G.J. and Agumya, A., 1998, "Liability Insurance and the Use of Geographic Information". *International Journal of Geographical Information Science*, 12, 3, pp. 203-214.

Fischhoff, B., Hope, C. and Watson, S., 1984, Defining Risk. *Policy Sciences*, 17, pp. 123-139.

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L. and Keeney, R. L., 1981, *Acceptable Risk*, (Cambridge: Cambridge University Press).

Frank, A. U., 1998, Metamodels for Data Quality Description. In *Data Quality in Geographic Information: From Error to Uncertainty*, Eds. Goodchild, M. F. and Jeansoulin, R., (Paris: Hermes), pp. 15-29.

Garvey, T. D., 1987, Evidential Reasoning for Geographical Evaluation for Helicopter Route Planning. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25 (3), pp. 294-304.

Goodchild, M. F., 1992, *Final Report of Research Initiative 1: Accuracy of Spatial Databases*. (Santa Barbara: National Center for Geographic Information and Analysis).

Goodchild, M. F., 1998, Uncertainty: The Achilles Heel of GIS? In *Geo Info Systems*, November 1998, pp. 50-52.

Gopal, S. and Woodcock, C., 1994, Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets. *Photogrammetric Engineering and Remote Sensing*, 60 (2), pp. 181-188.

Gratt, L. B., 1987, Risk Analysis or Risk Assessment: A Proposal for Consistent Definitions. In *Uncertainty in Risk Assessment, Risk Management and Decision Making*, Eds. Covello, V. T., Lave, L. B., Moghissi, A. and Uppuluri, V. R. R., (New York: Plenum Press), pp. 241-249.

Guptill, S. C. and Morrison, J. L. (Eds., 1995, *Elements of Spatial Data Quality*, (Oxford: Elsevier Science).

Haimes, Y. Y., 1989, Toward a Holistic Approach to Risk Assessment and Management. *Risk Analysis*, 9 (2), pp. 147-149.

Kaplan, S. and Garrick, J. B., 1981, On the Quantitative Definition of Risk. *Risk Analysis*, 1 (1), pp. 11-27.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X. and Ratick, S., 1988, The Social Amplification of Risk: A Conceptual Framework. *Risk Analysis*, 8 (2), pp. 177-187.

Knight, F., 1921, *Risk, Uncertainty, and Profit*, (Boston: Houghton Mifflin).

Laws, D., Gross, M. and Fabos, J., 1989, Information Resources and Public Decision Making. In *Proceedings of the Annual Conference of the Urban and Regional Information Systems Association*, Boston MA, pp. 160-174.

Lee, N. S., Grize, Y. L. and Dehnad, K., 1987, Probabilistic and Evidential Approaches for Multisource Data Analysis. *IEEE Transactions on Goescience and Remote Sensing*, GE-25 (3), pp. 283-293.

Lowrance, W. W., 1976, *Of Acceptable Risk: Science and the Determination of Safety*, (Los Altos CA: William Kaufmann).

McDonald, L. A., 1995, ANCOLD Risk Assessment Guidelines. In *Acceptable Risks for Major Infrastructure: Proceedings of the Seminar on Acceptable Risks for Extreme Events in the Planning and Design of Major Infrastructure*, Eds. Heinrichs, P. and Fell, R., (Rotterdam: A.A. Balkema), pp. 105-121.

Okrent, D., 1982, Comment on Societal Risk. In *Risk in the Technological Society*, Eds. Hohenemser, C. and Kasperson, J. X., (Boulder, CO: Westview Press), pp. 203-215.

Pidgeon, N., Hood, C., Jones, D., Turner, B. and Gibson, R., 1992, Risk Perception. In *Risk: Analysis, Perception and Management*, (London: The Royal Society), pp. 89-134.

Raftery, J., 1994, *Risk Analysis in Project Management*, (London: E & FN Spon).

Rasmussen, N., 1981, The Application of Probabilistic Risk Assessment Techniques to Energy Technologies. *Annual Review of Energy*, 6, pp. 123-138.

Reid, S. G., 1992, Acceptable Risk. In *Engineering Safety*, Ed. Blockley, D. I., (Berkshire: McGraw-Hill), pp. 138-166.

Rowe, W. D., 1977, *An Anatomy of Risk*, (New York: Wiley), 488 pp.

Sage, A. P., 1995, Systems Engineering for Risk Management. In *Computer Supported Risk Management*, Eds. Beroggi, E. G. and Wallace, W. A., (Rotterdam: Kluwer), pp. 3-31.

Schneider, S. H., 1987, Future Climatic Change and Energy System Planning: Are Risk Assessment Methods Applicable? In *Risk Analysis and Management of Natural and Man-made Hazards*, Eds. Haimes, Y. Y. and Stakhiv, E. Z., (New York: American Society of Civil Engineers), pp. 201-221.

Shi, W., 1995, Towards a Generic Theory for Handling Uncertainties in Spatial Data. In *GIS AM/FM Asia'95*, Bangkok Thailand, pp. I.3.1-1.3.9.

Smith, J. L., Prisley, S. and Weih, R. C., 1991, Considering the Effect of Spatial Data Variability on the Outcomes of Forest Management Decisions. In *Proceedings of GIS/LIS '91 Conference*, Atlanta, Georgia, pp. 286-292.

Starr, C., 1965, Social Benefit versus Technological Risk. *Science*, 165, pp. 1232-1238.

Starr, C., 1987, Risk Management, Assessment, and Acceptability. In *Uncertainty in Risk Assessment, Risk Management and Decision Making*, Eds. Covello, V. T., Lave, L. B., Moghissi, A. and Uppuluri, V. R. R., (New York: Plenum Press), pp. 63-70.

Stoms, D., 1987, Reasoning with Uncertainty in Intelligent Geographic Information Systems. In *Proceedings of the Second Annual International Conference, Exhibits and Workshops on Geographic Information Systems (GIS '87)*, San Francisco, pp. 693-700.

The Conservation Foundation, 1985, *Risk Assessment and Risk Control*, (Washington DC: Conservation Foundation).

Thomson, J. R., 1987, *Engineering Safety Assessment: An Introduction*, (New York: Wiley).

Veregin, H., 1989, Error Modelling for the Map Overlay Operation. In *Accuracy of Spatial Databases*, Eds. Goodchild, M. and Gopal, S., (London: Taylor and Francis), pp. 3-18.

Veregin, H. and Hargitai, P., 1995, An Evaluation Matrix for Geographical Data Quality. In *Elements of Data Quality*, Eds. Guptill, S. C. and Morrison, J. L., (Oxford, Elsevier Science), pp. 167-188.

Williams, T., 1995, A Classified Bibliography of Recent Research Relating to Project Risk Management. *European Journal of Operational Research*, 85, pp. 18-38.

Zadeh, L., 1965, Fuzzy Sets: Information and Control. *Information and Control*, 8, pp. 338-353.

Zwart, P., 1991, Some Indicators to Measure the Impact of Land Information Systems in Decision Making. *Proceedings of the URISA '91 Conference*, San Francisco, vol. 4, pp. 77-89.