Journal of Environmental Management (1997) **49**, 73–93

Using Dempster–Shafer Theory to Represent Climate Change Uncertainties

Wuben Ben Luo* and Bill Caselton†

**Hydrotechnical Division, BC Hydro and Power Authority, Burnaby, B.C., Canada V3N 4X8 and †Department of Civil Engineering, University of British Columbia, Vancouver, B.C., Canada V6T 1Z4*

When the prospect of climate change is admitted into the design and operation of long-lifetime water resource projects then the already high level of uncertainty is further increased. Currently, the longer-term climate is discussed in terms of a set of climate change scenarios. As a result of severe weakness in the knowledge and data pertaining to climate change, the relative probabilities of these scenarios are rarely indicated. This predominantly possibilistic rather than probabilistic view of climate change falls short of the needs of the established approaches to quantitative resource management.

Weak information and its associated uncertainties are easily distorted when quantified and this can lead to decision analysis results that are misleading. A promising scheme for dealing with such information, based on Dempster–Shafer (D–S) theory, is described in this paper. Through the use of probabilities that are assigned to freely defined intervals of values, D–S theory offers greater flexibility than the Bayesian approach when quantifying weak information and more faithfully reflects its consequences in the results of decision analysis. At the same time the Bayesian approach and D–S approach share many fundamental ideas and produce identical results when the uncertainties are less extreme.

This paper presents, along with some elementary examples, aspects of the D–S approach that contribute to its appeal when dealing with weak subjective and data-based information sources that have a bearing on climate change. The topics discussed include: the Basic Probability Assignment (BPA) and its visualization; combining BPAs from different sources; representing ignorance and near-ignorance; capturing subjective knowledge; inference; and decision analysis.

Keywords: Dempster–Shafer theory, Bayesian analysis, climate change, water resources, decision analysis.

1. Introduction

Water resource management decisions are routinely made in the face of substantial uncertainties, many hydrologic in origin and therefore climate related. When the prospect of climate change is admitted into the design and operation of long-lifetime water resource projects then the already high level of uncertainty is substantially

0301–4797/97/010073 + 21 $25.00/0/ev960117

increased. The increased uncertainty arises because more complex, non-stationary, climatic and hydrologic models must now be specified, yet the consequent increased demands for information must be met by essentially the same data base. Clearly, climate change cannot be ignored. Simply reverting back to our prior stationary view of climate and the perception of less uncertainty constitutes self-delusion. Now that the climate change genie is out of the bottle, its uncertainties have to be addressed. Equally important for professionals engaged in these activities, climate change and its associated uncertainty will have to be addressed in terms that are sufficiently quantitative to be incorporated into a formal decision analysis.

Currently the longer-term climate trend is discussed in terms of a set of climate change scenarios tied to CO_2 levels but the relative probabilities of these scenarios are rarely indicated. This reflects a purely possibilistic rather than probabilistic view of climate change and one that falls short of the needs of present-day quantitative approaches to resource management. It arises from having substantial amounts of information on global climatic processes but a scarcity of information concerning likelihoods of the causative factors. While in the long term there will undoubtedly be improvement in this situation it is quite possible that progress will be slow and may even regress for a time as research reveals new complexities and uncertainties in factors governing long-term climate change. For the present, though, we will have to use the very limited amounts of information on climate and hydrologic change as fully as possible.

We have categorized the problem of long-term water resource decision-making under climatic change as decision-making under "near-ignorance" conditions (Caselton and Luo, 1994). "Near-ignorance" is not used in a deprecatory way but to emphasize that the total information brought to bear on the decision problem is sufficiently weak to present difficulties with conventional uncertainty analyses. We find that concerns about expressing weak information in the form of conventional probability distributions have appeared in the statistical literature for well over 100 years. A paper by Boole (1854) is an early example. Weak information is fragile and when quantified is easily distorted. This distortion may lead to misleading conclusions and be misleading about the degree of uncertainty attached to these conclusions.

The search for better methods for dealing with uncertainty has intensified over the past 50 years and a variety of new approaches have been proposed. Some, such as Bayesian Robustness (Berger, 1984), involve only minor variations on the Bayesian scheme, while more radical alternatives such as Fuzzy Logic represent complete departures from statistical methods. Recently, the term "imprecise" probability has been used to describe any scheme that involves less than precise specifications of probabilities to reflect the uncertainties (Walley, 1991). The scheme we describe in this paper derives from work by A. P. Dempster (Dempster, 1967, 1968a,b, 1969) that was subsequently extended by Glen Shafer (Shafer, 1976, 1972) to form Dempster–Shafer (D–S) theory.

Undoubtedly, our choice of D–S theory was influenced by the fact that it was probability based and already enjoyed some prominence in the applied field of artificial intelligence. We eventually recognized that the needs there were somewhat different from those in engineering decision-making and returned to Dempster's earlier work for the necessary theory. Our interest in D–S theory has been sustained by the richness of its uncertainty representation scheme and its ability to reflect, through the use of intervals of values, the consequences of weak information on the results of a decision analysis.

Our purpose in this paper is not to deny the benefits of the Bayesian approach in general. We take considerable reassurance from the fact that the Bayesian approach and D–S approach share many concepts that are central to both and Bayesian analysis even arises as a special case within the D–S scheme. In the case of inference from sample data, the results from both D–S and Bayesian schemes are essentially identical down to just modest sample sizes. It is the kind of disagreement that occurs between these two schemes with small to non-existent samples, a situation that is analogous to the weak information and near-ignorance situations of climate change, that interests us most of all. But, ironically, the greater flexibility of the D–S scheme when dealing with near-ignorance conditions comes at the expense of greater dimensionality and consequently greater complexity than the Bayesian approach. Whether this added complexity is justified remains to be seen but, at the very least, the D–S approach does provide a valuable new perspective on the Bayesian approach and on near-ignorance.

The paper by Caselton and Luo (1992) was intended to bring the D–S alternative to the attention of decision-makers and decision analysts in water resource management when dealing with rare design events and longer-term perspectives. The greater weakness of the data supporting decisions in this context when climate change is considered means that there will be more dependence on any subjective and experiential information that bears on the problem. In this paper we present aspects of D–S theory that we believe increase the opportunities for accurately capturing information from both weak data and weak subjective sources.

2. Dempster–Shafer (D–S) theory

2.1. THE BASIC PROBABILITY ASSIGNMENT (BPA)

Both D–S and Bayesian theories adopt similar perspectives by focusing on the uncertainty concerning the value of a quantity that is generically labeled "a state of nature". In the context of water resource management and climate change an example of a state of nature might be the global surface warming by the year 2050. Both D–S and Bayesian theories represent information or knowledge (henceforth we will use these words interchangeably) concerning this state of nature by assigning probabilities. The interpretation of probability is, from an applications standpoint, essentially the same in both theories. For the more statistically inclined, Wasserman (1988) provides us with some technical assurance on that score. It follows that the probability assignments in either approach must always sum to 1·0.

However, the two theories differ quite fundamentally in the allowable subjects of the probability assignments. With a state of nature that can take on only discrete values, the Bayesian probability mass assignments must be made to individual discrete values only, while the D–S assignments may also be made to unions of discrete values. Furthermore, these subsets are allowed to have elements in common; that is, they may overlap. In the continuous state of nature case, the D–S assignments are of probability density and may be made to intervals of values as well as to point values. The intervals of values may also overlap. In both discrete and continuous cases, the D–S assignment of probability to a subset or interval says nothing about the way that probability is allocated to the values within the subset or interval. It is this freedom to choose the intervals and thereby vary the resolution of the probability assignments that gives the D–S scheme its flexible information representation capabilities.

In order to state these probability assignments more formally let Θ represent a set of n elements, ordered by magnitude, each element representing a possible discrete value θ_i of the state of nature θ, so that

$$\Theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_n\}$$

Henceforth, the words "element" and "value", implying a possible numerical value for a measurable state of nature, are used interchangeably in discussing the elements of Θ.

The D–S scheme will allow us to assign probabilities to the unions of subsets of values in Θ, such as $\{\theta_1, \theta_2\}$, $\{\theta_2, \theta_3, \theta_4\}$, etc., to Θ itself, and to the individual values of θ_i. When we refer to subsets of possible values their union will always be implied. Thus, for example,

$$\{\theta_2, \theta_3, \theta_4\} \quad \text{always implies} \quad \theta_2 \cup \theta_3 \cup \theta_4$$

In applications of decision analysis in resource management the state of nature will often be a physical quantity that is measurable on a continuum. We therefore confine our interest to subsets of values that, although possibly discretized for convenience, still represent measurable values. Information concerning a subset like $\{\theta_2, \theta_4, \theta_5\}$ that skips the intermediate value θ_3 is considered to be unlikely in our context so that discontinuous subsets or intervals like this are excluded entirely in the rest of this paper.

Using A to denote any one of the subsets of Θ, a probability assigned to A is an expression of support that the truth lies within A. Significantly, this says nothing about the way that this probability is apportioned amongst the values within A. In addition, it does not preclude other expressions of support that are directed to subsets of A, including to its individual values, or to other subsets that have some values in common with A. Probability assignments within this new framework are denoted as $m(\ )$ to distinguish them from conventional probability assignments $p(\ )$ that are always entirely confined to individual values. The specification of the $m(\ )$ for all of the subsets of Θ, including Θ itself, is called a Basic Probability Assignment or BPA. This collection of all the possible subsets and elements of Θ, but excluding discontinuous subsets in our case, is called the "frame" of Θ while the individual values θ_i are often called the "singleton" values of the frame of Θ.

A conventional Bayesian probability assignment or distribution is a special case of a BPA where the non-zero probability values are assigned solely to the singleton values. This is called a "Bayesian" BPA.

2.2. Visualization of a BPA

A secondary benefit resulting from avoiding discontinuous subsets of Θ is a very simple diagrammatic representation of a BPA. This is analogous to a conventional probability mass or density function plot but requires a third dimension to allow the various subsets or intervals to be represented.

Figure 1(a) is an example of a BPA for a case where the state of nature is discretized into four possible ordered values a, b, c, d. The coordinate values of each box, measured on the horizontal and vertical axes, identify the smallest and largest values, respectively, of the subset represented while its numerical entry is the probability mass assignment for that subset. For example, the box in the upper left-hand corner (coordinates: column "a"; row "d") represents the union of the entire set of values $\{a, b, c, d\}$ while boxes

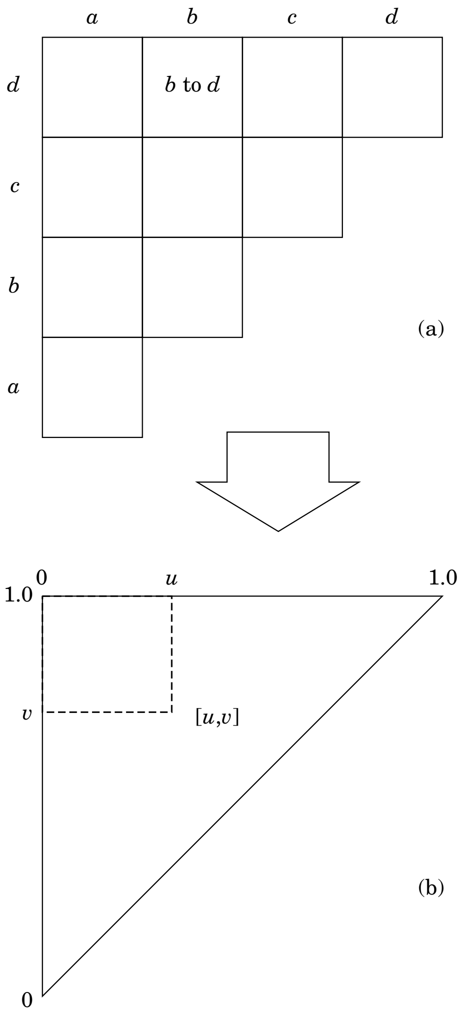


Figure 1. Representation of a BPA (a) discrete case, (b) continuous case.

on the hypotenuse have the same coordinate values on both axes and represent the singleton values. Probabilities will always sum to 1·0. When represented on this diagram, the probability assignments of a Bayesian BPA appear as non-zero entries in boxes along the hypotenuse of the triangle.

Figure 1(b) is the extension of Figure 1(a) to a continuous variable that, in this particular example, lies in the interval [0, 1·0] but in general could be any interval or even unbounded, i.e. interval $[+\infty, -\infty]$. The continuous BPA is now prescribed in the form of a continuous density surface over the triangular area. The height of this surface at any point $[u, v]$ represents the probability density assigned to the interval bounded by u and v. Densities on the hypotenuse of the triangular area represent assignments to point values. Again, Bayesian BPA assignments greater than zero occur only on this hypotenuse. The integrated density over the triangular region will be 1·0 for continuous BPAs.

TABLE 1. BPA_I derived from information source I

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	0·10	0·30	0·60
0 to +2°C	0·00	0·00	
−2 to 0°C	0·00		

2.2.1. *Example 1*

The representational flexibility and some key aspects of the D–S theory are demonstrated in the following example. It concerns information on global surface warming by the year 2050. There are two independent sources of information that involve quite different subjective inputs. Source I is entirely subjective while source II information is the subjective interpretation of a computerized literature search. Both sources of information agree that the true value will not fall outside the range −2 to +4°C. For simplicity, this range will be discretized. The true global warming value may fall into one of the following three intervals: −2 to 0°C; 0 to +2°C; and +2 to +4°C.

Information source I expresses with some confidence that the interval +2 to +4°C is most likely and subjectively assigns to it a probability of 0·60. This particular probability assignment is identified as m_I(+2 to +4°C). A probability of 0·3 is assigned to the possibility that global warming will fall in the interval 0 to +4°C (i.e. into the union of the two discretized intervals (0 to +2°C) and (+2 to +4°C)), while a probability of 0·1 supports the idea that the temperature change lies somewhere in the range −2 to +4°C. The resulting BPA is shown in Table 1, and follows the framework suggested in Figure 1(a).

Information source II is based on a survey of the literature on global warming, simply noting any statements of global temperature change, whether in the form of a prediction or otherwise. The relative frequencies observed were: 20/100 for the interval −2 to 0°C; 50/100 for interval +0 to +2°C; and 30/100 for interval +2 to +4°C. Recognizing the rather unscientific nature of the survey but attaching some significance to the fact that no values were found outside the entire −2 to +4°C range, a probability of 0·5 was assigned to the union of the three intervals. The relative frequencies were still thought to be significant and the remaining 0·5 probability was therefore assigned to the singleton intervals on the basis of the relative frequencies (this type of assignment is discussed later under "contaminated" priors). The BPA in Table 2 presents this set of probability assignments identified, individually, as $m_{II}(\ )$.

In the following section we will outline the D–S procedure for combining the BPAs from two sources of information into a single resultant BPA.

TABLE 2. BPA_{II} derived from information source II

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	0·50	0·00	0·15
0 to +2°C	0·00	0·25	
−2 to 0°C	0·10		

2.3. COMBINING BPAs

If two BPAs on the same state of nature are obtained from two different and independent sources of information, they can be combined to yield a new Resultant BPA. The combination procedure is performed with Dempster's rule of combination (Shafer, 1976) that plays an analogous role to Bayes' equation in the Bayesian scheme. The Resultant BPA is analogous to the Bayesian posterior distribution.

Given two BPAs, labeled I and II with probability assignments within these two BPAs identified as $m_{\mathrm{I}}(\ )$ and $m_{\mathrm{II}}(\ )$, respectively, then the combination rule specifies the probability assignments $m_{\mathrm{III}}(\ )$ in the resultant $\mathrm{BPA}_{\mathrm{III}}$ as follows.

Let A represent any subset in $\mathrm{BPA}_{\mathrm{I}}$ and let B represent any subset in $\mathrm{BPA}_{\mathrm{II}}$. Note that singletons may also constitute subsets. As these two BPAs convey information about the same state of nature it follows that A and B will always be subsets of the same Θ. The probability assignment to a subset C or element in the resultant BPA is the sum of the products of the probability assignments of all of the A and B whose intersection corresponds with C, so that

$$m_{\mathrm{III}}(C)=(1-k)^{-1}\sum_{A\cap B=C} m_{\mathrm{I}}(A)m_{\mathrm{II}}(B)$$

The value of k is determined by summing the products of the probability assignments for all cases where the subsets A and B do not intersect (i.e. their intersection is the null or empty set, represented by $\varnothing$). The factor $(1-k)^{-1}$ in the above equation, in effect, compensates for the loss of non-zero probability assignments to non-intersecting subsets in BPAs I and II, and ensures that the probability assignments of the resultant BPA also sum to 1·0. Thus, k is given by

$$k=\sum_{A\cap B=\varnothing} m_{\mathrm{I}}(A)m_{\mathrm{II}}(B)$$

Note that D–S theory represents both sample-based information and subjective prior knowledge in exactly the same fashion and they are treated in identical fashion in the combination process. A criticism leveled at the Bayes approach is that it discriminates between these two information sources, representing the sample-based information with a likelihood function and the prior knowledge as a Bayesian distribution.

2.3.1. *Example 1 (continued)*

The two BPAs presented at the introduction to Example 1 are to be combined to obtain a single resultant BPA, namely $\mathrm{BPA}_{\mathrm{III}}$.

First, consider the value of the normalizing factor. Inspection of $\mathrm{BPA}_{\mathrm{I}}$ and $\mathrm{BPA}_{\mathrm{II}}$ in Tables 1 and 2 reveals that the pairs of subsets in these two BPAs that do not intersect (those involving zero-valued assignments can be ignored), and their probability assignments, are

$$m_{\mathrm{I}}(+2 \text{ to } +4°\mathrm{C})=0{\cdot}60 \text{ and } m_{\mathrm{II}}(-2 \text{ to } 0°\mathrm{C})=0{\cdot}10$$

$$m_{\mathrm{I}}(+2 \text{ to } +4°\mathrm{C})=0{\cdot}60 \text{ and } m_{\mathrm{II}}(0 \text{ to } +2°\mathrm{C})=0{\cdot}25$$

$$m_{\mathrm{I}}(0 \text{ to } +2°\mathrm{C}\cup+2 \text{ to } +4°\mathrm{C})=0{\cdot}30 \text{ and } m_{\mathrm{II}}(-2 \text{ to } 0°\mathrm{C})=0{\cdot}10$$

TABLE 3. BPA_{III}, the result of combining BPA_{I} and BPA_{II}

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	0·066	0·197	0·592
0 to +2°C	0·000	0·132	
−2 to 0°C	0·013		

Hence, the value of k, the sum of these products, is $0{\cdot}60\times0{\cdot}10+0{\cdot}60\times0{\cdot}25+0{\cdot}30\times0{\cdot}10=0{\cdot}24$ and the value of the normalizing factor $(1-k)^{-1}$ is 1·316.

Sample calculations of the combination procedure leading to just two of the non-zero probability assignments in BPA_{III} are provided below:

$$
\begin{aligned}
&m_{III}(0 \text{ to } +2°C \cup +2 \text{ to } +4°C)\\
&\quad=(1-k)^{-1}\{m_{I}(0 \text{ to } +2°C \cup +2 \text{ to } +4°C)\\
&\qquad\times m_{II}(-2 \text{ to } 0°C \cup 0 \text{ to } +2°C \cup +2 \text{ to } +4°C)\}\\
&\quad=1{\cdot}316\{0{\cdot}30\times0{\cdot}50\}\\
&\quad=0{\cdot}197
\end{aligned}
$$

$$
\begin{aligned}
&m_{III}(+2 \text{ to } +4°C)\\
&\quad=(1-k)^{-1}\{m_{I}(0 \text{ to } +2°C \cup +2 \text{ to } +4°C)\times m_{II}(+2 \text{ to } +4°C)\\
&\qquad+m_{I}(-2 \text{ to } 0°C \cup 0 \text{ to } +2°C \cup +2 \text{ to } +4°C)\times m_{II}(+2 \text{ to } +4°C)\\
&\qquad+m_{I}(+2 \text{ to } +4°C)\times m_{II}(+2 \text{ to } +4°C)\\
&\qquad+m_{I}(+2 \text{ to } +4°C)\times m_{II}(-2 \text{ to } 0°C \cup 0 \text{ to } +2°C \cup +2 \text{ to } +4°C)\}\\
&\quad=1{\cdot}316\{0{\cdot}30\times0{\cdot}15+0{\cdot}10\times0{\cdot}15+0{\cdot}60\times0{\cdot}15+0{\cdot}60\times0{\cdot}50\}\\
&\quad=0{\cdot}592
\end{aligned}
$$

The complete resultant BPA, BPA_{III}, for this example is given in Table 3.

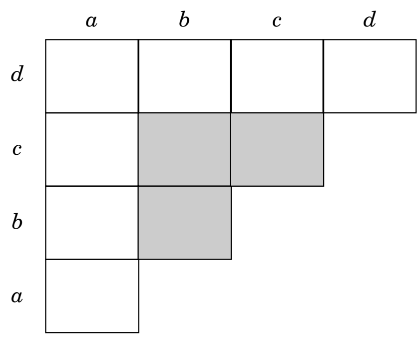
While tedious to formalize and process by hand, this combination procedure is relatively easy to automate in a computer program or spreadsheet.

2.4. BELIEF AND PLAUSIBILITY

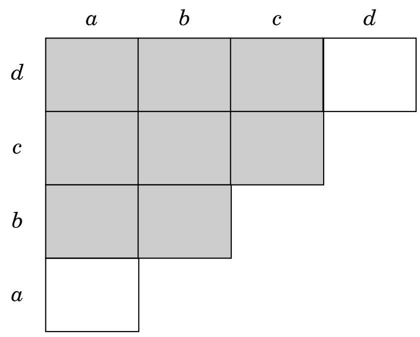
Once a BPA has been specified then additional quantities of both conceptual and computational importance can be obtained. The *belief* on subset A, $Bel(A)$, is obtained by gathering all of the probability values that have been assigned either directly to A, or to subsets of A, which therefore also imply A. Thus,

$$Bel(A)=\sum_{B\subset A} m(B) \quad A\subset\Theta$$

The *plausibility* of subset A, $Pl(A)$, is obtained by gathering all of the probability values on the subsets of Θ which do not exclude A. Thus,



$Bel(b,c) = m(b) + m(c) + m(b,c)$



$Pl(b,c) = m(b) + m(c) + m(a,b) + m(b,c) + m(c,d) + m(a,c) + m(b,d) + m(a,d)$

Figure 2. Computation of belief and plausibility for [*b*, *c*].

$$Pl(A) = \sum_{A \cap B = \varnothing} m(B) \quad A \subset \Theta$$

Figure 2 demonstrates, for the four-value discrete case presented in Figure 1(a), the locations (shaded) of the *m*( ) values that would be included in the above summations when *Bel*(*b*, *c*) and *Pl*(*b*, *c*) are computed.

In effect, the probabilistic nature of the assignment of probabilities to intervals permitted in a BPA allows two extreme interpretations. *Bel*(*A*) represents the minimum level of probability committed to the idea that the truth lies within *A*. On the other hand, *Pl*(*A*) represents the maximum probability that can be gathered for *A* when it is assumed to be the only possibility within any unions that include *A*. Therefore, one interpretation of the belief *Bel*(*A*) and plausibility *Pl*(*A*) is that they represent the lower and upper bounds of the conventional probability *p*(*A*) that *A* includes the true value of the state of nature (Yager, 1987). That is,

$$Bel(A) \leqslant p(A) \leqslant Pl(A) \quad A \subset \Theta$$

A BPA numerically specifies the above condition for every *A*. Of particular interest are the *A*s that correspond to the singleton values of the state of nature, i.e. for $A =$

TABLE 4. Beliefs *Bel*( ) derived from example BPA_{III}

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	1·00	0·60	0·60
0 to +2°C	0·15	0·13	
−2 to 0°C	0·01		

TABLE 5. Plausibilities *Pl*( ) derived from example BPA_{III}

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	1·00	0·87	0·86
0 to +2°C	0·41	0·40	
−2 to 0°C	0·08		

θ_i; $i=1, \ldots, n$. The conditions corresponding to these singleton *A*s can be interpreted as an envelope for all possible conventional Bayesian probability assignments, i.e. distributions, on θ. Probability distributions falling within this envelope are called the *compatible probability measures* of the BPA (Dempster, 1967) or, more loosely, the *compatible distributions.* This interpretation of a BPA is applicable to both discrete and continuous variable cases and is important if D–S theory is to be used for quantitative decision-making.

2.4.1. *Example 1 (continued)*

The *Bel*( ) and *Pl*( ) values can also be conveniently represented using the same triangular arrangement as for the BPA. The corresponding values extracted from the BPA_{III} obtained from combining the original two inputs are shown in Tables 4 and 5.

It can be deduced from the singleton entries in Tables 4 and 5, and the definition of compatible probability measures given above, that the envelope of probabilities for global warming values falling into each of the three temperature intervals are

$$0{\cdot}01 \leqslant p(-2 \text{ to } 0°\text{C}) \leqslant 0{\cdot}08$$

$$0{\cdot}13 \leqslant p(0 \text{ to } +2°\text{C}) \leqslant 0{\cdot}40$$

$$0{\cdot}60 \leqslant p(+2 \text{ to } +4°\text{C}) \leqslant 0{\cdot}86$$

This demonstrates that specifying probabilities on various intervals of global warming, in effect adopting a possibilistic specification for the state of nature, also creates a possibilistic specification of probabilities on the individual states of nature.

2.5. REPRESENTING IGNORANCE AND NEAR-IGNORANCE

Any scheme claiming to deal with weak information should provide a credible representation of complete ignorance. In a situation where there is no information about the state of nature then, in the D–S scheme, a probability value 1·0 would be assigned entirely to that subset which includes all of the possible values for the state of nature. That is, the probability of 1·0 would be assigned to Θ itself. This type of probability assignment is called an ignorance BPA.

The ignorance BPA can also be viewed from the perspectives of belief and plausibility. We find that any given singleton value has zero belief. This reflects that the smallest possible probability for that value of the state of nature, supportable with ignorance, is zero. But the singleton value also has a plausibility of 1·0, reflecting that the maximum possible probability of 1·0, implying certainty that it is the true value, cannot be excluded by ignorance. Thus, the ignorance BPA provides the largest range possible for the probability value of any singleton value of the state of nature.

It is interesting to contrast this representation of complete ignorance in the D–S scheme with its nearest equivalent in Bayesian theory—the non-informative prior. The non-informative prior must necessarily prescribe an assignment of probability to each of the singleton elements or point values. When the state of nature is treated as a continuous variable this mandatory precision is somewhat at odds with the concept of ignorance. It signifies that sufficient knowledge exists to make such an allocation and exclude all other possible distributions. This requires greater knowledge than is suggested by the ignorance BPA (Walley, 1991). Additionally, in a number of not uncommon situations, the non-informative Bayesian prior cannot be uniquely prescribed and must be arbitrarily chosen from a set of distinctly different distributions, all having equal claim to being the non-informative prior (Berger, 1984). Ambiguity in the choice of a non-informative prior, a choice that can alter the decision outcome, is a cause for concern. No such ambiguity in the choice of an ignorance BPA exists in D–S theory.

2.5.1. *Example 1 (continued)*

Let us imagine that there is a third information source pertaining to our global temperature change issue that contributes no information at all but must still be quantified. The ignorance BPA is shown in Table 6. Combining this ignorance BPA with, say, the example BPA_{III} produces a resultant BPA that is identical to BPA_{III}.

Contrasting the above ignorance BPA with the conventional uniform distribution is commonly, though far from universally, used in the Bayesian scheme to convey zero information. This is equivalent to a Bayesian BPA where equal probabilities are assigned only to the singletons (Table 7).

Combining BPA_{III} with the Bayesian (Uniform) BPA of Table 7 produces a result quite different from BPA_{III} (see Table 8).

The BPA provides considerable freedom for expressing information in the near-ignorance region. For example, the BPA in Table 9 represents a less precise commitment of probabilities, and therefore a less informative statement, than the Bayesian uniform BPA, but is still distinguishably different from a complete ignorance BPA.

2.6. THE UPPER AND LOWER COMPATIBLE DISTRIBUTIONS $f(\theta)$ AND $g(\theta)$

Amongst the compatible distributions of a BPA, as defined by beliefs and plausibilities in an earlier section, there are two distributions that are of particular interest. These

TABLE 6. Ignorance BPA

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	1·00	0·00	0·00
0 to +2°C	0·00	0·00	
−2 to 0°C	0·00		

TABLE 7. Bayesian (Uniform) distribution as non-informative BPA

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	0·00	0·00	0·333
0 to +2°C	0·00	0·333	
−2 to 0°C	0·333		

TABLE 8. Resultant BPA from combination of Bayesian (Uniform) BPA and example BPA_{III}

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	0·00	0·00	0·64
0 to +2°C	0·00	0·30	
−2 to 0°C	0·06		

TABLE 9. Example of a near-ignorance BPA

	−2 to 0°C	0 to +2°C	+2 to +4°C
+2 to +4°C	0·80	0·20	0·00
0 to +2°C	0·00	0·00	
−2 to 0°C	0·00		

are the lower and upper marginal distributions identified as $f(\theta)$ and $g(\theta)$. They are extracted, in the case of $f(\theta)$, by allowing the probability assignments (density or mass) to gravitate to the smallest value of θ in any interval to which they are attached. Similarly, $g(\theta)$ is obtained by allowing the probability assignments to gravitate to the largest values in the intervals. A mapping of the BPA function onto each of the two triangular BPA diagram axes, achieved by mathematically "integrating out" one axis variable at a time, is all that is necessary to obtain $f(\theta)$ and $g(\theta)$.

As will be demonstrated shortly, the shape of the $f(\theta)$ and $g(\theta)$ distributions will often approximate the shape of the posterior distribution obtained from the nearest equivalent Bayesian analysis and all three distributions will converge on the same distribution when the sample size becomes moderately large. The "distance" separating the $f(\theta)$ and $g(\theta)$ distributions is an additional component of uncertainty that is revealed

TABLE 10. Lower and upper distributions from BPA_{III}

	−2 to 0°C	0 to +2°C	+2 to +4°C
$f(\theta)$	0·08	0·33	0·59
$g(\theta)$	0·01	0·13	0·86

through the use of the D–S scheme. A statistical measure of this distance has not been proposed but, as will be seen in the subsequent section on decision analysis, it can be meaningfully translated into an expected utility interval.

2.6.1. *Example 1 (continued)*

BPA_{III} obtained previously in this Example is for a discretized state of nature and it follows that the $f(\theta)$ and $g(\theta)$ distributions extracted from it will also be discrete. Using the process described above, in this instance integrating out each axis variable in turn is achieved by simply adding rows and columns of BPA_{III}, produces the lower and upper marginal distributions shown in Table 10.

2.7. REPRESENTING SUBJECTIVE KNOWLEDGE

The importance of subjective knowledge in a decision analysis is greatest when the data-based information is weak. The use of probabilities on intervals opens up new opportunities for the elicitation and realistic quantification of subjective knowledge when it becomes important but is also weak. That is, under the conditions presently prevailing in climate change. The primary purpose here is to bring some of these opportunities to the reader's attention but it would be premature to suggest that they form an established D–S methodology for dealing with subjective information, though this prospect has received considerable attention in the probabilistic reasoning literature (Shafer and Tversky, 1985). The first two techniques described below show how a conventional Bayesian prior might be modified to more truly reflect weakness in the knowledge being represented, while the third technique decribes a framework for the probability assignments that may provide a natural representation of subjective knowledge derived from only very limited observations or very limited experience.

(i) *Contaminated prior.* Let us assume that it is convenient to first prescribe the knowledge in the form of a conventional Bayesian distribution $\pi(\theta)$ but, at the same time, some doubt or lack of confidence in the distribution because it originates from weak or suspect sources. For example, when Bayesian procedures have been applied to measurements records that are known to be incomplete or suspected of being fictionalized to some degree. Let α, where $0{\cdot}0 \leqslant \alpha \leqslant 1{\cdot}0$, represent the confidence we have in this Bayesian distribution. The value α might be interpreted as meaning that there is a probability α that the prior probability distribution is the correct one and a probability $(1-\alpha)$ that we know nothing at all. Thus, at its extremes, $\alpha = 0$ suggests that we have no faith in the Bayesian distribution, while the other extreme case, $\alpha = 1{\cdot}0$, indicates that we are completely confident that the knowledge is faithfully expressed by it. When α is $<1{\cdot}0$ this is known as a *contaminated* prior and is discussed by Berger (1984, 1985) and Wasserman (1990*a*,*b*). A BPA that reflects the implications of α has a probability $(1-\alpha)$ assigned to the entire set Θ (in effect, to ignorance) and the remaining probability assigned to the singleton values in proportion to the probability assignments of the Bayesian prior. That is, it is specified by

$$m(\theta_i)=\alpha\pi(\theta_i) \quad i=1,\ldots,n$$

and

$$m(\Theta)=(1-\alpha)$$

(ii) *Local perturbation intervals.* This approach diminishes the precision of the probability assignments that have been expressed in the form of a Bayesian prior $\pi(\theta)$ but the approach is quite different from the contaminated prior. In this case the probability values themselves are not modified in any way but the precision of their assignment is reduced by applying the probabilities to intervals of the state of nature of width $2w$ centered on the point, or singleton, values. The non-zero probability assignments of the resulting BPA are thus specified by

$$m(\theta-w,\theta+w)=\pi(\theta)$$

The actual value of w must be subjectively chosen but the degrading effect it has on a Bayesian prior is not difficult to grasp and has much in common with the idea of a dimensional tolerance in engineering. With this simplicity there is a reasonable prospect of eliciting an appropriate value for w. For example, a situation where the regional mean daily evaporation is to be estimated from a series of daily evaporation observations at a single monitoring site. An appreciation of the limitations of evaporation measuring devices might raise concerns that this data is prone to significant measurement error and that it may also have a tenuous relationship with regional values. In spite of these concerns, a conventional Bayesian inference of the mean daily evaporation is performed with the available evaporation data and the result is expressed in the form of a Bayesian posterior distribution. But the concerns expressed are at odds with the fact that Bayesian probability assignments are made to precise point values of evaporation. Although the true statistical nature of the measurement errors and discrepancies between point and regional evaporation, bias, etc. are not known there may be at least sufficient knowledge to place bounds on the possible error magnitude. That is, it may be possible to elicit a subjective estimate of the size of an interval, symmetrically placed around observed values, within which the true value would almost certainly fall. The size of this interval would provide the basis upon which the value of $2w$ is established.

(iii) *Consonant belief.* A BPA in which all of the subsets or intervals assigned non-zero probabilities are nested within each other is known as the *consonant* BPA. This is discussed in some detail by Shafer (1976), while the BPA for Example 1 shown in Table 1 provides a numerical example of a consonant BPA. In the discrete case, for example, the subscripting of the θ values can easily be revised to reflect the order in which the elements are added as the nested subsets get larger. The following condition then applies to the beliefs, as defined in an earlier section, associated with the subsets

$$Bel(\theta_1)<Bel(\theta_1,\theta_2)<\cdots<Bel(\theta_1,\theta_2,\ldots,\theta_n)$$

This is a form of probability assignment that contrasts with the Bayesian BPA where the probability assignments can only be made singleton values, in effect to only non-overlapping intervals or subsets. Thus, a conventional Bayesian probability assignment represents knowledge by decomposing it into probabilities assigned to mutually exclusive, and therefore clearly conflicting, conclusions. This mandatory fragmentation of the knowledge into conflicting components may be quite unnatural in some circumstances and, as a consequence, lead to distortion. In contrast, a consonant BPA offers the alternative of decomposing the knowledge into a set of conclusions that are all in general agreement with each other and differ only in their levels of precision. These distinctions may be of little consequence when dealing with knowledge derived from substantial experience but important when we are dealing with weak knowledge. It is interesting that consonant BPAs arise at a very fundamental level in Dempster's procedure for inference of the binomial parameter when the sample consists of just a single observation (Caselton and Luo, 1992). This adds some support to the idea that consonance is a natural characteristic of knowledge arising from a single opportunity to observe the phenomenon in question.

Consonance is justified and exploited by eliciting a series of BPAs, each one representing just a simple constituent piece of the overall knowledge. The overall knowledge BPA is then obtained by combining these BPAs. The elicitation process is facilitated because, with consonance, the intervals to which the probability assignments can be made are substantially reduced in number. The process for each piece of knowledge begins with an ignorance BPA which is then modified as each response is obtained. Various types of questions can be asked that reveal a subject's plausibility values as well as basic probability assignments and, although the BPA evolves incrementally, it should remain consonant at all stages of the process. Each modification to the BPA is therefore tested as it occurs to ensure that consonance is being maintained. The test involves the application of the following set of constraints on plausibilities and probability assignments that arise directly from the definition of consonance given above.

$$
\begin{aligned}
Pl(\theta_1) > Pl(\theta_2) > \cdots > Pl(\theta_n) \\
Pl(\theta_1) &= 1{\cdot}0 \\
m(\theta_1) &= 1{\cdot}0 - Pl(\theta_2) \\
m(\theta_1, \theta_2) &= Pl(\theta_2) - Pl(\theta_3) \\
m(\theta_1, \theta_2, \theta_3) &= Pl(\theta_2) - Pl(\theta_3) \\
\ldots \quad & \ldots \\
m(\theta_1, \theta_2, \ldots, \theta_n) &= Pl(\theta_{n-1}) - Pl(\theta_n)
\end{aligned}
$$

A simple computerized elicitation procedure that exploits this constraint set is described by Russell *et al.* (1988).

3. D–S inference

Central to decision analysis is a scheme for introducing the results of inference, that is, information derived from sampling or other forms of measurement. There are two possible avenues for inference in D–S theory. One is based on Dempster's early work on a "pure" D–S inference theory and the other developed by Shafer that first translates a conventional likelihood function into a BPA.

It is noteworthy that both of these approaches to inference produce identical results to the conventional Bayesian procedure when they are supplied with a Bayesian prior (expressed in the form of a Bayesian BPA) and the same sample information. Bayesian theory can therefore be viewed as a special case within the D–S scheme. In this way, D–S theory has also been claimed to be a Generalized Bayesian Theory (Dempster, 1968).

3.1. DEMPSTER'S APPROACH TO INFERENCE—SOME ILLUSTRATIVE CASES

In applications of Bayesian analysis the state of nature is often the unknown value of a parameter of a known type of density function. The Binomial case is mentioned here as it underlies the concept of return periods widely used in water resource engineering and is also fundamental to Dempster's more general approach to inference. The need to estimate the parameters of extreme value distributions also arises in resource engineering and management, and some typical results obtained with both D–S and Bayesian schemes are also presented.

(i) *Binomial parameter estimation.* Caselton and Luo (1992) describe the Dempster inference scheme for the binomial parameter in a more accessible but less rigorous form than the original presentation by Dempster (1968b). It explains the inference approach by induction from a single observed outcome of an elementary Monte Carlo simulation of a single Bernoulli trial. The information concerning the (unknown) binomial parameter value controlling the simulation, which is extracted from the single observed outcome, arises quite naturally in the form of probabilities on intervals.

The BPA obtained from each of a series of observations can be combined using Dempster's combination rule to produce a final posterior BPA. The inference results for single and very small samples, as reflected in the upper and lower compatible distributions of the binomial parameter, that is $f(\theta)$ and $g(\theta)$, are distinctly different from those obtained with Bayesian inference when the number of observations is less than about 30. Numerical examples demonstrating this are given by Caselton and Luo (1992).

(ii) *General distribution parameters.* Dempster extended his binomial parameter inference scheme first to the parameters of a multinomial distribution and then to individual parameters of general distributions such as the Normal. In turn we have applied Dempster's inference to the inference of single parameters of distributions of engineering interest, namely the lognormal and Type I Extreme distributions. The theoretical development is difficult and we are not aware of any published results for other distributions.

3.1.1. *Example 2*

A random event was assumed to be modeled by the Type I Extreme (largest) value distribution with known scale parameter but unknown location parameter. Sample values were first generated using the Type I distribution with both parameters specified.

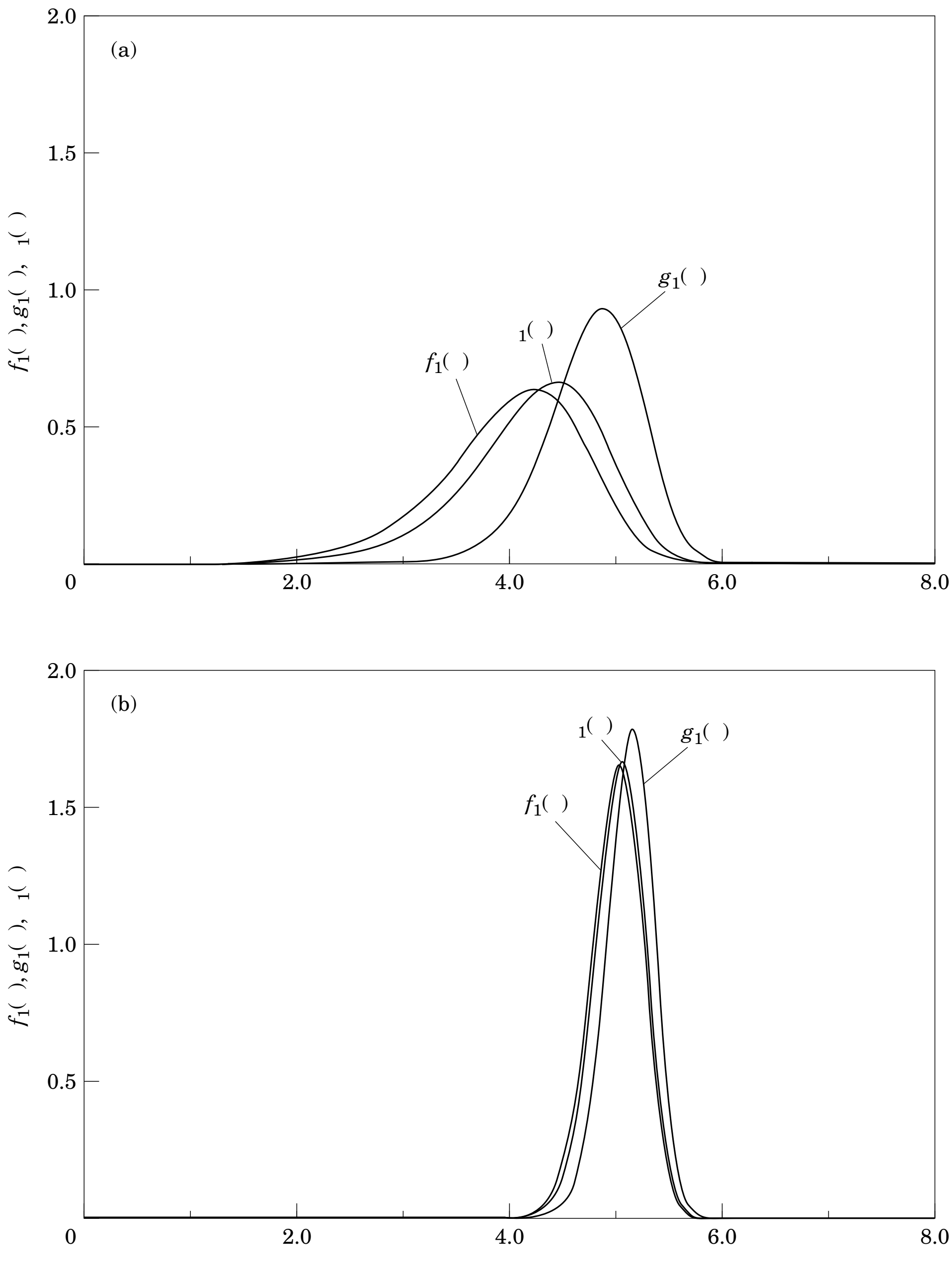


Figure 3. Type I Extreme (largest) distribution. Inference of location parameter μ. Compatible distributions $f(\mu)$, $g(\mu)$, together with Bayesian posterior $\pi(\mu)$, for sample size (a) 2, (b) 12.

The location parameter was then inferred from the sample using the D–S scheme, with ignorance prior, and also with the nearest equivalent Bayesian approach utilizing the appropriate non-informative prior. Plots of the lower and upper compatible distributions $f(\mu)$ and $g(\mu)$, together with the Bayesian posterior $\pi(\mu)$, are shown in Figure 3 for sample sizes 2 and 12. These plots demonstrate the increasing difference between these three distributions as the sample size diminishes. The D–S scheme reveals far greater uncertainty concerning the value of the unknown location parameter.

4. D–S theory and decision analysis

4.1. THE EXPECTED UTILITY INTERVAL

The principal inputs to a D–S decision analysis are the posterior BPA on the state of nature θ and a utility function $U(d_j, \theta)$ prescribing the utility for any particular values of decision choice d_j and θ. Upper and lower expected utilities $E^*[U(d_j)]$ and $E_*[U(d_j)]$ for decision choice d_j are determined by allowing probabilities assigned to an interval of θ in the BPA to be reassigned to a point within the interval so that the contributions to expected utility, and hence the overall expected utility, are first maximized and then minimized. Stating this formally:

$$E^*[U(d_j)] = \sum_{A \subset \Theta} m(A) \sup_{\theta_i \subset A} U(d_j, \theta_i)$$

and

$$E_*[U(d_j)] = \sum_{A \subset \Theta} m(A) \inf_{\theta_i \subset A} U(d_j, \theta_i)$$

When the utility function is monotone in θ, the points to which the probabilities must be assigned to obtain these upper and lower expected utilities will lie at the extreme upper and lower ends of the intervals. The compatible distributions $g(\theta)$ and $f(\theta)$ described earlier in this paper are constructed by reassigning the probabilities in precisely this fashion and, therefore, for monotone utility functions, can be used to compute the upper and lower expected utilities.

The width of the expected utility interval $E^*[U(d_j)] - E_*[U(d_j)]$ would be greatest when we have an ignorance posterior BPA, in which case the $f(\theta)$ and $g(\theta)$ distributions are, respectively, simply unit spikes at the smallest and largest values in Θ. The expected utility interval disappears whenever the posterior knowledge concerning θ is sufficiently strong to be specified in the form of a conventional probability distribution, i.e. when the D–S and conventional Bayesian results coincide. Between these two extremes, the width of the expected utility interval for a given decision is a combined reflection of the additional uncertainty exposed by the D–S scheme and the sensitivity of the utility function to the value of θ.

4.2. CONSEQUENCES OF THE EXPECTED VALUE INTERVAL IN DECISION ANALYSIS

The expected utility interval is a conspicuous feature of a D–S decision analysis. The consequence of this interval when it is large is that the utility intervals from the various decision alternatives substantially or even completely overlap. The ability to resolve decision choice by expected utility can then diminish to the point where the decision choice becomes indeterminate. This indeterminism in the decision choice does not mean that the approach is deficient in some respect. In fact just the opposite may be true. A

TABLE 11. Variation in expected utility interval with size of sample

Number of observations n	3	6	9	12	30
$E^*[\theta]$	0·87	0·83	0·80	0·79	0·74
$E_*[\theta]$	0·36	0·44	0·49	0·51	0·58
$\Delta U = c_i(E^*[\theta] - E_*[\theta])$	$0{\cdot}51c_i$	$0{\cdot}39c_i$	$0{\cdot}31c_i$	$0{\cdot}27c_i$	$0{\cdot}16c_i$

natural consequence of ever weakening inputs into a decision analysis should be that, at some point, the analysis becomes unable to draw any conclusion regarding decision choice.

4.2.1. *Example 3*

Consider a situation involving observation of annual exceedances of a prescribed flood level and the need to estimate the probability of exceeding this flood level in any future year. This probability is the parameter of the binomial distribution represented by θ, the unknown state of nature in this problem. As θ is a probability then it must lie in the interval 0·0 to 1·0. The utility is linearly proportional to θ with coefficient c_i for a decision choice d_i, thus the utility function is $c_i\theta$ and expected utility is given by $c_iE[\theta]$. To demonstrate how the expected utility interval arises and varies with the number of observations, n will be ranged from 3 to 30, but the ratio of number of exceedances observed to the number of observations made will be kept at 2/3. An ignorance prior BPA is assumed. The upper and lower expected values of θ, $E^*[\theta]$ and $E_*[\theta]$, together with the size of the expected utility interval $\Delta U = c_i(E^*[\theta] - E_*[\theta]$, computed for a range of values for n, are given in Table 11.

The extent to which these expected utility value intervals are representative of other situations will depend on a number of factors. The above results were obtained with an ignorance prior BPA. If more prior information was brought to bear on the problem, and this information was in general agreement with the information in the sample, then there would be some reduction in the size of the utility intervals. On the other hand, a different utility function that was non-linear and more sensitive to θ could substantially increase the relative sizes of the intervals.

The ability to resolve between decision choices is artificially improved if a decision criterion is adopted that, for example, minimizes just the upper expected utility value. This resembles the conventional minimax approach and permits the consequences of a decision to be represented by a single-valued quantity. Unfortunately, restoring the ability to resolve between decisions in this way suppresses indications that the decision choice under an expected utility criterion is indeterminate to some degree.

5. Conclusion

In view of the very large uncertainties surrounding predictions of climate change, and the important environmental and resource management decision problems that depend upon such predictions, there is clearly a need for a decision analysis scheme that is credible under near-ignorance conditions. Evidence can be found in the statistical literature indicating that the Bayesian approach, presently the most obvious candidate, has shortcomings under near-ignorance conditions. The shortcomings arise primarily in the restriction placed on Bayesian probability assignments, namely that they can only be made to mutually exclusive point values, or intervals, of the state of nature.

The D–S scheme is just one of a number of schemes that attempt to overcome this restriction. Its use of intervals of values, and the freedom of choice of interval size for all inputs, overlapping or otherwise, sets the D–S scheme apart. In our view it provides a natural and readily grasped basis for the expression of weak information. It is worth noting that intervals are not entirely excluded from the conventional Bayesian approach; they arise, for example, in the form of credibility intervals that can be derived from a posterior distribution. Furthermore, the ability of the D–S scheme to replicate Bayesian results when the prior information naturally conforms with Bayesian requirements, and moderate- to large-sized samples are involved, is most reassuring.

The cognitive and psychological aspects of capturing knowledge in the form of a BPA must also be considered. The BPA, and derivable quantities like belief and plausibility, provide a rich canvas upon which subjective information can be painted. Extensive discussions on the subject of elicitation can be found in the literature of artificial intelligence and statistics. There one can find papers by both supporters and opponents of almost every aspect of the D–S scheme. But controversy has been a characteristic of the Bayesian literature for over 200 years so that it should not be too surprising that this is also the case for post-Bayesian spin-offs, like D–S theory. From a user's point of view, though, it means that the literature does not yet presently provide a clear answer to the question of the D–S scheme's validity.

Professional resource managers dealing with climate change issues face unusually high levels of uncertainty in their decision-making along with ever-increasing demands for objectivity and justification. At the same time their decisions are subject to more and more public scrutiny and there is a growing need to formalize the decision-making process. Inevitably, this means quantifying the uncertainties, no matter how large. Any prospective uncertainty quantification and decision analysis schemes will have to be evaluated at the uncertainty extreme of near-ignorance. Given the limited number of options, studying the D–S scheme is a necessary part of the overall evaluation process but, as well, this particular scheme has a number of features that make it an attractive option.

References

Berger, J. O. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analysis* (J. Kadane, ed.), pp. 63–124. Amsterdam: North-Holland.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag.

Boole, G. (1854). *An Investigation of the Laws of Thought.* Reprint (1958). New York: Dover.

Caselton, W. F. and Luo, W. (1992). Decision making with imprecise probabilities—Dempster–Shafer theory and application. *Water Resources Research* **28**, 3071–3083.

Caselton, W. F. and Luo, W. (1994). Inference and decision under near ignorance conditions. In *Engineering Risk in Natural Resources Management* (L. Duckstein and E. Parent, eds), NATO Advanced Study Institute Series, pp. 291–304. Dordrecht: Kluwer Academic.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* **37**, 355–374.

Dempster, A. P. (1968*a*). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B* **30**, 205–247.

Dempster, A. P. (1968*b*). Upper and lower probabilities generalized by a random closed interval. *Annals of Mathematical Statistics* **39**, 957–966.

Dempster, A. P. (1969). Upper and lower probability inference for families of hypotheses with monotone density ratios. *Annals of Mathematical Statistics* **40**, 953–969.

Russell, A., Froese, T. and Caselton, W. (1988). A Dempster–Shafer based construction expert system. In *Proceedings, Third International Conference on Computing in Civil Engineering*, pp. 469–478. Vancouver, Canada.

Shafer, G. (1967). *A Mathematical Theory of Evidence.* Princeton, N.J.: Princeton University Press.
Shafer, G. (1982). Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B* **44**, 322–352.
Shafer, G. and Tversky, A. (1985). Languages and designs for probability judgment. *Cognitive Science* **9**, 309–339.
Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* London: Chapman & Hall.
Wasserman, L. A. (1988). Belief functions and likelihood. *Technical Report, No. 420.* Department of Statistics, Carnegie–Mellon University.
Wasserman, L. A. (1990*a*). Belief function and statistical inference. *Canadian Journal of Statistics* **18**, 197–204.
Wasserman, L. A. (1990*b*). Prior envelopes based on belief functions. *Annals of Statistics* **18**, 454–464.
Yager, R. R. (1987). On the Dempster–Shafer framework. *Artificial Sciences* **41**, 93–137.