Uncertain Reasoning and Forecasting

Paul Dagum* **Adam Galper†** **Eric Horvitz‡** **Adam Seiver§**

June 29, 1995

Abstract

We develop a probability forecasting model through a synthesis of Bayesian belief-network models and classical time-series analysis. By casting Bayesian time-series analyses as temporal belief-network problems, we introduce dependency models that capture richer and more realistic models of dynamic dependencies. With richer models and associated computational methods, we can move beyond the rigid classical assumptions of linearity in the relationships among variables and of normality of their probability distributions.

We apply the methodology to the difficult problem of predicting outcome in critically ill patients. The nonlinear, dynamic behavior of the critical-care domain highlights the need for a synthesis of probability forecasting and uncertain reasoning.

Keywords uncertainty, probability forecasting, Bayesian belief networks, critical care

*Section on Medical Informatics, Stanford University School of Medicine, Stanford, California 94305. Also at Palo Alto Laboratory, Rockwell International Science Center, 444 High Street, Palo Alto, California 94301.

†Section on Medical Informatics, Stanford University School of Medicine, Stanford, California 94305.

‡Microsoft Research, One Microsoft Way, Redmond, WA 98052

§Department of Surgery, Stanford University Medical Center, Stanford, California 94305.

1 Introduction

Classical time-series methodologies are restricted in their ability to represent the general probabilistic dependencies and the nonlinearities of many real-world processes. With traditional methods, investigators are often forced to wrestle complex problems into parametrized models that may sacrifice explicit knowledge of domain uncertainty for simplicity. In spite of prolific research in uncertain reasoning over the past decade, there has been relatively little interaction between the statisticians interested in time-series analysis and the computer scientists studying the representation of uncertain knowledge.

A *Bayesian belief network* is an expressive knowledge representation for uncertain reasoning that employs a graphical structure to capture explicit dependencies among domain variables [35, 36]. Belief networks are the basis of diagnostic systems for many real-world applications [2, 3, 7, 26, 28, 29, 42]. Techniques for probabilistic inference in belief networks [8, 12, 16, 27, 34] and for specification of belief networks [11, 37, 38, 43] have emerged in response to the diversity of belief-network applications. More generally, applied statisticians have recognized the importance of graphical structures and the fundamental notion of conditional independence in the multivariate analysis of categorical data [45].

Forecasting has been addressed only intermittently over the last decade by the uncertain reasoning community [19]. Early work on continuous-variable influence diagrams by Kenley [31] yielded an influence-diagram representation of the Kalman filter, a state-space forecasting model. Since then, few techniques have been developed to address the forecasting problem directly. Several approaches for probabilistic reasoning about change over time [18, 30] and for temporal reasoning with belief networks and influence diagrams [4, 44, 39] have been proposed. Real-world applications of forecasting with belief networks include forecasting crude-oil prices [1]. We have developed a probability forecasting model through a synthesis of belief-network models and classical time-series analysis [14, 15]. The *dynamic network model (DNM)* is based

on the integration of Bayesian time-series analysis with generalizations of belief-network representation and inference techniques [13].

DNMs inherit the expressiveness of belief networks and are thus suited ideally for domains with categorical variables. The graphical structure of dependencies in DNMs represents the direct causal relationships between these variables. Thus, by casting Bayesian time-series analyses as temporal belief-network problems, we can introduce explicit dependency models that capture richer and more realistic models of dynamic dependencies.

Whereas linear models for time-series analysis suggest that observations are samples from a normal distribution, direct use of the time-series data to quantify the probability distributions over domain variables often suggests that uncertainty in the model and the causal mechanisms leads to nonnormal distributions. Nonnormal and nonlinear dependencies arise naturally in DNMs from the quantification of the variable dependencies. Furthermore, we can exploit the structure of these models to design inference algorithms that generate probability forecasts efficiently [13]. Thus, the richer models and associated computational methods allow us to move beyond rigid classical assumptions of linearity in the relationships among variables and of normality of their probability distributions without an inordinate increase in the complexity of forecasting. The probability distributions over forecast variables summarize the uncertainty in the model due to the exogenous disturbances and the implicit uncertainty in the causal mechanisms. In contrast to point forecasts, with probabilities for the alternative outcomes, we can now consider the risks of alternative decisions.

In this paper, we use DNMs to predict the outcome in critically ill patients. The problem is of considerable interest due to the tremendous costs associated with prolonged hospital stays for critically ill patients. The nonlinear, dynamic behavior of the critical-care domain highlights the need for a synthesis of probability forecasting and uncertain reasoning.

2 Overview

Belief networks are powerful graphical representations of probabilistic dependencies among domain variables. Effective diagnostic reasoning systems use belief networks to assign probabilities to alternative hypotheses about a patient's health—for example, MUNIN [2], ALARM [3], Pathfinder [26], VPnet [40], and QMR-DT [42]—or about the source of failure in complex machinery, including jet engines, electric power generators, and copy machines [28, 29]. These applications of belief networks have focused almost exclusively on static domains. In many applications, however, we seek to model the dynamic behavior of complex nonlinear system. Belief networks cannot adequately model such domains because (1) we cannot update the belief network specification with new evidence to reflect changes in the domain, (2) for complex domains with multiple lagged dependencies among domain variables we must specify extremely large contingency tables to construct the belief network, and (3) *probabilistic inference*, the inferential mechanism of belief networks that computes the probability distribution of a domain variable conditioned on the current evidence, is intractable for belief networks with large contingency tables.

The latter two problems are not unique to dynamic domains, and researchers have developed models that overcome at least the problem of extremely large contingency tables. *Additive* belief-network models is one such approach that reduces the specification of an extremely large contingency table into the specification of a few very small tables [13]. These models belong to the more general class of *additive models* [5, 24, 25] that approximate multivariate functions by sums of univariate functions. Fortuitously, additive belief-network models also overcome the intractability of probabilistic inference, and thus they provide a reasonable starting point to design a knowledge representation for dynamic domains.

Additive belief-network models form the basis of DNMs. The additive decomposition not only simplifies the specification of the model, but also provides DNMs with a method to update

their forecasts based on new evidence. Probabilistic inference in DNMs generates probability distributions of the forecast values of domain variables conditioned on the time-series of current observations. The additive structure makes inference efficient, and therefore, it makes forecasting efficient.

In Section 3 we present the forecast model. We begin with a formal definition of belief networks and then introduce additive belief-network models. These models form the basis of DNMs which we introduce following additive belief-networks.

In Section 4 we discuss the specification of DNMs. We discuss the semantics of the DNM graphical structure that it inherits from belief networks and we discuss algorithms to infer this structure from time-series data. We show how changes in the parameters of the additive decomposition lead to different methods of pooling information about the domain. In DNMs, these different methods translate into methods for combining probability forecasts. We conclude this section by showing how the additive decomposition can make probabilistic inference tractable in complex DNMs.

Finally, in Section 5 we apply DNMs to generate probability forecasts of the deranged physiology in a critically-ill patient.

3 The Forecast Model

We first discuss Bayesian belief networks, a knowledge representation for uncertain reasoning. We next show how additive generalizations of these representations have desirable properties for time-series analysis. From these additive generalizations we construct *dynamic network models*, a probability forecasting model for domains with qualitative or categorical information.

3.1 Belief Networks

Belief networks are graphical models of conditional independence in multivariate systems. A belief network consists of a directed acyclic graph (DAG) and a set of conditional probability functions. Let $X_1, ..., X_n$ represent the nodes of the DAG, and let $\pi(X_i)$ denote the set of parents of X_i in the DAG. The nodes of the DAG represent the variables of the belief network. The directed arcs in the DAG represent explicit dependencies between the variables. To complete the definition of a belief network, we specify for each variable X_i, an associated conditional probability function (table) denoted $\Pr[X_i|\pi(X_i)]$. The full joint probability distribution is given by [35, 36]

$$\Pr[X_1, ..., X_n] = \prod_{i=1}^{n} \Pr[X_i|\pi(X_i)]. \tag{1}$$

Probabilistic inference in belief networks refers to the computation of an *inference probability*—that is, $\Pr[\mathbf{X} = x|\mathbf{E} = e]$ for any given set of nodes $\mathbf{X}$ instantiated to value x and conditioned on observation nodes $\mathbf{E}$ instantiated to value e. Probabilistic inference in large multiply-connected belief networks is difficult. Complexity analyses show that both exact and approximate algorithms pose intractable problems in the worst case [10, 17]. Nevertheless, for many problems, inference approximation procedures provide useful estimates of posterior probabilities in acceptable computation times.

3.2 Additive Belief-Network Models

There are several disadvantages to using a belief network to forecast: (1) if a node at time t has multiple parent nodes at times $t-1, t-2, ...$, then the storage requirement for the conditional probability table may be excessive; (2) if we have a multivariate time series, the belief-network model may be complex with multiple, lagged dependencies, and consequently, probabilistic inference to generate forecasts will be slow, or even intractable; and (3) apart from the storage and computational demands, for large conditional probability tables, the available training

set must be very large to obtain reliable maximum likelihood estimates of the conditional probabilities.

These problems are not unique to belief network forecast models; real-world applications of belief-network models also encounter these problems. These limitations prompted the development of models of *causal independence* and *disjunctive interactions* [23, 32, 36] that reduce substantially the storage requirements of conditional probability tables. These models, however, do not address the problem of intractable inference. Elsewhere, we develop *additive belief-network models* that address the intractable-inference problems researchers encounter in large belief network applications [13]. In this section, we discuss these models, and in Section 4.3, we show that an exact inference algorithm can exploit the properties of the model to reduce the complexity of inference.

Additive belief-network models, like causal independence models, inherit properties from a more general class of *separable models*. In separable models, we can express the joint effect of the set of causes $X_1, ..., X_p$ on the dependent variable Y in terms of the effects of each individual cause. We assume that for each cause X_i, there exists an "off" state—that is, a state in which X_i has no bearing on the value of Y. We denote these *distinguished* states by x_i^*. Thus, the conditional probabilities $\Pr[Y|X_i, X_{j\neq i} = x_j^*]$, for $i = 1, ..., p$, represent the isolated effects of each X_i on Y. More generally, in separable models, we express the joint effect of the causes on the dependent variable in terms of the effects of sets of causes $\mathbf{X}_i$, $i = 1, ..., k$, that partition the set $\{X_1, ..., X_p\}$. We represent the isolated effect of each set $\mathbf{X}_i$ on Y by the conditional probability $\Pr[Y|\mathbf{X}_i, \mathbf{X}_{j\neq i} = \mathbf{x}_j^*]$.

Let y^* denote the "off" state of the variable Y. An additive belief-network model is a separable model that satisfies

$$\Pr[Y = y|X_1, ..., X_p] = \begin{cases} \sum_{i=1}^{k} \phi_i \Pr[Y = y|\mathbf{X}_i, \mathbf{X}_{j\neq i} = x_j^*] & \text{if } y \neq y^* \\ 1 - \sum_{y' \neq y^*} \Pr[Y = y'|\mathbf{X}_1, ..., \mathbf{X}_k] & \text{if } y = y^* \end{cases} \quad (2)$$

The parameters $\phi_i \geq 0$, for $i = 1, ..., k$, must satisfy $\sum_{i=1}^{k} \phi_i \Pr[Y|\mathbf{X}_i, \mathbf{X}_{j\neq i} = x_j^*] \leq 1$. More specific restrictions on the parameters depend on the domain; we discuss these in Section 4.2.

Like other separable models, to specify the conditional probability $\Pr[Y|X_1, ..., X_p]$ of an additive belief-network model we need to specify only the conditional probabilities of the k isolated effects. Thus, for a binary valued belief network, we reduce the size of the conditional probability table from 2^{p+1} to $\sum_{i=1}^{k} 2^{|\mathbf{X}_i|+1}$. Unlike other separable models, however, we show in Section 4.3 that additive decomposition of conditional probabilities allows us to improve the efficiency of belief-network inference algorithms.

We conclude this section with a brief discussion of the theory of nonparametric *additive models* [5, 24, 25]. We show that the decomposition of the conditional probability $\Pr[Y = y|X_1, ..., X_p]$ into the additive terms $\Pr[Y = y|\mathbf{X}_i, \mathbf{X}_{j\neq i} = x_j^*]$ is a special case of such models.

Additive models maintain the properties that make linear regression models attractive—they are additive in the predictor effects—yet, they are not constrained by assumptions of linearity in the predictor effects. Let $\mathbf{X}_{t-i} = \{x_{1t-i}, ..., x_{mt-i}\}$. An additive model is defined by

$$\mathbf{E}(y_t|\mathbf{X}_t, ..., \mathbf{X}_{t-k}) = \sum_{i=0}^{k} f_i(\mathbf{X}_{t-i}), \tag{3}$$

where the functions f_i are arbitrary. Thus, Equation 3, which expresses the property of additive models, follows directly from Equation 2:

$$\mathbf{E}[Y|X_1, ..., X_p] = \sum_{i=1}^{k} f_i(\mathbf{X}_i),$$

where

$$f_i(\mathbf{X}_i) = \phi_i \mathbf{E}[Y|\mathbf{X}_i, \mathbf{X}_{j\neq i} = x_j^*].$$

3.3 Dynamic Network Models

In this section we introduce DNMs to model dynamic domains. These models are additive belief-network models with variables indexed by time. As in additive belief-network models we

use an additive decomposition to express the conditional probabilities of the model; after each new observation, however, we reestimate the parameters of the decomposition.

Similarly to belief networks, to construct a DNM we first identify the dependencies among domain variables in the model. We use these dependencies to specify the directed acyclic graph (DAG) of the model. For example, Figure 1 represents a DAG of dependencies of a single variable Y_t dependent on the sets of variables $\mathbf{X}_{t-i} = \{X_{1t-i}, ..., X_{mt-i}\}$. When we construct the DAG of a domain, we specify only the explicit dependencies between domain variables. To quantify the extent of these dependencies we must specify the conditional probability functions of the model.

We next specify the conditional probability for the DAG of the node Y_t shown in Figure 1. We use the additive decomposition to specify the conditional probability,

$$\Pr[Y_t = y|\mathbf{X}_t, ..., \mathbf{X}_{t-k}] = \begin{cases} \sum_{i=1}^{k} \phi_{ti} \Pr[Y_t = y|\mathbf{X}_{t-i}, \mathbf{X}_{t-j,j\neq i} = \mathbf{x}^*_{t-j}] & \text{if } y \neq y^* \\ 1 - \sum_{y'\neq y^*} \Pr[Y_t = y'|\mathbf{X}_t, ..., \mathbf{X}_{t-k}] & \text{if } y = y^* \end{cases} \quad (4)$$

To estimate the conditional probabilities $\Pr[Y_t|\mathbf{X}_{t-i}, \mathbf{X}_{t-j,j\neq i} = \mathbf{x}^*_{t-j}]$, we can use either expert assessment or maximum likelihood estimates computed from available time-series data.

Equation 4 is analogous to the additive decomposition of Equation 2. We discussed in Section 3.2 the advantages of the additive decomposition to specify conditional probabilities for belief networks. In addition, for DNMs this decomposition provides a means to update the conditional probabilities with new evidence through reestimation of the parameters $\phi_{t1}, ..., \phi_{tk}$.

3.4 Forecasting

We show that forecasting in a DNM reduces to probabilistic inference in the *forecast model*. Let DNM_t denote the DNM with leading time slice at time t. The forecast model is the union of DNM_t over all times t. Consider for example the DNM shown in Figure 2a. This figure shows the dependencies between nodes at time t, X_t, Y_t, and their causes at time $t-1$, X_{t-1}, Y_{t-1}.

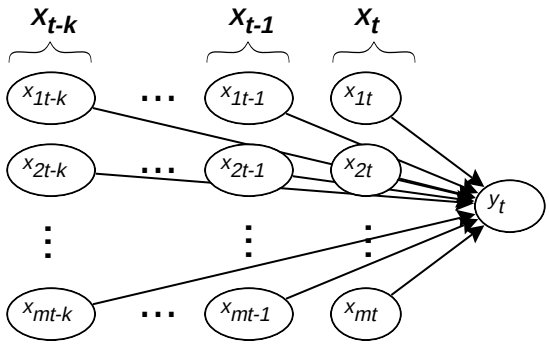

Figure 1: The dependency graph of a variable Y_t. Nodes denote model variables, and directed arcs between nodes denote dependencies.

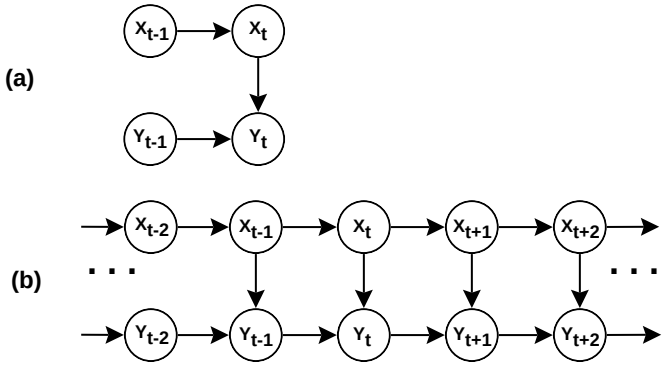

Figure 2: The DNM forecast model. (a) This structure represents a DNM with leading time slice at time t. (b) The forecast model is obtained from the union of the structures in (a) over all times t.

The forecast model for this DNM is the dependency structure shown in Figure 2b together with the conditional probabilities and parameters. Probabilistic inference in the forecast model yields the DNM forecasts.

Let $\mathcal{E}_t$ denote all the observations up to time t. In general, we allow missing observations in $\mathcal{E}_t$. To compute the forecasts for time $t+1$ given the evidence $\mathcal{E}_t$—that is, the probabilities $\Pr[X_{t+1}|\mathcal{E}_t]$ and $\Pr[Y_{t+1}|\mathcal{E}_t]$—we instantiate the forecast model with the observations $\mathcal{E}_t$. Nodes in the time slice $t+1$ represent the forecast nodes. We can truncate the forecast model at the

forecast nodes, since the uninstantiated future nodes do not affect the inference probabilities of the forecast nodes. Furthermore, if the observations $\mathcal{E}_t$ includes both nodes X_t and Y_t, then observations of previous nodes do not affect the forecast probabilities. In this case, the structure of the one-step-ahead forecast model reduces to the structure shown in Figure 2a. Probabilistic inference in this forecast model yields probability distributions for the forecast nodes. To compute the forecast probabilities for $t + k$, we proceed similarly. We truncate the forecast model at the $t + k$ forecast nodes. Probabilistic inference in this forecast model yields the probability distributions for forecast nodes for times $t + 1$ to $t + k$.

4 Model Specification

In Section 3 we discussed the DNM forecast model. We discuss in this section the specification of the model. In Section 4.1 we show that we can infer the DNM dependency structure from time-series data. In Section 4.2 we discuss the effect of the additive-decomposition parameters on the model's forecasts. Finally, in Section 4.3 we show how additive-decomposition of conditional probabilities reduces the complexity of probabilistic inference.

4.1 Dependency Structure Specification

Classical work in belief-network specification algorithms has focused explicitly on the construction of models for static domains [11, 37, 38, 43]. These algorithms automate the specification of belief networks from large databases of domain information. We can apply these same algorithms to specify models of dynamic domains. We decompose the task into two subtasks: (1) identification of the dependency structure of the model, (2) specification of the conditional probabilities. The second task is straightforward once we determine the dependency structure of the domain; we compute the conditional probabilities from maximum-likelihood estimates.

The dependency structure of a DNM, or a belief network, represents the set of causal

relationships between domain variables. These relationships differ significantly from the causal relationships suggested by classical time-series models of the domain. The latter are assessed from cross correlations between domain variables, whereas DNMs, or belief networks, use the probability distributions of the domain variables to infer causality.

We begin with a discussion of the different types of dependencies that can exist between variables in a domain. We distinguish among three types of dependencies: (1) explicit dependencies, (2) implicit dependencies, and (3) spurious correlations. Explicit dependencies comprise, for example, the directed arcs in a belief network. Similarly, the dependencies of a DNM represent explicit dependencies. These dependencies model a *direct* causal relationship between nodes.

Implicit dependencies model *indirect* causal relationships; they occur between nodes connected by a chain of explicit dependencies. Consider, for example, the belief network depicted in Figure 3a with three nodes X_{t-2}, X_{t-1}, Y_t. In this belief network, nodes X_{t-2} and Y_t are implicitly dependent. For these nodes, we observe that $\Pr[Y_t|X_{t-2}] \neq \Pr[Y_t]$, and therefore, they are dependent. We can distinguish this dependency from an explicit dependency, however, because if we observe node X_{t-1}, then $\Pr[Y_t|X_{t-1} = x, X_{t-2}] = \Pr[Y_t|X_{t-1} = x]$. Hence, a direct causal relationship cannot exist between these nodes.

Spurious correlations do not represent any causal relationship; they occur between nodes with a common parent. For example, in the belief network depicted in Figure 3b, nodes X_{t-1} and Y_t are correlated because of their common cause X_{t-2}. These nodes appear to be dependent because $\Pr[Y_t|X_{t-1}] \neq \Pr[Y_t]$. If, however, we observe node X_{t-2}, then $\Pr[Y_t|X_{t-1}, X_{t-2} = x] = \Pr[Y_t|X_{t-2} = x]$, demonstrating that the nodes are not explicitly dependent.

The preceding discussion suggests that we can distinguish explicit dependencies from other dependencies based on the conditional probabilities of the model. Two nodes are explicitly dependent if we cannot find a node, or more generally, a set of nodes, such that when we

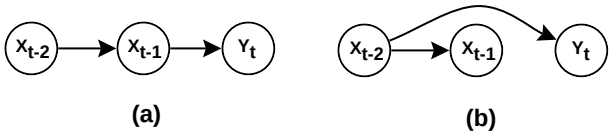

Figure 3: Dependency models. (a) This dependency structure exhibits an implicit dependency between X_{t-2} and Y_t. (b) This structure exhibits a spurious correlation between X_{t-1} and Y_t.

instantiate these nodes, we break their dependency [37, 38]. Thus, two nodes X and Y are explicitly dependent if and only if there does not exist a set of nodes $\mathbf{S}$, such that

$$\Pr[Y|\mathbf{S} = \mathbf{s}, X] = \Pr[Y|\mathbf{S} = \mathbf{s}]. \tag{5}$$

We can determine the dependency structure of a domain if we can compute the conditional probabilities between domain variables. We use the time-series data to derive maximum likelihood estimates of these probabilities. An alternative approach is to search over all possible dependency structures for the one that maximizes the likelihood of the observed data [11].

In contrast, classical approaches for specifying the domain dependencies, such as AR models, dynamic linear models, or transfer-function models, use cross correlations between the variables to construct the model. Thus, although the dependency structure of the dynamic linear model, shown in Figure 1, suggests explicit dependencies between the variables, these dependencies represent instead correlations between the variables. Consider, for example, the classical approach to constructing a single input-output transfer function model. Let x_t denote the input time series, and let y_t denote the output time series. For simplicity, we assume that the x_t time series is an uncorrelated process; thus, for all $i \neq j$, the cross correlations $\gamma_{xx}(k)$ between x_t and x_{t-k} are zero. To compute the transfer-function model for y_t,

$$y_t = v_0 x_t + \cdots + v_m x_{t-m},$$

we must estimate the impulse-response weights v_j. The classical approach estimates these weights from the cross correlations between y_t and the x_{t-k}. Because the x_t time series is un-

correlated, we get that $v_k = \gamma_{yx}(k)/\gamma_{xx}(k)$, where $\gamma_{yx}(k)$ denotes the cross correlation between y_t and x_{t-k}. Thus, the presence or absence of an arc in the dependency structure for this model depends on whether the response variable y_t is correlated with the predictor variable x_{t-k}. We cannot, however, use cross correlations to distinguish explicit dependencies from implicit dependencies or spurious correlations.

If the input time series x_t is correlated, then, provided that it is generated by a Gaussian stochastic process, a linear filter transforms it into a white-noise process. If, however, the stochastic process is not Gaussian, then a linear filter cannot uncorrelate the series. Computing the impulse-response weights for this process generally requires higher-order correlations. Like second-order correlations, higher-order correlations also cannot distinguish between the different types of dependencies. Therefore, these approaches to model construction cannot distinguish direct causality from indirect causality or spurious correlations.

4.2 Parameter Specification

Equation 2, which expresses the additive decomposition of the conditional probability between the outcome Y and the sets of causes $\mathbf{X}_1, ..., \mathbf{X}_k$, represents a model for combining probabilities. The theory of combining probability distributions has been extensively studied by the Bayesian community (eg., [22]). These combination rules range from linear combinations to multiplicative combinations, depending on the mutual dependence of the various alternative forecasts. This mutual dependence reflects the causal interactions among the causes of the common effect.

We can consider the set of causes $\mathbf{X}_i$ as the knowledge that an ith expert possesses about the domain. We express this expert's belief in the outcome $Y = y$ by the conditional probability

$$\Pr[Y = y | \mathbf{X}_i, \mathbf{X}_{j \neq i} = \mathbf{x}_j^*].$$

We would like to aggregate each agent's belief in the outcome into an overall belief. This aggregation depends on the degree of interdependency, or redundancy, between the various

sets of causes. Often, however, there exists complex dependencies between the different sets of causes, and determining the scheme that is appropriate for the degree of interdependency is the major difficulty encountered by aggregation schemes.

An additive decomposition expresses the composite belief in an outcome as the sum of each agent's belief weighted by some parameter ϕ_i. We show that these parameters can express diverse types of causal interactions between the different sets of causes. These different models of causal interactions lead naturally to different models for combining beliefs about the outcome Y.

We say that a set of causes $\mathbf{X}_i$ is *active* if $\mathbf{X}_i \neq \mathbf{x}^*$, and otherwise, the set is *inactive*.

4.2.1 Exclusive Causes

For exclusive causes, we show that if causes $\mathbf{X}_i$ are active, then the parameter $\phi_i = 1$, and if these causes are inactive, then $\phi_i = 0$. This result follows because, for exclusive causes, the contribution from causes $\mathbf{X}_i$ to the belief in an outcome $Y = y$ is independent of the contribution from other causes. Since the contribution to the belief from $\mathbf{X}_i$ is $\phi_i \Pr[Y_t = y|\mathbf{X}_i, \mathbf{X}_{j,j\neq i} = \mathbf{x}^*_j]$, it follows that the parameter ϕ_i is independent of the other causes. But when all the other causes are "off", by definition, the belief in an outcome $Y = y$ is $\Pr[Y = y|\mathbf{X}_i, \mathbf{X}_{j,j\neq i} = \mathbf{x}^*_j]$; that is, in this case the parameter $\phi_i = 1$. Therefore, $\phi_i = 1$ when other causes are active as well.

4.2.2 Alternative Causes

Suppose that different sets of causes of Y represent alternative explanations of the outcome of Y. These explanations typically share common background knowledge, or a common inferential mechanism, to arrive at some measure of support of the outcome of Y. Alternative explanations, therefore, do not increase our overall belief in an outcome; instead, they allow us to update our

belief by weighting the beliefs of each explanation according to its likelihood. Hence, we model the belief of Y by a convex combination of the beliefs of the alternative causes. This model, known as the *linear-opinion pool* [22], suggests that the parameters in Equation 4 must satisfy $\sum_{i=0}^{k} \phi_i = 1$.

4.2.3 Disjunctive Causes

Whereas each additional alternative cause of Y provides only an alternative explanation for, or degree of belief in Y, each additional disjunctive cause of Y increases monotonically our belief in Y. Disjunctive causes generalize exclusive causes because they model the information common to different sets of causes. We use the principle of inclusion-exclusion [41] to compute the parameters ϕ_i for disjunctive causes.

For disjunctive causes, if we further assume that the causes are mutually independent, then we derive the well-known *noisy-OR* model of causal interactions [23, 32, 36]. Noisy-OR models are models of *exception independence*; that is, the mechanism that may inhibit an outcome of Y when causes $\mathbf{X}_i$ are active, is independent of the mechanism that may inhibit this outcome when other causes $\mathbf{X}_j$ are active. Thus, noisy-OR is a product model of the inhibitors of an outcome.

To facilitate the exposition of disjunctive causes, we denote that causes $\mathbf{X}_i$ are active by the Boolean event $\mathbf{e}_i$. We let $\Pr[\mathbf{e}_i]$ denote this event's degree of support of the outcome $Y = y$; thus, $\Pr[\mathbf{e}_i] = \Pr[Y = y | \mathbf{X}_i, \mathbf{X}_{j,j\neq i} = \mathbf{x}_j^*]$. Thus, the belief in Y given the active disjunctive causes $\mathbf{X}_1, ..., \mathbf{X}_k$ is the degree of belief in the union of the events $\mathbf{e}_1, ..., \mathbf{e}_k$; that is, $\Pr[\mathbf{e}_1 \vee \cdots \vee \mathbf{e}_k]$. We can use the principle of inclusion-exclusion [41] to write

$$\Pr[\mathbf{e}_1 \vee \cdots \vee \mathbf{e}_k] = \sum_{i=1}^{k} (-1)^{i+1} \sum_{j_1,...,j_i} \Pr[\mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_i}],$$

where, for all i, the summation indices $j_1, ..., j_i$ are distinct. The probabilities $\Pr[\mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_i}]$ denote the degree of support of the outcome Y that is common to all the events $\mathbf{e}_{j_1}, ..., \mathbf{e}_{j_i}$.

Thus, the principle of inclusion-exclusion compensates the combined belief of the outcome by the degree of belief that is common to different sets of causes.

To compute the belief of Y, we must specify the probabilities $\Pr[\mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_i}]$. We explore two distinct specifications: (1) exclusive causes and (2) causal independence. If the causes are exclusive, then these probabilities are zero. In this case, $\phi_i = 1$, which is consistent with the previous analysis for exclusive causes.

The second specification, causal independence, leads to the noisy-OR model of causal interactions. Causal independence implies that $\Pr[\mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_i}] = \Pr[\mathbf{e}_{j_1}] \cdots \Pr[\mathbf{e}_{j_i}]$. Thus, in this model, we quantify explicitly the degree of dependency between alternative causes. The parameters for this model are

$$\phi_i = 1 + \sum_{l=1}^{m-1} (-1)^l \sum_{j_1,...,j_l \in \mathcal{S}_l} \frac{1}{l} \Pr[\mathbf{e}_{j_1}] \cdots \Pr[\mathbf{e}_{j_l}],$$

where $\mathcal{S}_l = \{1, ..., i-1, i+1, ..., l\}$, and the indices $j_1, ..., j_l$ in the summation are distinct. Thus, in the noisy-OR model of causal interaction, the parameter for the degree of belief in the proposition "$\mathbf{X}_i$ causes Y" depends on the degree of belief in the other causes.

More generally, the causal interactions between the domain variables may be intermediate between the extreme interactions we discussed. As more observations about the domain become available, Bayesian update of the parameters allows us to probe the range of causal interactions that can be expressed through an additive decomposition. If the causal interactions of the domain remain stable over time, and if they can be modeled by suitable choices of the parameters in an additive decomposition, then sequential Bayesian update of the parameters with new observations converges to these parameters.

4.3 Inference Algorithms

For complex applications, the size of belief-network models for dynamic domains prohibits tractable computation of inference probabilities [10, 17]. Dynamic network models, however,

inherit the properties of additive belief-network models [13]. As a result, dynamic network models benefit from an exact inference algorithm that exploits the additive decomposition to improve the efficiency of inference.

The inference algorithm for additive belief networks [13] is similar to Cooper's nested dissection algorithm for general probabilistic inference [9]. The algorithm decomposes the belief network into subnetworks using the additive decompositions. The Lauritzen-Speiglehalter (L-S) algorithm is used [34] to perform inference on the subnetworks. The decompositions render the subnetworks amenable to fast inference with the L-S algorithm. We then combine the results from each subnetwork inference to arrive at the desired inference probability. The run time of the algorithm depends on the decompositions in the model.

The L-S algorithm identifies subsets of related nodes called *cliques*. For each clique, the L-S algorithm computes the joint probability distribution over the nodes in the clique. Thus, to compute the joint distribution of a clique C, the L-S algorithm computes $\prod_{X_i \in C} |X_i|$ probabilities, where $|X_i|$ denotes the number of values of the categorical variable X_i. The running time of the L-S algorithm is exponential in the size of the largest clique. We show that in a dynamic network model, if the largest clique contains a node that has an additive decomposition, the running time to compute an inference can be reduced substantially.

Let Y be a node with parents $X_1, ..., X_{p-1}$. If a clique contains the node Y, then it must also contain the parents $X_1, ..., X_{p-1}$. Let C be the clique containing Y and its parents. If all the categorical variables have d values, then the size of the joint probability of the clique containing Y and its parents is d^p. However, if Y has an additive decomposition that partitions its parents into k sets $\mathbf{X}_i$, each of size m_i, then the size of the joint probability of the clique is $\sum_{i=1}^{k} d^{m_i+1}$. When Y can be partitioned into two parent sets of equal size, the size of the joint probability for the clique can be reduced from d^p to $2d^{p/2+1}$, a substantial reduction in running time when p or d is large.

In general, the additive belief-network inference algorithm selects a decomposable node Y contained in the largest clique and then partitions the belief network at the chosen node along its partitions $\mathbf{X}_i$ to generate multiple belief subnetworks and associated weights, ϕ_i. The ith belief subnetwork is obtained by setting the nodes $\mathbf{X}_{j \neq i}$ in the belief network to their distinguished values $\mathbf{x}^*_{j \neq i}$. The algorithm is invoked recursively on each belief subnetwork until the size of the largest clique of each subnetwork is smaller than some threshold size; this rule ensures that we do not generate a large number of sparse subnetworks. The resulting tree of belief subnetworks contains leaf subnetworks on which probabilistic inference can be performed. When inference probabilities are required, we instantiate the leaf subnetworks and sum their inference probabilities, each weighted by the appropriate ϕ_i.

5 Results

We apply the methods developed in Section 3 to generate a dynamic network model to predict the course of a patient in a surgical intensive care unit (SICU). We construct the model from patient-specific data acquired at the Veteran's Administration Medical Center (VAMC) in Palo Alto, California. We begin with a brief discussion of the VAMC ICU and the pertinent medical history of the patient.

5.1 Critical Care

Critical-care medicine is a multidisciplinary specialty based in the ICU, and concerned primarily with the care of the patient with a critical illness. The essence of critical-care medicine is *prompt recognition of rapid alterations in physiologic status and early intervention.*

Critical care is expensive and consumes many resources. Currently, ICUs consume 20% of total hospital charges. Furthermore, ICU beds are growing at approximately 6% annually while other hospital beds are diminishing. Efforts at ICU health-care cost containment have

focused on methods to decrease ICU admissions and length of stay. Length of hospital stay is the single most important variable that influences the overall cost of patient care. Systems to predict patient care would limit admission to the ICU to patients who would clearly benefit from critical care, and would withhold therapy from patients who would not benefit [21]. One system, the acute physiology and chronic health evaluation (APACHE) system, relates the severity of a patient's illness to the degree of physiologic derangement of a set of physiologic measurements [33]. Its successor, the APACHE III [20] system, is currently in use at many ICUs to aid in cost-benefit treatment decisions based on the mortality risk predictions of critically ill patients. Other scoring schemes have been developed based on similar principles; many of these are for specific critical illnesses—for example, the Glascow Coma Score.

None of the scoring systems, however, directly addresses acute management of the ICU patient. Models that aid in ICU decision making are well established. These models, however, neglect the temporal evolution of the domain and are too simplistic to be reliable. Models that simulate the temporal evolution of the physiology can be used to predict short-term patient outcome, as well as outcomes following interventions. However, current physiologic models employ close-form equations derived from basic principles, and are difficult to apply in real time to real-world scenarios. Furthermore, because of their inflexible structure, these models neither have the capacity to deal with uncertainty in the domain, nor the ability to adapt rapidly to changes in the environment. Physiologic models, therefore, are too restrictive to be the only model used in critical-care settings.

We developed a dynamic network model, constructed from patient-specific data, for automated monitoring of ICU patients. The principles discussed in Section 3 allow the model to handle uncertain information and to adapt to changes in the environment. For example, even a sedated patient may be stimulated by frequent interventions, with ensuing increases in heart rate, blood pressure, and oxygen consumption. In addition, interventions may disconnect or

disrupt temporarily various sensor devices, thereby distorting data.

The model tracks the beat-to-beat flow of data generated from an array of devices that monitor the patient continuously. The model is form-free, adapts quickly to changes, and makes reliable short-term projections of the physiologic status. Model projections replace early recognition of derangement with anticipation of derangement. Feed-forward and feedback systems have the potential to provide control or recommendations of oxygenation through ventilatory settings, fluid status through intravenous infusion rates, or pharmacologic titration with end points such as oxygen consumption, systemic and pulmonary pressures, venous pressures, and heart rate.

5.2 Analysis

CASE REPORT

A 69 year-old man with a long history of chronic obstructive pulmonary disease (COPD) was admitted to the Palo Alto VAMC for surgical staging of a known pancreatic cancer. On April 3, 1992, the patient underwent an exploratory laparotomy that revealed tumor extension into the porta hepatis with invasion into the portal vein. Based on these findings, the tumor was considered unresectable. The following day, the patient was admitted to the VAMC SICU because of hemodynamic decompensation. Clinical evaluation with invasive monitoring suggested sepsis, and appropriate measures were taken to stabilize the patient. In spite of these efforts, the patient expired two days later, on April 6.

While in the SICU, the patient was ventilated mechanically. Both arterial and venous oxygen saturations—SaO2 and SvO2, respectively—were monitored. Ventilator settings were constant and set to maintain the SaO2 in the 90s with a neutral pH. The SvO2, however, quickly degenerated into an oscillatory behavior with an 11-minute periodicity. The patient's hemodynamics were monitored invasively with a right, radial, arterial line, and a pulmonary

artery catheter. Systolic, diastolic, and mean pressures for both the systemic and pulmonary vasculatures were recorded. These pressure tracings also showed an oscillatory behavior with an 11-minute periodicity. A metabolic monitor recorded oxygen consumption (VO2) and carbon dioxide production (VCO2). The VO2 periodicity was similar to the SvO2 and the systemic and pulmonary pressures. Oscillations continued for 140 minutes and terminated when the patient was given a rapid intravenous infusion of 500 cc of 6% albumin.

DISCUSSION

The oscillation of pressures, VO2, and SvO2 is an unusual occurrence, first noted by Carroll [6]. Several other instances of this phenomenon have been discovered subsequently. The pathophysiology of the process has not yet been elucidated. Recent work, however, suggests a link between a sustained oxygen debt and a compensatory neuroendocrine response. Since this phenomenon appears to be a harbinger of shock, its identification and rapid termination are essential.

We constructed a dynamic network model for this domain using the methods discussed in Section 4.1. The additive decompositions used in this model reduce the joint probability space of the largest clique from 1306368 probabilities to 330 probabilities. Figure 4 illustrates the dependencies of the model generated from the multivariate time series. We focus on five physiologic measures: SaO2, SvO2, VO2, mean-arterial pressure (MAP), and heart rate (HR). In this case, all other physiologic measures are either static or can be derived from the five chosen measures and values for the static variables.

The model structure helps elucidate the pathophysiology of the oscillations. It establishes impulse-response relations between physiologic variables, and illuminates possible neuroendocrine efferent pathways responsible for the oscillations.

Figure 5 displays the observed oscillations, and the first moment of the one-step-ahead

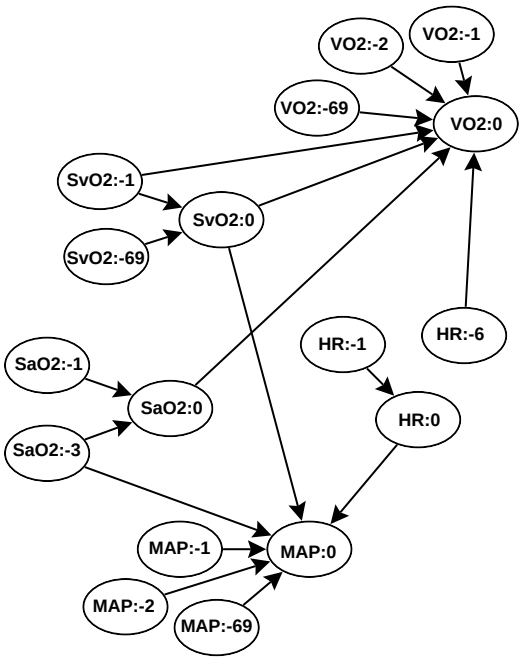

Figure 4: The dependency graph for the dynamic network model of the oscillatory pathophysiology of a critically ill ICU patient. Node label X:-i denotes the variable X_{t-i}.

forecast distributions of SvO2, VO2, and MAP, over a period of 35 minutes. The mean-prediction errors and the mean absolute prediction errors, shown in Table 5.2, suggest that the first moment of the one-step-ahead forecasts made by the model are in good agreement with the observed data, with the exception of the MAP predictions. In the case of MAP, the predictions show systematic bias in the region of the cycle troughs. We chose to focus only on the first moment of the predictions to allow comparisons with the observed time series. However, the model generates probability distributions for the predictions, and these can be used to provide confidence bounds on point estimates, or they can be used to hedge alternative outcomes.

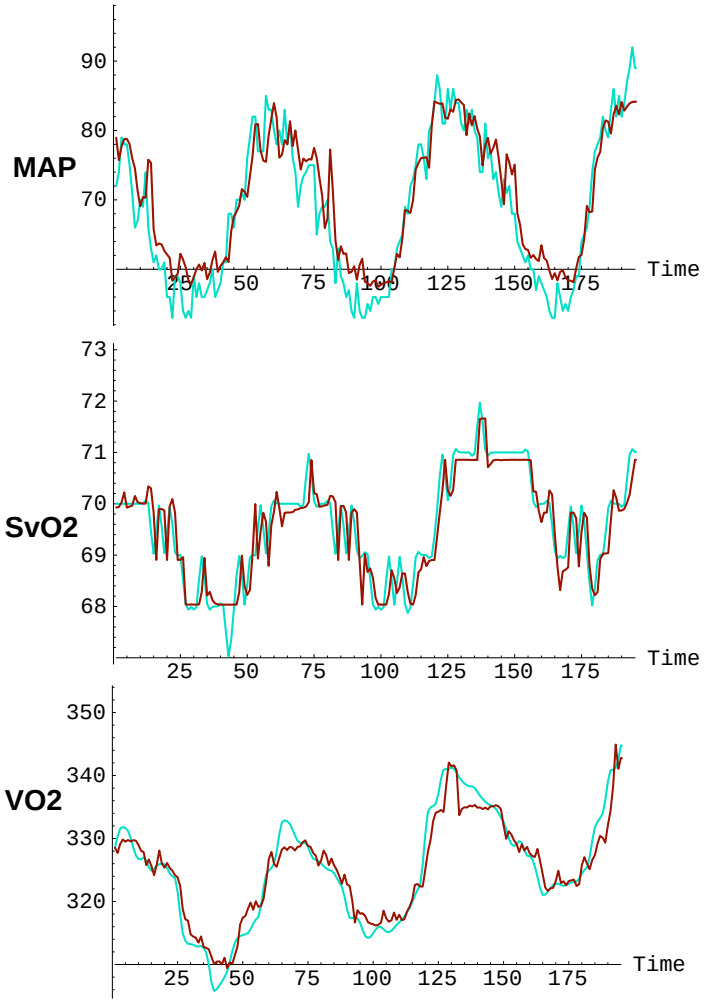

Figure 5: The one-step-ahead forecasts for MAP, SvO2, and VO2. The lighter curve depicts the time-series values; the darker curve depicts the forecasts made by the dynamic network model described in the text. We used the first 575 observations to identify the model; we generated forecasts for the subsequent 195 time steps.

	MAP	SvO2	VO2	HR	SaO2
MPE	-2.533%	0.062%	0.003%	0.015%	0.043%
MAPE	4.700%	0.443%	0.566%	0.220%	0.222%

Table 1: Table of Mean Prediction Errors (MPE) and Mean Absolute Prediction Errors (MAPE) for one-step-ahead forecasts between $t = 575$ and $t = 770$ of mean arterial pressure (MAP), venous O2 saturation (SvO2), O2 consumption (VO2), heart rate (HR), and arterial O2 saturation (SaO2).

6 Conclusions

We introduced a probability forecast model that integrates fundamental methods of Bayesian time-series analysis with additive generalizations of belief networks. The resultant nonnormal and nonlinear forecast model captures the explicit dependencies in the domain and is amenable to efficient model specification and inference.

We constructed a DNM for a multivariate time-series of physiologic measures taken from a critically ill patient in an ICU. The variables of the domain comprised both discrete and continuous variables that required quantification into categorical variables. We used the first moment of the probability forecasts to compute their mean-prediction errors and the mean absolute prediction errors. Quantification resulted in a small loss in the precision of the forecasts. This loss was evident predominantly in the predictions of MAP which showed a systematic bias in the region of the cycle troughs.

The main limitation of DNMs is quantification. If the domain consists of categorical variables only, then DNMs provide a robust and expressive model for forecasting. The presence of discrete or continuous variables, however, requires quantification, and consequently, a possible loss in precision, or an increase in the prediction bias. Thus, we benefit most from DNMs when

the domain is nonnormal and nonlinear, or when we require probability forecasts.

Acknowledgments

We are grateful to Estelle Bee Dagum and Camilo Dagum for sharing their insights with us and for their many contributions to this work. This work was supported by the National Science Foundation under grants IRI-9108385 and IRI-9311950, by Rockwell International Science Center IR&D funds, and by the Stanford University CAMIS project under grant IP41LM05305 from the National Library of Medicine of the National Institutes of Health.

References

[1] B. Abramson. ARCO1: An application of belief networks to the oil market. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pages 1–8, Los Angeles, CA, July 1991. Association for Uncertainty in Artificial Intelligence.

[2] S. Andreassen, M. Woldbye, B. Falck, and S.K. Andersen. Munin – a causal probabilistic network for interpretation of electromyographic findings. In *Proceedings of 10th International Joint Conference on Artificial Intelligence*, Milan, Italy, 1987.

[3] I. Beinlich, H. Suermondt, R. Chavez, and G. Cooper. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Proceedings of the Second European Conference on Artificial Intelligence*, Berlin, 1989. Springer-Verlag.

[4] C. Berzuini, R. Bellazzi, and S. Quaglini. Temporal reasoning with probabilities. In *Proceedings of the 1989 Workshop on Uncertainty in Artificial Intelligence*, pages 14–21, Windsor, Ontario, July 1989. Association for Uncertainty in Artificial Intelligence.

[5] A. Buja, T. Hastie, and R. Tibshirani. Linear smoothers and additive models (with discussion). *Ann. Statist.*, 17:453–555, 1989.

[6] G. C. Carroll. Fourth-order blood pressure waves. *Journal of the American Medical Association*, 263(6):856–855, 1990.

[7] E. Charniak. Bayesian networks without tears. *The AI Magazine*, 12(4):50–63, Winter 1991.

[8] R. Chavez and G. Cooper. A randomized approximation algorithm for probabilistic inference on bayesian belief networks. *Networks*, 20:661–685, 1990.

[9] G. Cooper. Bayesian belief-network inference using nested dissection. Technical Report KSL-90-05, Knowledge Systems Laboratory, Stanford University, Stanford, CA, February 1990.

[10] G. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[11] G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[12] P. Dagum and R.M. Chavez. Approximating probabilistic inference in Bayesian belief networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):246–255, 1993.

[13] P. Dagum and A. Galper. Additive belief network models. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 91–98, Washington, DC, July 1993. Association for Uncertainty in Artificial Intelligence.

[14] P. Dagum and A. Galper. Forecasting sleep apnea with dynamic network models. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 64–71, Washington, DC, July 1993. Association for Uncertainty in Artificial Intelligence.

[15] P. Dagum, A. Galper, and E. Horvitz. Dynamic network models for forecasting. In *Proceedings of the Eighth Workshop on Uncertainty in Artificial Intelligence*, pages 41–48, Stanford, CA, July 1992. American Association for Artificial Intelligence.

[16] P. Dagum and E. Horvitz. A Bayesian analysis of simulation algorithms for inference in belief networks. *Networks*, 23:499–516, 1993.

[17] P. Dagum and M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153, 1993.

[18] T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150, 1989.

[19] T. L. Dean and M. P. Wellman. *Planning and Control*. Morgan Kaufmann Publishers, San Mateo, CA, 1991.

[20] A. K. William et al. The APACHE III Prognostic System: Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 100:1619–1636, 1991.

[21] M. M. Fisher and R. F. Raper. Withdrawing and withholding treatment in intensive care. Part 2. Patient assessment. *Medical Journal of Australia*, 153(4):220–222, 1990.

[22] C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–148, 1986.

[23] J. Habbema. Models for diagnosis and detection of combinations of diseases. In F. de Dombal and F. Gremy, editors, *Decision Making and Medical Care*, pages 399–411, New York, 1976. North-Holland.

[24] T. Hastie and R. Tibshirani. Generalized additive models (with discussion). *Statistical Science*, 1:297–318, 1986.

[25] T. Hastie and R. Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82:371–386, 1987.

[26] D. Heckerman, E. Horvitz, and B. Nathwani. Toward normative expert systems: Part I. The Pathfinder project. *Methods of information in medicine*, 31:90–105, 1992.

[27] M. Henrion. An introduction to algorithms for inference in belief nets. In *Uncertainty in Artificial Intelligence* **5**, pages 129–138. Elsevier, Amsterdam, The Netherlands, 1990.

[28] M. Henrion, J.S. Breese, and E.J. Horvitz. Decision analysis and expert systems. *AI Magazine*, 12:64–91, 1992.

[29] E. J. Horvitz, J. S. Breese, and M. Henrion. Decision theory in expert systems and artificial intelligence. *Journal of Approximate Reasoning*, 2:247–302, 1988.

[30] K. Kanazawa. A logic and time nets for probabilistic inference. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 360–365. American Association for Artificial Intelligence, July 1991.

[31] C.R. Kenley. Influence diagram models with continuous variables. Technical Report LMSC-DO67192, Astronautics Division, Lockheed Missiles & Space Company, Sunnyvale, CA, June 1986.

[32] J. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference engines. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pages 190–193, Karlsruhe, West Germany, 1983. International Joint Conference on Artificial Intelligence.

[33] W. A. Knaus, J. E. Zimmerman, and D. P. Wagner et al. APACHE—acute physiology and chronic health evaluation:A physiologically based system. *Critical Care Medicine*, 9:591, 1981.

[34] S. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society* **B**, 50(19), 1988.

[35] R. Neapolitan. *Probabilistic Reasoning in Expert Systems.* John Wiley & Sons, Inc., New York, NY, 1990.

[36] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann, San Mateo, CA, 1988.

[37] J. Pearl and T. Verma. A theory of inferred causation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452. Morgan Kaufmann, April 1991.

[38] J. Pearl and T. Verma. A statistical semantics for causation. *Statistics and Computing*, 2:91–95, 1992.

[39] G. Provan. Tradeoffs in constructing and evaluating temporal influence diagrams. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 40–47, Washington, DC, July 1993. Association for Uncertainty in Artificial Intelligence.

[40] G. Rutledge, G. Thomsen, I. Beinlich, B. Farr, B. Sheiner, and L. Fagan. Combining qualitative and quantitative computation in a ventilator therapy planner. In *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care*, pages 315–319, Washington, D.C., 1989. The American Medical Informatics Association.

[41] H. Ryser. *Combinatorial Mathematics*. The Mathematical Association of America, Syracuse, NY, 1963.

[42] M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. A probabilistic reformulation of the QUICK MEDICAL REFERENCE system. In *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care*, pages 790–794, Washington, D.C., 1990. The American Medical Informatics Association.

[43] P. Spirtes, C. Glymour, and R. Scheines. *Causility, Prediction, and Search*. Springer-Verlag, New York, 1993.

[44] J. Tatman and R. Shachter. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetices*, 20:365–379, 1990.

[45] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, New York, NY, 1990.