Advances and Applications in Statistics

Published Online: April 2015
http://dx.doi.org/10.17654/ADASFeb2015_145_161
Volume 44, Number 2, 2015, Pages 145-161 **ISSN: 0972-3617**

OPTIMIZING PROBABILITY COMPUTATIONS ON DYNAMIC PROCESSES USING GRAPHICAL MODELS

Ali S. Gargoum

Department of Statistics
UAE University
Al Ain, United Arab Emirates
e-mail: alig@uaeu.ac.ae

Abstract

Graphical modeling (GM) plays an important role in providing efficient probability calculations in high dimensional problems (computational efficiency). In this paper, we address one of such problems where we discuss the fragmenting environmental Puff models and some distributional assumptions concerning models for the instantaneous, emission readings and for the fragmenting process. A graphical representation in terms of a junction tree of the conditional probability breakdown of puffs and puff fragments is proposed.

1. Introduction

Graphical models, as statistical models, embodying a collection of marginal and conditional independencies which may be summarized by means of a graph, are quickly becoming an integral part of modern statistics. The graphical representation of a statistical model can help in many ways: the graph provides an effective means for elicitation and simplification of a

Received: November 14, 2014; Accepted: November 27, 2014
2010 Mathematics Subject Classification: Primary 62-09; Secondary: 65C62.
Keywords and phrases: graphical models, influence diagrams, junction trees, probability propagation, Puff models.
Communicated by K. K. Azad

problem, it depicts the dependency structure posited in the model and it may be transformed into a structure that can be used for efficient calculations of various quantities of interest. Graphical methods have been used in the 1980s for the analysis of statistical problems where no decision variables or utilities are explicitly represented. In a series of papers by (Darroch et al. [1], Lauritzen et al. [6], Kliveri et al. [5], Lauritzen et al. [6] and Lauritzen and Wermuth [7]), the authors addressed the problem of how graphs such as influence diagrams can help in understanding the conditional independence properties that a given factorization of a probability density implies. Another issue of importance is how graphs can be used to perform efficient probability calculations in high dimensional problems (computational efficiency). This issue is discussed in a number of papers by (Kim and Pearl [4], Pearl [11], Lauritzen and Spiegelhalter [8], Spiegelhalter et al. [18] and Smith and Anderson [17]). In Section 2, we give some graph-theoretic results and a background material on graphs, which are necessary for the development of the paper. In Section 3, we show how to propagate information on junction trees. Section 4 describes an environmental application of a high dimensional process, namely, the atmospheric fragmenting Puff models. In this section, we propose a graphical representation of the conditional probability breakdown of puffs and puff fragments as a junction tree representation of a high dimensional problem. In Section 5, we give an example of clique representation for puff distributions. Section 6 concludes the paper.

2. Background Material

This section introduces some graph-theoretical terms, which will be used in the paper. A network or *graph* is a pair $G = (V, E)$ that consists of a finite set of vertices $V = 1, 2, ..., v$ and a set of edges (arcs) $E \subseteq V \times V$ of ordered pairs of distinct vertices. An edge from vertex *i* (parent) to vertex *j* (child) is a *directed* edge (arrow) denoted by $i \to j$ if $(i, j) \in E$ and $(j, i) \notin E$. If both (i, j) and (j, i) are $\in E$, then the edge between *i* and *j* is *undirected* (line). If the graph has only undirected edges, then it is *undirected*

graph and if all edges are directed, then the graph is said to be a *directed graph*. A *path* of length $m \geq 0$ from i to j is an ordered sequence $(i = i_1, i_2, ..., i_m = j)$ of distinct vertices $i_1, i_2, ..., i_m$ such that (i_l, i_{l+1}) is in E for each $l = 1, 2, ..., m$. If there is a path from i to j, then we say that i *leads* to j. A subset $C \subseteq V$ is said to be an (i, j) *separator* if all paths from i to j intersect C. The subset C is said to *separate* A from B if it is an (i, j) separator for every $i \in A$, $j \in B$. For $A \subseteq V$, the set of parents of A denoted by $P_a(A)$ is the set of all these vertices in V, but not in A that have a child in A. An *m-cycle* is a path of length m with the exception that the end points are equal; that is, $i = j$. A graph is *acyclic* if it has no cycles.

2.1. Influence diagrams

An influence diagram (ID) is a schematic representation of conditional independence relationships. It is used for deducing new independencies from those used in the construction of the diagram. Influence diagrams were first developed in the mid 1970s by Miller et al. [10]. Howard and Matheson [3] extended the theory to decision analysis. Shachter [13] gave a procedure for evaluating a decision problem using an influence diagram. In this section, we present a brief introduction on how to use influence diagrams, as a modeling framework, that underpins a probability distribution in order to learn about and calculate various quantities of interest efficiently. We begin by defining a chance influence diagram.

In graph-theoretic terms, a chance influence diagram or influence diagram (ID) is a directed graph $G = (V, E)$, where V is a set of nodes represented by circles and called *chance nodes* and E is the set of directed edges or arrows joining these nodes. Chance nodes label random variables (uncertain) quantities relevant to the problem being modeled and directed edges represent probabilistic dependencies.

A chance node labels a random variable X_1 must be a *parent* of a chance node labels a random variable X_2 if and only if the distribution of the random variable X_2 is calculated conditional on the value of the random

variable X_1 and X_1, X_2 are not independent. The generalization to higher dimensions is given below.

Let $\boldsymbol{X} = (X_1, ..., X_m)$ be an ordered set of m random variables with a joint probability function

$$p(\boldsymbol{x}) = p(x_1)\prod_{r=2}^{m} p(x_r \mid x_1, ..., x_{r-1}). \tag{1}$$

Suppose $p(x_r \mid x_1, ..., x_{r-1})$ is a function of x_r and the parent set $P(r) \subseteq \{x_1, ..., x_{r-1}\}$ only. This will imply that given $P(r)$, X_r is independent of $R(r)$, where

$$R(r) = \{X_1, ..., X_{r-1}\} \backslash P(r)$$

is the set of random variables listed before X_r, which do not appear explicitly in the conditional probability function $p(x_r \mid x_1, ..., x_{r-1})$. This can be expressed as in Dawid's [2] notation

$$X_r \perp\!\!\!\perp R(r) \mid P(r), \quad r = 2, ..., m. \tag{2}$$

Then the graph of an influence diagram over $X_1, ..., X_m$ is any directed graph with nodes representing random variables $X_1, ..., X_m$ satisfying property (2). Influence diagrams are clearly acyclic, because only nodes of lower index can be connected to nodes of higher index. As a simple illustration, suppose $\boldsymbol{X} = \{X_1, ..., X_8\}$. Then from (1),

$$p(\boldsymbol{x}) = p(x_1)\prod_{r=2}^{8} p(x_r \mid x_1, ..., x_{r-1}).$$

Suppose the parents are: $P(2) = \{X_1\}$, $P(3) = \{X_1, X_2\}$, $P(4) = \{X_3\}$, $P(5) = \{X_3, X_4\}$, $P(6) = \{\varnothing\}$ (the empty set), $P(7) = \{X_5, X_6\}$, $P(8) = \{X_7\}$. The influence diagram (G) of this example is given in Figure 1.

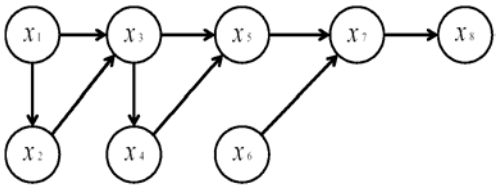

Figure 1. An influence diagram (ID) I.

2.2. Clique marginal representation

The clique marginal representation is one of many ways of specifying a joint probability distribution (see, for example, Lauritzen and Spiegelhalter [8] and Smith [14]). We start by identifying the cliques of an influence diagram G and $p(\boldsymbol{x})$ by looking at the small sets of variables called *precliques*, see Smith [15] of the form

$$\widetilde{C}(r) = \{X_r,\ P(r)\} \quad (P(1) = \varnothing), \quad 1 \le r \le m.$$

Then we delete from this collection any preclique $\widetilde{C}(r)$ for which there exists a $\widetilde{C}(k)$ $(k > r)$ such that

$$\widetilde{C}(r) \subseteq \widetilde{C}(k).$$

The remaining sets of variables after such deletions are called the *cliques* of $p(\boldsymbol{x})$ and G. This set of cliques will be denoted by $\mathcal{C} = \{C(1), ..., C(n)\}$, $1 \le n \le m - 1$.

After identifying the cliques, we can determine $p(\boldsymbol{x})$ in terms of the joint probability functions $p_1(\boldsymbol{x}), ..., p_n(\boldsymbol{x})$ over the cliques $\{C(1), ..., C(n)\}$. A sufficient condition for this is that $p(\boldsymbol{x} \in P(r)) > 0$ for each $\boldsymbol{x} \in P(r)$, $2 \le r \le m$ whenever $P(r) \ne \varnothing$. Then (1) can be expressed as:

$$p(\boldsymbol{x}) = \frac{\prod_{r=1}^{m} p(\boldsymbol{x} : \boldsymbol{x} \in \widetilde{C}(r))}{\prod_{r=2}^{m} p(\boldsymbol{x} : \boldsymbol{x} \in P(r))}, \tag{3}$$

where $p(\boldsymbol{x} \in P(r)) = 1$ if $P(r) = \varnothing$, the empty set.

Since by definition $p(\boldsymbol{x} : \boldsymbol{x} \in \widetilde{C}(r))$ and hence also $p(\boldsymbol{x} : \boldsymbol{x} \in P(r))$ can be obtained from $p(\boldsymbol{x} : \boldsymbol{x} \in C(k))$, where $C(k)$ is a clique of $p(\boldsymbol{x})$ such that $\widetilde{C}(r) \subseteq C(k)$, $2 \le r \le m$. Then (3) can be simplified to

$$p(\boldsymbol{x}) = \frac{\prod_{k=1}^{n} p_k(\boldsymbol{x})}{\prod_{k=2}^{n} q_k(\boldsymbol{x})}, \tag{4}$$

where $p_k(\boldsymbol{x})$ as defined above and $q_k(\boldsymbol{x}) = p(\boldsymbol{x} : \boldsymbol{x} \in P(r))$ for a $\widetilde{C}(r)$ remaining in the clique set such that $\widetilde{C}(r) = C(k), 1 \le k \le n$. A set of parents $P(r)$ associated with a clique $C(k)$ is called a *preseparator* and denoted by $\widetilde{S}(k)$, $2 \le k \le n$. The clique representation (4) of $p(\boldsymbol{x})$ has many computational advantages as we shall see later on.

2.3. Decomposable influence diagrams

An ID G is called *decomposable* if the set $P(X)$ of direct predecessors of X is completely connected (i.e., each node in $P(X)$ is connected by an edge to another node), this being true for all X in G. Figure 2 illustrates two graphs, one is decomposable and the other is not, since the parent nodes *a* and *b* are not joined.

Decomposable influence diagrams have several properties, which make them useful to study. One property is that their structure helps in propagating probabilities as the joint distribution of the system can be stored as margins of cliques. The cliques of a decomposable influence diagram can be ordered. Tarjan and Yannakakis [19] gave a simple technique for ordering nodes called the *maximum cardinality search* (*MCS*), so that in each of its disconnected subgraphs they satisfy the so called *running intersection property* (*RIP*) which states that *there exists an ordering* $C[1], ..., C[n]$ *of the cliques* $C(1), ..., C(n)$ *such that for all* $2 \le i \le n$

$$C[i] \cap \left[\bigcup_{j=1}^{i-1} C[j] \right] = S(i) \subseteq C(p_i),$$

for some p_i, $1 \le p_i \le i - 1$.

This means that the intersection of the *i*th clique with all the preceding ones is a subset of one of the preceding cliques.

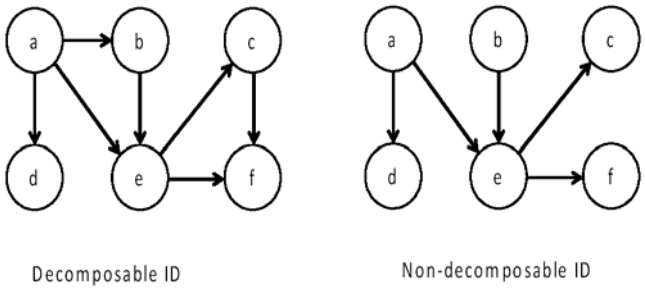

Figure 2. Graphs of decomposable and non-decomposable IDs.

3. Junction Trees and Probability Propagation

The clique representation (4) of $p(\boldsymbol{x})$ can be used efficiently to propagate information through the system, working indirectly with the margins $p_k(\boldsymbol{x})$ and $q_k(\boldsymbol{x})$ successively, updating them rather than updating the whole joint probability function $p(\boldsymbol{x})$ directly. This can be done by passing "simple messages" along the edges of a new graph called a *junction tree*, constructed from the influence diagram of $p(\boldsymbol{x})$. However, in the application cited below, distributions will not always remain decomposable. Because of this, we need to define a new graph called *junction graph*, which is an influence diagram on vectors of variables in the original influence diagram of the process. We then show that the definition of a junction tree is just a special case of the undirected version of a junction graph. The use of junction graphs will become apparent later in the paper. A formal definition of a junction graph follows.

A *junction graph* $\mathcal{G}$ of any density satisfying (4) is a directed graph with *n* nodes labeling the *n* cliques $C(1)$, ..., $C(n)$. There is an edge to node $C(i)$ from node $C(j)$, $i > j$ if and only if:

(i) $S(i) \bigcap C(j) \neq \varnothing$,

(ii) there exists no $j' < j$ such that

$$S(i) \cap C(j') \supseteq S(i) \cap C(j).$$

A *minimal junction graph* $\mathcal{G}$ is a junction graph which has no other junction graph $\mathcal{G}'$ as a proper subgraph.

In general, a joint probability function will have several junction graphs and minimal junction graphs over a chosen ordering of its cliques. An influence diagram and its junction graph are shown in Figure 3. The undirected versions of junction graphs are called *junction trees* when the separator of any clique is contained in exactly one previously listed clique or separator. Note that all junction graphs with no unmarried parents and the same undirected version (junction tree) embody an equivalent set of conditional independence statements.

In the case when $p(\boldsymbol{x})$ is decomposable, a collection of disconnected junction trees will be called a *junction forest*.

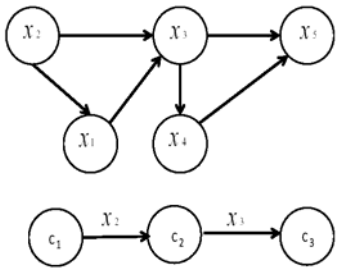

Figure 3. An influence diagram (ID) and its junction graph.

3.1. Propagation of information on junction trees

Let $\mathcal{C} = \{C(1), ..., C(n)\}$ denote the set of cliques of the joint probability function $p(\boldsymbol{x})$. Suppose we learn the values of some or all of the variables lying in some arbitrary clique $C(1) \in \mathcal{C}$ and we want to compute the conditional distribution of all variables in the system given a subset of variables in $C(1)$. To describe a propagation algorithm paralleling that given in Lauritzen and Spiegelhalter [8]. It is clear that we can obtain a new

probability function $p^*(\boldsymbol{x})$ of the variables $\boldsymbol{x}(1)$ in $C(1)$ from $p(\boldsymbol{x}(1))$ its original probability function using Bayes rule. Smith [15] showed how to update probabilities over the variables in the other cliques given the values of some of the variables in $C(1)$. The updating is possible using the junction tree of the system. For detailed discussions, see the above references.

4. Graphical Representation of Puff Models

4.1. Puff models

Most interest in the study of probabilistic networks has centered around problems where the junction trees (or variables in that tree) are fixed. However, there is a whole class of spatial temporal processes on which the efficient probability propagation algorithms developed for static networks can be used. For example, one of the methods of modeling atmospheric dispersion after an accidental release of radioactive pollutants is called the *Puff model*, Mikkelsen et al. [9]. According to this model, instead of assuming a continuous release from a source, it is assumed that the mass is released in a series of discrete puffs. These puffs can then be transported and dispersed around the local terrain based on the current wind field and local terrain. This method has been incorporated into the Risø-Meso-scale Puff model, RIMPUFF, Thykier-Nielsen and Mikkelsen [20]. To add to the accuracy of the RIMPUFF model, its designers added a further level of detail, puff splitting or *pentafurcation*. As puffs are released and transported over the local terrain, they grow in size. When their diameter reaches a chosen threshold, they can split into five smaller puffs. The mass associated with the parent puff is distributed amongst the children who are also smaller in size. In such examples, new variables (puffs) are being continually added so that, at any time in the process, the joint density of all variables up to that time satisfies equation (1).

4.2. Dynamic fragmenting of Puff models

The fragmenting Puff model described above can be reconstructed as a dynamic junction tree, Smith et al. [16]. In this section, we describe

briefly the reconstruction procedure starting with notation and distributional assumptions.

Let $m(t, \mathbf{l}) = m(t, l_1, ..., l_k)$ be the puff fragment which is the l_kth child of the l_{k-1}th child, ..., the l_2th of the l_1th child of the puff released at time t. In RIMPUFF, $1 \le l_i \le 5$, $1 \le i \le k$. The index k relates to the number of fragmentations that have taken place before fragment $m(t, \mathbf{l})$ appears. Let:

I_T denote the set of all puffs (puff fragments) appearing on or before time T,

$Q(\mathbf{l})$ denote the true mass under $m(t, \mathbf{l})$,

$\overline{\boldsymbol{Q}}(\mathbf{l})$ denote the vector of true masses under the set of the children of $m(t, \mathbf{l})$,

$\boldsymbol{Q}(\mathbf{l}) = (Q(\mathbf{l}), \overline{\boldsymbol{Q}}(t, \mathbf{l}))^T$.

Here we consider the following process.

The observation process. Let $\boldsymbol{Q}_T$ be the vector of masses of all puffs and puff fragments emitted on or before time T. Let $\boldsymbol{Y}(t, \boldsymbol{s})$ denote a vector of observations taken at time t at a selection of site(s) $\boldsymbol{s}$. Assume that $\boldsymbol{Y}(t, \boldsymbol{s})|\boldsymbol{\theta}(t, \boldsymbol{s})$ is independent of all other variables in the system. Here $\boldsymbol{\theta}(t, \boldsymbol{s})$ can be interpreted as a random vector relating to the actual mass at time t on site $\boldsymbol{s}$. As a simple process, $\boldsymbol{Y}(t, \boldsymbol{s})|\boldsymbol{\theta}(t, \boldsymbol{s})$ is defined to have a Gaussian distribution with mean $\boldsymbol{\theta}(t, \boldsymbol{s})$ and a fixed covariance matrix V. An important feature of Puff models is that at all points $(t, \boldsymbol{s})$ of the observation grid, $\boldsymbol{\theta}(t, \boldsymbol{s})$ can be written as

$$\boldsymbol{\theta}(t, \boldsymbol{s}) = F(t, \boldsymbol{s})\boldsymbol{Q}_t + \boldsymbol{\varepsilon}(t, \boldsymbol{s}).$$

The matrix $F(t, \boldsymbol{s})$ is a very complicated but known function of $(t, \boldsymbol{s})$, which defines the density of contamination contributed at sites $\boldsymbol{s}$ by each puff or puff fragment at time t. Each row of this matrix corresponds to the

weighings used in a dispersal model at a site which is a component of the vector of sites. Notice that $F(t, \boldsymbol{s})$ has non-zero components only on fragments that still exist and have not fragmented further. In practice, it is found that only a few puff fragments will be observed at a site at a given time, which implies that for most $(t, \boldsymbol{s})$ many components of each row of $F(t, \boldsymbol{s})$ will be zeros. The error process $\boldsymbol{\varepsilon}(t, \boldsymbol{s})$ will be Gaussian with zero mean and fixed covariance matrix U. In the particular case of observations at source $\boldsymbol{s} = \mathbf{0}$, where $\boldsymbol{\theta}(t, \boldsymbol{s})$ is a scalar, we set $\boldsymbol{\theta}(t, \boldsymbol{s}) = Q(t)$ and hence $\boldsymbol{\varepsilon}(t, \boldsymbol{s}) = 0$. To specify the joint distribution of $\boldsymbol{Q}_t$ at any time T, we need to specify the following processes.

The fragmentation process. This process assumes that a vector of mass fragments (children) $\overline{\boldsymbol{Q}}(\mathbf{l})$ of a parent $m(t, \mathbf{l})$ is independent of all masses $\boldsymbol{Q}_t$ given the mass $Q(\mathbf{l})$. This can be written as

$$\overline{\boldsymbol{Q}}(\mathbf{l}) \perp\!\!\!\perp \{\boldsymbol{Q}_t \backslash Q(\mathbf{l})\} \,|\, Q(\mathbf{l}).$$

Thus, the masses inherited by fragments depend only on the mass of the parent unfragmented puff and no other puff. Thus, to specify the joint distribution of puff fragments, it is only necessary to specify the conditional distribution of $\overline{\boldsymbol{Q}}(\mathbf{l}) \,|\, Q(\mathbf{l})$ for each puff/puff fragment $m(t, \mathbf{l})$. To model the dispersal of gas, these conditional distributions are usually chosen to conserve mass. For example, in RIMPUFF model, we set

$$E[\overline{\boldsymbol{Q}}(\mathbf{l}) \,|\, Q(\mathbf{l})] = \boldsymbol{\alpha} Q(\mathbf{l}),$$

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_5)^T,$$

$$\sum_{i=1}^{5} \alpha_i = 1, \quad \alpha_i > 0$$

and

$$Var[\overline{\boldsymbol{Q}}(\mathbf{l}) \,|\, Q(\mathbf{l})] = B^*,$$

where $\mathbf{l}^T B^* \mathbf{l} = 0$ and $\mathbf{l}$ denotes a vector of ones.

Obviously, if $\overline{\boldsymbol{Q}}(\mathbf{l})|Q(\mathbf{l})$ is chosen to be conditionally Gaussian, then this uniquely defines the joint distribution of $\boldsymbol{Q}_t$.

The emission process. The emission process is modeled as a dynamic linear model (DLM), West and Harrison [21], with state space $(Q(t), \boldsymbol{\psi}_t)^T$, where $\boldsymbol{\psi}_t$ is a vector of dummy variables. Special cases of these models set $\boldsymbol{\psi}(t)$ as null when the process becomes 1-dimensional; $Q(t)|Q(t-1) \sim N[Q(t-1) + \mu(t) - \mu(t-1), W]$, where W is a fixed variance and $\mu(t)$ is a trend term which is a function of time t. This is just a standard state space model on the univariate process $\{Q(t), t = 1, 2, ...\}$. Here, setting the conditional variance $V(t, 0)$ of $Y(t, 0)$, the source readings, given $Q(t)$ large relative to W gives a process, which after source readings are taken, still preserve strong relationship between masses $Q(t)$ and $Q(t-1)$. On the other hand, if $V(t, 0)$ is set to be negligible relative to W, then this assumes source readings $Y(t, 0), 1 \le t \le T$, are very accurate. As a consequence, it is not hard to prove that after observing $Y(1, 0), ..., Y(T, 0)$, $\{Q(1), ..., Q(T)\}$ are independent and future source emissions $Q(T + k), k = 1, 2, ...$ have expectation $\mu(T + k) - G^k[E[Q(T)] - \mu(T)]$ (say). When the shape of the emission profile is very vague, this can be modeled by setting $\mu(t) = 0$, $t = 1, 2, ...$ (a steady model). Here the forecast future emission $E[Q(T + k)] = E[Q(T)], k = 1, 2, ...,$ i.e., constant. If $Y(T, 0)$ is very accurate, i.e., $V(T, 0)$ is very small relative to W, then $E[Q(T + k)] \simeq Y(T, 0)$, the last observed emission.

4.3. Clique representation of Puff distributions

Let $\boldsymbol{X}_T$ denote a vector of state random variables of interest (vector of mass emissions and their fragments in our context) existing on or before time T. It is easy to check that because of the conditional independencies in the system, the joint density $p_T(\boldsymbol{x})$ of $\boldsymbol{X}_T$ can be written as

$$p_T(\boldsymbol{x}) = p(Q(1), \boldsymbol{\psi}(1)) \cdot \prod_{t=2}^{T} p(Q(t), \boldsymbol{\psi}(t) | Q(t-1), \boldsymbol{\psi}(t-1)) \cdot \prod_{I_T} p(\overline{\boldsymbol{Q}}(\mathbf{l}) | Q(\mathbf{l})), \tag{5}$$

where $Q(t)$, $\boldsymbol{\psi}(t)$, $\overline{\boldsymbol{Q}}(\mathbf{l})$, $Q(\mathbf{l})$ and I_T are as defined above. The density can be expressed in a suitable form, namely, the clique marginal representation form of equation (4). For an efficient propagation of probabilities, let

$$C^*(t) = \{Q(t), \boldsymbol{\psi}(t), Q(t+1), \boldsymbol{\psi}(t+1)\},$$

$$C(\mathbf{l}) = \{\boldsymbol{Q}(\mathbf{l}), Q(\mathbf{l}, l_1), ..., Q(\mathbf{l}, l_5)\}, \quad \mathbf{l} \in I_T,$$

where $C^*(t)$, $C(\mathbf{l})$ are cliques, $1 \le t \le T-1$.

Applying equation (4), $p_T(\boldsymbol{x})$ can be written as

$$p_T(\boldsymbol{x}) = \frac{\prod_{1 \le t \le T-1} p(C^*(t)) \prod_{\mathbf{l} \in I_T} p(C(\mathbf{l}))}{\prod_{2 \le t \le T-1} p(S(t)) \prod_{\mathbf{l} \in I_T} [p(Q(\mathbf{l}))]^{r_T(\mathbf{l})}}, \tag{6}$$

where $p(C^*(t))$ and $p(C(\mathbf{l}))$ denote, respectively, the joint densities of the variables in the cliques $C^*(t)$ and $C(\mathbf{l})$, $S(t) = \{Q(t), \boldsymbol{\psi}(t)\}$ and $r_T(\mathbf{l})$ is the number of offsprings of $Q(\mathbf{l})$ produced before or at time T. Using this simplified representation, the joint density $p_T(\boldsymbol{x})$ can be stored as a moderate number of joint densities of low dimension instead of a single density of a high dimension.

5. An Illustrative Example

The structure of the joint density $p_T(\boldsymbol{x})$ can be represented by a dynamic influence diagram, see, for example, Queen [12] and Smith et al.

[16]. The nodes of the ID are the random variables (or vectors) defined on the cliques. For example, the ID given in Figure 4 represents the conditional probability breakdown of puff and puff fragments in the early stages of an accidental release. As an example, let us assume that a source has emitted 4 puffs at time T, the first puff has pentificated, the 2nd and 5th fragments have then pentafurcated and further fragmentation has occurred on the 2nd offspring of the 2nd fragment. The second puff has also pentafurcated and its 2nd puff also split into 5. The 3rd and 4th puffs are yet to fragment.

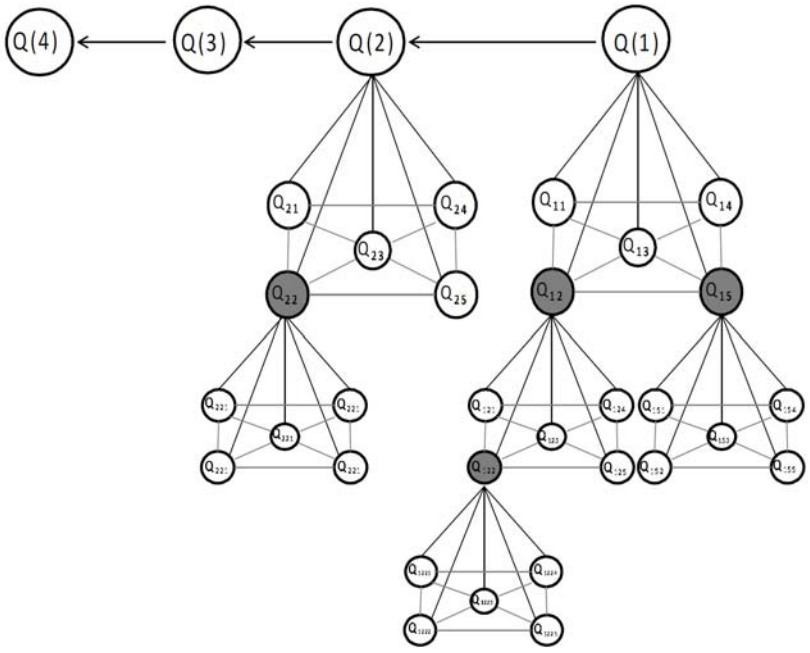

Figure 4. An ID of early emissions.

Here we note that it is easy to check that the ID of Figure 4 is *decomposable* (all parents of a given child are connected) with its cliques having the running intersection property (RIP), that is, at any time T, the cliques can be ordered as $C[1], ..., C[9]$ such that

$$C[i] \cap \left[\bigcup_{j=1}^{i-1} C[j] \right] = S[i] \subseteq C[p_i], \quad 2 \le i \le 9$$

for some $p_i, 1 \le p_i \le i-1$. Also, we note the following:

(i) If $C[i] = C^*(t)$, then $C[p_i] = C^*(t-1)$ and $S[i] = Q(t)$.

(ii) If $C[i] = C(\mathbf{l})$, if $\mathbf{l} = t$, then $C[p_i] = C^*(t)$ and $S[i] = \{Q(t)\}$, if $\mathbf{l} = (t, l_1, l_k)$, then $C[p_i] = C(t, l_1, l_{k-1})$ and $S[i] = \{Q(\mathbf{l})\}$.

Since the ID is decomposable, we can form a junction tree whose nodes are the cliques of $p_T(\boldsymbol{x})$ and whose node $C[i]$ is attached to node $C[p_i]$ by an edge which represents a separator $S[i]$. The junction tree which corresponds to the ID of Figure 4 is shown in Figure 5. A typical clique $\overline{C}[i]$ of this junction tree will have a probability defined conditionally in terms of a particular separator $\overline{S}[i]$ of the junction tree. That separator will take one of the forms:

(a) when $\overline{C}(i) = C^*(t)$, it will take the form $S(t)$ of equation (4),

(b) when $\overline{C}[i] = C(\mathbf{l})$, it will take the form $Q(\mathbf{l})$.

Now, an exact algorithm for quick absorption of information on such junction trees which evolve dynamically can easily be adopted.

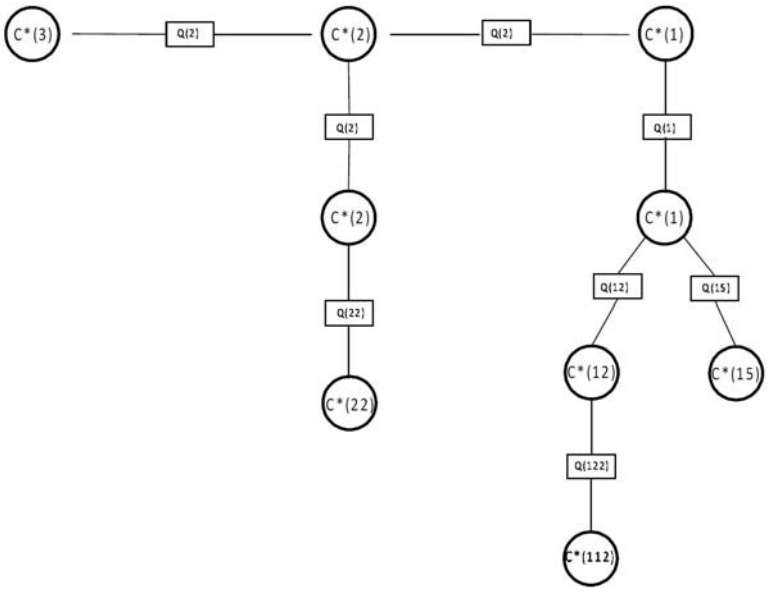

Figure 5. A junction tree of an ID of early emissions.

6. Conclusion

In this work, we showed how the continuous release of gas or radioactive material can be described as a series of puffs of contaminated mass emitted sequentially at discrete times and then dispersed and diffused (Puff models). We also described a stochastic version of these dispersal models. This version made it possible to incorporate and adjust to uncertain information about contamination readings at different sites. We then proceeded to show that all relevant uncertainties could be modeled by describing the evolution of puffs and puff fragments within the system by a high dimensional Gaussian process exhibiting many conditional independencies. Finally, a graphical representation (a clique representation) of these fragmentation processes was described. This representation is suitable for an efficient propagation of evidence as it arrives.

References

[1] J. N. Darroch, S. L. Lauritzen and T. P. Speed, Markov fields and loglinear interaction models for contingency tables, Ann. Statist. 8 (1980), 522-539.

[2] A. P. Dawid, Conditional independence in statistical theory, J. Roy. Statist. Soc. (Ser. B) 41 (1997), 1-31.

[3] R. A. Howard and J. E. Matheson, Influence diagrams, R. A. Howard and J. E. Matheson, eds., Reading on the Principles and Applications of Decision Analysis, Vol. II, Strategic Decisions Group, Menlo Park, California, 1981, pp. 719-762.

[4] J. Kim and J. Pearl, A computational model for combined causal and diagnostic reasoning in inference systems, Proc. 8th International Conference on Artificial Intelligence, 1983, pp. 190-193.

[5] H. Kliveri, T. P. Speed and J. B. Carlin, Recursive causal models, J. Aust. Math. Soc. (Ser. A) 36 (1984), 30-51.

[6] S. L. Lauritzen, T. P. Speed and K. Vijayan, Decomposable graphs and hypergraphs, J. Austral. Math. Soc. (Ser. A) 36 (1984), 12-29.

[7] S. L. Lauritzen and N. Wermuth, Mixed interaction models, Research Report R 84 - 8, Institute for Elektroniske Systems, Aalborg Universitetscenter, Denmark, 1987.

[8] S. L. Lauritzen and D. J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems (with discussion), J. R. Stat. Soc. Ser. B 50 (1988), 157-224.

[9] T. Mikkelsen, S. E. Larsen and S. Thykier-Nielsen, Description of RISO puff diffusion model, Nucl. Safety 67 (1984), 56-65.

[10] A. C. Miller, M. W. Merkhofer, R. A. Howard, J. E. Matheson and T. R. Rice, Development of automated aids for decision analysis, Stanford Research Institute, Menlo Park, California, 1976.

[11] J. Pearl, Fusion, propagation and structuring in belief networks, AI J. 29(3) (1986), 241-288.

[12] C. M. Queen, Bayesian graphical forecasting models for business time series, Ph.D. Thesis, Department of Statistics, University of Warwick, 1991.

[13] R. D. Shachter, Evaluating influence diagrams, Oper. Res. 34(6) (1986), 871-882.

[14] J. Q. Smith, Decision Analysis: A Bayesian Approach, Chapman and Hall, London, 1988.

[15] J. Q. Smith, Handling multiple sources of variation using influence diagrams, European J. Oper. Res. 86 (1995), 189-200.

[16] J. Q. Smith, S. French and D. Ranyard, An efficient graphical algorithm for updating the estimates of the dispersal of gaseous waste after an accidental release, Proceedings of Adaptive Computing and Information Processing, Unicom Seminar Ltd., 1995, pp. 583-610.

[17] J. Q. Smith and P. E. Anderson, Conditional independence and chain event graphs, Artificial Intelligence 172(1) (2008), 42-68.

[18] D. J. Spiegelhalter, A. P. Dawid, S. L. Lauritzen and R. G. Cowell, Bayesian analysis in expert systems, Statist. Sci. 8(3) (1993), 219-283.

[19] R. E. Tarjan and M. Yannakakis, Simple linear-time algorithms to test chordality of graphs, text acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, SIAM J. Comput. 13 (1984), 566-579.

[20] S. Thykier-Neilsen and T. Mikkelsen, RIMPUFF User Guide: Version 30, National Laboratory, Roskilde, Germany, 1991.

[21] M. West and P. J. Harrison, Bayesian Forecasting and Dynamic Linear Models, Springer-Verlag, 1997.