ADAPTIVE MANAGEMENT OF WATER RESOURCES
AWRA SUMMER SPECIALTY CONFERENCE
Missoula, Montana
June 26-28, 2006

AUTOMATED FAULT DETECTION: PREPARING REAL-TIME DATA FOR ADAPTIVE MANAGEMENT

David J. Hill, Barbara S. Minsker[*]

ABSTRACT: Real-time data provides an up-to-the minute characterization of environmental systems, which can be used to adapt management decisions to changing environmental conditions. The recent deployment of in-situ sensor arrays has increased the availability of real-time data to decision makers. When using sensor data for real-time decision-making, however, there is little time to verify its quality; therefore, a method for quickly and accurately identifying faults in the data collection process is needed. This study develops two fault detection strategies for in-situ sensors that are based on data-driven regression models of the sensor data. Sensor faults are determined by identifying anomalous measurements in the data stream, where anomalous measurements are operationally defined as measurements that fall outside of the bounds of an established prediction interval. Eight instantiations of each detection strategy are created, using different data-driven methods and either a 95% or a 99% prediction interval. The performance of these detectors for identifying data transmission faults is compared using windspeed data originating from Corpus Christi Bay, Texas. The basis for comparison is the number of false positives/negatives identified by each of the detectors. The results indicate that the strategies perform well for identifying data transmission faults.
KEY TERMS: coastal environment; data-driven modeling; fault detection; machine learning; real-time data; sensor networks

INTRODUCTION

Traditionally, decision-makers have relied on historical or averaged seasonal data to predict the response of environmental systems to events of interest, such as contaminant releases. Unfortunately, reliance on these types of data limits the value of these predictions for coordinating real-time responses to such events. For example, Elliott & Jones (2000) and Guinasso *et al.* (2001) present case studies in which real-time data successfully indicated the trajectory of oil slicks caused by tanker accidents, whereas the trajectory predicted using models relying on less current data differed significantly. In the latter case, decisions facilitated by the real-time data allowed the clean-up effort to be focused on specific areas, thus increasing clean-up efficiency.

The recent deployment of arrays of in-situ sensors into many environments provides a new opportunity to improve prediction accuracy and enable real-time adaptive management. For example, the Shoreline Environmental Research Facility (SERF) in Corpus Christi, Texas maintains a growing array of sensors that monitor estuarine and coastal environments in the Corpus Christi region. The value of the data collected by this sensor array for real-time decision-making has been demonstrated using a simulated oil spill (Bonner *et al.*, 2002), and continuing efforts are directed towards facilitating decision making using the real-time data collected by these sensors (Shah *et al.*, 2005).

In-situ sensors (sometimes called “embedded” sensors) are sensors that are physically located in the environment they are monitoring. These sensors collect time series data that flow from the sensor to the data repository continuously, creating a data stream. The sensors operate under harsh conditions, and the data they collect must be transmitted across wireless networks; thus, the data can easily become corrupted through data transmission or sensor faults. Undetected erroneous data can significantly affect the value of the collected data for applications such as environmental monitoring or real-time forecasting. For this reason, a method for detecting erroneous data before it is archived is necessary to ensure its quality. Due to the vast quantity of data being collected, this method must be automated in order for it to be practical. Since it is often difficult to determine whether an anomalous measurement has occurred due to a sensor or data transmission fault, or due to an unusual environmental system response, many fault detection techniques seek to identify anomalous measurements—measurements that do not fit the historical pattern of the data, but may not necessarily be caused by sensor or

[*] Respectively, Graduate Research Assistant, Associate Professor, University of Illinois Urbana-Champaign, Department of Civil and Environmental Engineering, 205 N. Mathews, Urbana, IL 61801, Phone: (217) 333-1657, Fax: (217)333-6968, E-Mail: djhill1@uiuc.edu.

data transmission faults. The causes of these measurements can then be investigated to determine whether or not the measurement actually represents the environmental system state.

This study develops two automated strategies for detecting anomalous data and compares the performance of eight instantiations of each strategy in detecting erroneous data caused by data transmission faults using a case study from an environmental sensor network deployed by SERF in Corpus Christi Bay, Texas. The following section outlines previous work on fault detection. Then, the two fault detection strategies and their instantiations are described. Application of the 16 instantiations to one of the Corpus Christi Bay sensors is presented, and the results of this case study are discussed. Finally, conclusions and suggestions for future work are given.

PREVIOUS RESEARCH

Previous work on fault detection has been conducted in the field of process control, and applications have traditionally been in the areas of manufacturing and electrical energy generation. Many of the published methods (Silvestri *et al.*, 1994; Upadhyaya *et al.*, 1990; Belle *et al.*, 1983) identify anomalous measurements based on the residual between a model of the time series and the actual values, while others (Ananthanaravanan & Hobert, 2004; DePold *et al.*, 2003; Mehanbod *et al.*, 2003) identify faulty sensors using other measurements of the system state. It is interesting to note that many of these techniques operate in controlled environments, such as manufacturing/power plants or machinery, where sensors are measuring process variables that are likely to be within a predefined range. In this research, however, the sensors are located in the natural environment, measuring very unpredictable process variables for which a predefined range is not known *a priori*. Furthermore, to the authors' knowledge, these studies did not consider errors caused by data transmission faults.

ANOMALY DETECTION STRATEGIES

Two strategies for detecting anomalous data are considered in this study: anomaly detection (AD) and anomaly detection and mitigation (ADAM). Both strategies consider the data stream sequentially and use a model to predict the next measurement and the bounds of the n% prediction interval (PI). The prediction interval gives the range of plausible values that the next measurement can take, and the prediction level (n) indicates the expected frequency with which measurements will actually fall in this range. If the new measurement falls within the bounds of the PI, then the measurement is classified as non-anomalous; otherwise, it is classified as anomalous. The AD strategy simply uses the previous measurements for future predictions, whether or not they were classified as anomalous, while the ADAM strategy replaces anomalous measurements with model predictions before making future predictions. In this study, four data-driven methods were used to create prediction models: naïve, clustering, perceptron, and artificial neural network (ANN). Data-driven methods like these develop models using sets of training examples. Each example contains a feature set (i.e. the set of variables used to make the prediction) and a target output. Training these models involves fitting their parameters to minimize error on the training examples, without overfitting that can lead to poor predictions. Since the PI requires the quantification of the standard deviation of the model error, 10-fold cross-validation (Duda *et al.*, 2001; Han & Kamber, 2000) was used to train the models. The remainder of this section summarizes these four modeling methods.

The naïve predictor is a nearest-neighbor approach (Duda *et al.*, 2001; Hastie *et al.*, 2001) that bases its prediction of an unseen event on the response of the system to the most similar historical event and defines similarity in terms of temporal distance. Thus, the naïve prediction of a measurement at time $t+\Delta t$ is equal to the value of the measurement at time t.

The clustering predictor (Vasquez & Fraichard, 2004) is slightly more sophisticated than the naïve method because, while it also predicts the value of an unseen event based on the observed responses of similar events, it defines similarity by mapping each measurement to a region of feature space. It then partitions the feature space into local regions (clusters) based on the training data and predicts the system response from each cluster to be the mean of the training data target values that mapped to each cluster. This method is similar to the one presented by Upadhyaya *et al.* (1988), except that no cluster dependent linear models are used to refine the prediction. The k-means clustering algorithm (Duda *et al.*, 2001; Hastie *et al.*, 2001; Han & Kamber, 2000) is used because it scales well to large quantities of data (Schütze & Silverstein, 1997). The number of clusters is specified using within cluster scatter (Hastie *et al.*, 2001), which indicates the similarity of the points to their assigned cluster center.

The linear perceptron model (Bishop, 1995) predicts the response of a system to be a linear combination of the input features describing the system state. The perceptron in this study is trained using the perceptron learning rule (Rosenblatt, 1958) with a user defined learning rate. This rate is selected using a trial-and-error approach.

Artificial neural networks (Duda *et al.*, 2001, Hastie *et al.*, 2001, Bishop, 1995) are networks of perceptron-like nodes that are capable of creating models of a system state that are non-linear combinations of the input features. The ANN

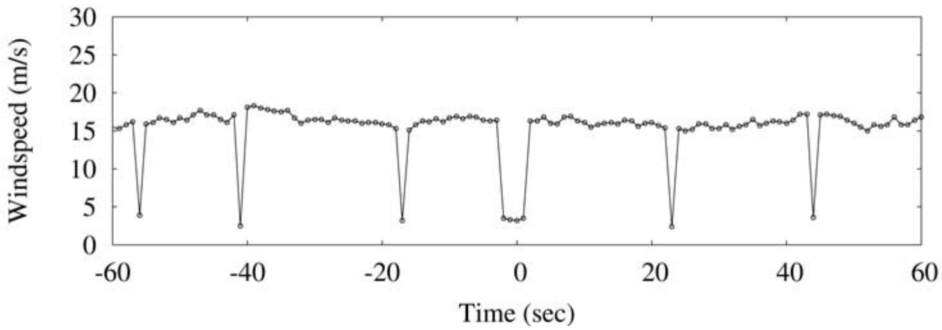

Figure 1. Data exhibiting errors resulting from short duration faults.

considered in this study is a feed-forward network that is trained using the standard backpropagation algorithm with gradient descent (Rummelhart *et al*, 1986). Training was terminated after a fixed number of training epochs were completed or when further training caused a decrease in the model performance on a testing set (Hastie *et al.*, 2001, Bishop, 1995). This latter condition discourages overtraining of the network. ANNs require that a learning rate, number of hidden layers, number of nodes in each hidden layer, and number of training epochs be specified. These values are selected using a trial and error approach (Bishop, 1995, Hecht-Nielsen, 1990).

CASE STUDY

To demonstrate and compare the efficacy of the fault detection strategies developed in this study, they were applied to a windspeed sensor data stream from Corpus Christi Bay. The AD and ADAM strategies were tested using all four modeling methods and 95% and 99% prediction intervals. These 16 combinations will hereafter be referred to as fault detectors. The features used to model the data stream were selected using correlation analysis, a common approach in time-series modeling (Box & Jenkins, 1970). This analysis indicated that the windspeed is strongly correlated with historical measurements as distant as 5.5 hours. However, because the measurement frequency is one second, it was necessary to reduce the number of descriptive features to the most recent 30 seconds of data. The windspeed models were developed using 30,000 training examples selected at random from the period of January–May 2004.The 16 resulting fault detectors were then compared, based on their ability to identify erroneous data caused by data transmission faults.

RESULTS

Since the data used in this study were subjected to manual quality control measures before they were archived, it was expected that the detectors would not identify many data anomalies in the archive. However, this was not the case. The detectors identified approximately 6% of the data during the month of June as anomalous. This result encouraged focused inspection of these data. For example, Figure 1 shows a two minute segment of the data stream in which six suspicious events can be easily identified. All six of these events had been classified as anomalous by one or more of the 16 detectors. Subsequent investigation of data, such as these six anomalous points, by the SERF data managers revealed that events such as these were most likely caused by wireless transmission errors. Further analysis of anomalous data revealed other suspicious events that were of significantly longer duration than those shown in Figure 1. For example, Figure 2 shows a 35 minute segment of the data stream during which a suspicious long duration event occurs. The windspeed between minutes 8 and 26 in the plot appear to have been offset by a constant 7 m/s. It is the sharpness of the transition from the slower (~5 m/s) to the faster (~12 m/s) windspeed and back, as well as the existence of data in both the slow and fast regimes that appear to correlate with data in the opposite regime, which suggests that a significant portion of the data presented in this figure do not represent the actual windspeed. This data segment is of particular interest, because its behavior is similar to the behavior of an offset bias sensor fault (Koushanfar *et al.*, 2003). Errors resulting from short duration faults, such as those shown in Figure 1, may not have a significant effect on the utility of the data if time averages are used (e.g. 2-minute averages). However, the high frequency with which these events are observed in this data stream may adversely affect time averages. Long duration errors, such as those shown in Figure 2, are more worrisome because their effect can only be mitigated if longer time averages are used.

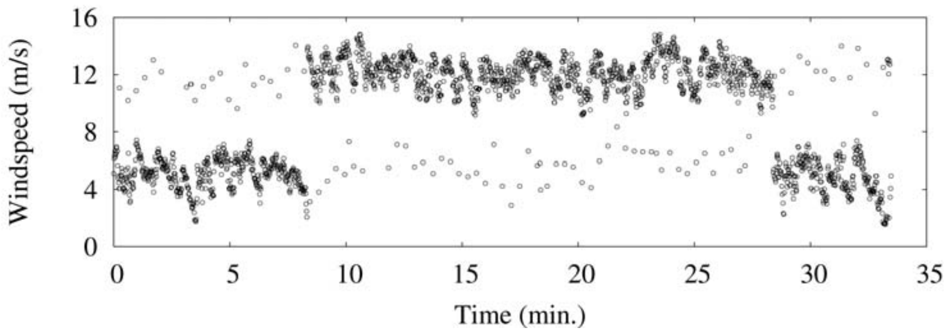

Figure 2. Data exhibiting errors resulting from long duration faults.

The existence of errors in June 2004 data indicated that the data from January-May (used for training) also contained errors. Thus, before proceeding with an assessment of the 16 detectors, it was necessary to clean the training data and retrain the regression models. Cleaning was performed using the Naïve-AD detector with a 95% PI. Records containing anomalous data were removed from the training set. The naïve detector was chosen because it performs well in identifying anomalous data and because it does not rely on a model of the data stream that could have been affected by errors in the training data. The AD strategy was used because, unlike the ADAM strategy, previous misclassifications do not affect its future performance.

Once the training data were cleaned, the performance of the 16 fault detectors for identifying additional data transmission faults was quantified using a sample of over 1200 other data points from the data archive. True/false positives were identified visually using domain knowledge provided by the data managers. Figure 3 shows the detectors' false positive rate for identifying erroneous data attributed to transmission faults. It can be seen that the ADAM strategy reduces the false positive rates of the perceptron and ANN-based detectors, whereas it increases the false positive rates of the naïve and clustering-based detectors. The use of mitigation decreases the number of false positives reported by the perceptron and ANN-based detectors because, without its use, transmission fault induced errors adversely affect future classifications by skewing input values to future prediction intervals. The naïve and clustering-based detectors, however, appear to be less sensitive to transmission fault induced errors in their input data. Perhaps this is due to the difference in how the perceptron and ANN-based detectors and naïve and clustering-based detectors predict the future windspeed: Unlike the perceptron and ANN detectors, the naïve and clustering-based methods do not predict the future windspeed using a function of the input values. Rather, these detectors predict the future windspeed using a similar previously observed example. Thus, measurements that vary significantly from the current locally averaged windspeed (e.g. errors due to transmission faults) in the input values to these detectors will not cause a future windspeed prediction that is vastly different from previously observed windspeeds.

However, the use of the ADAM strategy can negatively impact the performance of a detector when it misclassifies a

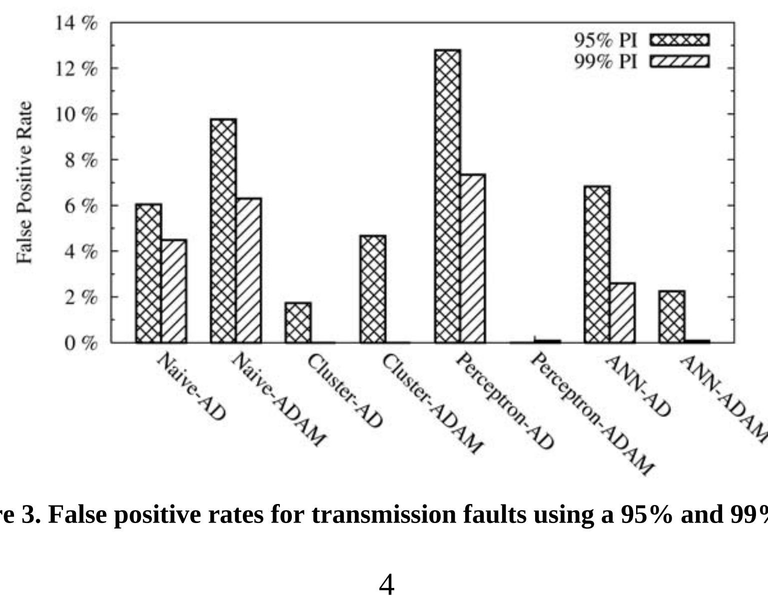

Figure 3. False positive rates for transmission faults using a 95% and 99% PI.

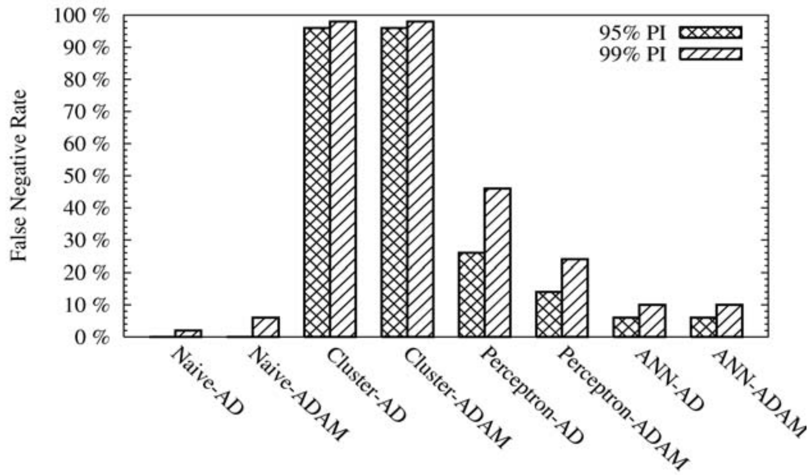

Figure 4. False negative rates for detecting transmission faults using a 95% and 99% PI.

non-anomalous point, because this causes the detector to replace the valid measurement with an incorrect value, which sometimes results in the detector continuing to make mistakes and replace valid measurements with incorrect values until the cycle is broken. In the case of the perceptron and ANN-based detectors, this behavior is uncommon and outweighed by the positive benefits of mitigation described previously, whereas in the case of the naïve and clustering-based detectors, this behavior significantly affects the detectors' performance. Since the naïve-based detectors use only one data point to predict the future windspeed, it is understandable that, when using the ADAM strategy, this method will perpetuate misclassifications because future predictions will reflect the incorrect values that replaced valid measurements. However, it is less clear why the clustering-based detectors behave in this manner. Figure 3 also indicates that an increase in the prediction level from 95% to 99% decreases the number of false positives, though this decrease is not as dramatic as the decrease/increase associated with the use of mitigation, indicating that the appropriate use of mitigation has the most significant effect on the detectors' false positive rate.

The false negative rate for the detectors is shown in Figure 4. It can be seen that the naïve, perceptron, and ANN-based detectors misclassify fewer erroneous data than the cluster-based detectors, which misclassify almost all of the erroneous data. Thus, the clustering-based detectors are not useful for detecting transmission fault induced data anomalies. It can also be seen that the use of the ADAM strategy significantly improves the ability of the perceptron-based detectors to correctly classify erroneous data, whereas it does not significantly affect the false negative rates of the other detectors. The decrease in the perceptron-based detectors' false negative rate, due to the use of the ADAM strategy, can again be attributed to transmission fault induced errors adversely affecting future classifications by skewing input values to future prediction intervals. Furthermore, an increase in the prediction level from 95% to 99% results in an increase in the false negative rate, which, for the best performing detectors, is larger than the corresponding decrease in the false positive rate.

CONCLUSIONS

This case study demonstrates the value and efficacy of the proposed fault detection strategies. Fault detectors using both the AD and ADAM strategies identified a significant number of previously unidentified errors in windspeed data from Corpus Christi Bay. The errors had durations ranging from 1 second to several minutes and affected approximately 6% of the data. After cleaning the errors in the training data, an assessment of 8 instantiations of both the AD and ADAM strategies indicated that the performance of the perceptron and ANN based detectors in detecting errors in the testing data was significantly improved by the use of the ADAM strategy, and that the naïve-AD, perceptron-ADAM, and ANN-ADAM detectors performed well. Furthermore, for these strategies, there is a larger increase in the false negative rate than decrease in the false positive rate when the prediction level is increased from 95% to 99%, indicating that a prediction level of 95% provides a reasonable tradeoff between misclassifying non-anomalous and anomalous points.

The results presented here focus on the ability of the proposed fault detection strategies to identify errors in the data resulting from transmission faults. Due to the harsh environment in which in-situ sensors must operate, sensor faults are also expected. At the conference, we plan to present additional results quantifying the performance of these fault detection strategies for identifying additional sensor faults.

ACKNOWLEDGMENTS

The authors would like to thank Desiree Trujillo, John Perez, Terry Riggs, and Jim Bonner of Shoreline Environmental Research Facility, of Corpus Christi, Texas, for providing data and knowledge regarding the sensors. This study was supported by the Office of Naval Research under grant N00014-04-1-0437.

REFERENCES

Ananthanarayanan, V. and K.E. Holbert, 2004. Power System Sensor Failure Detection and Characterization using Fuzzy Logic. Proc. of the Seventh IASTED International Conference on Power and Energy Systems, Clear Water Beach, FL, pp. 291-296.

Belle, R., B. Upadhyaya and M. Skorska, 1983. Sensor Fault Analysis using Decision Theory and Data-Driven Modeling of Pressure Water Reactor Subsystems. Nuclear Technology 64:70-77.

Bishop, C., 1995. Neural Networks for Pattern Recognition, Oxford University Press, New York.

Bonner, J.S., F.J. Kelly, P.R. Michaud, C.A. Page, J. Perez, C. Fuller, T. Ojo, and M. Sterling, 2002. Sensing the Coastal Environment. Proc. Third International Conference on EuroGOOS; Building the European Capacity in Operational Oceanography, pp. 167-173.

Box, G. and C. Jenkins, 1970. Time Series Analysis: Forecasting and Control, Holden-Day Inc., San Francisco.

DePold, H., A. Volponi, J. Siegel and J. Hull, 2003. Validation of Diagnostic Data with Statistical Analysis and Embedded Knowledge. Proc. American Society of Mechanical Engineers, International Gas Turbine Institute, Turbo Expo IGTI, Vol. 1 , pp 573-579.

Duda, R. P. Hart and D. Stork, 201. Pattern Classification, Wiley-Interscience, New York.

Elliott, A.J. and B. Jones, 2000. The Need for Operational Forecasting During Oil Spill Response. Marine Pollution Bulletin 40(2):110-121.

Guinaso, N.L., Jr.,J. Yip, R.O. Reid, L.C. Bender III, M. Howard, L.L Lee III, J.N. Walpert, D.A. Brooks, R.D. Hetland and R.D. Martin, 2001. Observing and Forecasting Coastal Currents: Texas Automated Buoy System (TABS), Proc. OCEANS 2001 MTS/IEEE, Honolulu, pp 1318-1322.

Han, J. and M. Kamber, 2000. Data Mining: Concepts and Techniques, Morgan Kaufmann, New York.

Hastie, T., R. Tibshirani and J. Friedman, 2001. The Elements of Statistical Learning, Springer-Verlag, New York.

Hecht-Nielsen, R., 1990. Neurocomputing, Adison-Wesley, New York.

Koushanfar, F., M. Potkonjak and A. Sangiovanni-Vincentelli, 2003. On-line Fault Detection of Sensor Measurements, Proc. IEEE Sensors, Vol. 2, pp 974-979.

Mehranbod, N., M. Soroush, M. Piovoso and B. Ogunnaike, 2003. Probabilistic Model for Sensor Fault Detection and Identification. AIChE Journal 49(7):1787-1802.

Rosenblatt, F., 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review 65:386-408.

Rummelhart, D.E., G. Hinton and R. Williams, 1986. Learning Internal Representations by Error Propagation. In D.E. Rummelhart, J.L. McClelland, and the PPD Research Group, eds. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press, Cambridge, MA, pp 318-62.

Schütze, H. and C. Silverstein, 1997. Projections for Efficient Document Clustering, Proc. of SIGIR ’97, Philadelphia, pp 74-81.

Shah, K., J. Bonner, D. Trujillo, C. Page and F. Kelly, 2005. Development of Real-time Data Monitoring System for Coastal Margin Research, Proc. 2005 International Oil Spill Conference, Miami.

Silvestri, G., F. Verona, M. Innocenti and M. Napolitano, 1994. Fault Detection using Neural Networks, Proc. IEEE International Conference on Neural Networks, Vol. 6, pp. 3796-799.

Upadhyaya, B., O. Glockler and J. Eklund, 1990. Multivariate Statistical Signal Processing for Fault Detection and Diagnostics. ISA Transactions 29(4):79-95.

Upadhyaya, B., G. Mathai and J. Green, 1988. Data Clustering and Prediction for Fault Detection and Diagnostics, Proc. American Control Conference, Atlanta, Vol. 1, pp 650-651.

Vasquez, D. and T. Fraichard, 2004. Motion Prediction of Moving Objects: a Statistical Approach, Proc. IEEE International Conference on Robotics and Automation, New Orleans, Vol. 4, pp 3931-3936.