Using graphical modeling and multi-attribute utility theory for uncertainty handling in large systems

Manuele Leonelli James Q. Smith
Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK.

Abstract *Although many decision-making problems involve uncertainty, uncertainty handling within large decision support systems (DSSs) is challenging. One domain where uncertainty handling is critical is emergency response management, in particular nuclear emergency response, where decision making takes place in an uncertain, dynamically changing environment. Assimilation and analysis of data can help reducing these uncertainties, but it is critical to do this in an efficient and defensible way. The paper, after briefly introducing the structure of a typical DSS for nuclear emergencies, sets up a theoretical structure that enables a formal Bayesian decision analysis to be performed for environments like this within a DSS architecture. In such probabilistic DSSs many input conditional probability distributions are provided by different sets of experts overseeing different aspects of the emergency. These probabilities are then used by the decision maker (DM) to find her optimal decision. But unless due care is taken in such a composite framework, coherence and rationality may be compromised in a sense made explicit in this paper. The technology we describe here provides a framework around which Bayesian data updating can be performed in a modular way, ensuring both coherence and efficiency, but nevertheless provides sufficient unambiguous information to enable the DM to discover her expected utility maximizing policy.*

Keywords: Uncertainty handling, graphical models, DSSs, utility factorizations.

1 Introduction

Most decision analyses require numerical inputs the DM is unsure about. This *uncertainty* can derive from physical randomness, which can be modeled using various methodologies, from judgmental estimates of quantities she is unsure about, or from many other sources as noted in [9]. In particular, in the case of an off-site nuclear emergency, uncertainty can derive from, for example, weather forecasts, observation errors in monitoring data, the degree of public agreement about the efficacy with any countermeasure, and the quality of the models used such as those for atmospheric dispersion. The need for uncertainty handling in any nuclear DSS has been recognized for some time [10]. In this work it was also demonstrated that, at least in this context, it is possible to propose an entirely *Bayesian* solution. However in [13] it was recognized that the development of methodologies for use in nuclear DSSs that are both formal and practical were still in their infancy for uncertainty handling. During the intervening years technologies have advanced sufficiently to ensure that if fully formal methods are developed then it will be possible to actually implement those. So it is now timely to revisit this problem. In more recent years some progress has been made in uncertainty handling for nuclear emergency management, for example by including statistical methodologies in components of DSSs that previously were deterministic, as shown in [4]. Also data assimilation using the appropriate statistical methodologies has been included in some analyses, as for the atmospheric dispersion of the radiation [17], with the main aim of reducing uncertainty. However, there still exists no complete formal theory about how this uncertainty should be propagated over the system. In this paper we describe the development of some new theoretical frameworks designated to provide a defensible as well as feasible methodology for uncertainty handling which we hope will make more possible the general implementation of Bayesian methods into DSSs.

In the context of nuclear emergency management the DSS is typically made up of several different components expressing best expert judgment often over very different domains. In such a situation the composite DSS needs to coherently network together judgments about the quantities that are necessary to perform the analysis, provided by different experts overseeing different areas of the problem. For example, an assessment of the safety of a nuclear plant will be based on expert judgments of nuclear physicists and engineers who build their own decision support model. However, the experts that are needed to assess what might happen *after* an accidental release has happened are different. There experts will be atmospheric physicists and local weather forecasters to inform the spread of the contamination, combining their knowledge into a different component. Yet another DSS will build on the *effect* this spread of contamination might have, because of the exposure of humans, animals and plants and because of the introduction of contaminated items into the food chain. At the very end of the chain of consequences doctors and geneticists can be asked to provide information about the impacts of the disaster to health on the population as a function of these different exposure patterns.

From the description above it is apparent how many heterogeneous aspects a DSS needs to consider, including judgments from a rather large set of people with an heterogeneous background. Because of the huge dimensionality of this class of problem, it would be highly desirable to obtain a *modular* structure in which every aspect of the overall problem was modeled and subsequently updated independently and in which the relevant experts could provide judgments autonomously. Critical elements of such a network would be to appropriately encode the relevant uncertainty associated with each component, to propagate these uncertainties over the system and so infer the best overall choice which accommodates all quantified uncertainties and structure in a typical user's utility function.

In the following sections we will show conditions when such modularity can be ensured both a priori and after the collection of data. We note that, assuming that the experts provide rational and coherent judgments about their area of jurisdiction, then, by standard decision theory (see [11] for example), when the system is modular, the decisions based on these judgments will be rational. In [32] it was noted that after a large scale disaster, for example after a nuclear accident, the lack of integration and collaboration among all the involved people threatens to cause further negative consequences to the event: in particular high levels of anxiety and stress. This kind of threat to rationality can be directly managed through developing the types of protocols we address here, permitting the people involved to be required only to provide quantitative judgments within their own area of expertise. At the same time they will be able to collaborate with experts with the same background and agreeing beforehand on how their diverse judgments should be a consensual joint approval of the potential risks.

To achieve this within a Bayesian paradigm, a DSS needs to be able to use agreed collective beliefs about the qualitative relationships in the process pasting together the outputs of the different modules, each of them having inputs from different sets of experts, into a single probability model. The DM would then be able to use these statements as her own. Then, when certain conditions hold, she would be able to behave as a single expected utility maximizer whose input probabilities are appropriately informed by the judgments of the relevant experts. Once in place she would also be able to argue that she has come to her decision drawing upon as fully as possible the data relevant to the decision making process in a way advised by these experts, and so defend her decisions to an external regulator, the general public and politicians.

The paper is structured as follows. Section 2, after briefly introducing the architecture of a typical DSS for nuclear emergencies, discusses the need of uncertainty handling in such a system. Section 3 then defines the notation, reviews the main background material and explains the statistical setting used to model uncertainty. In Section 4 an example representing a nuclear DSS is presented to illustrate the conditions necessary to ensure the modularity property needed for such a system. Section 5 presents a very simple scenario in which the omission of the uncertainty leads to wrong policies. We conclude the paper with a discussion of some ongoing developments in this area.

2 The needs of a nuclear DSS

In the last 25 years, subsequently to the Chernobyl accident, there has been an increasing number of European projects which have aimed to develop a joint DSS to improve the emergency management of any future accidental release of radioactivity. One of the DSSs developed during these collaborations was RODOS which, as other nuclear support systems, differed from traditional DSSs in the following aspects:

- It was designed to work in several different contexts, depending on where decisions needed to be made;
- It had a client/server architecture, built on a list of modules which are connected via a communication interface;
- It was designed to explicitly address and solve the issue of uncertainty handling.

Before examining how uncertainty can be incorporated in the system, we first present its conceptual architecture. As described in [5], this was split into three distinct subsystems:

- Analysing Subsytem (ASY) modules which processed incoming data and produce forecasts;
- Countermeasure Subsystem (CSY) modules which suggested possible countermeasures;

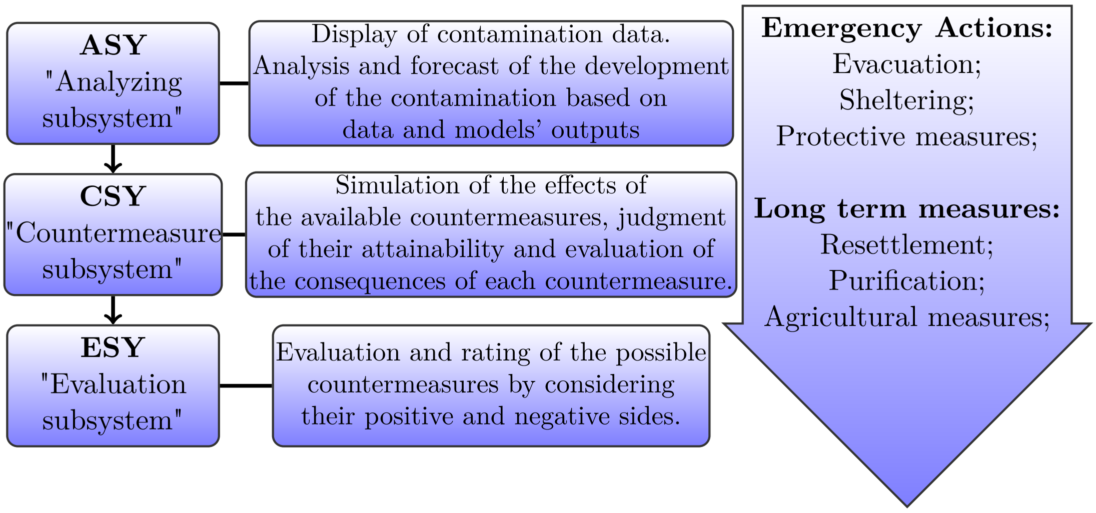

Figure 1: Conceptual structure of RODOS after [26].

- Evaluating Subsystem (ESY) modules which ranked these countermeasures.

In [13] it was argued that this architecture corresponds well with the Bayesian paradigm for decision-making, in the sense that the ASY modules build the statistical models underlying the system, the CSY modules identify actions and calculate their potential consequences and the ESY modules compute their expected utilities. Figure 1 summarizes the architecture of RODOS.

On the right of the diagram we can note two different sets of available actions: one for the short term and the other for the long one. Thus RODOS, like certain other nuclear DSSs, provided decision support for all the different phases of a nuclear accident. As time passes the system will arrive at different decision points where it must select three modules to form an ASY, CSY and ESY chain appropriate to the context of the decision. Thus it is helpful to recognize in this conceptualization the temporal control necessary to manipulate modules in response to the user's request or some pre-programmed sequence of desired analyses. These functions were performed by the so called Temporal Support Functions (TSF), providing the flexibility to change operating behavior according to the current phase of the accident, selecting the most suited models and processing and storing incoming data in a suitable way.

In an integrated DSS like the one described above, it is tempting solely to model expectations of inputs into such a system and treat them as known. However, if there is significant uncertainty associated with such input, then it is well known that such methods can grossly distort the assessment of appropriate policies. This will be demonstrated in Section 5 through a simple example. So to properly inform DMs it is actually necessary to incorporate uncertainty into such a DSS.

Sometimes this uncertainty will be intrinsic to a particular module *within* the panel of domain experts. For example, an estimate of the likely release of radioactivity before

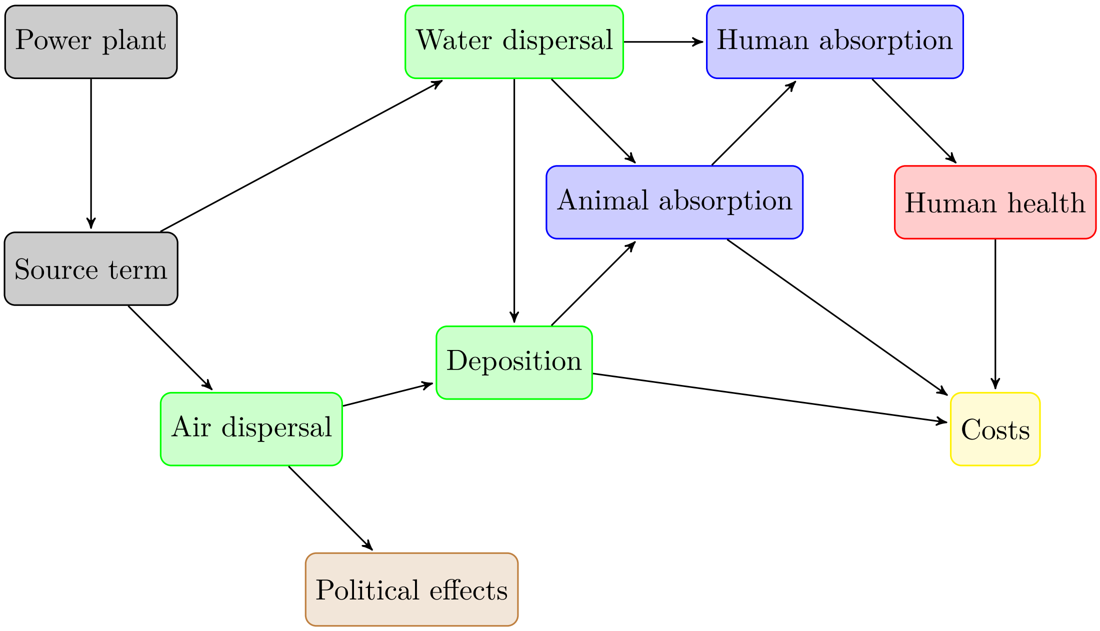

Figure 2: Plausible network for the modules of a nuclear DSS

it actually occurs is provided by the Source Term Module (STM), which consists of a directed statistical graphical model (see [19] and [22] for an introduction) as shown in [13], while one of the available atmospheric dispersion deposition modules was RIMPUFF, a puff diffusion model (see, for example [31]). In implemented systems the complexity of the domain demands that such methods are numerical, although occasionally it is possible to perform such uncertainty analytically (see below).

At other times the expert judgments are encoded in terms of a collection of sparse sets of simulators over an highly complex deterministic system. This is typical of modules associated with climate change, but in the context of systems appearing in nuclear emergencies, it can be associated, for example, to the modules predicting the effect of the nuclear accident on the food chain. In these cases uncertainty needs to be assessed by the appropriate panels of analysts using methodologies based on *emulators* which produce such uncertainty measures, analogous to those developed for climate change [34].

In other scenarios there is no model based methodologies possible, so that uncertainty measures need to be elicited *directly* from the panels.

However uncertainties are encoded by the single panels, these uncertainties will need to be communicated to panels using the outputs of that module as the inputs to their own forecasting module. For example uncertainties about the extent of release from a nuclear plant as well as point predictions of point releases will need to be communicated to the panel responsible for the prediction of the contamination spread across the locality. In reality the methodologies behind the calculation of these uncertainties can be very complex. In this paper we will therefore treat them as a given, centering the analyses on illustrations of what they might look like in very simple situations. This will allow us to concentrate more fully on how these assessments can be processed and propagated through the network of expert systems.

As we will formalize in the following section, the network of panels of experts can often be conveniently depicted through a *directed graphical model*, itself common knowledge to all involved people in the development of the DSS. With certain additional assumption this has a fully formal semantic, implying that certain analytical deductions can already be made by the group. A plausible example of such a network that may describe the whole system is presented in Figure 2, together with a division of the nodes into areas of interest for different sets of experts. Subsequently in this paper we will use this network to illustrate our methodology.

We will also illustrate below the issue of exactly which uncertainties are needed by the composite system to be determined by the purpose of the whole decision support. In particular we argue in this paper that ideally this should be assessed through the structure of the (multi-attribute) *utility function*, which is used to represent and analyze the preferences of the experts and consequently of the DM. There has been an increasing interest in using multi-attribute utility theory (MAUT) in DSSs as noted by the analyses drawing on [6] and [33]: the first studied the contributions of multi-criteria methods (among which MAUT) to DSSs, while the second evaluated the current state of implementation of these methods in DSSs. Both the articles noted the advantages in using such methodologies in the context of supporting decision-making. We here apply these ideas to this complex and necessarily distributed environment.

3 Setting and notation

We start by introducing the notation necessary for the following. Consider a set of decisions $d \in \mathcal{D}$ available at a certain decision point and let $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) \in \mathcal{Y} = \times_{i=1}^{n} \mathcal{Y}_i$ be a random vector of measurable quantities parametrized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ such that $f_{\mathbf{Y}}(\mathbf{y} \mid \boldsymbol{\theta})$ is a joint density of the model. Here $\mathbf{Y}_i$, for $i = 1, \ldots, n$, is one of the n modules that are included in the DSS. Let $\{G_1, \ldots, G_m\}$ denote m panels of domain experts, which input the relative information to the DSS necessary for the DM to make the optimal decision. Let also $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ such that $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$ and $\boldsymbol{\Theta} = \times_{i=1}^{n} \boldsymbol{\Theta}_i$. In a Bayesian framework we assume the parameter vectors $\boldsymbol{\theta}_i$ to be random vectors with density function $f_i(\mathbf{a}_i(d))$, in which $\mathbf{a}_i(d) \in \mathcal{A}_i$ are the hyper-parameters, for $i = 1, \ldots, m$. Note that we are assuming that the random variables of the systems are different for each decision available, with the hyper-parameters being allowed to be functions of d. Each module is under the sole jurisdiction of its associated panel, providing beliefs about the distribution of the appropriate vectors $\mathbf{Y}_i(d)|\mathbf{Z}_i(d)$, where $\mathbf{Y}_i(d)$ and $\mathbf{Z}_i(d)$ are disjoint ($\mathbf{Z}_i(d)$ possibly null) sub-vectors of $\mathbf{Y}(d)$. Depending on the situation, these sets of experts' beliefs may be expressed as conditional densities, as in a full Bayesian analysis, or alternatively, for example, as only some lower order moments of the conditional distributions.

With this notation we can denote the quantities of the network in Figure 2 as:

- G_1: engineering panel;
- G_2: environment panel;
- G_3: biological panel;
- G_4: political panel;
- G_5: medical panel;
- G_6: economical panel;

- $\mathbf{Y}_1$: power plant;
- $\mathbf{Y}_2$: source term;
- $\mathbf{Y}_3$: air dispersal;
- $\mathbf{Y}_4$: water dispersal;
- $\mathbf{Y}_5$: deposition;
- $\mathbf{Y}_6$: animal absorption;
- $\mathbf{Y}_7$: human absorption;
- $\mathbf{Y}_8$: human health;
- $\mathbf{Y}_9$: political effects;
- $\mathbf{Y}_{10}$: costs.

Panel G_1 (grey) oversees $\mathbf{Y}_1$ and $\mathbf{Y}_2$; panel G_2 (green) oversees $\mathbf{Y}_3$, $\mathbf{Y}_4$ and $\mathbf{Y}_5$; $\mathbf{Y}_6$ and $\mathbf{Y}_7$ are jurisdiction of G_3 (blue); G_5 (red), G_4 (brown) and G_6 (yellow) are responsible of $\mathbf{Y}_8$, $\mathbf{Y}_9$ and $\mathbf{Y}_{10}$, respectively. Assume also that the DM's decision space $\mathcal{D}$ for this example consists of:

- d_1: do nothing;
- d_2: tell the population to shelter;
- d_3: deliver protective measures;
- d_4: temporarily evacuate;
- d_5: leave the area forever.

Consider also $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_n) \in \mathcal{R}$ a set of attributes such that $\mathbf{R}_i = \mathbf{R}_i(\mathbf{Y}_i(\boldsymbol{\theta}), d)$ (note that also $\boldsymbol{\theta}$ depends on d) and a utility function $\mathrm{U} : \mathcal{R} \times \mathcal{D} \to [0, 1]$: thus, $\mathrm{U} = \mathrm{U}(d, \mathbf{R}_i(\mathbf{Y}_i(\boldsymbol{\theta}_i)), i = 1, \ldots, n)$. Also the utility function and the rewards can depend on the decision d. The attributes $\mathbf{R}_i$'s are introduced for the cases in which it is unclear or unfeasible to elicit preferences over some $\mathbf{Y}_i$'s. In the following we will suppress the dependence on d because the focus is the uncertainty handling and we are not studying yet the optimization problem that leads to the individuation of the optimal decision. The reader should remember though that, depending on the situation, some of the quantities involved may vary across the decisions available.

For the preferences of the DM about the composite system to be expressed in terms of the preferences over different domains of the problem, it is helpful for the composite utility function to factorize respecting the division of the vector $\mathbf{Y}$ into the sub-vectors overseen by the different panels: this concept will become clearer in the following section.

Moreover a system like the one above will usually be dynamic. However, again to limit the technicalities in this short paper we will restrict our discussion to the cross sectional/non-dynamical aspects of the support system. We therefore imagine the DM to be at a particular decision point, for which the vector $\mathbf{Y}$ represents all the relevant modules to the problem. Moreover, at this particular time point, we assume that a dataset $\mathbf{x}$ becomes available, collected from this non-dynamical system. We will discuss at a later point how the ideas we develop in this paper can be generalized into a dynamic framework.

3.1 The graphical input/output structure

In the previous section we noted that the outputs of some of the modules can be used as inputs for subsequent ones. This process would be handled by the temporal support, which follows the emergency during its developing. However, the modules can also all be depicted through a graphical model where, if the output of module j is used as input for module i, then a directed edge from the vertex representing j to the one of i is drawn. If the resulting graph fulfills some constraints then it can be shown that it will be a **Bayesian network (BN)** (see [27],[16]): now a widely used framework to express beliefs in terms of probabilities. In a Bayesian system this structure forms the framework over which a coherent Analyzing Subsystem can be build. The great success of these graphical models is due to the evocative and intuitive way in which they depict relationships between random variables and to the fast learning and inferential routines that can be performed over them. Let $\mathbf{Z}_i$ be the modules whose output is used as input for the module corresponding to $\mathbf{Y}_i$ and denote $\mathbf{Q}_i = \mathbf{Y} \setminus \{\mathbf{Y}_i \cup \mathbf{Z}_i\}$. With this further notation we can define a BN.

Definition 3.1. *A Bayesian network for the random vector $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ consists of a set of $n-1$ conditional independence statements of the form $\mathbf{Y}_i \perp\!\!\!\perp \mathbf{Q}_i|\mathbf{Z}_i$, together with a directed graph having no directed cycles $\mathcal{G}$. The directed graph has vertex set $V(\mathcal{G})$ and edge set $E(\mathcal{G})$, where*

$$V(\mathcal{G}) = \{\mathbf{Y}_1, \dots, \mathbf{Y}_n\} \qquad \textit{and} \qquad (\mathbf{Y}_i, \mathbf{Y}_j) \in E(\mathcal{G}) \iff i \in \mathbf{Z}_j, \qquad 1 \le i < j \le n.$$

Note the assumption of having a numerically ordered vertex set: that is, $(\mathbf{Y}_i, \mathbf{Y}_j) \in E(\mathcal{G})$ only if $i < j$. Note also that $\mathbf{Y}_i \perp\!\!\!\perp \mathbf{Q}_i|\mathbf{Z}_i$ reads that the vector $\mathbf{Y}_i$ is independent of $\mathbf{Q}_i$ given $\mathbf{Z}_i$: so that the only information to infer $\mathbf{Y}_i$ from $\mathbf{Q}_i$ & $\mathbf{Z}_i$ is from $\mathbf{Z}_i$. Given a directed acyclic graph a BN can be alternatively defined by the recursive factorization formula

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^{n} f(\mathbf{y}_i \,|\mathbf{Z}_i = \mathbf{z}_i),$$

where $\mathbf{y}_i$ and $\mathbf{z}_i$ are generic realizations of the vectors $\mathbf{Y}_i$ and $\mathbf{Z}_i$ respectively.

Every conditional independence statement of a BN can be deduced from the topology of its graph after a simple transformation, called *moralization*, as stated by the *d-separation* theorem [19]. Recall that a vertex $\mathbf{Y}_i$ is a *parent* of a vertex $\mathbf{Y}_j$, and $\mathbf{Y}_j$ is a *child* of $\mathbf{Y}_i$ in $\mathcal{G}$ if $(\mathbf{Y}_i, \mathbf{Y}_j) \in E(\mathcal{G})$. The set of all parents of a vertex $\mathbf{Y}_i$ is the *parent set*, denoted with $Pa(\mathbf{Y}_i)$, and, for our construction, it corresponds to $\mathbf{Q}_i$. The moralization process simply joins with an edge two vertices in the same parent set that are not connected for an appropriate subset of the graph $\mathcal{G}$. These conditional independences are usually called *Markov conditions* and they state that for every vertex $\mathbf{Y}_i$ in $V(\mathcal{G})$ we have that $\mathbf{Y}_i$ is independent of $V(\mathcal{G}) \setminus (Dn(\mathbf{Y}_i) \cup Pa(\mathbf{Y}_i))$ given $Pa(\mathbf{Y}_i)$, where $Dn(\mathbf{Y}_i)$ corresponds to the descendant set of $\mathbf{Y}_i$. The descendant set of $\mathbf{Y}_i$ consists of those vertices $\mathbf{Y}_j$ for which there is a directed path on the graph starting from $\mathbf{Y}_i$ and finishing in $\mathbf{Y}_j$.

In Section 2 we noted that uncertainty can be included into the DSS by introducing non-deterministic relationships between the modules. In the graphical context, one of the simplest such model defines a multivariate Gaussian distribution over the network, through a set of regression relationships. For example in the simplest case when the output variable Y_j is a univariate Gaussian variable then, for $j = 1, \dots, n$:

$$Y_j = \theta_{0,j} + \sum_i \theta_{i,j} Y_i + W_j, \qquad \forall\, i \in Pa(Y_j), \tag{1}$$

where $\theta_{i,j} \in \mathbb{R}$, for $i < j$, are regression parameter and $W_j \sim \mathcal{N}(0, \psi_j^2)$ independent of the Y_i's and of the other W_i's, for $i < j$.

In the Bayesian context each of these parameters is treated as a random variable with its own distribution expressing the prior knowledge and data provided by the associated expert panel. For the purposes of this paper, we define only some of the moments of the parameters. Let $\mathrm{E}(\theta_{i,j}) = a_{i,j}$ and $\mathbb{V}(\theta_{0,j}) = \tau_{0,j}$ for $i = 0, \dots, n-1$ and $j = 1, \dots, n$ with $i < j$. Let also $\mathrm{E}(\psi_i^2) = \lambda_i$ for $i = 1, \dots, n$. We note that in fact in the current applications of this technology it is usually appropriate to treat these parameters as random processes which change during the progress of the emergency. This generalization is in fact quite straightforward using the class of multiregression dynamic model [24]. However, for simplicity of notation and explanation here we will restrict ourselves only to consider problems where these relationships are fixed in time.

When a graphical model expresses the beliefs of a single expert the common and often default choices are the so called *global* and *local* independence assumptions (see [28],[2]). The global independence assumption states that, if $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ are the parameters of $\mathbf{Y}_i$ and $\mathbf{Y}_j$ respectively, then $\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_j$, for all the possible couples i, j. The local independence assumption on the other hand states that if $\boldsymbol{\theta}_i$ is a random vector, then its components are independent of each other, for $i = 1, \dots, n$. It can be shown that these two assumptions together imply that all the parameters are mutually independent. For the purposes of this paper we will also adopt these assumptions, but for the implications on the composite system see the discussion in [7].

Under these conditions, the following theorem, from [30], shows how to derive the full distribution of the system, when the variables are defined as in (1).

Theorem 3.2. *Let $Y = (Y_1, \dots, Y_n)$ be a random vector, where every Y_i's is defined as in (1). Let Ψ be the diagonal matrix $\Psi = diag(\psi_1^2, \dots, \psi_n^2)$ and let L be an $n \times n$ upper triangular matrix defined as*

$$L_{ij} = \begin{cases} \theta_{i,j}, & \text{if } \{i,j\} \in E(\mathcal{G}), \\ 0, & \text{otherwise.} \end{cases}$$

Set also $B = I - L$, where I is the $n \times n$ identity matrix. Then $Y \sim \mathcal{N}(B^{-T}\boldsymbol{\theta}_0, \Sigma)$, where $\boldsymbol{\theta}_0 = (\theta_{0,1}, \dots, \theta_{0,n})$ and $\Sigma = B^{-T}\Psi B^{-1}$.

Since the Gaussian distribution is fully described by its first two moments, an alternative way of stating Theorem 3.2 is in terms of recursive expressions of the expectations and the covariances/variances of the vertices. Then the mean relationships are simply obtained by plugging in the expected values for their associated variables. This gives us that

$$\mu_j = \mathrm{E}(Y_j) = \theta_{0j} + \sum_i \theta_{i,j}\mu_i, \qquad \forall\, i \in Pa(Y_j).$$

In order to compute the variances, which will capture the uncertainties about these variables, we first need to derive the covariance relationships between any two random variables. For $k < j$, call

$$\sigma_{k,j} = \mathrm{Cov}(Y_k, Y_j) = \sum_i \theta_{i,j}\sigma_{i,k}, \qquad \forall\ i \in Pa(Y_j). \tag{2}$$

It then follows that the required variances can be calculated using formula (2) as

$$\sigma_{j,j} = \mathbb{V}(Y_j) = \sum_i \sum_k \theta_{i,j}\theta_{k,j}\sigma_{i,k} + \psi_j^2, \qquad\qquad \forall i, k \in Pa(Y_j).$$

Note that since each parameter $\theta_{i,j}$ in these expressions is itself a random variable, it will also have its own associated uncertainty related, for example, to the quality of the estimate the panel can make of it. However, these other uncertainties can also be folded into these expressions in a straightforward way illustrated later in the paper. Our underlying methodology is based on discovering the appropriate algebraic relationships like the ones above associated both with the underlying graphical structure and the assumed form of utility functions.

3.2 Causality, learning and updating

In order to incorporate information deriving from experimental evidence each panel might have available, it has now been recognized that the network like the one described above is not simply a BN, but a causal Bayesian network (CBN) [16]. The relevant concept of causation leading to the CBN was defined in [21]. It has also been recognized that even when the network is not assumed to be causal, a CBN can be derived as a natural extension of a BN in which the global independence assumption is met, as shown in [3]. In a CBN each arc can be interpreted as a direct causal influence between a parent node and a child node, relative to the other nodes in the network: an assumption which has some plausibility within this type of applications (see [27]). The intrinsic hypothesis behind a causal model is the so-called *manipulation property* [29]. This states that if one of the variables/vectors of the system, $\mathbf{Y}_i$ say, is controlled and set to a particular value, then the outcomes of the remaining variables would occur as if the outcome of the i-th vertex was not controlled. There are often strong arguments for believing that this causal hypothesis will hold in a given scenario. It is this property that allows us at a preliminary stage to conduct experiments, manipulating some of the of the inputs to each module in a designed experiment and observing the results of these actions, which we assume would be just like what we might observe after a real accident. We note that in practice this assumption of perfect correspondence between designed experiments and observational studies - implicitly a causal hypothesis of the type discussed above - is almost always made in inferences about the inputs of emergency responses in contexts like these.

The databases resulting from these manipulations can be used to start populating the DSS with supporting experimental evidence and also to derive the topology of the network connecting the modules. In [1] a Bayesian method to learn CBNs is introduced, which, taking as input a dataset, produce then a causal graphical model based on the data, which can be both observational and experimental. Thus, using this algorithm, a network for the modules can be derived, over which the relative experts provide the required conditional judgments. Suppose henceforth we now assume that all the users of the DSS agree on the graphical statistical model which connects the modules. This, for example, will be implicit if they accept the network architecture of the composite system.

The collection of data can also reduce the uncertainty in the DSS by the combination of the information in the sample with the prior belief specification of the relative experts. In the full Bayesian framework, this updating is performed through *Bayes theorem* combining the prior distribution with the sample's likelihood to obtain the *posterior* distribution. All the information in the sample about the parameter of interest, $\boldsymbol{\theta}$ say, is contained in its likelihood function as assured by the *likelihood principle.* In Appendix B we show how this updating can be performed in a simple scenario of a conjugate analysis for a normal linear model. However, even when no such conjugate analysis is possible, current sampling methods allow us to calculate posterior distributions numerically from which various posterior moments we use later in this paper can be drawn. Another possibility when the quantities of interest are only some of the moments is to follow a linear Bayes methodology [15] while performing the updating. The effect of relevant experimental evidence, used with the necessary causal assumptions, then reduces the variance of output variables, used as inputs to other modules in an exactly analogous way.

3.3 The combination of beliefs

One of the main issues for maintaining the integrity of large DSSs is the fact that many different sets of people introduce information into the system. There exists a quite large literature on how to combine beliefs provided from different people into a single coherent statement. See [8] for a recent review. However, it mostly focus on the case in which everyone provides information about the whole system, which is a different setting from the one considered in this paper. We now however review the main results in the area, since in the next section we will show that with the technology we are developing, some of the properties can hold also in this more composite case.

When requiring that the DM acts as if she is a single expected utility maximizer who accommodates all information provided by the involved experts it is necessary for the DSS to be externally Bayesian [20]. Loosely speaking External Bayesianity (EB) demands that an aggregated posterior distribution is the same either if every expert updates via Bayes Formula his prior distribution and then the aggregation is performed and if the prior distributions are combined and then Bayes Theorem is applied to obtain a single posterior distribution. We have argued above that this property is critical if the decision of the DM are going to be defensible to an outside auditor.

In [14] it is shown that under a certain set of conditions the only non-dictatorship EB pooling operator, which performs the aggregation of a set of densities, is the Logarithmic Opinion pool (LogOp). If the i-th member of a pool of k members provides a probability distribution $f_i(\boldsymbol{\theta})$, then the LogOp is defined as

$$\bar{f}(\boldsymbol{\theta}) \propto \prod_{i=1}^{k} f_i^{w_i}(\theta)$$

where $w_i \geq 0$, for $i = 1, \ldots, m$, are arbitrary constants such that $\sum_{i=1}^{m} w_i = 1$. The constant w_i is a weight representing the level of trust of the expert that delivered the density f_i.

In [7] a new class of logarithmic operators were introduced, which possesses a generalized version of EB, called Conditional External Bayesianity (CEB). Roughly speaking, CEB requires the same indifference about the order of aggregation and prior-to-posterior analysis for the conditional distributions associated to a BN $\mathcal{G}$. This can be obtained if the likelihood factorizes in the same way as the density function associated to $\mathcal{G}$. Then a generalization of the LogOp was introduced which, instead of pooling marginal distributions, pools conditional ones. Without considering the details related to measure theory, this generalization is quite straightforward and leads to the *conditional LogOp* defined as

$$T_j(f_{1j}, \ldots, f_{mj})(\mathbf{y}_j\,|A_j) \propto \prod_{i=1}^{m} [f_{ij}(\mathbf{y}_j\,|A_j)]^{w_{ij}(A_j)},$$

where w_{ij} are the weights, possibly function of A_j, adding up to one, $A_j = \{Pa(\mathbf{Y}_j) = Pa(\mathbf{y}_j)\}$ and $f_{i,j}$ is the conditional distribution relative to $\mathbf{Y}_j$ provided by the i-th expert. Recall that in the framework of DSSs most of this quantities can be a function of the decisions d, and consequently be elicited for each decision available.

Two other works are worth noting here. In [23] the use of graphical models to aid the process of aggregating beliefs is investigated. In particular, it was proved there that a pooled distribution obtained with the LogOp reflects any shared Markov independence by the group and for this reason, it can be represented in a concise and natural manner with a BN as well. However, it was also noted that any pooling operator cannot maintain all independences representable in a BN. On the other hand in [35] a new framework of opinions' aggregation for Linear Bayes Theory [15] was developed, based on expectations rather then densities. After redefining variations of the main properties present in the literature about the aggregation of densities, it was shown there that an expectation aggregated through a linear combination of expectations is coherent if each of the experts has given a coherent specification.

3.4 Multi-attribute utility theory

Multi-attribute utility theory provides independence concepts for preferences over a large number of attributes. Again in any Bayesian DSS the typical form of the user's utility function will provide the framework from which the Evaluation Subsystem can be quantified.

When no preferential independences hold, it is difficult for a DM to express her preferences coherently. On the other hand, the recognition of independences reduces the dimensionality of the spaces over which she has to express preferential judgments, simplifying considerably the elicitation of her utility function. For simplicity here we assume that $\mathbf{R}_i \equiv \mathbf{Y}_i$, for $i = 1, \ldots, n$. We also suppress the dependence on d. Denote also with $\boldsymbol{Y}_A$ the subset of the component attributes whose indices lie in a subset $A \subseteq [n] = \{1, \ldots, n\}$ and let also $\bar{A} = [n] \setminus A$ denote the complement in $[n]$ of the set A.

In utility theory there are many different definitions of independence and we define them here through the factorization of the utility function that they induce (for more references see [18]).

Definition 3.3. *The set of components attributes $\mathbf{Y}_A$ is said to be **utility independent** of the other attributes $\mathbf{Y}_{\bar{A}}$ ($\mathrm{UI}(\mathbf{Y}_A, \mathbf{Y}_{\bar{A}})$) if the joint utility function can be factorized as*

$$\mathrm{U}(\mathbf{y}) = f(\mathbf{y}_{\bar{A}}) + g(\mathbf{y}_{\bar{A}})\,\mathrm{U}(\mathbf{y}_A), \qquad g(\cdot) > 0,$$

where f and g depend only on their arguments and $U(\mathbf{y}_A)$ is a utility function not depending on the attributes $\mathbf{y}_{\bar{A}}$, which are fixed to a reference value.

The utility independence condition $\mathrm{UI}(\mathbf{Y}_A, \mathbf{Y}_{\bar{A}})$ demands that the preferences over the attributes $\mathbf{Y}_A$ are the same for every level at which the attributes $\mathbf{Y}_{\bar{A}}$ can be fixed.

When UI can be verified for every attribute in the system or for every subset of the attributes' set, then the joint utility function simplifies remarkably.

Theorem 3.4. *Let $2^{[n]}$ be set of all subsets of $[n]$. Given attributes $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, with $n \geq 2$ and values $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, a **multi-linear** utility function*

$$\mathrm{U}(\mathbf{y}) = \sum_{A \in 2^{[n]}} k_A \prod_{a \in A} \mathrm{U}(\mathbf{y}_a),$$

*exists if and only if $\mathbf{Y}_a$, $a = 1, \ldots, n$, is utility independent of its complement $\mathbf{Y}_{\bar{a}}$, where the utilities U are functions of their arguments only and the k's are the **criterion weights**.*

The criterion weights represent the importance of the terms in the previous utility functions (for a definition see [27] and [11]). For example, if a high value is assigned to k_i, it follows that a value of this attribute with a high utility is desirable. Note that, calling $\mathbf{y}^0$ and $\mathbf{y}^*$ the worst and the best possible values of $\mathbf{y}$ respectively, if one assigns $\mathrm{U}(\mathbf{y}^0) = 0$ and $\mathrm{U}(\mathbf{y}^*) = 1$, then all the criterion weights are in $(0, 1]$.

In the multi-linear factorization multiplicative terms are included: these are able to represent the interactions among some of the attributes. If these interaction terms are not considered relevant, that is the overall utility depends only on the marginal values, then the attributes verify a different class of independence, called *additive independence*. This assumption is the most used because it is simplest to explain and enact.

Definition 3.5. *Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_r$ be disjoint subsets of $\mathbf{Y}$, such that $\cup_{i=1}^r \mathbf{Y}_i = \mathbf{Y}$. We say that $\mathbf{Y}_1, \ldots, \mathbf{Y}_r$ are additive independent (AI) if the joint utility function can be written as*

$$U(\mathbf{Y}) = \sum_{i=1}^{r} f_i(\mathbf{Y}_i).$$

4 Modularity over the network

Expected utility theory tells us that the optimal decision for a DM is the one maximizing her expected utility. We note that this expected utility can often be expressed as a function of various moments of the random variables. For example, in the linear case for which $U(y_i) = y_i$, the expected utility of Y_i corresponds to the expectation of Y_i. In the following section we will show some results relating the shape of the utility function and the moments necessary to derive its expectation. Moments are functions of the parameters of these random variables and in the multi-attribute case we need conditions on the interactions between these parameters if we want to obtain a modular structure for the DSS. In particular we now require that global and local independence assumptions hold and that the preferences respect an additive factorization.

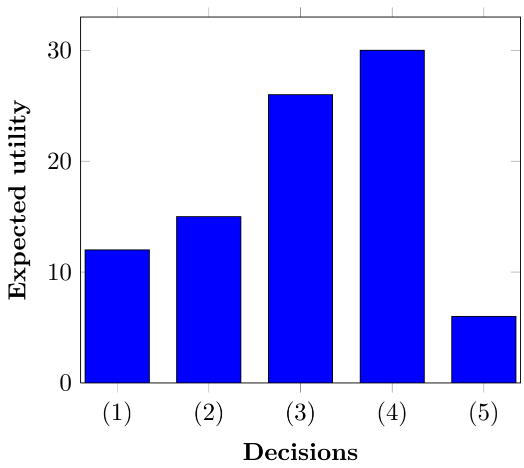

Figure 3: *A priori expected utility scores*

This kind of assumptions is usually discussed and, eventually, agreed during *Facilitated Workshops* and *Decision Conferences* (see [12]), in which representatives of different stakeholders, appropriate scientists and authorities involved meet. The prime purpose of these meetings is the elicitation of an agreed utility function between potential users in an emergency. This is elicited using some simple examples of hypothetical scenarios to get the group's preferences appropriately calibrated. Once this utility function, and in particular the criterion weights, are fixed, the different groups of experts individually continue to draw out their beliefs about the process, as described in the previous sections.

As a further starting assumption, consider the panel to consist of one member only. When these assumptions hold, the relevant attributes are informed by a particular panel who will ideally provide all involved in the decision making with the posterior probability distributions or summaries thereof associated with the outputs of their particular domain. Within the context of the regression models these assessment would include posterior distributions on the regression parameters or, more generally, on the parameters of the conditional distributions of the attributes under their jurisdiction.

Figure 3 shows the expected utility scores of the different decisions available to the DM in the nuclear example introduced with the network in Figure 1, under the conditions just introduced. Decisions d_3 (protective measure) and d_4 (evacuate) obtained the highest scores, with a significant distance from the other decisions available. In this situation the overall expected utility of a decision is a function of the panels' expected utilities and consequently one can analyze the scores associated with the different panel domains for the single decisions.

Figure 4 shows the panels' expected utility scores for the decisions of evacuating and delivering protective measures. Recall that the overall expected utility obtained by a decision is a weighted average of the relative scores shown in this graph. We can note that providing protective measures obtained a way higher score in the economical area comparing it to the evacuation option, because of the high costs consequently to transfer a high number of people. However, the decision of evacuating performed far better in the biological and the medical modules, because of the smaller intake of radiations for the population. On the other modules, the results are the same for the two options and from these considerations (we assume the weights to be the same for the different areas) follows the fact that d_4 obtained an higher overall expected utility score.

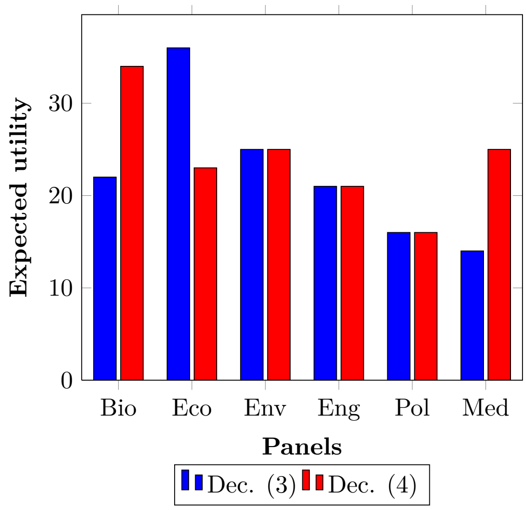

Figure 4: A priori panel scores for the two best decisions

The type of analysis illustrated through this simple example relies on various assumptions. Prior to an actual accident different panels will bring the posterior distributions associated with their domain of expertise. These can also be updated in the light of further experiments before the actual emergency starts. When an emergency occurs, these judgments are fed into the DSS and combined with the utility function elicited during the Decision Conferences to help elucidating *real* scenarios faced by the group's directly countermeasures. We have assumed that these judgments about parameters are mutually independent of each other: otherwise any formula associated with uncertainty handling may no longer be valid. There are two simple conditions that can be imposed on the system that will allow this assumption. The first is that any experimental data used by the different panels is associated with experiments that are mutually independent of each other. A second kind of legitimate data set is one which is observational but complete. Both these scenarios lead to a likelihood which separates over the different panel parameter vectors. This assumption will then continue to be valid as the actual incident proceeds, provided that information about the progress of the incident is complete until that point. For a discussion of this and some related issues, see [7].

There are two other assumptions. The demand that there is only one expert for each panel can be initially generalized if the various panel members can agree to deliver their opinions in terms of a *single* probability distribution. In practice this may be difficult to achieve, or even inappropriate to the context. However recent developments in Decision support theory have made such delivery at least plausible. In particular, the elicitation of the quantities necessary to compute the scores of the individual panels can be performed during other Facilitated Workshops, in which only the experts concerning that particular

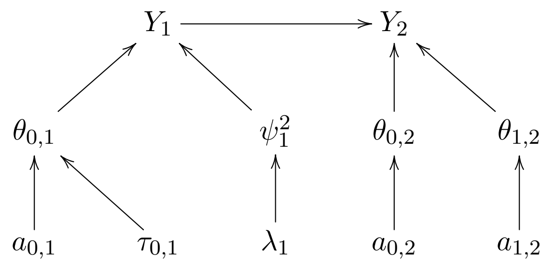

Figure 5: Network for the simplified example of a nuclear DSS

module meet and discuss the features of the models available. We note that the recent explosion of available Bayesian software has made this argument much more achievable, because scientists are now much more familiar with such probabilistic outputs.

Lastly, if we move our discussion from the additive case to the multi-linear case for utility functions, we add interactions between the preferences of different panels. Sadly, often to be realistic it is necessary to consider this situation. In this case the expected utility might no longer lead to the useful factorization of the expected utility we employed in this example. Fortunately in the next subsection we will show the polynomial associated to a simplification of the running example for a multi-linear factorization and the issues that may arise from the misuse of uncertainty in this scenario.

5 Consequences of the omission of uncertainty

Consider a simplification of the example in Figure 2, in which only the human health and costs modules are included. Assume also that each of the two modules consists of a single random variable. The variable for human health is Y_1, say, while the one for costs is Y_2: they can correspond for example to the number of people experiencing negative symptoms and the financial implications to the government, respectively. Their relationship is described by the network in Figure 5: in this case the graph includes the relationships among the parameters that will appear in the expect utility function when the variables are define as in (1).

Assume that $\mathrm{U}(y_1, y_2) = y_1 y_2$ in this case (a multilinear factorization without the additive component) and that the decisions available are either to evacuate (d_1) or to provide protective measures (d_2). It can be shown that in this case the expected utility is of the form

$$a_{0,2}a_{0,1} + a_{1,2}[\lambda_1 + \tau_{0,1} + a_{0,1}^2], \tag{3}$$

where all the quantities in the previous expression depend on d. The optimal decision will be then the one minimizing expression (3) (since the attributes represent negative consequences, the optimal decision is the one minimizing the expected utility). Now assume that the values provided for the decision of evacuation are:

$$a_{0,1}(d_1) = 2, \quad a_{0,2}(d_1) = 2.5, \quad a_{1,2}(d_1) = 2, \quad \lambda_1(d_1) = 4, \quad \tau_{0,1}(d_1) = 4,$$

while for the protective measures option we have that:

$$a_{0,1}(d_2) = 2.5, \quad a_{0,2}(d_2) = 2, \quad a_{1,2}(d_2) = 2 \quad \lambda_1(d_2) = 2, \quad \tau_{0,1}(d_2) = 1.$$

For the purposes of this simple example, which aims to show the consequences of omitting the uncertainty of the estimates, we will not focus on the heterogeneity of the experts that provide these estimates and we will simply assume that they are given by a single person. The hyper-parameter $a_{0,1}$ corresponds to the expectation of the expected number of people with adverse symptoms, while $a_{0,2}$ represents the expectation of expected amount of financial cost. It is then reasonable to provide a higher value for $a_{0,1}$ in the case of protective measures than in the evacuation scenario. By the same logic, even though protective measures may lead to more expenses for medical cures, the costs of an evacuation are incredibly high and significantly greater than for the other available option. The hyper-parameter $a_{1,2}$ represents the strength of the causal relationship between the two variables. A priori, the expert(s) may not be very well informed about this quantity and decide to provide the same value for the two scenarios available. The expectation of the variance of the number of people with negative symptoms is λ_1: in the case of delivering protective measures this value is definitively lower, since it is easier to provide shelter to most of the population. On the other hand during an evacuation, it may not be possible to help all the people in danger and only a subset of them may be moved to safer areas: as a consequence, the expectation of the variance of Y_1 for decision d_1 is higher. The last hyper-parameter $\tau_{0,1}$ corresponds to the variance of the expectation of Y_1 and, by the same argument, it is higher in the evacuation case than in the protective measures one.

In Section 2 we noted that, once the expectations are provided, then it is possible to treat them as known quantities, so that it is very close to a deterministic one. In this example the hyper-parameters expressing expectations are $a_{0,1}$, $a_{0,2}$ and $a_{1,2}$, while $\tau_{0,1}$ and λ_1 refer to variances. To show the issues deriving from the suppression of the uncertainty of the estimates, imagine that the values for $\tau_{0,1}$ and λ_1 are set to zero, so that the expression to minimize is now

$$a_{0,1}a_{0,2} + a_{1,2}a_{0,1}^2.$$

In this case, plugging in the values provided by the appropriate experts, the expected utility for the evacuation scenario is 13, while the value for protective measures delivering is 17.5: the optimal choice, when no variation in the estimates is allowed is to evacuate the population. Consider now the actual expected utility expression as in (3), including the second-order uncertainty. In this situation the expected utility of d_1 is 29, while the expected utility under d_2 is 23.5, from which it follows that the optimal decision consists of delivering protective measures. The misjudgment of uncertainty in this case has led to an indefensible decision, highlighting the need to handle uncertainty at every level.

6 Discussion

The implementation of Bayesian decision theoretic techniques into DSSs has been in many cases challenging and often considered too complex to be actually developed. In this paper we have shown that there are some conditions where the non-deterministic framework do not bring too many issues in the modeling of decision-making problems. In these cases software, ad hoc or already implemented methods, can be used to introduce uncertainty in the single modules. There are also several pieces of available software which encodes statistical graphical modeling that can be embedded into the DSS to describe the relationships among the modules. Moreover, we have shown through simple examples the threats to rationality that can arise from an omission of the uncertainty component in a decisional process. We note that, although uncertainty handling concerning most nuclear DSSs has vastly improved after recent years [25], the formal incorporation of these methods within the composite whole is still in its infancy. The methodological developments illustrated in this paper are now being used to structure protocols for Decision Conferencing of DSSs improvements in this important area.

Due to the space constraint we have had to restrict ourselves to the discussion of non dynamical systems at a single decision point. However, the technologies we are developing do not require such non-dynamic systems to be valid. In fact a technology that generalizes the apparatus discussed here and incorporates these necessary features is straightforward to specify and will reported in a later paper. For example, the multi-regression dynamical model [24] guarantees under a certain set of conditions the modularity of the probability distribution over the network at every time point. Moreover, the study of decisional processes, such as the Markov decision process, can also be incorporated into the modeling of several time decision points.

To aid the efficient application of more general techniques, the authors are also currently investigating the use of computer algebra to simplify the polynomials resulting from the expected utilities when the relationships through the modules are defined as in (1). In the last few years computer algebra and algebraic geometry have been used with success in different areas of statistics and we believe that such a technology can solve many of the problems faced in DSSs. Indeed, the expression resulting from the expected utility computations are often intractable and a higher level of abstraction seems to be necessary in order to be able to identify the relevant quantities for the individuation of an optimal policy.

Finally we note that probabilistic DSSs that use the technologies we employ here are not restricted to the management of nuclear emergency response. The authors are also currently applying their methods to network of expert judgments associated with food security. Again details of this new application will be reported in a future paper.

Acknowledgment

The authors want to thank Simon French for the suggestions and the references pointed out throughout the writing of this paper.

References

[1] G. Cooper and C. Yoo. "Causal discovery from a mixture of experimental and observational data." In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 1999, pp. 116–125.

[2] R. Cowell et al. *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer, 2007.

[3] A. Daneshkhah and J. Smith. "Multicausal Prior Families, Randomisation, and Essential Graphs." In: *Probabilistic Graphical Models* (2002).

[4] J. Ehrhardt. *Improvement, extension and integration of operational decision support systems for nuclear emergency management (DSSNET)*. Tech. rep. Report FZKA-7122, 2005.

[5] J. Ehrhardt et al. "Development of RODOS, A Comprehensive Decision Support System for Nuclear Emergencies in Europe-An Overview." In: *Radiation Protection Dosimetry* 50.2-4 (1993), pp. 195–203.

[6] S. Eom and H. Min. "The contributions of multi-criteria decision making to the development of decision support systems subspecialties: an empirical investigation." In: *Journal of multi-criteria decision analysis* 8.5 (1999), pp. 239–255.

[7] Á. Faria and J. Smith. "Conditionally externally Bayesian pooling operators in chain graphs." In: *The Annals of Statistics* 25.4 (1997), pp. 1740–1761.

[8] S. French. "Aggregating expert judgement." In: *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas* 105.1 (2011), pp. 181–206.

[9] S. French. "Uncertainty and imprecision: Modelling and analysis." In: *Journal of the Operational Research Society* (1995), pp. 70–79.

[10] S. French. "Uncertainty modelling, data assimilation and decision support for management of off-site nuclear emergencies." In: *Radiation protection dosimetry* 73.1-4 (1997), pp. 11–15.

[11] S. French and D. Insua. *Statistical decision theory. Kendall's Library of Statistics 9*. Arnold Publishers, London, 2000.

[12] S. French, J. Maule and K. Papamichail. *Decision behaviour, analysis and support*. Cambridge University Press Cambridge, UK, 2009.

[13] S. French, D. Ranyard and J. Smith. *Uncertainty in RODOS*. University of Leeds, School of Computer Studies, 1995.

[14] C. Genest. "A characterization theorem for externally Bayesian groups." In: *The Annals of Statistics* (1984), pp. 1100–1105.

[15] M. Goldstein and D. Wooff. *Bayes Linear Statistics, Theory & Methods*. Vol. 635. Wiley, 2007.

[16] F. Jensen and T. Nielsen. *Bayesian networks and decision graphs.* Springer Verlag, 2007.

[17] J. Kaiser et al. "Data assimilation approaches in the EURANOS project." In: *Radioprotection* 45.5 (2010), pp. 123–131.

[18] R. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge University Press, 1993.

[19] S. Lauritzen. *Graphical models.* Vol. 17. Oxford University Press, USA, 1996.

[20] A. Madansky. *Externally Bayesian groups.* Tech. rep. Rand Corporation, 1964.

[21] J. Pearl. *Causality: models, reasoning and inference.* Vol. 29. Cambridge Univ Press, 2000.

[22] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, 1988.

[23] D. Pennock and M. Wellman. "Graphical models for groups: Belief aggregation and risk sharing." In: *Decision Analysis* 2.3 (2005), pp. 148–164.

[24] C. Queen and J. Smith. "Multiregression dynamic models." In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1993), pp. 849–870.

[25] W. Raskob. "Overview on the model development carried out in the EURANOS project." In: *Radioprotection* 45.5 (2010), pp. 47–54.

[26] W. Raskob, F. Gering and V. Bertsch. "Approaches to visualisation of uncertainties to decision makers in an operational Decision Support System." In: *Proceedings of the 6th International Conference on Information Systems for Crisis Response and Management.* 2009.

[27] J. Smith. *Bayesian decision analysis: principles and practice.* Cambridge University Press, 2010.

[28] D. Spiegelhalter et al. "Bayesian analysis in expert systems." In: *Statistical science* (1993), pp. 219–247.

[29] P. Spirtes, C. Glymour and R. Scheines. *Causation, prediction, and search.* Vol. 81. MIT press, 2001.

[30] S. Sullivant, K. Talaska and J. Draisma. "Trek separation for Gaussian graphical models." In: *The Annals of Statistics* 38.3 (2010), pp. 1665–1685.

[31] S. Thykier-Nielsen, S. Deme and T. Mikkelsen. "Description of the atmospheric dispersion module RIMPUFF." In: *Riso National Laboratory, PO Box* 49 (1999).

[32] B. Walle and M. Turoff. "Decision support for emergency situations." In: *Handbook on Decision Support Systems 2* (2008), pp. 39–63.

[33] R. Weistroffer, C. Smith and S. Narula. "Multiple criteria decision support software." In: *Multiple criteria decision analysis: state of the art surveys* (2005), pp. 989–1009.

[34] D. Williamson and M. Goldstein. "Bayesian policy support for adaptive strategies using computer models for complex physical systems." In: *Journal of the Operational Research Society* 63.8 (2011), pp. 1021–1033.

[35] B. Wisse, T. Bedford and J. Quigley. "Expert judgement combination using moment methods." In: *Reliability Engineering & System Safety* 93.5 (2008), pp. 675–686.

A Bayesian updating

We will demonstrate in this appendix how the updating is performed for one of the regressions that define the Gaussian distribution over the graph $\mathcal{G}$ as in (1). The equation for the i-th variable can be rewritten as

$$Y_i = \mathbf{X}_i \boldsymbol{\theta}_i + \varepsilon, \tag{4}$$

where $\mathbf{X}_i$ is a row vector of $k-1$ explanatory variables (we are including a costant term), $\boldsymbol{\theta}_i$ is a column vector of k parameters and $\varepsilon \sim N(0, \psi_i^2)$. Thus, the $k-1$ explanatory variables are the parents in the network and $\boldsymbol{\theta}_i$ contains all the regression parameters.

In a full Bayesian framework the experts need to define prior distributions for the parameters. One common choice is to pick *conjugate* distributions, that is ones for which the resulting posterior distribution is in the same family. For a Bayesian Gaussian linear model, as the one we are considering in (4), it is possible to find a conjugate prior, which is the normal-inverse gamma, as we will show.

Assume that, after an observational study or a manipulation (we have seen that in the causal framework there is no distinction among the two), a sample consisting of n observations is available, where now $\mathbf{Y}_i$ is a $n \times 1$ column vector, $\mathbf{X}_i$ is a $n \times k$ matrix and $\varepsilon \sim N(0, \psi_i^2 I)$. Also recall that in this setting the likelihood is defined to be

$$L(\boldsymbol{\theta}_i, \psi_i^2 | \mathbf{X}_i, \mathbf{Y}_i) \propto \left(\psi_i^2\right)^{-n/2} \exp\left[-\frac{1}{2\psi_i^2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)'(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)\right].$$

We also define the priors. Recall that, if $(\boldsymbol{\theta}_i, \psi_i^2) \sim NIG(\boldsymbol{\mu}_i, \Sigma_i, a_i, b_i)$, where $\Sigma_i = \psi_i^2 V_i$ is positive-semidefinite and symmetric matrix, then its density is

$$f(\boldsymbol{\theta}_i, \psi_i^2) = \frac{1}{2\pi|\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)'V_i^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)\right) \frac{b_i^{a_i}}{\Gamma(a_i)}(\psi_i^2)^{-(a_i+1)} \exp\left[-\frac{b_i}{\psi_i^2}\right].$$

We now show that this a conjugate prior for a normal linear model. Using Bayes theorem we have that

$$\begin{aligned} f(\boldsymbol{\theta}_i, \psi_i^2 | \mathbf{Y}_i, \mathbf{X}_i) \propto & \left(\psi_i^2\right)^{-n/2} \exp\left[-\frac{1}{2\psi_i^2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)'(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)\right]|\Sigma_i|^{-1/2} \\ & \exp\left[-\frac{1}{2\psi_i^2}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)'V_i^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)\right]\left(\psi_i^2\right)^{-a_i-1} \exp\left[-\frac{b_i}{\psi_i^2}\right] \\ \propto & \left(\psi_i^2\right)^{-n/2-1/2-a_i-1} \exp\left[-\frac{Q}{2\psi_i^2}\right], \end{aligned}$$

where Q is equal to

$$\begin{aligned} Q =& (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)'(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\theta}_i) + (\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)'V_i^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i) + 2b_i \\ =& \boldsymbol{\theta}_i'(\mathbf{X}_i'\mathbf{X}_i + V_i^{-1})\boldsymbol{\theta}_i + \boldsymbol{\theta}_i'(\mathbf{X}_i'\mathbf{Y}_i + V_i^{-1}\boldsymbol{\mu}_i) \\ &+ (\mathbf{Y}_i'\mathbf{X}_i + \boldsymbol{\mu}_i'V_i^{-1})\boldsymbol{\theta}_i + (\mathbf{Y}_i'\mathbf{Y}_i + \boldsymbol{\mu}_i'V_i^{-1}\boldsymbol{\mu}_i + 2b_i) \\ =& (\boldsymbol{\theta}_i - \boldsymbol{\mu}_i^*)'(V_i^*)^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i^*) + 2b_i^{**}. \end{aligned}$$

Note that we have defined

$$V_i^* = (V_i^{-1} + \mathbf{X}_i^{'}\mathbf{X}_i)^{-1},$$

$$\boldsymbol{\mu}_i^* = V_i^*(V_i^{-1}\boldsymbol{\mu}_i + \mathbf{X}_i^{'}\mathbf{Y}_i),$$

$$b^{**} = \mathbf{Y}_i^{'}\mathbf{Y}_i + \boldsymbol{\mu}_i^{'}V_i^{-1}\boldsymbol{\mu}_i + 2b_i - (\boldsymbol{\mu}_i^*)^{'}(V_i^*)^{-1}\boldsymbol{\mu}_i^*.$$

We can consider again the posterior distribution to derive that

$$\begin{aligned} f(\boldsymbol{\theta}_i, \psi_i^2|\mathbf{X}_i, \mathbf{Y}_i) \propto \left(\psi_i^2\right)^{-n/2-a_i-1} \exp\left[-\frac{b_i^{**}/2}{\psi_i^2}\right] \times \left(\psi_i^2\right)^{-1/2} \\ \exp\left[-\frac{1}{2\psi_i^2}((\boldsymbol{\theta}_i - \boldsymbol{\mu}_i^*)^{'}(V_i^*)^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i^*))\right] \\ \sim NIG(\boldsymbol{\mu}_i^*, \Sigma_i^*, a_i^*, b_i^*), \end{aligned}$$

where $a_i^* = a_i + n/2$, $\Sigma_i^* = \sigma_i^2 V_i^*$ and $b_i^* = b_i^{**}/2$.