# Real-time Bayesian anomaly detection in streaming environmental data

David J. Hill,[1] Barbara S. Minsker,[2] and Eyal Amir[3]

Received 28 February 2008; revised 23 November 2008; accepted 7 January 2009; published 4 March 2009.

[1] With large volumes of data arriving in near real time from environmental sensors, there is a need for automated detection of anomalous data caused by sensor or transmission errors or by infrequent system behaviors. This study develops and evaluates three automated anomaly detection methods using dynamic Bayesian networks (DBNs), which perform fast, incremental evaluation of data as they become available, scale to large quantities of data, and require no a priori information regarding process variables or types of anomalies that may be encountered. This study investigates these methods' abilities to identify anomalies in eight meteorological data streams from Corpus Christi, Texas. The results indicate that DBN-based detectors, using either robust Kalman filtering or Rao-Blackwellized particle filtering, outperform a DBN-based detector using Kalman filtering, with the former having false positive/negative rates of less than 2%. These methods were successful at identifying data anomalies caused by two real events: a sensor failure and a large storm.

**Citation:** Hill, D. J., B. S. Minsker, and E. Amir (2009), Real-time Bayesian anomaly detection in streaming environmental data, *Water Resour. Res.*, *45*, W00D28, doi:10.1029/2008WR006956.

## 1. Introduction

[2] Recently, there have been efforts to make use of streaming data from environmental sensors for real-time applications. For example, draft plans for the Water and Environmental Research Systems (WATERS) network, a proposed national environmental observatory network, have identified real-time analysis and modeling as a significant priority [*National Research Council*, 2006]. Furthermore, the National Science Foundation–sponsored workshop on sensors for environmental observatories has indicated the need for automated quality assurance and control (QA/QC) [*National Science Foundation*, 2005].

[3] Anomaly detection is the process of identifying data that deviate markedly from historical patterns [*Hodge and Austin*, 2004]. Anomalous data can be caused by sensor or data transmission errors or by infrequent system behaviors that are often of interest to scientific and regulatory communities. Anomaly detection performed in real time has many practical applications for environmental sensors such as real-time QA/QC, adaptive sampling, and anomalous event detection. Successful real-time anomaly detection in environmental streaming data must surmount four key challenges: (1) continuous collection of streaming data results in a large volume of data, so the entire data set cannot usually be held in memory nor can all existing data be reprocessed when new measurements become available; (2) real-time decisions can only use previous observations, so future observations cannot be used for anomaly classification; (3) environmental sensors go off line frequently because of the harsh environments in which they are deployed, so if a significant number of specific historical measurements are necessary to process a new measurement, then many measurements will not be able to be processed; and (4) sensors deployed in the natural environment behave in unexpected ways, so no a priori definition of the types of anomalies that may be encountered is available.

[4] Several methods have been suggested for addressing the problem of real-time detection of anomalies in streaming data. These methods can be divided into three categories: (1) redundancy-based approaches, (2) Bayesian filtering approaches, and (3) rule-based approaches. Redundancy-based approaches can be further divided into two subcategories: physical and analytical. Physical redundancy-based approaches employ two or more identical sensors at a particular location, resulting in multiple coincident measurements which can be directly compared. If the measurements deviate significantly, then at least one of the measurements can be deemed erroneous. However, if only one redundant sensor is used (two total sensors), then it is impossible to determine which measurement is erroneous; thus, it has been suggested [*Mourad and Bertrand-Krajewski*, 2002] that at least two redundant sensors should be used. However, because environmental sensors are often deployed in areas with limited power supplies (e.g., solar arrays), physical redundancy-based methods are undesirable because they take power that could be used for measuring additional processes. Furthermore, physical redundancy methods are only useful for identifying sensor faults and thus cannot be used for adaptive sampling or event detection.

[5] Analytical redundancy-based methods remove the burden of operating redundant sensors. Instead, a model of

[1] National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

[2] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

[3] Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

0043-1397/09/2008WR006956$09.00

the sensor data stream is used to simulate a redundant sensor. Since the model will reflect the "expected" behavior of the system given a record of historical measurements, analytical redundancy-based methods are suitable for identifying anomalies caused by sensor errors as well as by infrequent events of scientific interest. The classification of a measurement as anomalous is based on the difference between the model prediction and the sensor measurement. *Krajewski and Krajewski* [1989] present a method that employs model error standard deviations to set the threshold for identification of errors in stream gauge data. This method, however, relies on a physically based real-time model of the natural system, a tool that may not always be readily available. *Hill and Minsker* [2006] and D. J. Hill and B. S. Minsker (Anomaly detection in streaming environmental sensor data: A data-driven modeling approach, submitted to *Environmental Modelling and Software*, 2007) present an analytical redundancy method for wind speed data that employs the model error distribution to set the threshold and compares several data-driven modeling methods for simulating the redundant sensor. However, because of the complexity of modeling time series with frequent missing values, this method employs univariate autoregressive models and thus does not permit coupled anomaly detection on multiple data streams at once. Additionally, this method requires a fixed set of historical measurements to be available to classify a new measurement: thus, it is incapable of classifying measurements that immediately follow the missing values. The method presented by *Krajewski and Krajewski* [1989] may also have this limitation, but the exact details of the model used in the study were not provided.

[6] Bayesian filtering approaches operate similarly to analytical redundancy methods except that they use Bayesian filtering to determine the likelihood of a particular measurement, given all previous measurements in the sensor data stream. Filtering-based methods have been used for error detection in robotic sensors *Goel et al.* [2000], *Nicholson and Brady* [1994], and *Lerner et al.* [2000] demonstrate a filtering error detection method using a hypothetical process control case study. Extending these methods to the data collected by environmental sensors requires that the methods be modified such that they are robust to the large quantities of missing values encountered in environmental data streams, and such that they can be deployed without a well-defined model of the processes controlling anomalous data. Recently, *Dereszynski and Dietterich* [2007] presented a dynamic Bayesian network–based method for anomaly detection in environmental sensors. This method, however, only models individual sensor streams; thus, this method is unable to identify anomalies that only become apparent when information from other sensors are added to the anomaly detection process.

[7] Rule-based approaches use knowledge regarding normal sensor operation and likely data sequences to indicate data that may be anomalous. For example, *Ramanathan et al.* [2006] explored the use of rule-based anomaly detection for identifying sensor errors in a large array of groundwater quality sensors deployed in a rice field. Unfortunately, they were unable to verify their system's classifications because of deployment conditions. However, *Krajewski and Krajewski* [1989] investigated rule-based approaches, including several of those employed by *Ramanathan et al.* [2006], and found that analytical redundancy methods had greater potential for real-time QA/QC based on a comparison of the methods' abilities to detect errors in stream gauge data. Furthermore, since no model of the processes being measured is used, estimates of the erroneous data values cannot be suggested, a valuable feature if the anomaly detection is used as part of a real-time QA/QC system that prepares the data for real-time forecast models.

[8] This paper proposes three real-time anomaly detection methods that employ dynamic Bayesian networks (DBNs) to identify anomalies in streaming environmental data. DBNs are artificial intelligence techniques that model the evolution of discrete- and/or continuous-valued states of a dynamic system by tracking changes in the system states over time (K. P. Murphy, Dynamic Bayesian networks, unpublished manuscript, 2002). The methods developed in this study use three different DBN implementations: the well-known Kalman filter; the robust Kalman filter, which, to the authors' knowledge, has not found wide application in the field of environmental engineering; and the Rao-Blackwellized particle filter, which has only recently been developed [*Doucet et al.*, 2000a]. Unlike the method presented by *Dereszynski and Dietterich* [2007], all of the methods developed in this study can treat multiple sensor data streams and thus can take advantage of correlated information from different sensors. Furthermore, because of the nature of environmental streaming data, it was necessary for the implementations of each of these DBNs to be modified such that they were robust to missing values in the sensor data. The following section describes these methods in more detail. The DBN-based methods are then tested using a case study, where they are used to identify anomalous measurements in eight meteorological data streams from the WATERS Network Corpus Christi testbed. Finally, implications of the results are discussed.

## 2. Methodology

[9] This study develops three DBN-based methods for real-time detection of anomalies in environmental sensor data streams. Bayesian networks represent a set of variables and their dependencies as directed acyclic graphs, whose nodes represent the variables and whose arcs represent the conditional dependencies between the variables. DBNs are Bayesian networks that grow dynamically by adding new variables incrementally at each time step according to a template that defines the conditional dependencies among the new variables and between the new variables and the existing network. Thus, DBNs are well suited for modeling data sequences because new variables can be added incrementally to the network to represent new members of the sequence. Additionally, DBNs are well suited for time series modeling because of their flexibility in handling multivariate data and nonstationary processes and their ability to capture the uncertainty in expected state variables and measurements arising from uncertain process dynamics and noisy or missing sensor measurements [*Spall*, 1988].

[10] The DBN time series models used in this research take the general form

$$X_{t+1} = A(X_t) + \varepsilon_a \tag{1a}$$

$$Z_t = C(X_t) + \varepsilon_c \tag{1b}$$

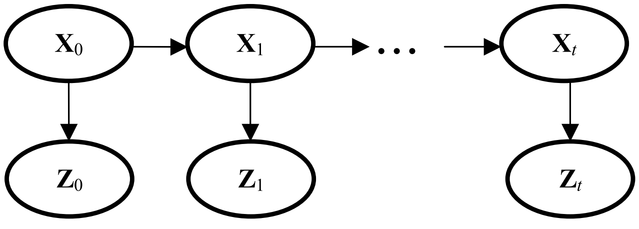

**Figure 1.** Schematic of DBN used in the BCI-kf method. Vector **X** is a random variable representing the continuous-valued system variables, and vector **Z** is a random variable representing the continuous-valued observations. Subscripts indicate time. Oval nodes represent Gaussian distributions.

where $X_t$ is a random variable describing the true state of the system at time $t$, $Z_t$ is a random variable describing the sensor measurement corresponding to the system state at time $t$, $A$ is an operator describing the evolution of the state variable distribution from time $t$ to $t + 1$ (transition model), $\varepsilon_a$ is the error of model $A$, $C$ is an operator describing the relationship between the state variables and the observed variables (observation model), and $\varepsilon_c$ is the error of model $C$. Since the future state (i.e., at time $t + 1$) depends only on the previous state (i.e., at time $t$), the model in equation (1) assumes that the time series is generated by a first-order Markov process.

[11] Filtering refers to the process of inferring the current system state ($X_t$) in equation (1) given all of the measurements up to and including the measurement at time $t$. The system state at time $t$ can also be inferred using both previous and future measurements, in a process known as smoothing. However, since smoothing requires future data (and complete reprocessing of the data when new measurements are taken), it is less suitable for real-time applications and thus is not considered in this study. Solving equation (1) through filtering has a long history, with the most notable early work being attributed to *Kalman* [1960], who developed a filtering algorithm for continuous linear-Gaussian variables. Since then, many researchers have generalized this work and associated it with state space models and hidden Markov models (e.g., M. I. Jordan, An introduction to probabilistic graphical models, unpublished manuscript, 2002). The result of this study is a well-known recursive algorithm for filtering DBNs, the state variables of which possess the Markov property (i.e., are generated by finite-order Markov processes) (Murphy, unpublished manuscript, 2002). In a recursive algorithm, data are processed sequentially, rather than in batch form; the complete set of data never needs to be stored prior to calculating the filtered estimate; and reprocessing of existing data is unnecessary if new measurements become available. Thus, recursive algorithms are well suited for real-time forecasting of streaming data. The next three subsections describe the three recursive DBN-based anomaly detectors developed in this study, while the final subsection describes a method based on a conventional time series model that will be used for comparison of the DBN-based methods.

### 2.1. BCI-kf

[12] In the BCI-kf method, Kalman filtering is used to track the system state according to equation (1). Kalman filtering is an analytical method for filtering a DBN in which the system state is a continuously valued vector or scalar, the state transition and observation models are linear, and the errors in the state and measurement models are considered to be Gaussian. A graphical representation of this DBN is shown in Figure 1. In this model, the operators $A(X_t)$ and $C(X_t)$ in equation (1) are equal to the products of a matrix **A** and the vector $\mathbf{X}_t$ and a matrix **C** and the vector $\mathbf{X}_t$, respectively. The terms $\varepsilon_a$ and $\varepsilon_c$ are represented as $N(0,Q)$ and $N(0,R)$, respectively, where the operator $N(\mu,\Sigma)$ indicates a normal distribution with mean $\mu$ and covariance matrix $\mathbf{\Sigma}$, and **Q** and **R** are the system and observation noise covariance matrices, respectively. Because the state and observation vectors are normally distributed, and because the state transition and observation models are linear, the exact prior and posterior distributions can be calculated analytically. This results in a compact representation of the belief state of the DBN (i.e., mean and covariance matrix), as well as in efficient updates to accommodate new measurements [*Maybeck*, 1979]. Additionally, these updates can accommodate missing or partially missing measurements in the vector $Z_t$ by treating the missing measurements as uninformative [*Liu and Goldsmith*, 2004; *Shumway and Stoffer*, 1982]. These equations can be found in section 1 of Text S1 in the auxiliary material.[1]

[13] In this study, the transition and observation models used for Kalman filtering are learned from sensor data using the expectation maximization (EM) method [*Ghahramani and Hinton*, 1996; *Digalakis et al.*, 1993; *Shumway and Stoffer*, 1982]. Since the measurements correspond directly to the state variables tracked by the DBN, the observation model is constrained such that the observation matrix is the identity matrix and the observation covariance matrix is diagonal. These constraints incorporate domain knowledge into the DBN, indicating that the measurements directly correspond to the process being measured and that measurement errors are not correlated between sensors, respectively. Because of the tendency of the EM method to converge to suboptimal solutions as the joint probability distribution narrows to a degenerate distribution centered on one of the training data points [*Hastie et al.*, 2001], it was necessary to perform nine EM trials starting with different initial conditions before a suitable parameterization was found. In four of the trials EM converged to the same parameterization, while the other five trials converged to degenerate solutions. The parameters common to the four nondegenerate solutions were used in the study.

[14] Using the Kalman filter estimate of the current system state, anomalies can be detected sequentially as new measurements become available from the sensors by constructing the Bayesian credible interval (BCI) for the current measurement. Since failure of one sensor should not affect the capabilities of other sensors (i.e., the measurement capabilities of different sensors are independent), the marginal posterior distributions of the observed variables can then be used to construct BCIs for the most recent set of measurements from each of the sensors. The $p$% BCI indicates that the posterior (i.e., adjusted for the available observations) probability of the observed state variables falling within the interval is $p$; thus, the BCI delineates the range of plausible values for sensor

[1] Auxiliary materials are available in the HTML. doi:10.1029/2008WR006956.

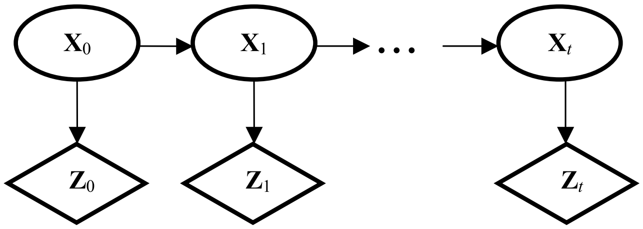

**Figure 2.** Schematic of DBN used in the BCI-rkf method. Vector **X** is a random variable representing the continuous-valued system variables, and vector **Z** is a random variable representing the continuous-valued observations. Subscripts indicate time. Oval nodes represent Gaussian distributions, and diamond nodes represent MOG distributions.

measurements. For this reason, any measurements that fall outside of the *p*% BCI will be classified as anomalous. The 100(1 − $\alpha$)% BCI for a new measurement can be calculated as

$$\bar{x} \pm z_{\alpha/2}\sqrt{\sigma^2} \quad (2)$$

where $z_{\alpha/2}$ is the 100(1 − $\alpha$/2)th percentile of a normal distribution, and $\sigma^2$ is the variance of the marginal posterior distribution of the measurement prediction. This method of classifying anomalous data from a DBN model will hereafter be referred to as the BCI method. The BCI method of anomaly detection is similar to the prediction interval–based method presented by *Hill and Minsker* [2006, also submitted manuscript, 2007]; however, because the BCI is based on the posterior distribution (whereas the prediction interval is not), the width of the *p*%-BCI changes dynamically with the uncertainty of the modeled system.

### 2.2. BCI-rkf

[15] The BCI-rkf method is similar to the BCI-kf method, except that robust Kalman filtering is used to track the state variables. The Kalman filter used in the BCI-kf method requires the strong assumption of linear Gaussian distributions; thus, it is very sensitive to outlying observations. This can adversely affect inferences about the belief state [*Schick and Mitter*, 1994; *Meinhold and Singpurwalla*, 1989] because, upon receiving an observation that deviates significantly from the expected observation specified by the process and observation models, the Kalman filter expands the variance of the state variable such that it includes the anomalous measurement. Thus, the Kalman filter will include the observations in future detections even if they are erroneous. This limitation is a particular concern for anomaly detection, which seeks to identify outlying observations.

[16] The BCI-rkf method uses a DBN in which the system state is a continuously valued vector or scalar, the state transition and observation models are linear, the errors in the state model are considered to be Gaussian, and the errors in the measurement models are considered to be a mixture-of-Gaussian (MOG) distribution. Thus, like the Kalman filter, the operators $A(X_t)$ and $C(X_t)$ in equation (1) take the products of a matrix **A** and the vector $\mathbf{X}_t$ and a matrix **C** and the vector $\mathbf{X}_t$, respectively, and $\varepsilon_a$ is represented as $N(0,Q)$. The observation model error term $\varepsilon_a$, however, is represented as a MOG distribution. MOG distributions have been demonstrated to approximate many heavy-tailed distributions (i.e., distributions that contain outliers) with high fidelity [*McLachlan and Peel*, 2000; *Blum et al.*, 1999; *Efron and Olshen*, 1978], and thus this model should be much less sensitive to outliers than Kalman filtering. A graphical representation of this DBN is shown in Figure 2.

[17] The probability density function of the MoG distribution is represented as a weighted average of the probability distribution functions of Gaussian mixture components:

$$f(x) = \sum_i \alpha_i f_i(x) \quad (3)$$

where $\alpha_i$ is the mixture proportion, the weight of the *i*th mixture component, $\sum_i \alpha_i = 1$, and $f_i(x)$ is the probability density function of the *i*th mixture component. Heavy-tailed distributions can be modeled by MOG distributions consisting of some components with large variance and small mixture proportions and some components with small variances and large mixture proportions [*Blum et al.*, 1999]. For example, a univariate MOG distribution with two components with means and variances of (0, 1) and (0, 10), respectively, will have heavy-tail behavior because of the second mixture component.

[18] To filter the DBN shown in Figure 2, robust Kalman filtering [*Frühwirth*, 1995; *Peña and Guttman*, 1988] is used, which is discussed in detail in section 2 of the auxiliary material. Robust Kalman filtering using MOG distributions for the observation error is a special case of an assumed density filter/moment matching filter (Murphy, unpublished manuscript, 2002); hence, it can be envisioned as using a mixture of Kalman filters, each specific to one Gaussian mixture component. Since the number of components in the MOG distribution grows exponentially with the number of time steps evaluated by the robust Kalman filter, it is necessary to limit the number of components. In this research, this is done by approximating the MOG posterior distribution as a Gaussian. Thus, robust Kalman filtering is an approximate filtering algorithm.

[19] In the BCI-rkf method, the distribution of the sensor measurements at each time step is represented as a mixture of $2^N$ Gaussians, where $N$ is the number of sensors, each corresponding to a unique combination of normal and anomalous measurements. For example, if there are two sensors, then the Gaussian mixture will have four components corresponding to the cases (normal, normal); (normal, anomalous); (anomalous, normal); and (anomalous, anomalous). The transition and observation models learned for the Kalman filter via EM are adapted for use in the robust Kalman filter as follows. The state transition model for each mixture component in the robust Kalman filter is equivalent to that of the Kalman filter. For the observation model, the observation matrix of the robust Kalman filter is the same as that of the Kalman filter, but the observation covariance matrix changes for each of the mixture components. For the case in which all of the measurements are nonanomalous, the observation covariance matrix is equivalent to that of the Kalman filter, whereas for the cases in which one or more measurements are anomalous, the parameter specifying the measurement variance of the anomalous measurement is set to be a large number (e.g., 1,000), indicating that regardless of the true state of the system, the measurement could take any real value with approximately equal probability. This description of anomalous measurements is used because it

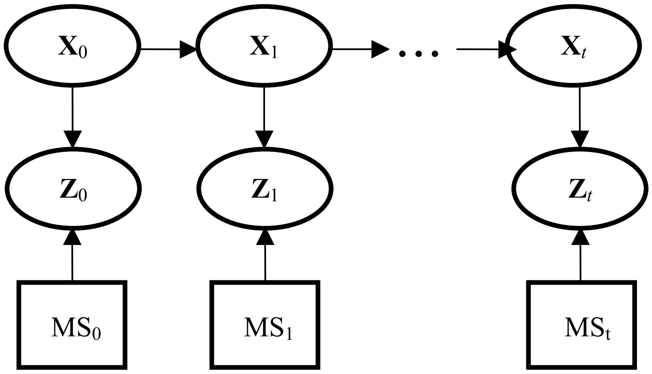

**Figure 3.** Schematic of DBN used in MAP-ms anomaly detection. Vector **X** is a random variable representing the continuous-valued system variables, vector **Z** is a random variable representing the continuous-valued observations, and scalar MS is a random variable indicating the discrete measurement status. Oval nodes represent Gaussian distributions, rectangular nodes represent discrete distributions, and subscripts indicate time.

indicates that an anomalous measurement is more likely to fall outside the range of plausible measurements than is a nonanomalous measurement, without requiring a priori knowledge of the types of anomalies that can occur. The mixture proportion prior used by the robust Kalman filter is set manually, using domain knowledge regarding the frequency of measurement anomalies. Manually setting the parameters for the cases in which one or more measurements are anomalous is necessary because anomalous measurements are, by definition, infrequent; as such, sufficient information may not be available for learning these parameters from the data. Furthermore, learned parameters may define anomalies too narrowly to identify the range of anomalies that may be encountered. Because the BCI-rkf method does not explicitly track the anomalies through time, it cannot represent any dependency relationships between anomalies; thus, the assumption of time independence of the anomalies is implicit in this method.

### 2.3. MAP-ms

[20] The MAP-ms anomaly detection method uses a DBN that is very similar to the BCI-rkf method, in which the system state is a continuously valued vector or scalar, the state transition and observation models are linear, the errors in the state model are considered to be Gaussian, and the errors in the measurement models are modeled with a mixture of Gaussians. However, unlike the BCI-rkf method, the distribution of the measurement error is constructed using a discrete variable to indicate whether or not each sensor measurement at time *t* is anomalous and a conditional linear Gaussian distribution for the measurement error, rather than combining these variables to create a MOG distribution. A graphical representation of this DBN is shown in Figure 3. If there are *N* measurements within each time step, and if there are only two possible measurement states (i.e., normal/anomalous), then this variable will have $2^N$ values, each corresponding to a unique combination of measurement classifications. For example, if there are two measurements, then the measurement status variable will have four values: (normal, normal); (normal, anomalous); (anomalous, normal); and (anomalous, anomalous); thus, it can be seen that with a binary (normal/anomalous) classification strategy, the DBNs considered in the BCI-rkf and MAP-ms method are similar because each represents the belief state as a mixture of $2^N$ Gaussian components. Thus, this DBN will also be able to represent heavy-tailed distributions well.

[21] Since there are no known exact algorithms for performing inference on DBNs that contain both discrete and continuous variables [*Lerner and Parr*, 2001], Rao-Blackwellized particle filtering is used to sequentially infer the posterior distribution of the state variables and their observations as new measurements become available from the sensors. The Rao-Blackwellized particle filter is a special type of particle filter that uses the Rao-Blackwell formula [*Casella and Robert*, 1996] to separate inference on the continuous and discrete state variables. This separation of variables requires only that the discrete variables be conditionally independent of the continuous variables; however, in this research the discrete variable indicating the measurement status is considered to be independent in time. Thus, to propagate each particle in the filter to the next time step, the value of the discrete variable is sampled from its prior, and the continuous variables are propagated using the Kalman filter parameters associated with the value of the discrete variable. Thus, the Rao-Blackwell division results in significantly increased computational efficiency over traditional particle filters [*Doucet et al.*, 2000a, 2000b]. The mathematical formulation of the Rao-Blackwellized particle filter can be found in section 3 of the auxiliary material.

[22] The MAP estimate (i.e., the most likely value, given the posterior distribution) of the state variable that indicates the measurement status can then be used to classify the sensor measurements as normal or anomalous. As in the case of the robust Kalman filter, the transition and observation models learned via EM for the Kalman filter are used to describe the case in which all measurements are normal; for the cases in which one or more measurements are anomalous, the parameter specifying the measurement variance of the anomalous measurement is set to be a large number (e.g., 1,000). Like the MOG mixture proportions used by the BCI-rkf method, the probability of different combinations of measurement anomalies is specified using domain knowledge.

### 2.4. AR_ADET

[23] In order to illustrate the performance of the DBN-based detectors compared to anomaly detection with an

autoregressive time series model, the AR_ADET detector of *Hill and Minsker* [2006, also submitted manuscript, 2007] is used. In order to avoid complications arising from missing measurements, this method considers each data stream independently [*Hill*, 2007]. This detector uses an autoregressive model of the sensor data stream incrementally for one-step-ahead prediction, and measurement anomalies are classified on the basis of the deviation of the observed sensor measurement from its predicted value. Specifically, this detector uses a neural network that takes a moving window of historical measurements as input to predict the expected value of the sensor measurement at time $t + 1$. It then calculates the prediction interval (PI) of the new measurement. When the measurement at time $t + 1$ arrives from the sensor, it is compared to the PI: if it falls outside of the PI it is classified as anomalous; otherwise it is classified as nonanomalous. This method was demonstrated to outperform several other autoregressive model-based anomaly detectors [*Hill*, 2007], and it performed well on identifying real anomalies in a wind speed data stream from Corpus Christi Bay [*Hill and Minsker*, 2006]. (Note that this data stream was different from the one addressed in the case study below.)

## 3. Case Study

[24] To test the efficacy of the three Bayesian anomaly detectors developed in this study, each approach was applied to eight meteorological data streams from Corpus Christi Bay collected by the Shoreline Environmental Research Facility (SERF) (available at http://www.serf.tamus.edu/). The data streams measure wind speed, wind direction, air temperature, and barometric pressure at 2-min intervals at two SERF sensor platform locations: CC003 and CC009.

[25] Because the sensor network addressed in this case study is a research network, there are many missing measurements because of sensor outages in the historical data record. Additionally, even though the historical data were subjected to manual quality control measures before they were archived, an initial application of the anomaly detection algorithms identified several anomalous events that were subsequently confirmed, through investigation by the SERF data managers, to be the result of sensor failures. Data from November 2006 were used for parameterizing the DBNs because these data do not appear to contain sensor failure errors and because they are reasonably complete (during this time period, there were approximately 10 and 800 missing measurements in the CC003 and CC009 data streams, respectively). This time period does, however, reflect the effects of two storm events on 11 and 31 November. During November, the approximate ranges of the wind speed, wind direction, air temperature, and barometric pressure at both platform locations were 0–40 knots (kt), −180–180 degrees from north (dfn), 5°–28°C (°C), and 1,000–1,040 mbar, respectively. The performance of the anomaly detectors was then evaluated using data from early December. This time period was chosen because the CC003 wind direction sensor went offline around 15 December. To demonstrate the efficacy of these anomaly detection methods, their performance is evaluated using both synthetic anomalies and actual data anomalies that were identified within the December meteorological data.

### 3.1. Detector Parameterization

[26] Two of the data types addressed in this study represent measurements of processes with nonlinear dynamics: wind speed, which cannot be negative, and wind direction, which has a discontinuity between −180 and 180 dfn. Despite the nonlinearity of wind speed, preliminary work [*Hill et al.*, 2007] demonstrated that good anomaly classifications could be achieved using a linear dynamics model. A linear dynamics model, however, is not sufficient for anomaly classification in wind direction data; thus, this study uses a transformation that converts wind direction into a two-component vector composed of the cosine of the angle (with respect to north) of the wind and the sine of the angle (with respect to north) of the wind. Preliminary work also investigated the transformation of the wind speed and wind direction data streams into a two-dimensional wind velocity vector. This transformation addressed not only the nonlinearity of wind direction, but also removed the nonnegativity condition of the wind speed data. However, it rendered the detectors unable to distinguish between anomalies in the wind speed sensor, the wind direction sensor, or both sensors because the transformation fused the data from the two different sensors. For this reason, this latter transformation was not pursued further.

[27] Because of the transformation of the wind direction data, the DBNs in all three anomaly detection methods tracked 10 state variables (one each for wind speed, temperature, and pressure at each location and two each for wind direction at each location). Thus, the vector $X$ in equation (1) has 10 dimensions referring to the true system states: wind speed, sin(wind direction), cos(wind direction), air temperature, and barometric pressure at both the CC003 and CC009 locations. The vector $Z$ in equation (1) also has 10 dimensions referring to the measurements of these 10 states.

[28] Parameters (i.e., transition model and error covariance matrix and observation noise covariance matrix) for the Kalman filter were learned from the approximately 21,600 measurements made by each of the eight sensors during the month of November 2006, using EM. Four versions of the BCI-kf method were created using an 80%, 90%, 95%, and 99% BCI for anomaly classification to test the sensitivity of the method to this parameter.

[29] The transition model and error covariance matrix and observation model of the robust Kalman filter were specified to be equivalent to the transition model and error covariance matrix and observation model used in the Kalman filter. Since the failure of a sensor rarely affects the capabilities of other sensors, it was assumed that concurrent anomalies in different data streams were independent. Thus, the robust Kalman filter used a mixture of $2^8 = 256$ Gaussians in the observation model to represent all of the possible anomaly combinations from the eight sensors. The observation covariance matrix of the Gaussian component representing no failure was equivalent to the observation covariance matrix used in the Kalman filter model. The covariance matrix of each of the components corresponding to a combination of failed sensors was equivalent to the measurement covariance matrix of the no anomaly case, except that the variance of the anomalous measurements was set to a large value. In this study we compared three versions of the BCI-rkf detector with values of 100, 500, and 1,000 for the covariance of an anomalous measurement to test the sensi-

tivity of the method to this parameter. The mixture proportion was set to the probability of the combination of anomalies in the data stream. The probability of an anomaly was specified to be 5%, which was selected to be representative of the frequency of data anomalies because of measurement or data transmission errors in the Corpus Christi sensor array, based on the results of previous work conducted with the Corpus Christi meteorological sensors [*Hill and Minsker*, 2006, also submitted manuscript, 2007]. Thus the mixture proportion of components representing 0, 1, 2, 3, 4, 5, 6, 7, and 8 failures were $0.95^8$, $(0.95^7 \times 0.05)$, $(0.95^6 \times 0.05^2)$, $(0.95^5 \times 0.05^3)$, $(0.95^4 \times 0.05^4)$, $(0.95^3 \times 0.05^5)$, $(0.95^2 \times 0.05^6)$, $(0.95^1 \times 0.05^7)$, and $0.05^8$, respectively. Again four levels of 80%, 90%, 95%, and 99% BCI were used for anomaly classification to test the sensitivity of the BCI-rkf method to this parameter.

[30] The Rao-Blackwellized particle filter used a discrete variable with $2^8 = 256$ values to represent the status of the eight sensors at each time step. For every value of the discrete variable, the transition model and error covariance matrix and observation model of the continuous variables were equivalent to the transition model and error covariance matrix and observation model used by the Kalman filter. For the value of the discrete variable representing no sensor failures, the observation covariance matrix of the continuous variables was equivalent to the observation covariance matrix used in the Kalman filter model, while the covariance matrix of each of the discrete variable values corresponding to a combination of failed sensors was equivalent to the measurement covariance matrix of the no anomaly case, except that the variance of the anomalous measurements was set to a large value. Again, three different versions of the MAP-ms detector with values of 100, 500, and 1,000 for the covariance of an anomalous measurement were compared to test the sensitivity of the method to this parameter.

[31] Although the MAP-ms method can model the time dynamics of the sensor status variable, there was no specific information regarding the persistence of anomalies caused by sensor failures. Thus, anomalies in the sensor data stream were assumed to be time independent with a likelihood of 5%. Additionally, it was assumed that concurrent anomalies in different data streams were independent because failure of a sensor rarely affects the capabilities of other sensors. Thus, the probability of the discrete variable taking a value corresponding to 0, 1, 2, 3, 4, 5, 6, 7, or 8 sensor failures during any time step is $0.95^8$, $(0.95^7 \times 0.05)$, $(0.95^6 \times 0.05^2)$, $(0.95^5 \times 0.05^3)$, $(0.95^4 \times 0.05^4)$, $(0.95^3 \times 0.05^5)$, $(0.95^2 \times 0.05^6)$, $(0.95^1 \times 0.05^7)$, and $0.05^8$, respectively. The particle filter was specified to use 50,000 particles because *Hill* [2007] showed that this number of particles resulted in the best trade-off between accuracy and computational demand based on a comparison of the performance of particle filters with 1,000, 10,000, 50,000, and 100,000 particles, using the synthetic anomaly data set described in the next section. The synthetic data set was used instead of actual anomalies because too few real anomalies exist in the data to provide a suitable comparison. Figure 3 shows a graphical representation of the DBN used in the MAP-ms method.

[32] The parameters for the AR_ADET detector were selected using correlation analysis to determine the salient autoregressive variables, as described by *Hill* [2007]. The resulting autoregressive models for wind speed and wind direction use the most recent 30 measurements to predict the next measurement, while the resulting models for air temperature and barometric pressure use the most recent two measurements. The models were then trained using 30,000 randomly selected data points from October through November 2006. These data were selected for training because data prior to October were not available and data from December were to be used for demonstration of the detectors. Since neural networks are suitable for modeling nonlinear data, the wind direction was modeled directly (i.e., the cosine/sine transformation described in the Detector Parameterization section was not used). A 95% PI was used for anomaly classification, as suggested by *Hill* [2007].

### 3.2. Detection of Synthetic Anomalies

[33] Synthetic anomalies affecting data from 2 to 5 December are used to compare the performance of the anomaly detection methods developed in this study with each other, as well as with the AR_ADET method. Synthetic anomalies are used for this comparison because there are not enough known anomalies in the historical sensor data to evaluate the relative performance of the methods and because it is difficult to know the true classification of observed data, which is necessary for false positive/negative calculations.

[34] The synthetic errors were specified to be transient errors (i.e., they did not persist) and were generated according to the following equation:

$$M^* = M \pm \Delta \tag{4}$$

where $M^*$ is the anomalous measurement, $M$ is the true measurement, and $\Delta$ is an offset. The range of the offsets for wind speed, wind direction, air temperature, and barometric pressure were selected on the basis of judgment, to be [4–12] kt, [45–180] dfn, [3–9]°C, and [10–30] mbar, respectively. Synthetic anomalies of this type were randomly introduced into each of the eight data streams independently with a frequency of 5%. For each anomaly instance, addition or subtraction of the offset was chosen at random with equal probability (i.e., 50% addition and 50% subtraction), and the value of the offset was selected by sampling a uniform distribution over the offset range. Since the anomalies in one sensor data stream were independent of other previous and current anomalies, both concurrent anomalies in different data streams and multiple sequential anomalies in a single data stream could occur sequentially.

[35] The false positive rates of the BCI-kf detector using a 99% BCI, the BCI-rkf detector using a 99% BCI and anomalous measurement variance of 1,000, the MAP-ms detector using an anomalous measurement variance of 1,000, and the AR_ADET detector are shown in Figure 4. Here the false positive rate is the ratio of the number of data misclassified as anomalous to the total number of non-anomalous data, where misclassification is based on data points that were randomly selected to be modified according to equation (4). Only one version of each of the DBN-based detectors is shown in Figure 4 because the BCI-rkf and MAP-ms detectors were reasonably insensitive to the setting of the BCI and/or anomalous measurement variance, while the BCI-kf detector was sensitive to the selected BCI, but performed poorly overall. The false positive rates of the 12 versions of the BCI-rkf detector using BCIs of 80%, 90%,

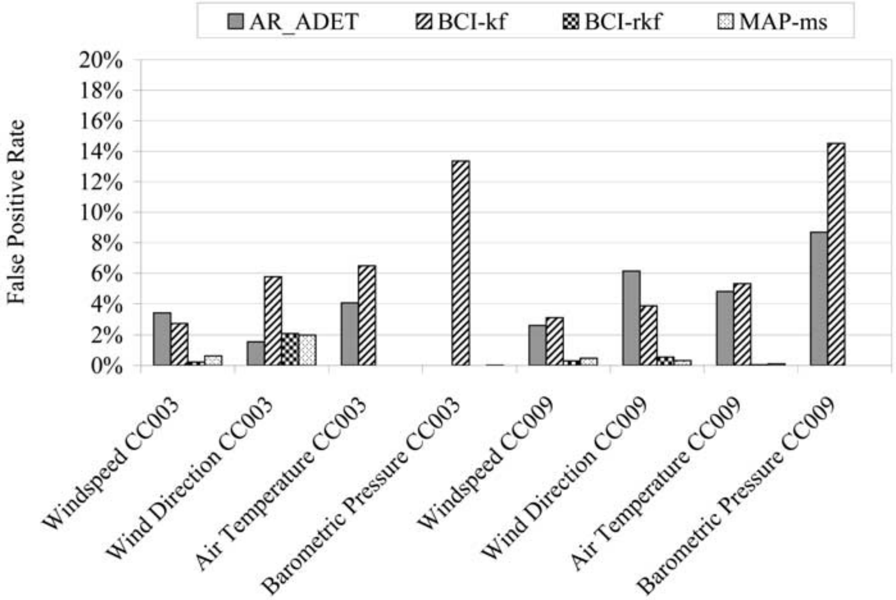

**Figure 4.** False positive rates for the AR_ADET, BCI-kf, BCI-rkf, and MAP-ms detectors for classifying transient synthetic anomalies.

95%, and 99% and anomalous measurement variances of 100, 500, and 1,000 varied by 0.14% on average (1.3% maximum), while the false positive rates of the three versions of the MAP-ms detector using anomalous measurement variances of 100, 500, and 1,000 varied by 0.19% on average (0.82% maximum). The false positive rate of the four versions of the BCI-kf detector varied by 5.3% on average (maximum 9.5%).

[36] Because of the stochastic nature of particle filtering, the results for the MAP-ms detector were averaged over five replicates. As can be seen in Figure 4, not only do the BCI-rkf and MAP-ms detectors perform well (with false positive rates in the range of 0 to 2% for the different data streams), but they also performed significantly better than the BCI-kf and AR_ADET detectors, which had false positive rates in the ranges of 2 to 15% and 2 to 8%, respectively. This result is not surprising, because both the BCI-kf method (as discussed previously) and the AR_ADET method are adversely affected by anomalous measurements, and because the AR_ADET method cannot take advantage of information in other data streams that may help the detector discern between an anomalous and nonanomalous measurement. There is little difference between the false positive rates of the BCI-rkf detector and the MAP-ms detector. These results indicate that a conceptual model that accounts for outlying measurements is better suited for describing systems in which measurements can be anomalous.

[37] Figure 5 illustrates the false negative rates of the Bayesian and AR_ADET anomaly detection methods, where the false negative rate is the ratio of the number of data misclassified as nonanomalous to the total number of anomalous data, and misclassification is based on the original data points that were selected to be modified according to equation (4). Again the results of the MAP-ms method were averaged over five replicates, and only one version of each of the detectors is shown because the BCI-rkf and MAP-ms detectors are reasonably insensitive to the setting of the BCI and/or anomalous measurement variance, and the BCI-kf method, while sensitive to the BCI, performs poorly overall. The false negative rates of the 12 versions of the BCI-rkf detector varied by 0.065% on average (0.74% maximum), the false negative rates of the three versions of the MAP-ms detector using anomalous measurement variances of 100, 500, and 1,000 varied by 0.18% on average (1.3% maximum), and the false negative rates of the four versions of the BCI-kf detector varied by 3.8% on average (maximum 18%).

[38] As can be seen in Figure 5, the BCI-rkf and MAP-ms detectors again behave similarly, resulting in false negative rates of at most 1.5%. On the other hand, the BCI-kf method has low false negative rates (<2%) for some data streams (generally those data streams on which the method had a high false positive rate), and high false positive rates (2–16%) for others (generally, those data streams on which the method had a low false negative rate). This result is caused by adverse effects of anomalous measurements on the quality of the Kalman filter estimate of the BCI. Finally, the AR_ADET method has the highest false negative rates of all the detectors.

[39] In the above analysis, the DBNs considered all eight sensor data streams at once to perform coupled anomaly detection; thus, the detectors could take advantage of correlated data being measured by other sensors. To demonstrate the beneficial effect of considering multiple data streams in the DBN model, eight additional detectors using the MAP-ms method of anomaly detection and an anomaly variance of 1,000, but addressing only one of the data streams each, were created. Parameterization of these DBNs was performed using the same method used to parameterize the DBNs addressing all eight data streams; however, because of the reduced dimensionality, 1,000 particles were sufficient for the MAP-ms detector. Figure 6 shows the false positive and false negative rates of these detectors for identifying synthetic anomalies in the 2–5 December data and illustrates that coupling the anomaly detection process significantly reduces the false negative rate.

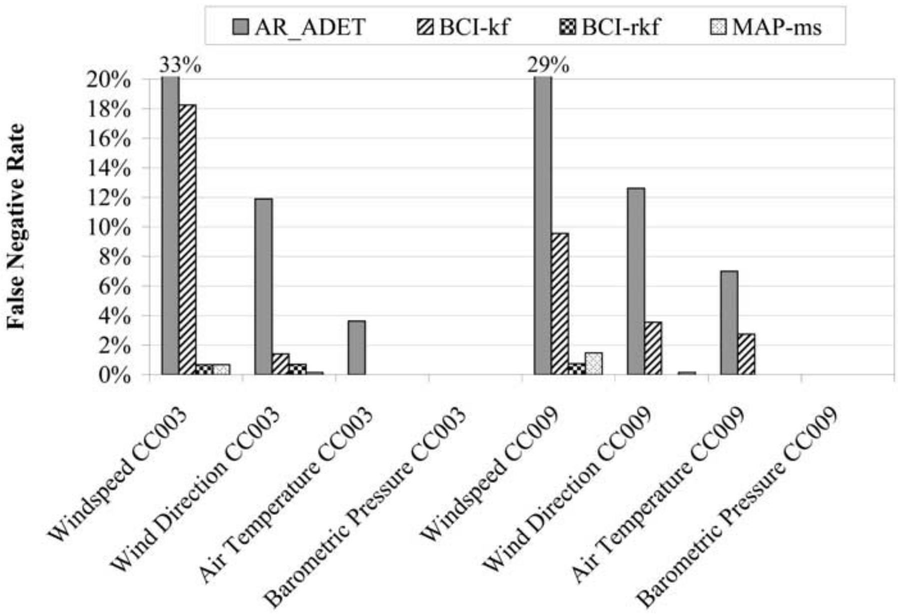

**Figure 5.** False negative rates for the AR_ADET, BCI-kf, BCI-rkf, and MAP-ms detectors for classifying transient synthetic anomalies. The bars for the AR_ADET wind speed data from both platforms are truncated.

[40] During the period of 2–5 December, there were 14 missing measurements in the data from the CC009 sensors. Because the AR_ADET method requires a particular number of previous measurements to be available in order to process a new measurement, these missing data rendered the AR_ADET detectors for the CC009 sensors unable to classify 427, 427, 44, and 107 measurements from the wind speed, wind direction, air temperature, and barometric pressure data streams, respectively. Since the Bayesian anomaly detection methods do not require any fixed set of measurements to be available, the BCI-kf, BCI-rkf, and MAP-ms methods were able to classify all of the available measurements.

[41] The time required by the AR_ADET, BCI-kf, BCI-rkf, and MAP-ms detectors to classify a new measurement was 0.004, 0.002, 0.49, and 32 s, respectively. Timing was performed on a Suse Linux workstation equipped with AMD dual core Opteron 1.8 GHz processors and 7 GB of memory. Since the AR_ADET detector only operates on one data stream at a time, the time reported here is equal to

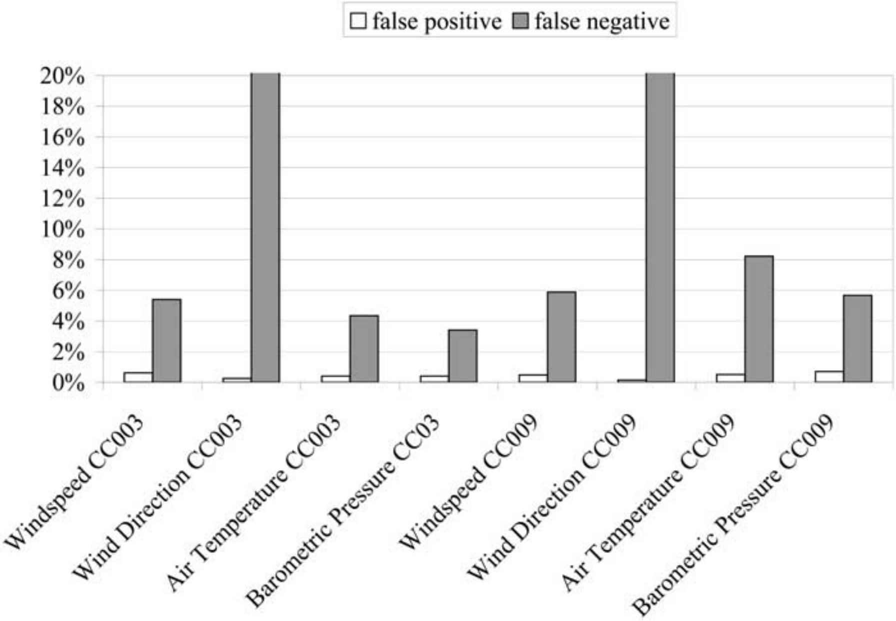

**Figure 6.** Performance of uncoupled MAP-ms anomaly detection method for classifying transient synthetic anomalies. The bars for the false negative rates for the wind direction data from both platforms are truncated.

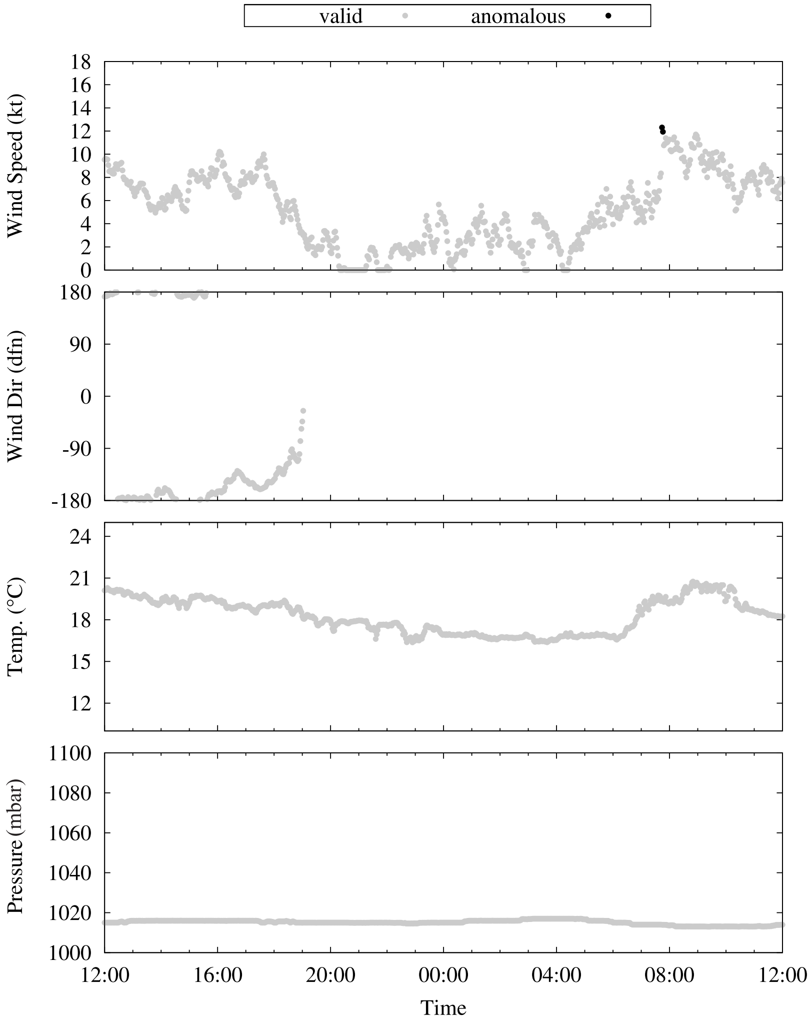

**Figure 7.** Classification of the 15–16 December 2007 sensor measurements from platform CC003 by the MAP-ms detector.

the time needed to process a new measurement in a single data stream multiplied by eight (the number of data streams concurrently processed by the Bayesian anomaly detectors). Since the measurement frequency of the sensors considered in this study is 2 min, all of the anomaly detection methods are quite viable for real-time use.

### 3.3. Detection of Observed Anomalies

[42] This subsection describes how the MAP-ms method (using an anomalous measurement variance of 1,000), which performed best on the synthetic anomalies, performs on two real data anomalies observed within the December meteorological data. The first anomalous event occurs around midnight on 16 December. The anomalous measurements caused by this event were first identified by the anomaly detectors developed in this study and subsequently brought to the attention of the data managers, who suggested that the anomalous data were errors caused by the failure of the CC009 barometer. The second anomalous event, which the SERF data managers attributed to the arrival of a storm front, occurs around 0400 on 1 December.

[43] Figures 7 and 8 show a 24-h segment of data spanning from 1200 UT 15 December through 1200 UT 16 December and their classifications (discussed below) by the MAP-ms detector, from the CC003 and CC009 detectors, respectively. Figures 7 and 8, it can be seen that the CC003 wind direction sensor goes off line at approximately 1930 UT 15 December and that all of the CC009 sensors go off line at approximately 0600 16 December. Furthermore, the CC009 barometer reports a large transient deviation at 2100 15 December, as well as a rapid decrease followed by a rapid increase of pressure starting at 0200 15 December and continuing until the sensor goes off line. This behavior, according to the SERF data managers, is indicative of a barometer failure on platform CC009. Further evidence that this event was caused by a sensor failure, rather than by a system anomaly, is found by considering the behavior of the CC003 barometer, which does not echo the behavior of the CC009 barometer. A similar event also occurred at approximately 0900 on 2 October. Neither of these events was identified during SERF's manual QA/QC regimen, but both were identified by all three of the detection methods presented in this paper during preliminary work.

[44] The classification (shown in Figures 7 and 8) of the meteorological data from 1200 15 December to 1200 16 December from the CC003 and CC009 detectors indicates that the MAP-ms detector can effectively identify the anomalous measurements caused by the CC009 barometer failure.

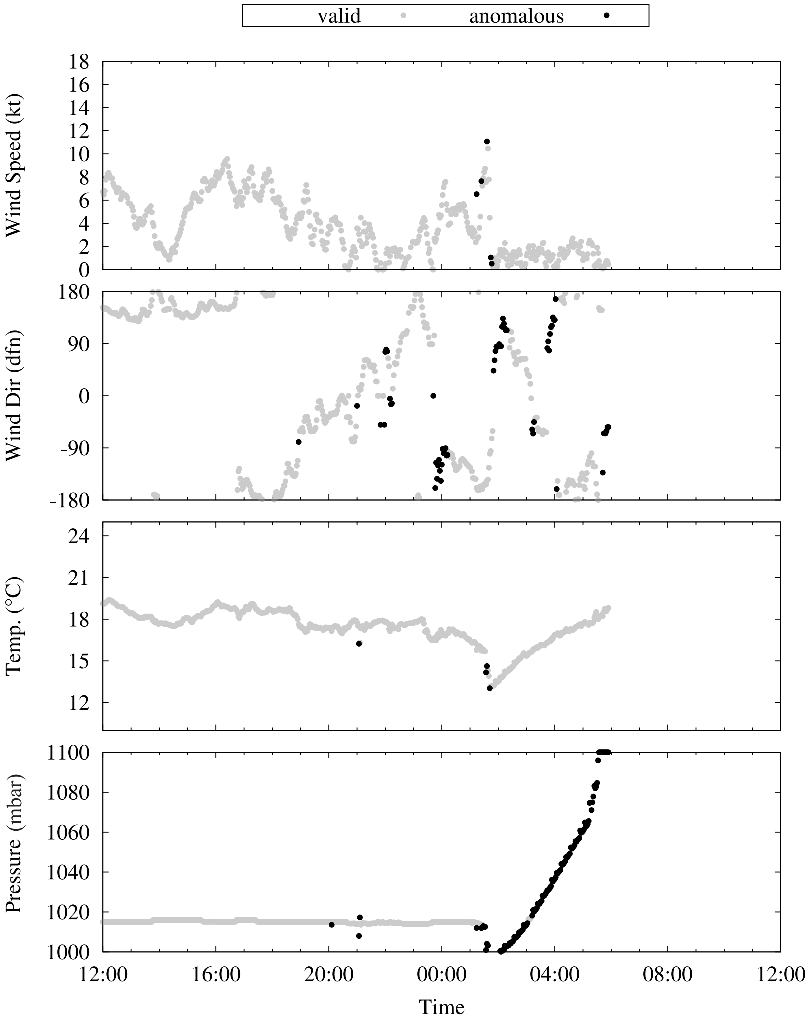

**Figure 8.** Classification of the 15–16 December 2007 sensor measurements from platform CC009 by the MAP-ms detector.

This result indicates that the assumption of temporal independence used in the anomaly definition of the MAP-ms method does not adversely affect the detector's ability to identify persistent failures. The assumption of independence does, however, cause the detector to classify a couple of CC009 barometric pressure data points at approximately 0300 as normal, because these data fall into the expected range of the barometric pressure, given previous CC009 measurements and the current CC003 measurement. Figures 7 and 8 also show that the detector makes only a few false positive classifications (i.e., normal data classified as anomalous), though it does appear that the false positive rate of the CC009 wind direction data has increased slightly from what was expected, given Figure 4. Further inspection reveals that this slight increase in false positive rate occurs after the wind direction sensor on the other platform (CC003) has gone off line and after the barometer on the same platform (CC009) has begun malfunctioning. These sensor failures, however, do not appear to increase the false positive rate in the other seven data streams; thus, since these failures only resulted in a marginal increase in the number of false positives on one data stream, this method appears to be robust to the failure of up to two sensors for this case study. Though the results are not shown, the BCI-rkf detector also exhibited behavior similar to that shown in Figures 7 and 8.

[45] Figures 9 and 10 show a 24-h segment of data corresponding to 1 December from the CC003 and CC009 detectors and their classifications (discussed below) by the MAP-ms detector, respectively. From Figures 9 and 10, it can be seen that at approximately 03:40, the wind speed increases dramatically (maximum rate of change of approximately 15 kt over 2 min), the wind direction changes from southerly to northerly, the temperature drops dramatically (maximum rate of change of approximately 5°C over 2 min), and the barometer rises. Furthermore, the corresponding sensors on both sensor platforms report similar observations. This behavior, according to the SERF data managers, is indicative of the arrival of a severe storm front (an infrequent event).

[46] The classification (shown in Figures 9 and 10) of the meteorological data from 1 December from the CC003 and CC009 detectors indicates that the MAP-ms detector identifies data corresponding to the severe changes in the wind speed and temperature as anomalous, indicating that the anomaly detectors are able to identify anomalies caused by infrequent system behaviors. Figures 9 and 10 also show that the MAP-ms detector behaves differently on the wind

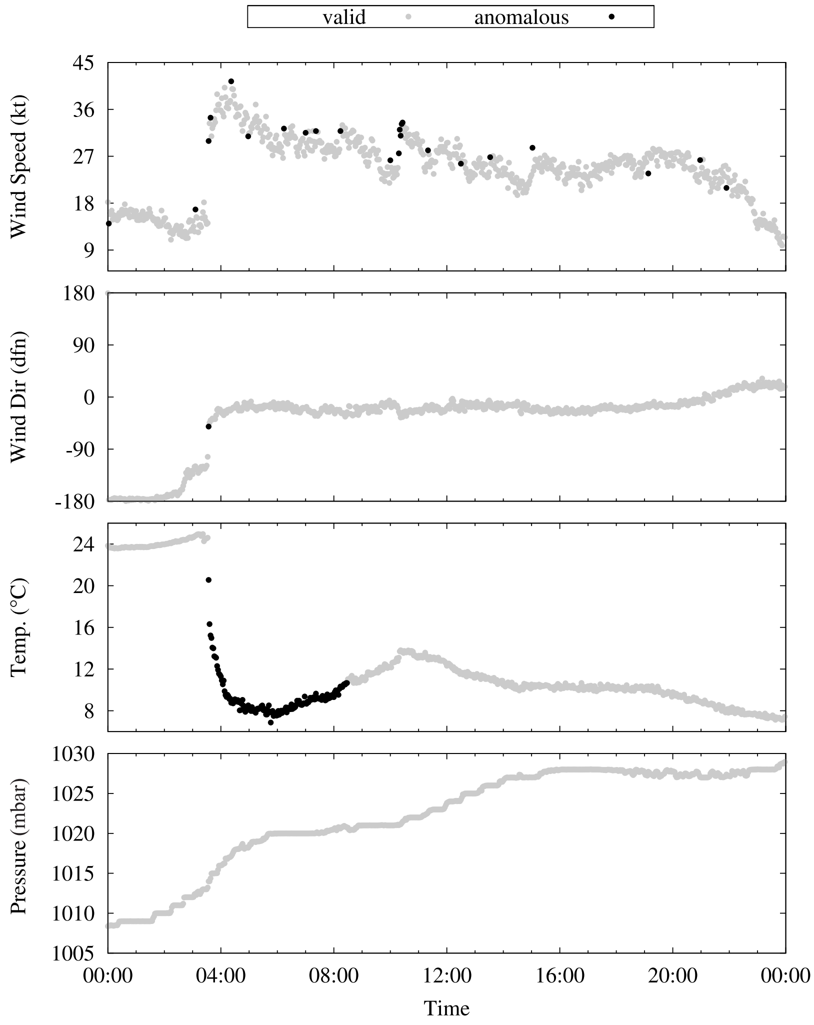

**Figure 9.** Classification of the 1 December 2007 sensor measurements from platform CC003 by the MAP-ms detector.

speed data than on the temperature data. Following the large change in wind speed, the MAP-ms detector quickly returns to classifying the majority of the wind speed data as normal; however, it continues to classify wind speed data as anomalous for approximately 12 h following the initial large increase in wind speed, at a rate higher than the expected false positive rate. Following the initial wind speed increase, the wind speed variability is higher than usual (as should be expected during a storm), and the data that are classified as anomalous appear to represent more extreme deviations from the general wind speed pattern than the data that are classified as nonanomalous. Following the large drop in temperature, however, the MAP-ms detector does not return to classifying the temperature measurements as nonanomalous for approximately 5 h. This difference in behavior is related to the state transition model used by the MAP-ms detector, as well as the assumption in the MAP-ms method that the anomalies are independent in time. Because the historical data indicate that the wind speed changes more rapidly than the temperature, the MAP-ms detector requires more evidence (in the form of measurements) to change its belief state about the air temperature than to change its belief state about the wind speed. Once the belief state of the MAP-ms detector has changed to reflect the decrease in temperature caused by the storm, it ceases to classify new measurements as anomalous because, following the initial decrease in temperature, the observed air temperature does not exhibit higher variability than usual. On the other hand, after the initial increase in wind speed, the variability of the wind speed remains larger than usual; thus, even though the MAP-ms detector quickly reflects the increased wind speed, it continues to classify data that exhibit large deviations from the general wind pattern as anomalous. If the MAP-ms method had assumed that the anomalies were correlated in time, rather than being time independent, then the detector would likely have continued to classify the wind speed and temperature data as anomalous for a longer period of time following the initial increase/decrease than was observed in Figures 9 and 10. Though the results are not shown, the BCI-rkf detector also exhibited behavior similar to that shown in Figures 9 and 10.

## 4. Discussion

[47] The previous section demonstrated that the BCI-rkf and MAP-ms anomaly detection methods developed in this study can reliably identify anomalies in the SERF meteorological data collected at two spatial locations within Corpus Christi Bay. Through the use of simple data transformations

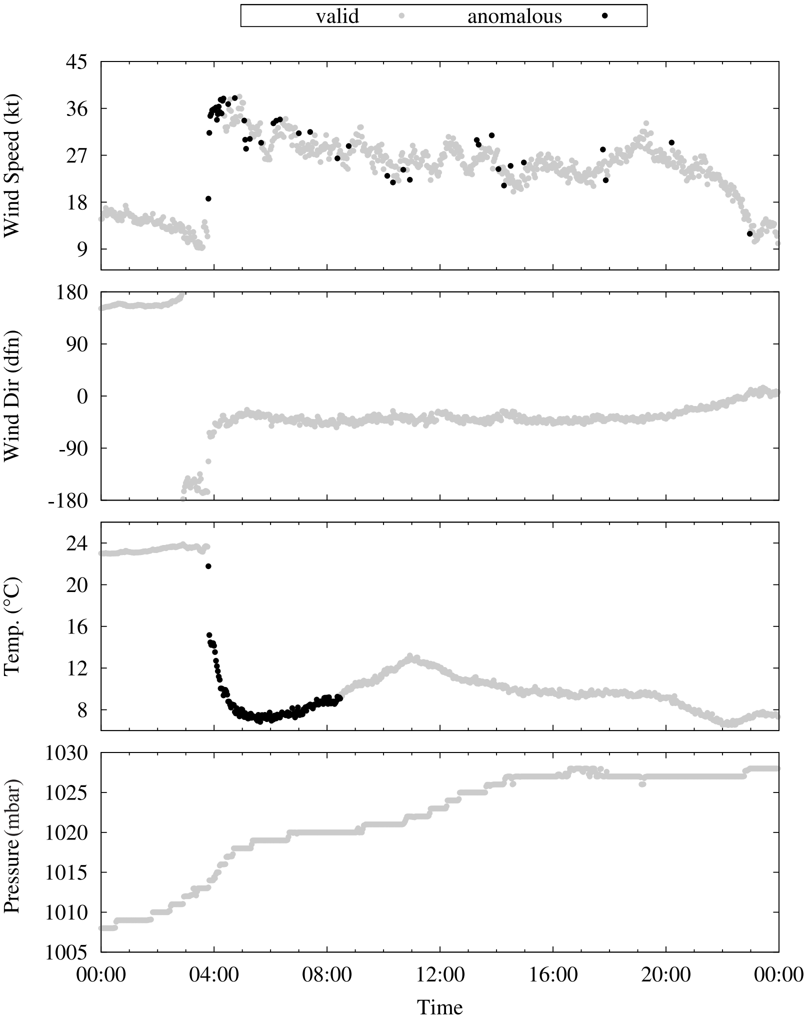

**Figure 10.** Classification of the 1 December 2007 sensor measurements from platform CC009 by the MAP-ms detector.

(e.g., transforming the wind direction variable into two variables that varied smoothly between −1 and 1 by calculating the cosine and sine of the wind direction), filtering algorithms like robust Kalman filtering or Rao-Blackwellized particle filtering, which achieve significant computational efficiency through the use of linear transition models, can be successfully applied for anomaly detection. For highly nonlinear processes, transformations that project the data into a different, possibly higher-dimensional space, in which it behaves linearly, should be considered [e.g., *Liu and Motoda*, 1998]. For processes whose measurements do not vary smoothly, more complex DBNs should be explored [e.g., *Lerner*, 2002].

[48] The Bayesian framework of the anomaly detection methods presented in this study renders them well suited for concurrently processing multiple nonstationary data streams that may contain many missing values. Since these methods consider multiple correlated data streams at the same time, they can process data immediately following one or more missing values in a particular data stream. Furthermore, information from other sensors improves the classification accuracy of the detectors, as demonstrated in Figure 6, and may be instrumental in identifying certain types of anomalies, such as those caused by the slow drift of one sensor.

[49] For this case study, the BCI-rkf and MAP-ms anomaly detectors were able to evaluate new measurements more quickly than the measurement frequency of the sensors, so they can be used to detect anomalies in real time for this case study. Other sensor arrays may contain substantially more sensors than the Corpus Christi Bay meteorological sensor array; hence, it is valuable to consider how the time complexity of the anomaly detection methods would scale as a function of the number of sensors. Since robust Kalman filtering and Rao-Blackwellized particle filtering have commonalities with Kalman filtering, the analysis will begin with Kalman filtering. Assuming that there are $n$ sensors, each measuring one process variable, and that each DBN state variable corresponds to one of the $n$ sensor measurements, the system noise covariance matrix (matrix $\mathbf{Q}$ in equations (4) and (5) in the auxiliary material) and the measurement noise covariance matrix (matrix $\mathbf{R}$ in equation (5) in the auxiliary material) will have dimensions $n \times n$. Propagating the state distribution forward to the next measurement time and updating for the new measurements via the Kalman filter equations (see equations (3)–(5) in the auxiliary material) requires a finite number of matrix multiplications and inversions of $n \times n$ matrices. Using the Coppersmith-Winograd algorithm, each multiplication or inversion is an

O($n^{2.376}$) process [*Coppersmith and Winograd*, 1990]; thus, the time complexity of the BCI-kf detector will scale approximately with the square of the number of sensors. As discussed previously, robust Kalman filtering is a weighted average of $k$ Kalman filters, where $k$ is the number of Gaussian components in the MOG distribution. Since the number of Gaussian components of the BCI-rkf detector is related to the number of sensors as $2^n$, the BCI-rkf detector will scale as O($2^n n^{2.376}$). As discussed previously, the Rao-Blackwellized particle filter is essentially a population of $p$ Kalman filters, where $p$ is the number of particles; thus, the MAP-ms detector will scale as O($pn^{2.376}$). This result is confirmed by *Hill*'s [2007] observation that the time complexity of the Rao-Blackwellized particle filter scales linearly with the number of particles. As discussed in the previous section, the upper limit for $p$ is inversely proportional to the frequency of the least likely measurement status (in this study, the failure of all eight sensors, which has probability of $0.05^8$), though to achieve a performance similar to that of the BCI-rkf method, the number of particles, $p$, needs to be much larger than $2^n$. These results suggest that $p$ increases faster than an exponential function of the number of sensors. From this analysis, it appears that the time complexity of both the BCI-rkf and MAP-ms detectors scales faster than an exponential function of the number of sensors, thus suggesting that these methods would be intractable for a very large number of sensors. However, this analysis assumes that all of the sensor data streams modeled by the DBNs are highly correlated, thus requiring a fully coupled model of the processes being measured. For processes that are marginally correlated, however, a fully coupled model would not be necessary. Decoupling weakly correlated processes within the DBN framework does not significantly affect the quality of the DBN model [*Boyen and Koller*, 1998, 1999] and would result in substantial computational economy.

[50] Because DBNs do not require that the processes they model be stationary, but only that the process dynamics be stationary, the DBNs employed by the BCI and MAP-ms detectors would only have to be retrained if the dynamics of the system were to change. While it is unlikely that system dynamics will change quickly, periodic reparameterization of the DBN may be desirable, so that new data streams or new information about the types of anomalies that may be encountered can be included in the DBN. Parameter learning is somewhat time consuming, requiring several minutes on a RedHat Linux workstation equipped with an Intel Xeon 2.4 GHz processor and 1 GB of memory. However, a dual model approach can be used, in which a new model is trained while the previous model is being used for anomaly detection, such that data to be processed do not accumulate while reparameterization of the DBN is occurring. Finally, since the BCI-rkf method is reasonably insensitive to the anomalous measurement variance and BCI, and the MAP-ms method is reasonably insensitive to the anomalous measurement variance, selecting a large BCI and/or a large-valued anomalous measurement variance will suffice (as discussed in section 2).

## 5. Conclusions

[51] This paper presents three DBN-based anomaly detection methods employing the well-known Kalman filter, the robust Kalman filter, and the recently developed Rao-Blackwellized particle filter, which have not yet found wide application in environmental research, and compares these methods to each other and to an autoregressive data-driven anomaly detector. The DBNs were implemented such that they are robust to missing values in the sensor data streams by adaptively modifying the filtering method to use only the available measurements. The Bayesian anomaly detection methods perform fast, incremental evaluation of data as they become available, can scale up to large quantities of data, and require no a priori information regarding process variables or the types of anomalies that may be encountered. Furthermore, these methods can process data from multiple sensors at the same time and thus, as demonstrated, can be applied to a network of heterogeneous sensors.

[52] The value and efficacy of the BCI-kf, BCI-rkf, and MAP-ms anomaly detection methods are illustrated using a case study involving eight data streams, including wind speed, wind direction, air temperature, and barometric pressure, at two spatial locations within Corpus Christi Bay. In this case study, the performance of these detectors was evaluated using a suite of synthetic and actual data anomalies. The synthetic anomaly results indicated that all of the methods require less time to evaluate a new measurement than the frequency at which the measurements are collected and are thus suitable for real-time anomaly detection. Furthermore, the BCI-rkf and MAP-ms methods had false positive and negative rates of less than 2% and 1.5%, respectively, and misclassified significantly fewer data points than did the BCI-kf or AR_ADET methods, both of which had false positive and negative rates in excess of 10%. The BCI-rkf and MAP-ms methods outperformed the BCI-kf and AR_ADET methods because the DBNs employed by these detectors can explicitly account for outliers in the sensor measurements and thus are not as adversely affected by measurement outliers as are the BCI-kf and AR_ADET methods. Additionally, the AR_ADET method only considers a single data stream and thus cannot take advantage of information from other sensors, which improved the classification accuracy of the DBN-based detectors. The BCI-rkf and MAP-ms methods also performed well at identifying anomalous data caused by two real anomalous events that manual QA/QC had failed to detect. The detection of both system anomalies, such as the storm front in the first event, and measurement anomalies caused by failing sensors in the second event, indicates that even with no a priori information about the types of anomalies that could be encountered, the Bayesian anomaly detectors were effective at identifying real anomalies in the data.

[53] These anomaly detection methods could be incorporated into an adaptive sampling method that would trigger further sampling in real time. For this type of application, the nonspecific identification of both types of anomalies might be acceptable (since it is often desirable to know under what conditions the sensors fail). However, if anomaly detection were to be incorporated into a real-time QA/QC algorithm or into a data cleaning system, or if it were to be used for real-time modeling, then developing detectors capable of making distinctions among system and measurement anomalies would clearly be beneficial. While there are existing models [e.g., *Koushanfar et al.*, 2003] that explain how sensor failures affect data, the parameters of these

models are specific to each deployed sensor, and sufficient instances of sensor failure have not been observed to parameterize these models, especially for new sensor deployments (such as the Corpus Christi meteorological array). Deployment of one of the DBN-based anomaly detectors presented here on a sensor array would make possible the accumulation of a labeled set of data corrupted by sensor failure that could be used to create better models of sensor failures. These models could then be incorporated into the DBN-based anomaly detector, such that the detector could recognize known types of sensor failures. This feature would be beneficial for suggesting remedial action on the sensor, as well as for aiding the detector in discriminating between measurement and process anomalies.

[54] There are at least two other extensions of this research that could be explored. The first extension would be to investigate DBNs that can represent a wider variety of distributions than the linear Gaussian, mixture of Gaussian, and conditional linear Gaussian variables considered here. For example, perhaps the circular Gaussian (von Mises) distribution [*Best and Fisher*, 1979] could be used to represent the wind direction variable. The second extension would be to test the detectors on other types of sensors and evaluate whether errors in these sensors have correlations and, if so, whether the anomaly detectors need to be modified to reliably identify correlated errors.

[55] **Acknowledgments.** The authors would like to thank Desiree Trujillo, John Perez, Terry Riggs, and Jim Bonner of Shoreline Environmental Research Facility, Corpus Christi, Texas, for providing data and knowledge regarding their sensors. The authors gratefully acknowledge the help and valuable suggestions of Praveen Kumar, University of Illinois at Urbana-Champaign. This study was supported by the Office of Naval Research under grant N00014-04-1-0437.

## References

Best, D., and N. Fisher (1979), Efficient simulation of the von Mises distribution, *Appl. Stat.*, *28*, 152–157, doi:10.2307/2346732.

Blum, R. S., Y. Zhang, B. Sadler, and R. Kozick (1999), On the approximation of correlated non-Gaussian noise pdfs using Gaussian mixture models, paper presented at 1st Conference on the Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics, Am. Stat. Assoc., Washington, D. C.

Boyen, X., and D. Koller (1998), Tractable inference for complex stochastic processes, in *Proceedings of the Conference on Uncertainty in AI*, edited by G. Cooper and S. Moral, pp. 33–42, Morgan Kaufmann, San Francisco, Calif.

Boyen, X., and D. Koller (1999), Exploiting the architecture of dynamic systems, in *Proceedings of the Sixteenth National Conference on Artificial Intellitence (AAAI-99)*, pp. 313–320, AAAI Press, Menlo Park, Calif.

Casella, G., and C. P. Robert (1996), Rao-Blackwellisation of sampling schemes, *Biometrika*, *83*(1), 81–94, doi:10.1093/biomet/83.1.81.

Coppersmith, D., and S. Winograd (1990), Matrix multiplication via arithmetic progressions, *J. Symbolic Comput.*, *9*, 251–280, doi:10.1016/S0747-7171(08)80013-2.

Dereszynski, E. W., and T. G. Dietterich (2007), Probabilistic models for anomaly detection in remote sensor data streams, in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, edited by R. Parr and L. van der Gaag, pp. 75–82, AUAI Pres, Arlington, Va.

Digalakis, V., J. R. Rohlicek, and M. Ostendorf (1993), ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition, *IEEE Trans. Speech Audio Process.*, *1*(4), 431–442, doi:10.1109/89.242489.

Doucet, A., N. de Freitas, K. Murphy, and S. Russell (2000a), Rao-Blackwellised particle filtering for dynamic Bayesian networks, in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, edited by C. Boutilier and M. Goldszmidt, pp. 176–183, Morgan Kaufmann, San Francisco, Calif.

Doucet, A., S. Godsill, and C. Andrieu (2000b), On sequential Monte Carlo sampling methods for Bayesian filtering, *Stat. Comput.*, *10*(3), 197–208, doi:10.1023/A:1008935410038.

Efron, B., and R. A. Olshen (1978), How broad is the class of normal scale mixtures?, *Ann. Stat.*, *6*(5), 1159–1164, doi:10.1214/aos/1176344318.

Frühwirth, R. (1995), Track fitting with long-tailed noise: A Bayesian approach, *Comput. Phys. Commun.*, *85*, 189–199, doi:10.1016/0010-4655(94)00121-H.

Ghahramani, Z., and G. E. Hinton (1996), Parameter estimation for linear dynamical systems, Tech. Rep. CRG-TR-96-2, Dep. of Comput. Sci., Univ. of Toronto, Toronto, Ont., Canada.

Goel, P., G. Dedeoglu, S. I. Roumeliotis, and G. S. Sukhatme (2000), Fault detection and identification in a mobile robot using multiple model estimation and neural network, in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2302–2309, Inst of Electr. and Electron. Eng., New York.

Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning*, Springer, New York.

Hill, D. (2007), Data mining approaches to complex environmental problems, Ph.D. thesis, Univ. of Ill. at Urbana-Champaign, Champaign.

Hill, D. J., and B. S. Minsker (2006), Automated fault detection for in-situ environmental sensors, in *Hydroinformatics 2006: Proceedings of the 7th International Conference on Hydroinformatics*, edited by P. Gourbesville, J. Cunge, and S.-Y. Liong, Res. Publ. Serv., Chennai, India.

Hill, D. J., B. S. Minsker, and E. Amir (2007), Real-time Bayesian anomaly detection for environmental sensor data, paper presented at 32nd Congress, Int. Assoc. of Hydraul. Eng. and Res., Venice, Italy.

Hodge, V. J., and J. Austin (2004), A survey of outlier detection methodologies, *Artif. Intell. Rev.*, *22*, 85–126.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D*, *82*, 35–45.

Koushanfar, F., M. Potkonjak, and A. Sangiovanni-Vincentelli (2003), On-line fault detection of sensor measurements, in *Second IEEE International Conference on Sensors*, vol. 2, pp. 974–979, Inst. of Electr. and Electron. Eng., New York.

Krajewski, W. F., and K. L. Krajewski (1989), Real-time quality control of streamflow data—A simulation study, *Water Resour. Bull.*, *25*(2), 391–399.

Lerner, U. N. (2002), Hybrid Bayesian networks for reasoning about complex systems, Ph.D. thesis, Stanford Univ., Stanford, Calif.

Lerner, U., and R. Parr (2001), Inference in hybrid networks: Theoretical limits and practical algorithms, *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence*, edited by J. Breese and D. Koller, pp. 310–318, Morgan Kaufmann, San Francisco, Calif.

Lerner, U., R. Parr, D. Koller, and G. Biswas (2000), Bayesian fault detection and diagnosis in dynamic systems, in *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pp. 531–537, AAAI Press, Menlo Park, Calif.

Liu, H., and H. Motoda (1998), Feature transformation and subset selection, *IEEE Intell. Syst.*, *13*(2), 26–28, doi:10.1109/MIS.1998.671088.

Liu, X., and A. Goldsmith (2004), Kalman filtering with partial observation losses, in *43rd IEEE Conference on Decision and Control*, vol. 4, pp. 4180–4186, Inst. of Electr. and Electron. Eng., New York.

Maybeck, P. S. (1979), *Stochastic Models, Estimation, and Control*, 2nd ed., Academic, San Diego, Calif.

McLachlan, G., and D. Peel (2000), *Finite Mixture Models*, John Wiley, New York.

Meinhold, R. J., and N. D. Singpurwalla (1989), Robustification of Kalman filter models, *J. Am. Stat. Assoc.*, *84*(406), 479–486, doi:10.2307/2289933.

Mourad, M., and J. L. Bertrand-Krajewski (2002), A method for automatic validation of long time series of data in urban hydrology, *Water Sci. Technol.*, *45*(4–5), 263–270.

National Research Council (2006), *CLEANER and NSF's Environmental Observatories*, Natl. Acad. Press, Washington, D. C.

National Science Foundation (2005), Sensors for Environmental Observatories Rrport of the NSF sponsored workshop December 2004, Arlington, Va.

Nicholson, A. E., and J. M. Brady (1994), Dynamic belief networks for discrete monitoring, *IEEE Trans. Syst. Man Cybern.*, *24*(11), 1593–1610, doi:10.1109/21.328910.

Peña, D., and I. Guttman (1988), Bayesian approach to robustifying the Kalman filter, in *Bayesian Analysis of Time Series and Dynamic Models*, edited by J. C. Spall, pp. 227–253, Marcel Dekker, New York.

Ramanathan, N., L. Balzano, M. Burt, D. Estrin, E. Kohler, T. Harmon, C. Harvey, J. Jay, S. Rothberg, and M. Srivastava (2006), Monitoring a toxin in a rural rice field with a wireless sensor network, Tech. Rep. 62, Cent. for Embedded Network Syst., Univ. of Calif., Los Angeles, Calif.

Schick, I. C., and S. K. Mitter (1994), Robust recursive estimation in the presence of heavy-tailed noise, *Ann. Stat.*, *22*(2), 1045–1080, doi:10.1214/aos/1176325511.

Shumway, R. H., and D. S. Stoffer (1982), An approach to time series smoothing and forecasting using the EM algorithm, *J. Time Ser. Anal.*, *3*(4), 253–264, doi:10.1111/j.1467-9892.1982.tb00349.x.

Spall, J. C. (1988), An overview of key developments in dynamic modeling and estimation, in *Bayesian Analysis of Time Series and Dynamic Models*, edited by J. C. Spall, pp. xv–xxvii, Marcel Dekker, New York.

E. Amir, Department of Computer Science, University of Illinois at Urbana-Champaign, 3314 Siebel Center, MC-258, 201 North Goodwin Road, Urbana, IL 61801-2302, USA. (eyal@cs.uiuc.edu)

D. J. Hill, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 4018 NCSA, MC-257, 1205 West Clark Street, Urbana, IL 61801, USA. (djhill1@uiuc.edu)

B. S. Minsker, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 3230d NCEL, MC-250, 205 North Mathews Avenue, Urbana, IL 61801, USA. (minsker@uiuc.edu)