Studies on Development of Efficient State Estimation Models, Time Synchronization Techniques for Discrete-Time Wireless Sensor Network Applications

THESIS SUBMITTED BY

ADITI CHATTERJEE

DOCTOR OF PHILOSOPHY (ENGINEERING)

DEPT. OF ELECTRONICS & TELECOMMUNICATION ENGG.

FACULTY COUNCIL OF ENGINEERING & TECHNOLOGY

JADAVPUR UNIVERSITY

KOLKATA-700032

INDIA

2018

JADAVPUR UNIVERSITY

KOLKATA-700032, INDIA

INDEX NO. 102/15/E

1. Title of the Thesis:

Studies on Development of Efficient State Estimation Models, Time Synchronization Techniques for Discrete-Time Wireless Sensor Network Applications.

2. Name, Designation & Institution of the Supervisor:

Dr. P. Venkateswaran, Professor

Department of Electronics & Telecommunication Engineering,

Jadavpur University, Kolkata-700032.

3. List of Publications:

Journal Articles

[1] **A. Chatterjee**, P. Venkateswaran, and D. Mukherjee, "A unified approach of simultaneous state estimation and anomalous node detection in distributed wireless sensor networks," *International Journal of Communication Systems*, Vol. 30, No. 9, June 2017.

[2] **A. Chatterjee**, P. Venkateswaran, and K. Das, "Simultaneous State Estimation of Cluster-Based Wireless Sensor Networks," *IEEE Transactions on Wireless Communications*, Vol. 15, No. 12, pp. 7985-7995, December 2016.

[3] **A. Chatterjee**, P. Venkateswaran, "An efficient statistical approach for time synchronization in wireless sensor networks," *International Journal of Communication Systems*, Vol. 29, No. 4, pp. 722-733, March 2016.

Conference Proceedings

[1] **A. Chatterjee**, P. Venkateswaran, "A Novel Efficient State Estimation based Relocation Algorithm for Wireless Sensor Networks," *Applications and Innovations in Mobile Computing*, pp. 1-5, February, 2016.

[2] **A. Chatterjee**, and D. Mukherjee, "Variety event detection in Wireless Sensor Networks through single hop cluster topology," *Tenth International Conference on Wireless and Optical Communications Networks*, pp. 1–5, May 2013.

[3] **A. Chatterjee**, and D. Mukherjee, "Reliable Multipath Wireless Sensor Network Routing Protocol Scheme for Network Lifetime Maximization," *Second International Conference on Business and Information Management*, pp. 64–68, January 2014.

4. List of Patents: Nil

5. List of Presentations in International Conferences:

[1] **A. Chatterjee**, P. Venkateswaran, "A Novel Efficient State Estimation based Relocation Algorithm for Wireless Sensor Networks," *Applications and Innovations in Mobile Computing*, pp. 1-5, February, 2016.

[2] **A. Chatterjee**, and D. Mukherjee, "Reliable Multipath Wireless Sensor Network Routing Protocol Scheme for Network Lifetime Maximization," *Second International Conference on Business and Information Management*, pp. 64–68, January 2014.

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled ***"Studies on Development of Efficient State Estimation Models, Time Synchronization Techniques for Discrete-Time Wireless Sensor Network Applications"*** *submitted by* ***Smt. Aditi Chatterjee****, who got her name registered on* ***20.05.2015*** *for the award of Ph.D. (Engg.) degree of Jadavpur University is absolutely based upon her own work under the supervision of* ***Prof. Palaniandavar Venkateswaran*** *and that neither her thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.*

Signature of the Supervisor
and date with official seal

ACKNOWLEDGEMENTS

First of all, I express my sincere gratitude to my supervisor Prof. P. Venkateswaran, without whose continuous support, kind help and indomitable spirit I could not have completed my thesis. Ever since I entered Jadavpur University, he constantly guided me which culminates in the completion of my thesis work. I have learnt a lot from him during the course of this Ph.D. thesis work, and without his continuous support this thesis would not have been possible.

I am also indebted to Prof. R. Nandi, Prof. B. Gupta, Prof. S. K. Sarkar, Prof. C. Bose of Department of Electronics & Telecommunication Engineering (ETCE), Jadavpur University (JU) and Prof. S. Neogi of Department of Computer Science & Engineering, JU for their constant support and constructive suggestions at the various stages of my Ph.D. thesis work.

I am indeed grateful to my parents, my elder-sister, and brother-in-law for their immense support and inspiration. They motivated me a lot for my higher studies and research activities. They provided me enough strength to overcome the difficulties I faced during my student life. I also owe a lot to my in-laws for their inspiration and support.

Finally, I must acknowledge the endless support I received from my husband and our dearest son. Their deep understanding and encouragement has inspired me a lot, and helped me to complete the thesis work.

(Aditi Chatterjee)

Date:

Computer Networking Laboratory, T-3-16A

Dept. of Electronics & Telecommunication Engineering,

Jadavpur University, Kolkata-700032, India.

Contents

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
AMSE	Average Mean Squared Error
BIC	Bayesian Information Criterion
CPU	Central Processing Unit
CI	Confidence Interval
DIC	Deviance Information Criterion
DP	Dirichlet Process
EKF	Extended Kalman Filter
FPR	False Positive Rate
FTSP	Flooding Time Synchronization Protocol
GLS	Generalized Least Squares
IG	Inverse Gamma
KF	Kalman Filter
LASSO	Least Absolute Shrinkage and Selection Operator
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MCMC	Markov Chain Monte Carlo
MSBP	Matrix Stick-Breaking Process
MSE	Mean Squared Error
MH	Metropolis-Hastings
MCSE	Monte Carlo Standard Error
OLS	Ordinary Least Squares
ROS	Receiver Only synchronization
SE	Standard Error
TSPN	Time Synchronization Protocols for Sensor Networks
WSNs	Wireless Sensor Networks

Chapter 1

INTRODUCTION

The last decade witnessed huge advancement and wide applications of Wireless Sensor Networks (WSNs). WSNs are used in a variety of areas, including but not limited to, military surveillance [1], battlefield monitoring [2], health care [3], geological survey [4], criminology [5] and environment monitoring [6]. WSNs are mainly task-oriented networks with the objective of obtaining and transmitting information of interest. WSNs consist of ten to thousands of sensor nodes, which are low powered, and have limited processing and storage capacity. Sensor nodes sense and obtain relevant information (e.g. temperature, humidity, event occurrence etc.) and then transmit it to the sink of the network via wireless multi-hop routing. The sink is also known as base station, and it is a high-powered device linked to databases via satellite links [7].

WSNs can be used either for discrete time, or for continuous time monitoring. For a discrete time WSN, the time points $t_1, t_2, \ldots, t_k$, are integers, and for a continuous time WSN the time points are real numbers. The quantity being measured is usually continuous, e.g. room temperature, humidity etc. However, sensor nodes consume more energy for collecting, storing, and relaying continuous measurements over time. Since energy conservation is always desirable [8], sometimes sensor nodes are configured in such a way that they collect continuous measurements, and then based on some prefixed threshold value, the nodes store binary responses only (e.g. normal/abnormal). The nodes transmit these binary outcomes to the base station dynamically. Sometimes sensor nodes collect measurements at each time point, but store and transmit "important" information (e.g. event based, or query based) only to the base station [9,10]. WSNs that employ such protocols are energy efficient.

Reliable data delivery is the fundamental goal of WSNs [11]. However, reliability of data delivered can be affected by several factors, for example, node failure, presence of anomalous node(s), lack of proper communication, defective or disturbed communication mechanism etc. Such challenges motivated the researchers to explore further works in the areas of efficient network design, reliability checking, state estimation, and anomaly detection. The issues and challenges related to network design for WSNs are discussed in Hac [12], Fischione et al. [13], Lazaropoulos [14], Gupta and Sikka [15] and the references therein. A number of recent articles proposed reliability analysis of WSNs under different model settings. Examples include Zhu et al. [16], Mahmood et al. [17], Silva et al. [18], Damaso et al. [19], Zhu et al. [20], and the references therein. Because of their wide area of applications, it is extremely important to develop energy efficient and reliable WSNs.

Many real applications of WSNs are based on state estimation. However, most of the works on state estimation are based on the traditional state-space models [21]. State estimation is a statistical problem and has wide applications in a variety of areas (such as physiological signal processing [22] and target tracking [23] etc.). The need for powerful statistical models, which can be used for efficient state estimation and for subsequent applications in WSNs has motivated the author to carry out the work presented in this thesis. In this thesis work, novel linear statistical models are proposed for state estimation, anomaly detection, and time synchronization for discrete-time WSNs. The author would like to emphasize that the proposed approach of anomaly detection and time synchronization is based on statistical models and can be used in a variety of other applications of WSNs as well. Some of the fundamental concepts of state estimation and time synchronization methods used in WSNs are discussed below.

For any dynamic system, 'state' refers to the smallest vector that fully summarizes the "past" of the system. In the context of WSNs, it may be noted that for a particular sensor node one may have measurements say, $x_1, x_2, \ldots, x_t$, till time point t. Based on these t

measurements, the measurement for time point $(t+1)$ is predicted, and the predicted value, denoted by $\hat{x}_{t+1}$, will be called 'state' of the sensor node at time $(t+1)$. For estimating state of a dynamic system, traditionally state-space models are used in the literature. State-space models assume that 'true' state of the system at time t, denoted by x_t, is latent (i.e. unobserved), and can be modelled as the following:

$$x_t = f_t(x_{t-1}, v_{t-1}) \tag{1.1}$$

where v_{t-1} denotes the random noise and f_t is a time-dependent mathematical function (possibly non-linear) describing the evaluation of states. Since 'true' state is latent, a measurement process is assumed in which 'true' state is obtained from noisy measurements z_t as the following:

$$z_t = h_t(x_t, n_t) \tag{1.2}$$

where the time-dependent function h_t defines measurement process and n_t denotes random noise. Under linearity of f_t and h_t, and under the assumption of Gaussian distribution for noise terms v_t and n_t (at each time point t), the estimation method becomes simpler and is known as Kalman Filter (KF). Thus, the state-space model for KF can be written as the following [24]:

$$x_t = F_t x_{t-1} + v_{t-1}; \quad z_t = H_t x_t + n_t$$

Here F_t and H_t are process and measurement matrices, respectively; and the noise terms v_{t-1} and n_t are assumed to be Gaussian with mean=0, and unknown but fixed variances.

In this thesis work, dynamic linear statistical model for state estimation is proposed, and efficiency of this model is compared to the traditional KF based state space models. An alternative Bayesian model is also proposed, which can efficiently estimate the state values over time. In recent years, Bayesian statistics is playing a key role in various disciplines. The popularity of Bayesian statistics is due to the specification of "prior" information available from the previous study or provided by some domain experts. Such prior data are combined with observed measurements from the designed experiment, and then "posterior" inference is made. Posterior inferences are powerful since the prior and the experimental data are used together for such inferences. Detailed discussion on Bayesian modelling and estimation method can be found in Robert and Casella [25]. Further, in this thesis, some in-depth discussion of Bayesian models in the context of state estimation and time synchronization techniques has been provided in subsequent chapters.

The work presented in this thesis is organized in different chapters, and is summarized as the following:

Chapter 2 presents a novel anomaly detection method based on the estimated state values of the sensor nodes. The proposed model is fundamentally different from the previous works and powerful for state estimation. In the proposed method, a novel dynamic regression model which allows "exchange of the relevant information" among the spatially close nodes belonging to the same cluster has been used. The author developed an algorithm for locating the possible anomalous node(s) in the network and thus enhanced the reliability of the network. Detection of anomalous node is important in many applications of WSNs, since an anomalous node may destroy the network communication. Because the state estimation and anomaly detection are closely related, a powerful model for state estimation is arguably very important for intruder detection. A simple algorithm of anomaly detection through splitting and merging of the clusters is illustrated.

Next, an alternative Bayesian model is proposed for state estimation. Then performances of (i) Maximum Likelihood (ML) based regression model (ii) Proposed Bayesian method and (iii) KF based state-space model are compared. Based on the simulation studies, it is found that proposed Bayesian method is more appealing since it is computationally faster (less CPU time for state estimation) and provides the smallest Average Mean Squared Error (AMSE) compared to the other two methods.

The proposed approach is powerful since it considers the effect of the nearest neighbours on the current state values and then detects the anomalous nodes based on the estimated state values. Additionally, the proposed method is "energy efficient" in the sense that some sensor nodes can be kept in sleep node for sometime but still one can impute those missing values quite accurately for effective inference. This process of imputation has been numerically assessed and presented in this chapter.

Chapter 3 presents a non-parametric Bayesian approach for simultaneous state estimation and anomaly detection. Here also, a cluster-based WSN, similar to the setting in Chapter 2, is used. However, in Chapter 2, it is assumed that the clusters are independent and do not share any information among them and the information exchange takes place only among the sensor nodes within a particular cluster. However, more realistic and powerful approach is to consider information exchange among the sensor nodes belonging to different clusters as well. Hence, in Chapter 3, a dynamic model is proposed for simultaneous state estimation of all sensor nodes across different clusters over time.

In non-parametric Bayesian literature, Dirichlet Process (DP) priors are often used for information sharing [26,27,28]. While DP can allow information exchange among different clusters effectively, one of the recent developments of DP is Matrix Stick-Breaking Process (MSBP) proposed by Dunson et al. [29]. MSBP considers all the model parameters from all clusters, and then 'group' them based on their numerical values estimated from the available data. The author first considered cluster-specific dynamic models for state estimation across different clusters of the network, and then assumed MSBP prior distribution on the model parameters. The author also assessed the similarity of different parameters across the clusters. This approach can locate anomalous node in the network more effectively since all the clusters are properly linked to each other. Through simulation studies, the usefulness of the proposed approach has been compared and assessed with the traditional approach i.e. state-space model via KF [24]. Finally, application of the proposed approach in anomaly detection is demonstrated through simulation studies.

Chapter 4 presents a powerful statistical model for time synchronization in a discrete-time WSN. Time synchronization is extremely important in WSNs for efficient network communication. There is a rich literature on this, and different protocols (for example, Receiver-Only Synchronization (ROS) [30], Reference Broadcast Synchronization (RBS) [31], Flooding Time Synchronization protocol (FTSP) [32] have been proposed in the literature for efficient time synchronization in WSNs. Most of these protocols are based on linear statistical model proposed in Noh et al. [30], where Ordinary Least Squares (OLS) estimates are used for estimating the clock-offset, and clock-skew parameters. All the existing approaches assume that the set of time readings between a pair of nodes are uncorrelated, which might not be true in real applications. Since the readings are taken from the same pair of nodes at different time points, it is expected that these time readings will be correlated over time. Hence, the author considered the model proposed in Noh et al. [30], but additionally considered an auto-regressive model for capturing dependence among the random components over different time points. In the proposed approach, model parameters are estimated iteratively by Generalized Least Squares (GLS) method.

Further, an alternative Bayesian model is proposed for time synchronization of WSNs. The effective use of the prior distributions makes Bayesian approach powerful and computationally fast. Through simulation studies, the accuracy of the proposed estimation method is assessed. Then the three methods: (i) traditional approach of OLS as proposed in Noh et al. [30], (ii) GLS method, and (iii) Proposed Bayesian method, are compared in terms of computational time and AMSE through computer simulations. It is inferred that GLS method works better than OLS method, however Bayesian method outperforms other two methods since it provides the smallest AMSE, and works much faster.

It may be noted that the works presented in this thesis are carried out with the objective of anomaly detection through state estimation for discrete-time WSNs. Anomaly detection can be performed in many different approaches in practice, but the author has proposed a Statistical method, which detects anomaly through dynamic state estimation. For a cluster-based network, one has to estimate the state value for each sensor node, and also has to estimate the state value for each cluster. Mathematical models have been proposed for such estimation in this thesis. For the state estimation of the sensor nodes, time-synchronization is an important aspect. Otherwise, the estimated state values will not provide necessary information for detecting the anomalous node dynamically. Therefore, a Statistical model for time-synchronization has also been proposed. Thus, for all practical applications of anomaly detection, time synchronization, state estimation and anomaly detection happen in order.

Finally in **Chapter 5**, the whole work carried out in this thesis is summarized in a concise manner with some concluding remarks. The citations of the research paper published in support of various chapter of the thesis are included as a footnote at the relevant chapters. Finally, the list of publications consulted in carrying out the present research work and also in preparing the thesis is included in the Bibliography.

Chapter 2

STATE ESTIMATION AND ANOMALOUS NODE DETECTION IN WIRELESS SENSOR NETWORKS

2.1 Preamble

Wireless Sensor Networks (WSNs) consist of a mass of sensor nodes distributed over a physical space for monitoring the environmental conditions, such as, temperature, pressure, humidity, acoustics, resonance etc. In recent years, researchers are interested in developing mathematical models for WSNs due to extensive applications of WSNs in various fields including but not limited to habitat monitoring [33], object tracking [34], event detection [35], intruder locating [36] etc. In medical science [37], security surveillance [5], pattern recognition [39] and many other fields, WSNs are used successfully for inference and decisions making.

Most of the above applications of WSNs mainly depend on the accuracy and precision related to the estimation of the "state values" of different sensor nodes. It may be noted that the term "state" in the context of WSNs is quite subjective in the sense that it depends on the ultimate goal of the study, and the quantity being measured from the sensor nodes over time. Specifically, the state of a sensor node at time t is estimated based on the available measurements till time $(t-1)$. "State" of a sensor node thus can be continuous (e.g. air pressure, humidity etc.) as well as binary (e.g. occurrence of an event). However, WSNs with continuous state values are considered in this study.

In a purely probabilistic framework, Liang et al. [40] proposed models for distributed state estimation of discrete-time WSNs. Similar models are proposed by Xu and Li [41]. Sun et al. [42] proposed a very powerful model for state estimation of multiple mobile targets. Quevedo et al. [43] proposed models which can efficiently estimate the states of WSNs with correlated wireless fading channels. Mo et al. [31] proposed a model which can handle the false data injection attacks in state estimation of sensor networks. Since there are several resource constraints in WSNs, anomaly detection using state estimation becomes very essential for reliable networks.

This chapter focuses on a statistical model based state estimation approach, which can be used effectively in detecting one or more possible anomalous nodes in WSNs. Anomalous nodes are those nodes, which behave differently from the majority of the sensor nodes within a cluster. Detection of anomalous node is important because such nodes might have detrimental effects on the surrounding sensors and thus may affect the performance of the entire network over time [44,45]. Analysis of sensor measurements is very important for anomaly detection in WSNs. The presence of anomaly is confirmed when one (or more) sensor node(s) behaves differently from the majority of the sensor nodes [46]. However, sensor measurements contain spatio-temporal correlations. Since the sensor nodes are densely deployed, spatial correlation exists among the neighbouring sensor nodes. Temporal correlation occurs due to predictable relationship that exists in sequential measurements of the sensor nodes.

Detection of anomalous node in wireless body area networks enable real time global patient and health care monitoring [47]. Sun et al. [48] used Extended Kalman Filter (EKF) for detecting the false injected data from the anomalous behaviour of the sensor nodes. Rajasegarar et al. [49] used distributed one-class quarter-sphere support vector machines to distinguish anomalous measurements from the obtained data.

There is a rich literature on object tracking using WSNs, one such example is discussed in [34]. But relatively few papers are available on anomalous node detection. Almost all these papers [49-51] detect an anomalous node based on its relative performance compared to the neighbouring sensor nodes. A novel anomaly detection method based on the estimated state values of the sensor nodes has been proposed in this chapter. Most of the state estimation methods rely on state-space models based on Kalman Filter (KF) or stochastic differential equations [43,52,53]. The proposed model is fundamentally different from the previous works and powerful for state estimation of discrete-time WSNs. In the proposed scheme, a novel dynamic regression model has been used which allows the "exchange of the relevant information" among the spatially close nodes. Since, information exchange is inevitable in any network, the models for state estimation should consider this appropriately for reliable state estimation. Maximum Likelihood (ML) estimation method has been used for estimation and related inferential properties.

In recent years, there is a growing interest in Bayesian models and estimating the model parameters by Markov Chain Monte Carlo (MCMC). Hence, Bayesian model is also used in this study for state estimation using MCMC. Based on the simulation studies, it is found that Bayesian estimation method is more appealing because of its ability to handle complex models in relatively less time. Further, computations based on MCMC are much faster than the traditional ML estimation [54]. We also assess the effectiveness of our proposed anomaly detection technique by computing False Positive Rate (FPR) [55] as shown in section 2.4.3.

The current chapter has two major contributions:

- First, a novel dynamic statistical model has been proposed for state estimation, which allows appropriate information exchange among spatially close sensors. An algorithm has been developed for locating the possible anomalous node(s) in the network which enhances reliability of the WSN.
- Second, in addition to the traditional ML method for state estimation, a Bayesian model has been proposed as an alternative method for estimating the state values. Then the performances of (i) regression model based on ML estimation (ii) Bayesian model based

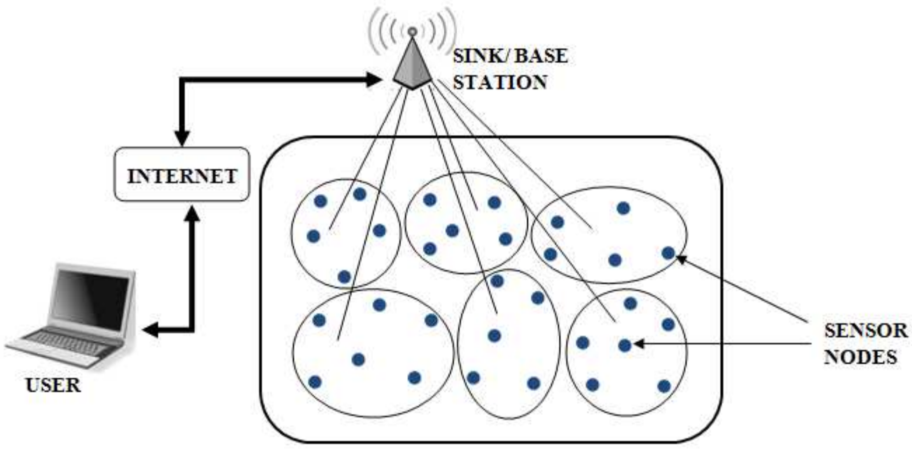


Figure 2.1: Cluster based distributed wireless sensor network

on MCMC and (iii) state-space model based on KF are compared. It has been found that MCMC based Bayesian method takes less CPU time for state estimation and provides less Average Mean Squared Error (AMSE).

The findings of this chapter has been published in a research paper. [1]

2.2 Method of Longitudinal State Estimation

Consider a particular WSN consisting of N sensor nodes with one or more anomalous nodes. For each sensor node, the state values are estimated longitudinally at T different discrete time points $1, 2, \ldots, T$. At each time point t, the state value of the i-th sensor is denoted by $X_i(t)$, which is communicated to the 'sink' of the network. Let (θ_i, δ_i) be the coordinates of the i-th sensor node, $i = 1, 2, \ldots N$. The Euclidean distance between the sensor i and the sensor j is denoted by D_{ij}. Based on the D_{ij} values, the sensor nodes are "grouped" into several clusters so that the sensors belonging to same cluster are spatially close to each other. In Figure 2.1, a cluster based distributed WSN is shown.

[1] A. Chatterjee, P. Venkateswaran, et.al, "A unified approach of simultaneous state estimation and anomalous node detection in distributed wireless sensor networks", International Journal of Communication Systems, Vol 30, Issue 9, June 2017.

2.2.1 Linear Statistical Model for State Estimation

Traditionally, state space models are used in the state estimation of WSNs [43,52,53]. In state-space models, the state values y_i's are modelled as a function of the unknown states x_i's and then auto-regressive models are used for the unknown states. Mathematically, one can summarize a state-space model as :

$$y_i|x_i = f(x_i) + d_i; \;\; x_i|x_{i-1} = g(x_{i-1}) + e_i$$

where d_i and e_i denote the observation error and system error, respectively. The functions f and g are assumed to be linear and errors are assumed to be Gaussian for estimating the parameters via KF. However, such models do not consider information exchange among the sensor nodes which are spatially close to each other. In any network, information exchange is inevitable and must be taken into account. Considering this, the author proposes the following linear model for estimating the state values of the sensor nodes at different discrete time points:

$$X_i(t) = f(t) + \alpha X_i(t-1) + \beta Z_i(t-1) + \epsilon_i(t) \tag{2.1}$$

where $X_i(t)$ denotes the state value of the i-th sensor node at time t belonging to a particular cluster. The smooth function (at least twice differentiable) f measures the general effect of time on the current state value and $Z_i(t-1)$ is the average state values of all the other sensors belonging to that cluster at time $(t-1)$. Also, $X_i(t-1)$ is the observed state value of the i-th sensor at time $(t-1)$. The regression coefficients α and β measure the effect of the immediate predecessor state value and the effect of the nearest neighbours on the current state value, respectively. The residual errors $\epsilon_i(t)$'s are assumed to be independent and follow a Gaussian distribution with mean=0 and unknown variance=σ_ϵ^2.

It may be noted that the proposed model in equation (2.1) is fundamentally different from the commonly used state-space model like KF described earlier in Chapter 1 using equations (1.1) and (1.2), and also in the previous paragraph [21-24]. In state-space models there are two errors: observation error and system error. But in the proposed model there is only one error component, i.e. the residual error, which is basically the measurement error.

Also in the proposed model the coefficient β measures the effect of the nearest neighbours on the current state value and the covariate Z allows the information sharing among the sensor nodes belonging to the same cluster. However, traditional state-space models do not allow this.

The motivation of the model given in equation (2.1) is as follows: The state value of a particular sensor is updated over time. The general effect of time, $f(t)$, explains how much of this change is due to time. It is noted that the state values of a particular sensor node can possibly be affected by many unobserved time-varying covariates. The effects of those covariates are captured in the smooth function f. Also the current state value (at time t) of a sensor node should depend on its state value at time $(t-1)$, and this makes the model dynamic in nature. The state value of a particular sensor node at time t is expected to be affected by the state values of its nearest neighbours at time $(t-1)$. This effect will be reflected in the estimate of β. A statistically significant value of β indicates that the sensor nodes share information at each time point and accordingly update the state values at the next time point.

The significance for using the above linear regression model is summarized below:

- To capture the neighbourhood effect which allows exchange of relevant information among the spatially close nodes.
- To incorporate suitable prior structure for the coefficients as discussed in section 2.3.1.
- To estimate the regression coefficients in much shorter time as shown in section 2.4.3.

The proposed model can also estimate the missing observations in WSNs. In WSNs, sometimes few sensor nodes are reserved in "sleep mode" for a certain period for energy conservation [38]. This model can be used for estimating the missing state values and thus effectively estimates the state of the network. This property of the proposed model is demonstrated in Section 4.

2.2.2 Joint Likelihood and Parameter Estimation

The smooth function f can be modelled in various parametric or non-parametric approaches like polynomial functions, wavelets, B-splines, Penalized splines, orthogonal Legendre polynomials etc. [56-57]. Here for the sake of simplicity, the author uses polynomial function for modelling f. In other words, it is assumed that $f(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_p t^p$.

The optimal order (p) of the polynomial function is traditionally obtained from the information criteria like Akaike Information Criterion (AIC)$=-2logL + 2P^*$ and Bayesian Information Criterion (BIC)$=-2logL + log(n)P^*$, where L denotes the joint likelihood function discussed later, P^* denotes the total number of model parameters that need to be estimated and n denotes the total number of measurements (in this case $n = NT$). The author fits the model given in equation (2.1) for $p = 1, 2, 3, 4$, and chooses the optimal order p for which the smallest AIC and/or BIC values are obtained [56].

Note that because of the Markovian assumption of the model in equation (2.1), the conditional distribution of $X_i(t)$ given all the previous time points can be expressed as: $l(X_i(t)|t-1, \ldots, 1) = l(X_i(t)|t-1)$. Here by 'given time $(t-1)$' the author essentially means that 'given the measurements of all sensors at time $(t-1)$'. Thus the conditional distribution of $X_i(t)$ given all the previous measurements is expressed as: $X_i(t)|t-1 \sim \text{Gaussian}(f(t) + \alpha X_i(t-1) + \beta Z_i(t-1), \sigma_\epsilon^2)$. Hence,

$$l(X_i(t)|t-1) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left[-\frac{(X_i(t) - f(t) - \alpha X_i(t-1) - \beta Z_i(t-1))^2}{2\sigma_\epsilon^2}\right] \tag{2.2}$$

Note that given all the previous time points, the state values of different sensor nodes at a fixed time point are independently distributed. The author exploits this conditional independence property to formulate the joint likelihood function. However, it is noted that the state values of different sensor nodes at a fixed time point are not marginally independent.

The joint likelihood function of all the N sensors for T time points can be expressed as:

$$L = \prod_{i=1}^{N} [l(X_i(1)l(X_i(2)|1) \dots l(X_i(T)|T-1)] \tag{2.3}$$

Hence the joint log-likelihood function is given as:

$$logL = \sum_{i=1}^{N} \left[logl(X_i(1)) + \sum_{t=2}^{T} logl(X_i(t)|t-1) \right] \tag{2.4}$$

Model parameters are to be estimated by maximizing the joint log-likelihood function given in equation (2.4). The author solves the following equations simultaneously to get the maximum likelihood estimates (MLEs) of the model parameters:
$\frac{\partial logL}{\partial \alpha} = 0, \; \frac{\partial logL}{\partial \beta} = 0, \; \frac{\partial logL}{\partial a_j} = 0, \; \forall j = 0, 1, \dots, p$

2.2.3 Anomalous Node Detection

Once the model parameters are estimated, one can compute the average state value of each cluster. Suppose, there are c sensor nodes in a certain cluster, then the average state value of that cluster at time t will be $V(t) = \frac{1}{c} \sum_{i=1}^{c} \hat{X}_i(t)$, where $\hat{X}_i(t) = \hat{f}(t) + \hat{\alpha} X_i(t-1) + \hat{\beta} Z_i(t-1)$. Note that "hat" indicates the estimate of the respective parameter or function obtained in Section 2.2. If the absolute difference between the average state values at two consecutive time points is consistently below a pre-specified threshold (δ), i.e. if $|V(t) - V(t-1)| < \delta$, for all time points, then it is concluded that the cluster under consideration is less dynamic and hence possibly does not contain the anomalous node. In such situations, one can keep some of the sensor nodes of this "less dynamic" cluster in sleep mode for a certain period of time for energy conservation. However, if one observes $|V(t) - V(t-1)| > \delta$, for few consecutive time points, it is inferred as a dynamic cluster and possibly contains anomalous node(s). Note that the choice of the threshold value depends on the desired accuracy level and application. The author divides a dynamic cluster into two "sub-clusters" and adds some new sensor nodes to each sub-cluster. These new sensor nodes can be borrowed from the "less dynamic" clusters or one may redeploy some new sensor nodes. This process is

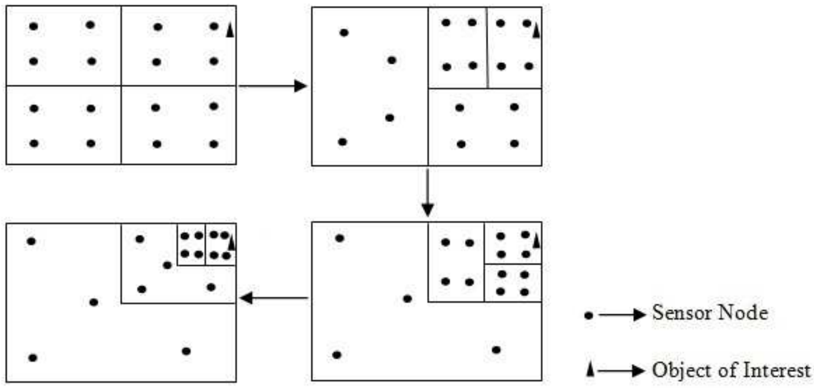


Figure 2.2: Locating the anomalous node by repeated splitting and merging of the clusters and redeployment

continued until one obtains the target node and the state trajectory of this node will be significantly different from the trajectories of the other nodes. The author detects this node as an "anomalous node". Figure 2.2 summarizes this process of anomaly detection and the same is numerically illustrated in Section 2.4.2. If a network contains more than one anomalous node, then one simply has to repeat this splitting process for multiple dynamic sub-clusters and will eventually detect all the anomalous nodes. The author considers this as a major application of the proposed model.

2.3 Bayesian Approach and MCMC

2.3.1 Model and Priors

Next, the author proposes a Bayesian approach for the state estimation and detection of the anomalous nodes. The following Bayesian approach is employed and the model parameters are estimated by MCMC. Consider the linear model for state estimation given in equation (2.1) where the function f is modelled as a p-th degree polynomial function of time ($f(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_p t^p$), and the author takes the following prior distributions for the model parameters:

$$a_j \sim N(\mu_j, \sigma_j^2),\ j = 0, 1, \ldots, p;\ \ \alpha \sim N(\mu_\alpha, \sigma_\alpha^2),\ \beta \sim N(\mu_\beta, \sigma_\beta^2)\ \sigma_\epsilon^2 \sim IG(\nu_1, \nu_2) \tag{2.5}$$

where N stands for normal (Gaussian) and IG stands for inverse gamma distribution. Such priors will result in conjugacy, i.e. the posterior and prior distribution of each parameter will belong to the same family.

The joint likelihood function can be written as the following:

$$L = (2\pi\sigma_\epsilon^2)^{-\frac{NT}{2}} \exp\left[-\frac{\sum_{i=1}^{N} X_i^2(1) + E}{2\sigma_\epsilon^2}\right] \tag{2.6}$$

where $E = \sum_{i=1}^{N}\sum_{t=2}^{T} [X_i(t) - f(t) - \alpha X_i(t-1) - \beta Z_i(t-1)]^2$

By using Bayes theorem and assuming the priors for different parameters are independent, the joint posterior distribution is given by

$$\pi(\boldsymbol{a}, \alpha, \beta, \sigma_\epsilon^2|X) \propto L \times \prod_{j=0}^{p} \pi(a_j) \times \pi(\alpha) \times \pi(\beta) \times \pi(\sigma_\epsilon^2) \tag{2.7}$$

where $\pi(.)$ denotes the respective prior distribution and $\boldsymbol{a} = [a_0, a_1, \ldots, a_p]^T$.

2.3.2 Full conditionals and Gibbs sampling

For MCMC iterations, the goal is to simulate from the joint posterior distribution and then estimate the model parameters based on the simulated values. Following Robert and Casella [18], the author computes the full conditional distribution of the model parameters and then simulates the parameters from their respective full conditionals.

The full conditional distribution of α is given by:

$\pi(\alpha|\boldsymbol{a}, \beta, \sigma_\epsilon^2, X) = N\left(\frac{B-\beta C-A+D}{V_1}, V_1^{-1}\right)$; where B$=\frac{\sum_{i=1}^{N}\sum_{t=2}^{T} X_i(t)X_i(t-1)}{\sigma_\epsilon^2}$; C$=\frac{\sum_{i=1}^{N}\sum_{t=2}^{T} X_i(t-1)Z_i(t-1)}{\sigma_\epsilon^2}$; $A = \frac{\sum_{j=0}^{p} a_j\left(\sum_{t=2}^{T} t^j X_i(t-1)\right)}{\sigma_\epsilon^2}$; D$=\frac{\mu_\alpha}{\sigma_\alpha^2}$ and $V_1 = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} X_i^2(t-1)}{\sigma_\epsilon^2} + \frac{1}{\sigma_\alpha^2}$

The full conditional distribution of β is given by:

$\pi(\beta|\boldsymbol{a},\alpha,\sigma_\epsilon^2,X) = N\left(\frac{B_1-\alpha C-A_1+D_1}{V_2}, V_2^{-1}\right)$; where $B_1 = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} X_i(t)Z_i(t-1)}{\sigma_\epsilon^2}$; $A_1 = \frac{\sum_{j=0}^{p} a_j\left(\sum_{t=2}^{T} t^j Z_i(t-1)\right)}{\sigma_\epsilon^2}$; $D_1 = \frac{\mu_\beta}{\sigma_\beta^2}$ and $V_2 = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} Z_i^2(t-1)}{\sigma_\epsilon^2} + \frac{1}{\sigma_\beta^2}$

The full conditional distribution of a_j is given by:

$\pi(a_j|\boldsymbol{a}-a_j,\alpha,\beta,\sigma_\epsilon^2,X) = N\left(\frac{A_j}{V_j^*},(V_j^*)^{-1}\right)$; where $A_j = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(X_i(t)-\alpha X_i(t-1)-\beta Z_i(t-1)+\sum_{j'\neq j} a_{j'}t^{j'}\right)}{\sigma_\epsilon^2}$; and $V_j^* = \frac{\sum_{t=1}^{T} t^j}{\sigma_\epsilon^2} + \frac{1}{\sigma_j^2}$

And finally the full conditional distribution of σ_ϵ^2 is given by:

$$\pi(\sigma_\epsilon^2|\alpha,\beta,X) = IG\left(\nu_1 + \frac{NT}{2}, \nu_2 + \frac{1}{2}(\sum_{i=1}^{N} X_i^2(1) + E)\right)$$

It is noted that all the full conditional distributions are all known densities and one can directly simulate from those densities. So, the author employs a Gibbs sampler technique [46] and directly simulates from the posterior distributions. MCMC is run for 100,000 iterations and the first 20,000 "burn-in" iterations are discarded to remove the effect of the starting values. Also following the tradition, to remove the effect of the auto-correlation among the successive iterations, the author thins the chains by keeping every 10-th iteration. Starting values are simulated from the respective prior distributions.

2.4 Performance Evaluation

In this section, the author evaluates the performance of the proposed method of state estimation and anomalous node detection through simulation studies. The author has performed two simulation studies. In the first simulation study, the accuracy of the proposed method is investigated in estimating the model parameters. Further in the presence of missingness in the state values, the proposed model is used to estimate the missing state values and then using those state values the model parameters are estimated and compared to the previously estimated parameters with no missing values. In the second study, the author investigated the effectiveness of the proposed algorithm for detecting anomalous node in WSNs.

2.4.1 Simulation studies for the estimation of the model parameters and the missing state values

Consider a distributed WSN consisting of multiple clusters. The number of clusters, however, depends on the geographical area to be covered and the number of sensor nodes in each cluster depends on the desired accuracy level and cost. For each cluster, the author separately fits a model given in equation (2.1) for estimating the state values of the sensor nodes belonging to that particular cluster. However, for the sake of illustration, the author considers one cluster containing 10 sensor nodes. For each sensor node, the state values are measured at 8 discrete time points ($t = 1, 2, \ldots, 8$). Note that the sensor nodes are densely deployed and thus information exchange happens among these nodes belonging to a common cluster. In practice, for a more complex network one should 'group' the sensor nodes based on the relative Euclidean distances.

The author uses the following regression model for simulating the state values of the sensor nodes:

$$X_i(t) = a_0 + a_1 t + a_2 t^2 + \alpha X_i(t-1) + \beta Z_i(t-1) + \epsilon_i(t) \tag{2.8}$$

The values of (a_0, a_1, a_2), α, and β are chosen as $(a_0, a_1, a_2) = (2, -0.6, 1.4)$, $\alpha = 2.5$, $\beta = 1.5$. With respect to Table 2.1, these values are considered as true value for the purpose of simulation. Note that this proposed estimation method does not depend on any specific set of values for these parameters. For the simulation purpose, the author considers one set of parameter values and shows the effectiveness of the proposed approach of estimating the state values corresponding to this particular set of the parameter values. One can consider a completely different set of parameter values, but still can efficiently estimate the corresponding state values. This fact has been verified by the author by choosing different sets of values for (a_0, a_1, a_2), α, and β. The residual errors $\epsilon_i(t)$'s are simulated from Gaussian with mean=0, and variance=1.44; so that the residual error is minimal. Z_i values for each sensor node are obtained by averaging the state values for all the other nodes at the previous time point. Note that for t=1, there is no X_i and Z_i values. For the simulated data, the author fits the model given in equation (2.1) and selects the optimal order of the the function f by computing BIC. The smallest BIC is obtained for the second order polynomial function and then the author estimates the model parameters using the ML estimation discussed in Section 2.2. The estimated parameters and the corresponding standard errors are shown in column 3 of Table 2.1.

Table 2.1: Results on the parameter estimation with no missing data and imputed data for Simulation 1.

Model parameter	True value	Estimate (no missing data) (SE)	Estimate (imputed data) (SE)
a_0	2	1.88 (0.98)	1.76(1.12)
a_1	-0.6	-0.63 (0.57)	-0.54(0.85)
a_2	1.4	1.37 (1.06)	1.46(1.33)
α	2.5	2.56 (0.72)	2.1(1.16)
β	1.5	1.48 (0.64)	1.43(1.08)
σ_ϵ^2	1.44	1.40 (1.16)	1.32(1.68)

Next, artificial "missingness" is created in the simulated data. At time $t = 4$, the author only considers the simulated state values for the first 6 sensor nodes and last 4 nodes are assumed to give missing values. Also, for the last time point ($t = 8$), the state values are simulated for sensors 1,3,6,8 and 10; and for the other nodes the author considers missing values. For this simulated data, the model parameters are estimated based on the available data and then the missing values are imputed. After imputation, model (1) is fitted again and the parameters are re-estimated. The estimated parameters and the corresponding standard errors are now shown in column 4 of Table 2.1.

It may be noted that the parameter estimates obtained by using the "imputed" missing values slightly deviate from the true values. In Figure 2.3, the true mean trajectory and the

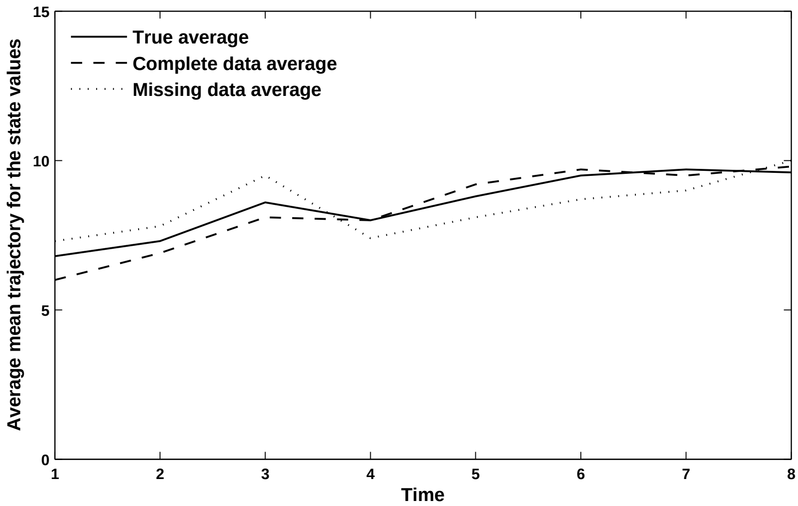


Figure 2.3: True mean trajectory, estimated mean trajectory for the complete data and the imputed data in Simulation 1

estimated mean trajectories from the complete data and the "imputed" data are plotted. It is noted that for the "imputed" data, the estimated mean trajectory closely follows the true trajectory. This illustrates the effectiveness of the proposed model and the estimation method even when some sensor nodes are kept in sleep mode for certain time points.

2.4.2 Simulation study for the detection of anomalous node in WSN

Here the author numerically illustrates the effectiveness of the proposed model for detecting the anomalous node in WSNs. Similar to the previous simulation study, the author considers one cluster of a distributed WSN consisting of 4 sensor nodes denoted by node 1, 2, 3 and 4, respectively. Consider node 3 is an anomalous node and simulate the state values for the other nodes at 5 consecutive time points from the following model:

$$X_i(t) = a_0 + a_1 t + \alpha X_i(t-1) + \beta Z_i(t-1) + \epsilon_i(t) \tag{2.9}$$

where $(a_0, a_1) = (2, -0.6)$, $\alpha = 1.8$, $\beta = 2.3$. The values chosen for (a_0, a_1), and (α, β) are as per the assumptions for simulation using equation (2.8) given in Section 2.4.1. The residual errors $\epsilon_i(t)$'s are simulated from Gaussian distribution with mean=0, and variance=1.21, to ensure that the residual errors close to 0.

However, for node 3, the state values are simulated from the following model:

$$X_i(t) = \alpha X_i(t-1) + \epsilon_i(t) \tag{2.10}$$

where $\alpha = 0.85$, and the residual errors $\epsilon_i(t)$'s are simulated from Gaussian distribution with mean=0, and variance=25. Note that this setting ensures that the node 3 is behaving differently from the rest of the sensor nodes in the cluster with larger residual errors over time.

For the above simulated dataset, the author first fits the model as proposed in equation (2.1) and then computes the average state values at each time point. In Table 2.2, absolute differences $|V(t)-V(t-1)|$ are shown and it is noted that the network is dynamic in nature. The author has chosen the acceptance limit $\delta = 4.0$ in this model. It may be noted that the value of δ depends on desired accuracy and the nature of application.

As explained in Section 2.2.3, in order to identify the anomalous node, the author splits the cluster into two sub-clusters by keeping nodes 1 and 2 in one group and nodes 3 and 4 in the other. In Table 2.2, it is noticed that sub-cluster 2 (consisting of nodes 3,4) is more dynamic in nature and hence it is decomposed into two finer sub-clusters again, each containing only one node. Now, it is noticed that node 3 is giving more dynamic state values than node 4 and hence it is marked as the "anomalous node". Thus it is seen that the method "correctly" identifies the anomalous node in the network and hence can be used in real applications.

Table 2.2: Absolute mean difference $|V(t)-V(t-1)|$ at successive time points for detecting anomalous node in Simulation 2.

		after the first splitting		after the second splitting	
Time	Before splitting	cluster 1 (node 1,2)	cluster 2 (node 3,4)	cluster 1 (node 3)	cluster 2 (node 4)
1	-	-	-	-	-
2	4.35	0.254	4.72	7.54	0.523
3	4.98	0.326	6.59	8.63	0.761
4	6.72	0.188	7.38	4.99	0.482
5	8.94	0.427	5.92	8.84	0.307

2.4.3 Simulation Studies for Model Comparisons

Parameter Estimation

Consider a simple WSN with 15 sensor nodes where all these sensors are spatially close to each other and are kept in a single cluster. The continuous state values for each sensor node is measured at 10 different discrete time points. The author uses the following expanded regression model as per equation (2.1) for simulating the state values of the sensor nodes:

$$X_i(t) = a_0 + a_1 t + a_2 t^2 + \alpha X_i(t-1) + \beta Z_i(t-1) + \epsilon_i(t) \tag{2.11}$$

where $(a_0, a_1, a_2) = (1.5, -0.78, 2.18)$, $\alpha = 1.6$, $\beta = 2.33$. The residual errors $\epsilon_i(t)$'s are simulated from Gaussian distribution with mean=0, variance=0.9. Z_i values for each sensor node are obtained just by averaging the state values for all the other nodes at the previous time point. The values chosen for (a_0, a_1, a_2), and (α, β) are as per the assumptions for simulation using equation (2.8) as well as (2.9), given in Section 2.4.1 and 2.4.2.

Once the dataset is simulated, the author first employs the proposed approach of estimating the model parameters by maximizing the joint log-likelihood function as discussed in Section 2.2. Next, the author uses the traditional state-space model and estimates the model parameters using KF. Specifically, the following KF model is fitted to the simulated data:

$$X_i(t) = \alpha_t + d_i(t), \quad \alpha_t = \alpha_{t-1} + e(t)$$

where the measurement error $d_i(t)$ and the system error $e(t)$ are independently distributed, following Gaussian (0, ν^2) and Gaussian (0, κ^2) respectively.

Finally, the proposed Bayesian method is used for estimating the parameters through MCMC. For α, β and σ_ϵ^2, the author considers Normal (1,4), Normal (1.5,3.8), and Inverse Gamma (2.6,4.2) prior, respectively. The author also considers the standard normal N(0,1) priors for the coefficients a_j, as these prior distributions result in the normal full conditional posterior distribution for a_j. The choice of the prior distributions for a_j , α, β and σ_ϵ^2 results in the minimal effects on the final estimates. Gibbs sampling algorithm is used for simulating from the full conditional densities as given in Section 2.3.2. Below, a step-by-step standard algorithm is [25] used for implementing the Gibbs sampler in this problem:

1. Generate the starting values of the parameters from their respective prior densities. Denote the starting values by $\boldsymbol{a}^{(0)}$, $\alpha^{(0)}$, $\beta^{(0)}$ and $\sigma_\epsilon^{2(0)}$

2. Generate the first iteration $\boldsymbol{a}^{(1)}$ for given $\alpha^{(0)}$, $\beta^{(0)}$ and $\sigma_\epsilon^{2(0)}$ using the full conditional densities from Section 2.3.2.

3. Generate the first iteration $\alpha^{(1)}$ for given $\boldsymbol{a}^{(1)}$, $\beta^{(0)}$ and $\sigma_\epsilon^{2(0)}$

4. Similarly, generate $\beta^{(1)}$ for given $\boldsymbol{a}^{(1)}$, $\alpha^{(1)}$ and $\sigma_\epsilon^{2(0)}$; and also generate $\sigma_\epsilon^{2(1)}$ for given $\boldsymbol{a}^{(1)}$, $\alpha^{(1)}$ and $\beta^{(1)}$. This completes the first iteration.

5. Repeat the steps from (2 to 4), 100,000 times which ensures adequate number of iterations. The first 20,000 iterations are discarded to remove the effects of the initial choice of the parameter values.

6. Thin the chain by keeping every 10-th iteration, this will give 8,000 simulated values for each parameter.

7. Estimate the parameters by the respective sample means.

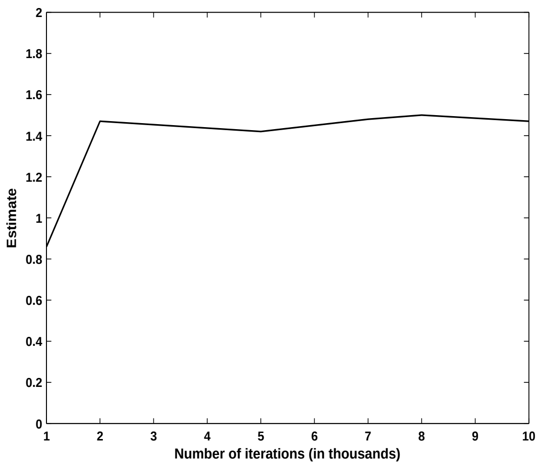
In Figures 2.4 and 2.5, the MCMC estimates of different model parameters are shown for different number of iterations and observed that the estimates converge to certain fixed value for each parameter.

Once the model parameters are estimated by the all three methods mentioned above, the performance of different methods are assessed in terms of the Mean Squared Error (MSE) defined as:

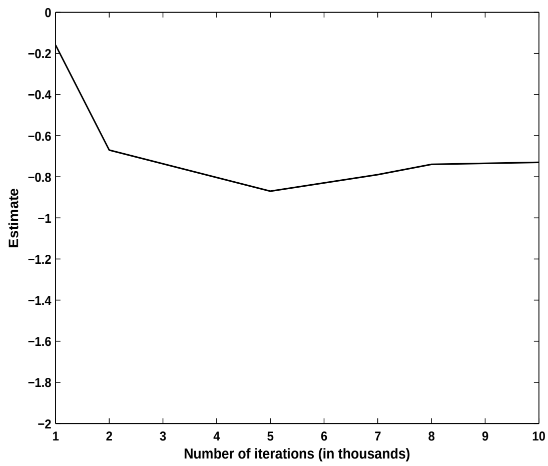
$$\text{MSE} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}\left(X_i(t)-\hat{X}_i(t)\right)^2}{NT}$$

where $\hat{X}_i(t)$ denotes the estimated state value of the i-th sensor node at time t.

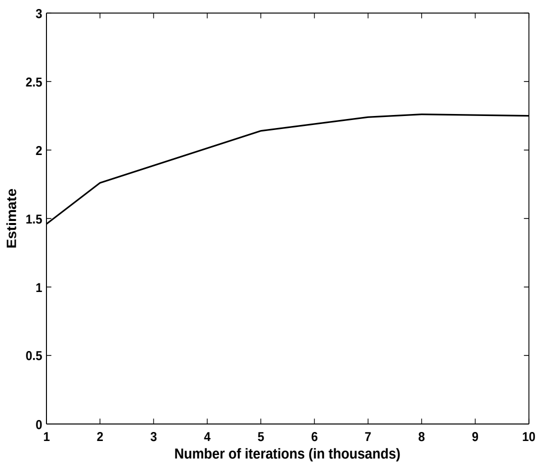
The author considers 50 replicates of the data simulated above and computes the average MSE (AMSE) for three different methods. Next, 100, 1000 and 10000 replicates are considered and the average MSE are computed. In Table 2.3, the average MSE values are shown and it is noted that the state space model based on KF provides the largest average MSE. The proposed Bayesian model results in the smallest AMSE. This demonstrates the



(i)

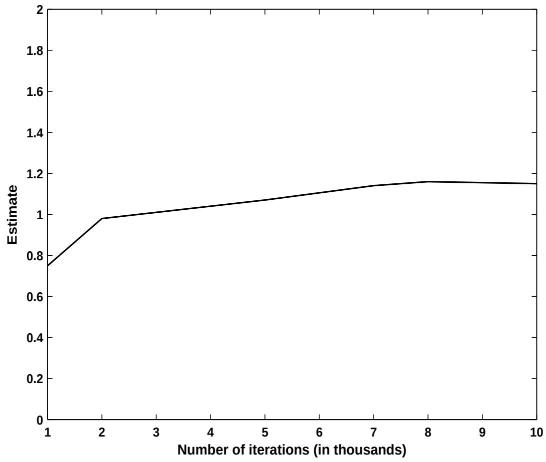


(ii)

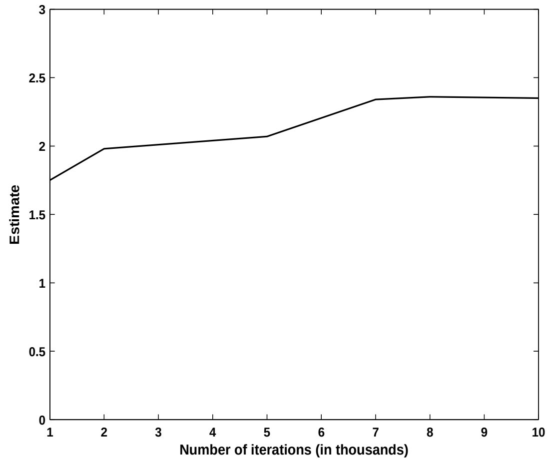


(iii)

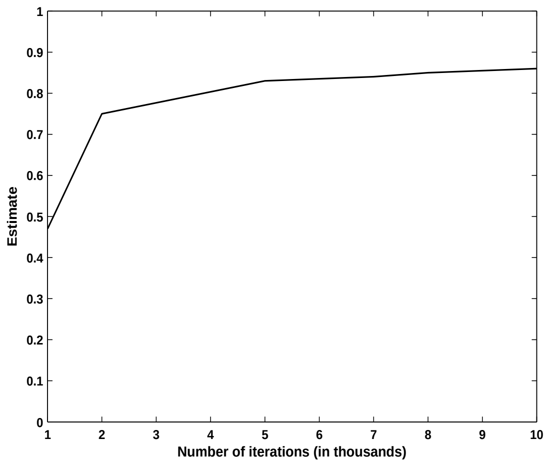
Figure 2.4: MCMC estimates for different number of iterations for (i) a_0 (ii) a_1 and (iii) a_2 respectively.



(i)



(ii)



(iii)

Figure 2.5: MCMC estimates for different number of iterations for (i) α (ii) β and (iii) σ_ϵ^2 respectively.

better predictive power of the MCMC based Bayesian method and the poor performance of the KF based state-space model in terms of the prediction.

Table 2.3: Average MSE values for different methods in model comparison

Number of Replicates	ML based method	MCMC based method	KF based method
50	5.48	3.81	6.37
100	6.41	4.98	7.13
1000	10.23	7.28	11.84
10000	11.62	9.44	15.79

In Table 2.4, the author shows the CPU time (in seconds) required for state estimation using three different methods with different number of sensor nodes (N) in the network. It is noticed that the computational cost for the Bayesian method is much lower than the other methods. The ML based approach and the KF based approach are comparable in terms of the computational cost. This illustrates the fastest computational response of the proposed Bayesian approach. All computations are performed in WINDOWS 8, Intel Core i7 Processor.

Table 2.4: CPU times (seconds) for the three competing models with different number of sensor nodes (N)

N	ML method	MCMC method	KF method
5	78	56	75
10	92	67	96
20	104	74	108
30	123	86	118
50	134	95	130
100	156	103	159

Anomaly Detection

The author considers the same WSN but now 15 sensors are grouped into three clusters; nodes 1 to 5 belong to the first cluster, nodes 6 to 10 belong to the second cluster and the rest of the nodes are kept in the third cluster. The first node in third cluster (i.e. the 11-th sensor in the previous network) is considered as an anomaly and the author simulates the state values for this node from the model given in equation (2.10) with $\alpha = 0.45$ and the residuals are generated from Gaussian distribution with higher variance, i.e. Gaussian distribution with mean=0, and variance=10. For all the other nodes from all the three clusters, the state values are simulated using the model given in equation (2.11) with the same values used as before.

The author uses KF, ML method and MCMC for parameter estimation and the anomalous node detection with a threshold $\delta = 3.5$. As mentioned earlier, the value of δ depends on the desired accuracy and the nature of application. Here 100 datasets are replicated from the same set of models and note the average time to detect the anomalous node. The average computational times for KF, ML method and MCMC are 163 seconds, 152 seconds and 129 seconds, respectively. The computed False Positive Rates (FPR) [55] (defined as the rate of detecting a normal sensor node as an anomalous node) based on 100 replications are 0.13, 0.09 and 0.04, respectively. Hence it is observed that MCMC based Bayesian approach is computationally faster and more reliable (less false positive rate) than the other two methods. All computations are performed and verified using the software R. This software is freely available and there are in-built functions for ML estimation, MCMC, and state-space models.

2.5 Summary

In this chapter, the author has proposed a powerful linear statistical model for state estimation and anomaly detection.

- The proposed approach is powerful since it considers the effect of the nearest neighbours on the current state values and then detects anomalous nodes based on the estimated state values. The proposed method can also estimate the missing state values of the sensor nodes which are kept in sleep mode for energy conservation.
- A Bayesian model has been proposed which is computationally faster for state estimation and anomaly detection.

The effectiveness of the proposed model is investigated through extensive simulation studies and the performance of the proposed approach is compared to that of the traditional approaches.

Chapter 3

SIMULTANEOUS STATE ESTIMATION OF CLUSTER BASED WIRELESS SENSOR NETWORKS

3.1 Preamble

Many of the practical applications of Wireless Sensor Networks (WSNs) depend on the accuracy and precision related to the state estimation of the network over different time points [33-39]. In Chapter 2, the author proposed a statistical model for state estimation of a single cluster in WSNs, and considered "neighbourhood" effect which estimates the amount of information exchange among the sensor nodes within a single cluster. In reality, WSNs consist of multiple clusters, and in this chapter a model has been proposed, which can simultaneously consider information exchange among the sensor nodes belonging to different clusters of the same network. Thus, dependence of the sensor nodes within the same cluster as well as between the clusters of the same network has been considered. For this, the state values of each sensor node are estimated longitudinally and the state values of the entire network is estimated by combining the individual state values of the sensor nodes.

Information exchange among the sensor nodes at different time points is fundamentally important in WSNs and hence, statistical models for state estimation must be different from the traditional regression models where subjects are assumed to be independent [58]. In dynamic state-space models, the observations y_i's are modelled as a function of unknown state values x_i's and then Markov (auto-regressive) models are used for the unknown states. Mathematically one can write,

$$y_i|x_i = f(x_i) + d_i; \quad x_i|x_{i-1} = g(x_{i-1}) + e_i$$

where d_i and e_i denote the measurement error and system error, respectively. The functions f and g are assumed to be linear and the errors are assumed to be Gaussian for estimating the parameters via Kalman-Filter (KF). Although there is a rich literature on state estimation using KF, which enable information exchange among the sensor nodes [59,60], relatively little attention has been given for information exchange among different clusters of a network. The linear Markov model proposed in Section 3.2 is different from the existing works as:

(i) the author specifies a single regression model with only one error component and

(ii) the information exchange is allowed among the sensors belonging to the same cluster by considering the "neighbourhood effect" and among different clusters of the same WSN by appropriately sharing the model parameters in a non-parametric Bayesian approach.

Dirichlet Process (DP) priors were originally proposed by Blackwell & MacQueen [26], and Ferguson [27]. This prior have been used in non-parametric Bayesian literature for classification and information sharing. The popularity of DP prior is mainly due to its computational ease and powerful inferential properties due to stick breaking formulation of Dirichlet Process proposed by Sethuraman [28]. More recently, Dunson et al. [29] formulated Matrix Stick Breaking Process (MSBP) priors, which can handle the information sharing across the model parameters for the datasets coming from different related "groups". Gaskins and Daniels [61], Das and Daniels [57] extended the MSBP for sharing the large covariance parameters for different related groups. Figure 3.1 illustrates the DP and MSBP in the context of parameter estimation. The proposed work in this chapter is based on Dunson et al. [29] which used MSBP for sharing the model parameters across different groups.

In this chapter, the cluster topology for a general WSN has been considered, as shown in Figure 2.1. The sensor nodes are "grouped" into "clusters" based on their relative Euclidean distances. Thus the nodes which are spatially close to each other are kept in the same cluster.

- The proposed linear model considers the effect of the nearest neighbours on the current state value for each sensor node and thus allows information exchange within cluster.
- Also, a non-parametric MSBP has been proposed for the cluster specific model parameters, which takes care of the possible information exchange between the clusters in the network under consideration.
- An application of the proposed approach in detecting immobile anomalous node has been illustrated.

An anomalous node may be a foreign object, a selfish node or a malicious node, which might have detrimental effect on the surrounding sensor nodes in course of time. The proposed approach can accurately detect such nodes. Removal of such nodes makes the network more effective.

The findings of this chapter has been published in a research paper.[1]

3.2 Proposed Model

3.2.1 Linear Markov Model

Consider a WSN with K clusters and the k-th cluster consists of n_k sensors with a total of $n = \sum_{k=1}^{K} n_k$ sensors. Note that clusters are formed by considering the Euclidean distance between the sensors, i.e. for some fixed δ, all the sensors which are in the δ-neighbourhood of each other are kept in a single cluster. Figure 2.1 demonstrates the process of cluster formation, as described in Section 2.2. Note that δ should be chosen such that the number of clusters is neither too small nor too large (typically less than 10 in this setting). However,

[1] A. Chatterjee, P. Venkateswaran et al., "Simultaneous State Estimation of Cluster-Based Wireless Sensor Networks", IEEE Transactions on Wireless Communications, Vol 15, Issue 12, 2016.

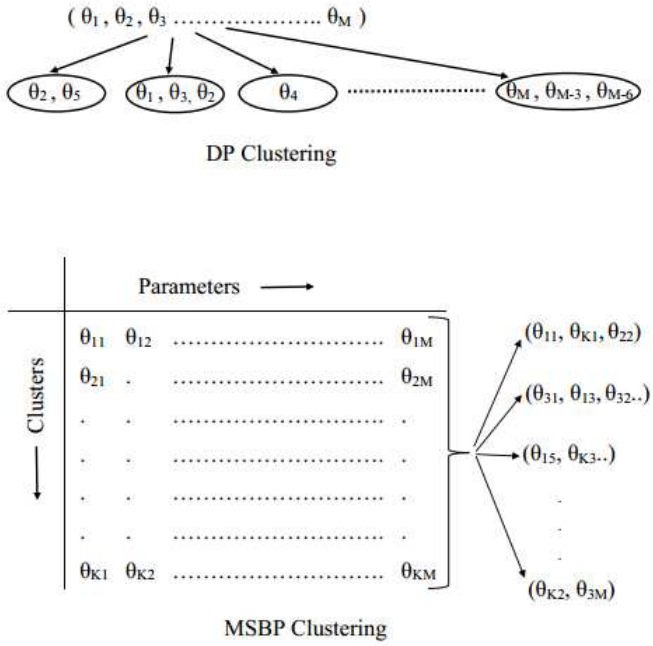


Figure 3.1: DP and MSBP for clustering/grouping the model parameters

in reality, the number of clusters (and hence the value of δ) depends on the coverage area and the nature of the experiment.

In this chapter, also a discrete-time state estimation procedure has been considered. For making an energy efficient protocol, it is assumed that all the sensors within a cluster are not necessarily measured exactly at the same time points. Some sensors might be kept in the sleep mode for some time as the sensors consume very little battery power in sleep mode, which makes the network energy efficient [38]. Let us assume that the i-th sensor belonging to the k-th cluster ($k = 1, 2, \ldots, K$) is measured at T_i^k different time points and $X_{ik}(t_{ij})$ denotes its state value at time t_{ij} ($j = 1, 2, \ldots, T_i^k$). Also assume that at each time point at least one sensor from each cluster is measured to keep the cluster active. Proposed linear Markov model for estimating the state value of the i-th sensor at time t_{ij} based on the observed measurements till time $t_{i(j-1)}$ can be expressed as:

$$X_{ik}(t_{ij}) = f_k(t_{ij}) + \theta_{k1} X_{ik}(t_{i(j-1)}) I(|t_{ij} - t_{i(j-1)}| < p) \\ + \theta_{k2} Z_{ik}(t_{i(j-1)}) + e_{ijk} \tag{3.1}$$

where f_k is the cluster specific general effect of time, which is modelled using Penalized splines. The effect of time on the state values is possibly different for different clusters, and hence, the subscript k in the function $f(.)$ is included. Here the indicator function $I(|t_{ij} - t_{i(j-1)}| < p)$ takes value 1, if $|t_{ij} - t_{i(j-1)}| < p$ and 0, otherwise. Note that θ_{k1} is the cluster specific effect of the previous available measurement on the current state value and will be estimated based on the available data. The previous available measurement will influence the current state value only when the time difference is below a fixed (known) threshold p, typically p=3 in this case. This is based on the fact that measurements corresponding to the closer time points are more related than those from the further time points.

In equation (3.1), $Z_{ik}(t_{i(j-1)})$ denotes the average measurement from all the sensors belonging to the k-th cluster (except the i-th sensor) which are measured at time $t_{i(j-1)}$. Since it is assumed that at each time point, some sensors only are measured from each cluster, based on the available data for the k-th cluster $Z_{ik}(t_{i(j-1)})$ can be easily obtained. Hence θ_{k2} basically denotes the "neighbourhood effect" on the state value and needs to be estimated from the available data. Note that by introducing Z variable, the author essentially incorporates the information sharing among the sensors within a single cluster. The residual errors e_{ijk}'s are assumed to be independently normally distributed with mean=0 and unknown variance=σ^2. For all K clusters, σ^2 need to be estimated from the available data. The above model is Markovian as the current state value depends only on the immediate (available) predecessor values.

The above model is novel in the context of state estimation problem in WSNs. The model is dynamic in nature and considers adequately the neighbourhood effect for estimating sensor-specific state values. It allows information sharing within the same cluster. In latter sections, the author will allow information sharing (in terms of the parameter values) between clusters and thus proposing a simultaneous state estimation of cluster based WSNs.

3.2.2 Penalized Splines for the General Effect of Time

In equation (3.1), f_k basically explains how much of the response $X_{ik}(t_{ij})$ is due to time and the fact that this taken into account is the effect might be different for different clusters. In other words, state values of the sensors within a certain cluster might vary more than the sensors in a different cluster. In fact, there might be some unobserved time-varying factors, which influence the observed state values and all these hidden factors are combined into this general effect of time. Ideally, the function f_k should be smooth (at least twice differentiable).

There is a rich literature in statistics for efficiently modelling f_k. Different smoothing techniques like smoothing splines, B-splines, Penalized splines, wavelets are typically used for this purpose. Traditionally, smoothing splines are preferred since they give the smallest Mean Squared Error (MSE). However fitting of smoothing splines is expensive in terms of the computational cost. Hence, Penalized splines (P-splines) ([63]) has been preferred to avoid the so called "curse of dimensionality" (large number of model parameters for high dimensional data).

With a polynomial spline of degree r with knots $(\mathcal{T}_1, ..., \mathcal{T}_S)$, the general effect of time as described in equation (3.1) can be expressed as the following non-parametric regression model:

$$f_k(t_{ij}) = b_{k0} + b_{k1}t_{ij} + b_{k2}t_{ij}^2 + ... + b_{kr}t_{ij}^r + \sum_{s=1}^{S} c_{ks}(t_{ij} - \mathcal{T}_s)_+^r \tag{3.2}$$

where $(x)_+^r = x^r$, for $x > 0$, and 0 otherwise; and $(\mathcal{T}_1 < \mathcal{T}_2 < ... < \mathcal{T}_S)$ is a fixed (known) set of knots. In practice, it is desirable to keep the degree of the spline (r) relatively small, between 1 and 3, so that one need to estimate only a few parameters for the above functional relation. Here the parameters $b_{k0}, b_{k1}, \ldots, b_{kr}$; $c_{k1}, c_{k2}, \ldots, c_{kS}$; r and S are to be estimated from the data.

Note that the model in equation (3.2) has two components: a polynomial component and a spline component. Knots are the points where the function changes its form. Knots divide the time domain into different pieces. For each piece, a polynomial of order r has been fitted but these polynomials are changed for different pieces. Observe that the function $f_k(.)$ has continuous derivatives at all the points except the knots. Thus, the coefficients of the spline part (c_{ks}) represent the measure of roughness at the respective knots. Typically the evenly spaced sample quantiles of time are taken as the knots [62,63]. The optimal number of knots (S) and the optimal degree (r) are to be obtained from the information criteria like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Deviance Information Criterion (DIC) [61,57] etc.

Define $b_k = [b_{k0}, b_{k1}, \ldots, b_{kr}]^T$, $c_k = [c_{k1}, c_{k2}, \ldots, c_{kS}]^T$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{iT_i^k}]^T$.
Then in matrix notation, equation (3.2) can be expressed as:

$$f_k(t_i) = U_i^k b_k + V_i^k c_k \tag{3.3}$$

where

$$\mathbf{U_i^k} = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & \ldots & t_{i1}^r \\ 1 & t_{i2} & t_{i2}^2 & \ldots & t_{i2}^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{iT_i^k} & t_{iT_i^k}^2 & \ldots & t_{iT_i^k}^r \end{bmatrix}$$

and

$$\mathbf{V_i^k} = \begin{bmatrix} (t_{i1} - \mathcal{T}_1)_+^r & (t_{i1} - \mathcal{T}_2)_+^r & \ldots & (t_{i1} - \mathcal{T}_S)_+^r \\ (t_{i2} - \mathcal{T}_1)_+^r & (t_{i2} - \mathcal{T}_2)_+^r & \ldots & (t_{i2} - \mathcal{T}_S)_+^r \\ \vdots & \vdots & \ddots & \vdots \\ (t_{iT_i^k} - \mathcal{T}_1)_+^r & (t_{iT_i^k} - \mathcal{T}_2)_+^r & \ldots & (t_{iT_i^k} - \mathcal{T}_S)_+^r \end{bmatrix}$$

The curve can be made smoother by placing a penalty on the roughness parameters. In other words, the author would like to minimize the following expression:

$$\mathcal{E} = \sum_{i,j,k} (X_{ik}(t_{ij}) - f_k(t_{ij}))^2 + \lambda^* \sum_{s=1}^{S} \sum_{k=1}^{K} |c_{ks}|^d \tag{3.4}$$

where λ^* is the smoothing parameter having significant control over the smoothing process. Note that if $\lambda^* \to 0$, the author simply get the ordinary least squares (OLS) estimates. Again if $\lambda^* \to \infty$, then the author have the polynomial part only (and no spline part) in equation (3.2). For d=2, the ridge regression is obtained and d=1 gives Least Absolute Shrinkage and Selection Operator (LASSO), which has a literature of its own [64]. However, d=2 has been considered in this presentation to get the ridge regression.

In a Bayesian framework, ridge regression estimates can be obtained by considering a normal prior for c_k followed by a gamma prior for λ^*. It is noted that the optimal number of knots and the locations of the knots can be left as variables and can be estimated from the data [42] using reversible jump MCMC. This is more flexible but computationally expensive because the reversible jump MCMC algorithms are complicated in general. The author avoids such methods for computational ease.

3.3 Prior Structure and Parameter Estimation

3.3.1 Non-parametric Dirichlet Process (DP) Prior

In the non-parametric Bayesian literature, Dirichlet process priors are used quite frequently for clustering, classification and information sharing [26,27,28,57]. A Dirichlet process is a distribution over distributions having two parts, the base distribution and the concentration parameter. Symbolically, the author say $G \sim DP(G_0, \gamma)$ to denote that G follows a Dirichlet process distribution with base measure G_0 and concentration parameter γ. Below, the clustering nature of Dirichlet Process has been discussed briefly through finite mixture model.

Suppose that random variables $y_1, \ldots, y_n$ are drawn independently from some unknown density. Then one can model the density of y as a mixture of C (finite) distributions where the density of the c-th ($c = 1, 2, \ldots, C$) distribution is $F(\phi_c)$. Thus the density of $y_1, \ldots, y_n$

has been modelled as the following:

$$P(y) = \sum_{c=1}^{C} p_c F(\phi_c) \tag{3.5}$$

where p_c's are the mixing probabilities and ϕ_c's are the parameters of the respective density functions. Considering finite C, a symmetric Dirichlet prior has been taken with the following density function for the mixing probabilities p_c:

$$P(p_1, \ldots, p_C) = \frac{\Gamma(\gamma)}{\Gamma(\gamma/C)^C} \prod_{c=1}^{C} p_c^{(\gamma/C)-1} \tag{3.6}$$

Further, it is assumed ϕ_c's are independently distributed with distribution G_0.

Using the mixture identifiers $c_i$this model can be represented as follows:
$y_i|(c_i, \phi) \;\; \sim F(\phi_{c_i}),\;\; c_i|(p_1, \ldots, p_C) \;\; \sim Discrete(p_1, \ldots, p_C),\;\; p_1, \ldots, p_C \;\; \sim \text{Dirichlet}(\frac{\gamma}{C}, \ldots, \frac{\gamma}{C})$, $\phi_{c_i} \;\; \sim G_0$.

By integrating over the Dirichlet prior, the following conditional distribution is obtained: $P(c_i = c|c_1, \ldots, c_{i-1}) = \frac{n_{ic}+\gamma/C}{i-1+\gamma}$, where n_{ic}= Number of data points previously assigned to component c. Also when $C \to \infty$, $P(c_i = c|c_1, \ldots, c_{i-1}) \to \frac{n_{ic}}{i-1+\gamma}$; and $P(c_i \neq c_j$ for all $j < i|c_1, \ldots, c_{i-1}) \to \frac{\gamma}{i-1+\gamma}$.
Consequently, the conditional probability of θ_i, where $\theta_i = \phi_{c_i}$, becomes
$\theta_i|(\theta_1, \ldots, \theta_{i-1}) \;\; \sim \frac{1}{i-1+\gamma} \sum_{j<i} \delta_{\theta_j} + \frac{\gamma}{i-1+\gamma} G_0$, where δ_θ is a point mass at θ.

Note that the above formulation gives:

$$y_i|\theta_i \;\; \sim F(\theta_i);\;\; \theta_i|G \;\; \sim G;\;\; G \;\; \sim DP(G_0, \gamma)$$

Thus, each data point i has its own parameter θ_i drawn independently from a distribution which is again drawn from a DP prior. Because of the clustering property of DP, θ_i for different data points might be the same.

Sethuraman [28] provided the well-known stick-breaking formulation of Dirichlet Process, which is very useful for simulation and mathematical treatments. In this formulation, if $G \sim DP(G_0, \gamma)$, then

$$G = \sum_{h=1}^{\infty} \left[V_h \prod_{l<h} \bar{V_l} \right] \delta_{\Theta_h}, \;\; V_h \overset{\text{iid}}{\sim} Beta(1,\gamma), \; \Theta_h \overset{\text{iid}}{\sim} G_0, \tag{3.7}$$

where $\mathbf{V} = \{V_h, h = 1, 2, \ldots, \infty\}$ is an infinite sequence of stick-breaking weights with $\bar{V_h} = 1 - V_h$ and $\Theta = \{\Theta_h, h = 1, 2, \ldots, \infty\}$ is an infinite sequence of random atoms from the base distribution G_0. In practice, for the ease of computation, typically the infinite sum in equation (3.7) is truncated upto finite N with $V_N = 1$ by making the expected approximation error arbitrarily small, in general, smaller than 0.01 [29].

3.3.2 Matrix Stick-breaking Process (MSBP) for related groups

Sometimes in real applications, data are collected from multiple related groups and then the focus is on borrowing information over the groups. Dunson et al. [29] proposed MSBP which allows information sharing across groups. It may be noted that by "information sharing", the author means similarity of the model parameters for different groups. MSBP allows the model parameters to be similar (or same) for the related groups by considering a modified stick-breaking prior on the parameters. The author first give a brief review of MSBP in the context of proposed linear Markov model given in equation (3.1).

Recall the setting in section 3.2.1, there are total K clusters and the k-th cluster consists of n_k sensors. Assume that the sensors from the k-th cluster are drawn from a parametric model characterized by the L-dimensional parameter vector Ψ_k. MSBP considers a random probability measure F_{kl} such that $\psi_{kl} \sim F_{kl}; \;\; k = 1, 2, \ldots, K; l = 1, 2, \ldots, L$. The matrix of random probability measures $F = \{F_{kl}; \;\; k = 1, 2, \ldots, K; l = 1, 2, \ldots, L\}$ will have a distribution which will induce the correlation among different F_{kl}'s. The form of each F_{kl} is given as the following:

$$F_{kl} = \sum_{h=1}^{H} \pi_{klh} \delta_{\epsilon_{lh}}, \;\; \epsilon_{lh} \overset{\text{iid}}{\sim} F_{0l}$$

where $E = \{\epsilon_{lh}\}$ is $L \times H$ matrix and δ_x is a point mass at x. Note that the rows of E correspond to each model parameter with base distribution F_{0l} and the columns correspond to H clusters.

The stick-breaking weights π_{klh} control the dependence among the different distribution in the following way:

$$\pi_{klh} = V_{klh} \prod_{r<h} (1 - V_{klr}), \;\; V_{klh} = U_{kh}Y_{lh}, \;\; U_{kh} \overset{\text{iid}}{\sim} Beta(1, \alpha), \; Y_{lh} \overset{\text{iid}}{\sim} Beta(1, \beta)$$

Note that here V is expressed as the product of U and Y, where U_{kh} controls the likelihood that parameter from k-th group (cluster of sensors, in this case) belongs to the h-th cluster induced by MSBP and Y_{lh} controls the likelihood that the l-th parameter is drawn from the h-th cluster. Note that U's are shared across the parameters and Y's are shared across the clusters, MSBP allows the possible correlation among the distributions in F. In order to make F_{kl} a proper probability distribution, one need to take $U_{kH} = 1$ and $Y_{lH} = 1$.

Note that for $H \to \infty$, the above formulation is an actual MSBP and for finite H, it is called truncated approximation of MSBP. Following Dunson et al. [29], Das and Daniels [57], the expected truncation error is given as:

$$E(\sum_{h=H}^{\infty} \{\pi_{klh}\}) = \left[1 - \frac{1}{(1+\alpha)(1+\beta)}\right]^{H-1} \tag{3.8}$$

Thus for a prefixed δ^* (≤ 0.01) one can choose H so that the expression in the equation (3.8) is below δ^*. Further properties of MSBP and the truncation are given in Das and Daniels [57], Dunson et al. [29], Gaskins and Daniels [61].

3.3.3 Proposed Prior Structure

The proposed prior structure is based on the MSBP as outlined in the previous section. Consider the Markovian model as given in equation (3.1). It is noted that the proposed approach will allow information sharing (i.e. same and/or similar parameter values) across the clusters of sensors and this is achieved by considering MSBP priors on the model parameters.

First, the parameters given in the equation (3.2) has been considered. Let $\mathbf{b}_k =$

$[b_{k1}, b_{k2}, \ldots, b_{kr}]^T$, $\mathbf{c}_k = [c_{k1}, c_{k2}, \ldots, c_{kS}]^T$ and $\boldsymbol{\theta} = [\theta_{k1}, \theta_{k2}]^T$.

For the parameters $\mathbf{b}_k$, the following prior structure has been proposed:

$$
\begin{aligned}
&b_{kl} \sim F^b_{kl} = \sum_{q=1}^{N_b} \pi^b_{klq} \delta_{\epsilon^b_{lq}}(.); \;\; k = 1, \ldots, K; \;\; l = 0, \ldots, r; \\
&\epsilon^b_{lq} \sim (1 - p^b)\delta_0(.) + p^b N(0, \sigma^2_b); \;\; q = 1, ..., N_b; \;\; l = 0, \ldots, r; \\
&\pi^b_{klq} = U^b_{kq} W^b_{lq} \prod_{h<q} (1 - U^b_{kh} W^b_{lh}); \\
&U^b_{kq} \overset{\text{iid}}{\sim} Beta(1, \alpha_b), \;\; W^b_{lq} \overset{\text{iid}}{\sim} Beta(1, \beta_b); \;\; q = 1, \ldots, N_b - 1.
\end{aligned}
$$

Note that here $\delta_0(.)$ denotes a point mass at zero. For the atoms ϵ^b_{lq}, a slab and spike distribution has been considered and this ensures a positive probability at zero. A zero value basically indicates the absence of the respective coefficient in the polynomial part of equation (3.2). The above prior structure allows the sharing of the parameter values across various clusters and thus making the network more informative. For the mixing proportion p^b, a beta prior has been considered following the usual convention [57].

For the parameters $\mathbf{c}_k$, proposed prior structure is the following:

$$
\begin{aligned}
&c_{ks} \sim F^c_{ks} = \sum_{q=1}^{N_c} \pi^c_{ksq} \delta_{\epsilon^c_{sq}}(.); \;\; k = 1, \ldots, K; \;\; s = 1, \ldots, S; \\
&\epsilon^c_{sq} \sim (1 - p^c)\delta_0(.) + p^c N(0, \frac{1}{\lambda^*}); \;\; q = 1, ..., N_c; \;\; s = 1, \ldots, S; \\
&\lambda^* \sim Gamma(\alpha, \beta); \\
&\pi^c_{ksq} = U^c_{kq} W^c_{sq} \prod_{h<q} (1 - U^c_{kh} W^c_{sh}); \\
&U^c_{kq} \overset{\text{iid}}{\sim} Beta(1, \alpha_c), \;\; W^c_{sq} \overset{\text{iid}}{\sim} Beta(1, \beta_c); \;\; q = 1, \ldots, N_c - 1.
\end{aligned}
$$

The above prior structure for $\mathbf{c}_k$ allows information sharing (by MSBP property) across the clusters and it also employs a Ridge regression technique for minimizing the expression as given in equation (3.4). Here also, a slab and spike distribution has been taken for the atoms which includes the possibility of the values to be zero. A zero value indicates the absence of the respective knot in the spline part of equation (3.2). For the non-null case, by taking a normal distribution whose variance follows an inverse gamma distribution, the author essentially do the standard Ridge regression [63]. A beta prior is taken for the mixing proportion p^c [57].

And for $\boldsymbol{\theta}$, the proposed prior is the following:

$$\theta_{kx} \sim F^{\theta}_{kx} = \sum_{q=1}^{N_\theta} \pi^{\theta}_{kxq} \delta_{\epsilon^{\theta}_{xq}}(.); \;\; k = 1, \ldots, K; \;\; x = 1, 2;$$
$$\epsilon^{\theta}_{xq} \sim (1 - p^{\theta})\delta_0(.) + p^{\theta} N(0, \sigma^2_{\theta}); \;\; q = 1, ..., N_\theta; \;\; x = 1, 2;$$
$$\pi^{\theta}_{kxq} = U^{\theta}_{kq} W^{\theta}_{xq} \prod_{h<q} (1 - U^{\theta}_{kh} W^{\theta}_{xh});$$
$$U^{\theta}_{kq} \overset{\text{iid}}{\sim} Beta(1, \alpha_\theta), \;\; W^{\theta}_{xq} \overset{\text{iid}}{\sim} Beta(1, \beta_\theta); \;\; q = 1, \ldots, N_\theta - 1$$

Here again, a beta prior is considered for p^{θ}. Just like the previous parameters, a slab and spike distribution is considered for allowing a zero value. Note that a zero value for θ_{k1} in equation (3.1) reflects that the available state value of the immediate predecessor time point has no effect on the current state value. Similarly, it is inferred no neighbourhood effect on the current state value if θ_{k2} is found to be zero. One can perform a formal statistical hypothesis test for the significance of the neighbourhood effect and the predecessor effect.

The above priors for the model parameters ensure the fundamental goal of this chapter. By considering MSBP for the model parameters, the parameters are allowed to be same or similar (in their numerical values) across the clusters. For example, the neighbourhood effect for two (or more) clusters might be exactly the same. The proposed approach also provides a numerical measure of similarity for each parameter across the clusters (shown in Section 4.1). In reality, if the clusters do not have the similar (or same) parameter values or all the clusters have exactly the same parameter values, the performance of the proposed model will be close to the truth [27, 61]. However, if some parameters are similar across the clusters but the other parameters are not so similar (or different), then the proposed approach provides better estimate of the model parameters with relatively small standard errors. Such operating characteristics and practical applications of the proposed approach has been demonstrated through simulation studies in Section 4.

3.3.4 Posterior Estimation and MCMC

It may be noted that in a Bayesian approach, the parameters are estimated from the joint posterior distribution which is the product of the likelihood function and the prior distributions of all the model parameters. For the sake of simplicity in posterior inference, it is assumed that the priors are independent of each other.

Consider the linear model in equation (3.1). It may be noted that, the author assumed variance$(e_{ijk}) = \sigma^2$, and considered an inverse gamma (κ_1, κ_2) prior for σ^2.

Note that from equation (3.1) of Section 3.2.1, the residual
$e_{ijk} = X_{ik}(t_{ij}) - f_k(t_{ij}) - \theta_{k1} X_{ik}(t_{i(j-1)}) I(|t_{ij} - t_{i(j-1)}| < p) - \theta_{k2} Z_{ik}(t_{ij} - 1)$. Thus the joint likelihood of the vector of state values $\mathbf{X}_{ik}$ can be expressed as the joint likelihood of the residuals as the following:

$$L \propto \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left[\sigma^{-T_i^k} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{e}_{ik}^T \mathbf{e}_{ik}) \right) \right] \tag{3.9}$$

It has been noted that our proposed prior structure is very similar for $\mathbf{b}_k$, $\mathbf{c}_k$ and $\boldsymbol{\theta}$ and hence the posterior inferences will also be similar. Here the full conditional distributions of $\mathbf{b}_k$ obtained from the joint posterior distribution has been provided. One can derive the full conditional distributions of $\mathbf{c}_k$ and $\boldsymbol{\theta}$ accordingly.

Following Dunson et al. [29], latent variables R_{lk}^b are introduced from multinomial distributions with respective probabilities π_{klq}^b. Let us consider the following binary dummy variables for $k = 1, \ldots, K;\ l = 0, \ldots, r$ and $q = 1, \ldots, N_b$
$u_{lkq}^b \sim$ Bernouli(U_{kq}^b), \$$w_{lkq}^b$ $\sim$Bernouli(W_{lq}^b).

Now, let us define $R_{lk}^b = min\left(q : 1 = u_{lkq}^b = w_{lkq}^b\right)$. Note that R_{lk}^b's are distributed as multinomial distributions.

1. The full conditional distribution of ϵ_{lq}^b is given as:

$$\epsilon_{lq}^b | - \propto \prod_{i=1}^{n_k} \left[\sigma^{-T_i} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{e}_{ik}^T \mathbf{e}_{ik}) \right) \right] \times \left((1 - p_b)\delta_0(.) + p_b \exp\left(\frac{-(\epsilon_{lq}^b)^2}{2\sigma_b^2} \right) \right) \tag{3.10}$$

2. The full conditional distribution of R^b_{lk} is given as:

$$P(R^b_{lk} = q|-) \propto \pi^b_{klq} \times \prod_{i=1}^{n_k} \left[\sigma^{-T_i} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{e}^T_{ik}\mathbf{e}_{ik}) \right) \right] \tag{3.11}$$

R^b_{lk} has been sampled from the multinomial distribution with the above probabilities (normalized to sum to one). Given the values of R^b_{lk}, one can sample u^b_{lkq} and w^b_{lkq} accordingly.

3. The full conditional distributions for U^b_{kq} for $q < N_b$ are given as:

$$U^b_{kq}|- \sim Beta\left(1 + \sum_{l=0}^{r} u^b_{lkq}, \alpha_b + \sum_{l=0}^{r}(1 - u^b_{lkq})\right) \tag{3.12}$$

Similarly for $q < N_b$,

$$W^b_{lq}|- \sim Beta\left(1 + \sum_{k=1}^{K} w^b_{lkq}, \beta_b + \sum_{k=1}^{K}(1 - w^b_{lkq})\right) \tag{3.13}$$

$U^b_{kN_b}$ and $W^b_{lN_b}$ are drawn from distribution degenerate at 1.

4. By considering Gamma(1,1) priors for α_b and β_b, the following full conditionals are obtained:

$$\alpha_b|- \sim Gamma\left(K(N_b - 1) + 1, 1 - \sum_{k=1}^{K}\sum_{q=1}^{N_b-1} log(1 - U^b_{kq})\right) \tag{3.14}$$

Similarly,

$$\beta_b|- \sim Gamma\left(r(N_b - 1) + 1, 1 - \sum_{l=0}^{r}\sum_{q=1}^{N_b-1} log(1 - W^b_{lq})\right) \tag{3.15}$$

Finally, the full conditional distribution of σ^2 is given as :

$$\sigma^2|- \sim \prod_{k=1}^{K}\prod_{i=1}^{n_k}\left[(\sigma^2)^{-T_i/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{e}^T_{ik}\mathbf{e}_{ik})\right)\right] \times \pi(\sigma^2) \tag{3.16}$$

where $\pi(\sigma^2)$ is the prior for σ^2. If an Inverse Gamma prior for σ^2 is taken, the posterior will also be an Inverse Gamma distribution (conjugate prior).

A hybrid combination of Gibbs sampler and Metropolis-Hastings (MH) algorithm has been implemented. The initial values of the model parameters are taken from their respective prior distributions. Following the tradition, the first few "burn-in" iterations are discarded to remove the effect of the starting values and also thin the chains by keeping every 10-th iteration. The convergence of the chains are monitored graphically and also by computing multivariate potential scale reduction factor as proposed by Brooks and Gelman [65].

3.4 Simulation Studies

The performance of the proposed modelling approach has been assessed through extensive simulation studies. Also, the proposed approach has been compared to similar other models in terms of the parameter estimates and predictability. In the first simulation study, the performance of the model has been investigated while in the second simulation, a comparative study has been performed.

3.4.1 Simulation I: Performance Evaluation

A network with several clusters has been considered where the clusters do share the model parameters. The proposed WSN consists of three clusters (k=1,2,3) with 40, 45 and 50 sensors, respectively. For simplicity, it has been assumed that each sensor is measured exactly at 10 different time points (i.e. consider a case of regular longitudinal measurement). The state values are simulated from the following model:

$$X_{ik}(t) = f_k(t) + \theta_{k1} X_{ik}(t-1) + \theta_{k2} Z_{ik}(t-1) + e_{ik}(t) \tag{3.17}$$

where the errors $e_{ik}(t)$'s are Gaussian (0, 1.5) and the general effect of time has been taken as the following quadratic spline function of time, with two evenly spaced knots:
$f_k(t) = b_{k0} + b_{k1}t + b_{k2}t^2 + c_{k1}(t-\tau_1)_+^2 + c_{k2}(t-\tau_2)_+^2$

Denote $\mathbf{b}_k = [b_{k0}, b_{k1}, b_{k2}]$, $\mathbf{c}_k = [c_{k1}, c_{k2}]$ and $\boldsymbol{\theta}_k = [\theta_{k1}, \theta_{k2}]$. For simulation, the following values has been taken of $\mathbf{b}_k$, $\mathbf{c}_k$ and $\boldsymbol{\theta}_k$, k=1,2,3;

$\mathbf{b}_1 = (2.3, 1.45, 2.17),\ \mathbf{b}_2 = (2.26, 1.48, 2.55),\ \mathbf{b}_3 = (3.4, 1.48, 2.23)$

$\mathbf{c}_1 = (1.5, 0.96)$, $\mathbf{c}_2 = (1.6, 1.12)$, $\mathbf{c}_3 = (0.49, 1.10)$

$\boldsymbol{\theta}_1 = (1.8, 2.5)$, $\boldsymbol{\theta}_2 = (2.8, 3.6)$, $\boldsymbol{\theta}_3 = (2.0, 3.9)$

Note that the proposed estimation method does not depend on any specific set of values for these parameters. For the simulation purpose, one set of parameter values has been considered which shows the effectiveness of the proposed approach of estimating the state values corresponding to this particular set of the parameter values. One can consider a completely different set of parameter values, but still can efficiently estimate the corresponding state values.

At first, the state values for each sensor node belonging to different clusters at 10 different time points are simulated.

Then the proposed approach has been used and the appropriate MSBP priors are taken as mentioned in the earlier sections for estimating the model parameters. The author run 100,000 iterations of the Markov Chains and the first 20,000 iterations were discarded as "burn-in" to remove the effect of the starting values of the parameters. Also to thin the chains by saving every 10-th iteration to remove the auto-correlation among the estimates from the successive iterations. Under the squared error loss function, the model parameters has been estimated by the respective posterior sample means and also the Monte Carlo Standard Errors (MCSE) is estimated.

Table 3.1, shows the estimated parameters, the true parameter values and 95% Confidence Interval (CI) of the parameters computed as estimate $\pm$ 2 $\times$ MCSE. It has been noted that in the Bayesian framework, the true parameter belongs to the corresponding confidence intervals with probability=0.95. Hence the width of the 95% confidence interval reflects how precisely one estimates the model parameter. It has been observed that the estimated parameter values are very close to the true values and the width of the confidence intervals are small. Thus the proposed approach estimates the parameters accurately and precisely and hence one can use these parameters to estimate the state values of the nodes for valid inference.

Next, a sensitivity analysis has been performed to investigate the effect of the priors with

Table 3.1: Estimates and 95% CI of the model parameters in Simulation 1

Parameter	True Value	Estimates	95% CI
b_{10}	2.3	2.27	(1.86, 2.73)
b_{11}	1.45	1.46	(1.35,1.58)
b_{12}	2.17	2.15	(2.03,2.39)
b_{20}	2.26	2.24	(2.11,2.45)
b_{21}	1.48	1.50	(1.23,1.88)
b_{22}	2.55	2.53	(2.32,2.96)
b_{30}	3.4	3.36	(3.19,3.78)
b_{31}	1.48	1.52	(1.26,1.80)
b_{32}	2.23	2.20	(2.03,2.49)
c_{11}	1.5	1.38	(1.15, 1.56)
c_{12}	0.96	1.03	(0.73,1.44)
c_{21}	1.6	1.54	(1.28, 1.83)
c_{22}	1.12	1.17	(1.04,1.36)
c_{31}	0.49	0.43	(0.23,0.74)
c_{32}	1.10	1.18	(0.92, 1.33)
θ_{11}	1.8	1.66	(1.45,1.91)
θ_{12}	2.5	2.52	(2.27,2.77)
θ_{21}	2.8	2.76	(2.44,3.12)
θ_{22}	3.6	3.55	(3.36,3.88)
θ_{31}	2.0	1.88	(1.54,2.39)
θ_{32}	3.9	3.84	(3.48,4.36)
σ^2	1.5	1.31	(1.11,1.58)

different parameter values on the estimates. This is a routine work in Bayesian analysis. As one starts with some prior distribution in a Bayesian analysis, it is important to show that the final estimate of the parameters do not depend much on the choice of such prior distributions. In this study, for the parameters p^b, σ_b^2, α_b and β_b, priors have been taken with three different parameter combinations. Table 3.2, summarizes the results and it has been observed that the final estimates do not depend on the parameter values of the prior distributions. Ideally, the inference should not be sensitive to the choice of the parameter of the priors and this holds in this analysis. Similar results are obtained for the other parameters as well.

In Figure 3.2, the grouping nature of the proposed priors has been demonstrated for the parameter **b, c** and $\boldsymbol{\theta}$. The posterior probabilities shows: $Pr(b_{kl} = b_{k'l})$; for each $(k = 1,2,3; k' = 1,2,3; l = 1,2,3)$, $Pr(c_{kl} = c_{k'l})$; for $(k = 1,2,3; k' = 1,2,3; l = 1,2)$, and $Pr(\theta_{kx} = \theta_{k'x})$ for $(k = 1,2,3; k' = 1,2,3; x = 1,2)$. Such probabilities are computed based on MCMC iterations. For example, 8,000 iterations are obtained for b_{kl} and $b_{k'l}$ for $k = 1,2,3$; $k' = 1,2,3$ and $l = 1,2,3$. $Pr(b_{kl} = b_{k'l})$ is computed as the proportion of times

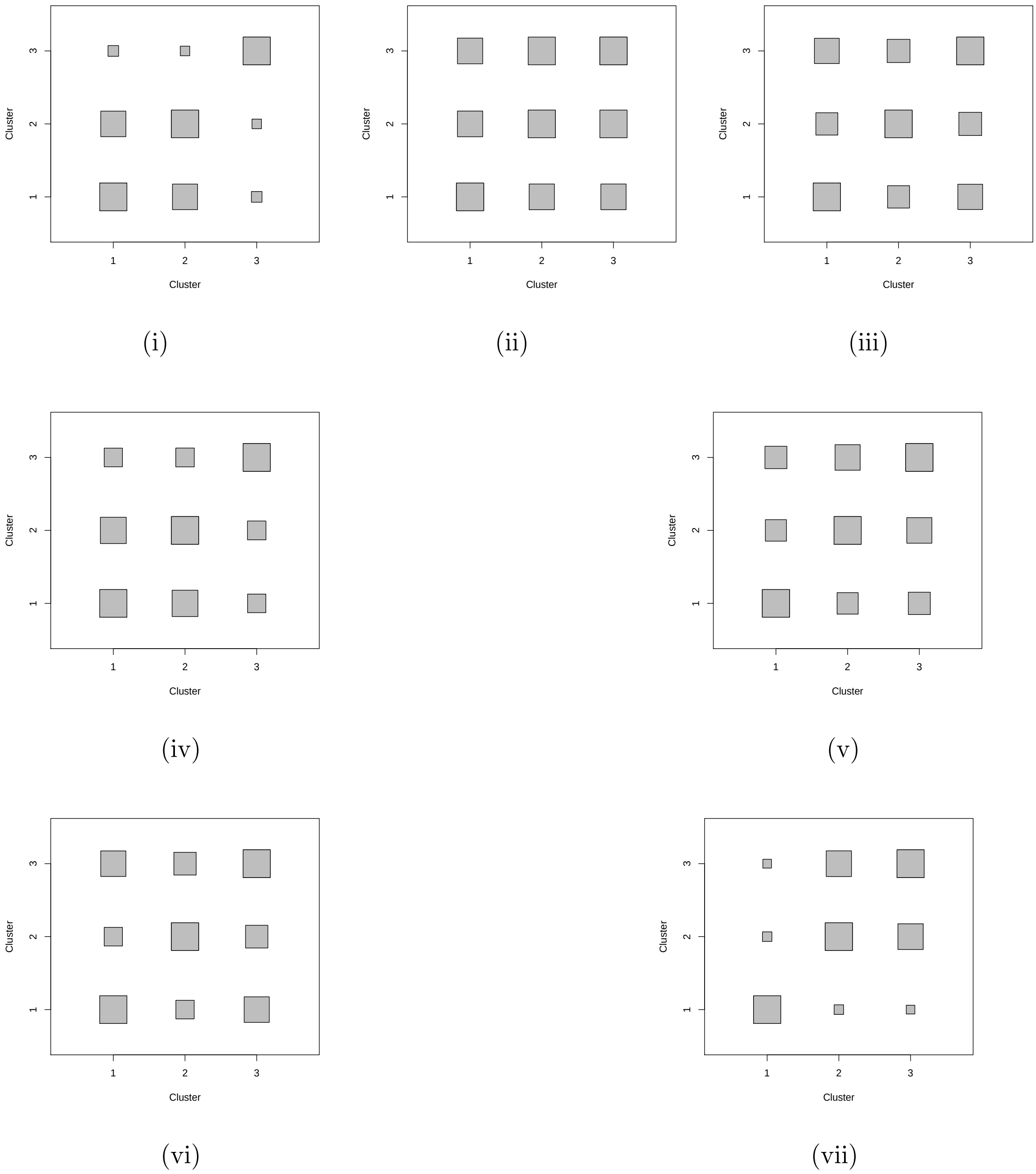


Figure 3.2: The posterior probabilities of matching for (i)b_0 (ii)b_1 (iii)b_2 (iv)c_1 (v)c_2 (vi)θ_1 and (vii)θ_2 respectively for Simulation I

Table 3.2: Results from the sensitivity analysis in Simulation 1

Parameter	Prior Choice	Estimate
p^b	Beta(1,2)	0.43
–	Beta(1.5,3.6)	0.44
–	Beta(2.5,4)	0.43
σ_b^2	inverse gamma (2,3)	2.48
–	inverse gamma (3,5)	2.46
–	inverse gamma (1.8,4.7)	2.47
α_b	gamma (2,3.5)	1.78
–	gamma (3.4,4.9)	1.77
–	gamma (1.5,4.5)	1.78
β_b	gamma (2,3.5)	1.63
–	gamma (3.4,4.9)	1.65
–	gamma (1.5,4.5)	1.65

$b_{kl} = b_{k'l}$ (upto two decimal places). Larger boxes indicate higher grouping probabilities, with the largest boxes on the diagonal corresponding to probability=1. It is noted that for b_0, the box corresponding to the cluster 1 and cluster 2 is larger than the boxes corresponding to cluster 1 and cluster 3; and cluster 2 and cluster 3. This reflects the fact that for this parameter, cluster 1 and cluster 2 are similar with higher probability. It may be noted that this is actually the case as per simulation. Similarly for θ_2, cluster 2 and cluster 3 are similar with higher probability. As seen from Figure 3.2 for other parameters also, similar results are obtained. Thus the proposed approach automatically provides a measure of similarity for different clusters on each model parameter.

3.4.2 Simulation 2: Model Comparison

The performance of the proposed non-parametric MSBP based approach has been compared to the other traditionally used methods. First, data (state values) for an artificially created WSN are simulated with three clusters (k=1,2,3) containing 40, 45 and 50 sensors, respectively; using model given in equation (3.17) with the errors $e_{ik}(t)$'s distributed as Gaussian (0, 1.5) and $f_k(t) = b_{k0} + b_{k1}t + b_{k2}t^2 + c_{k1}(t-\tau_1)_+^2 + c_{k2}(t-\tau_2)_+^2$. Note that similar to Section 3.4.1 one can consider different prior distributions for the model parameters, and errors. But for the purpose of demonstration, one set of distributions has been considered.

The following priors has been considered for the proposed model parameters. For $\mathbf{b}_k$, trivariate normal $(0,\Sigma_b)$ prior has been taken where Σ_b is a diagonal matrix with (2,2.5,3.5) as the diagonal elements. For $\mathbf{c}_k$, bivariate normal $(0, \frac{1}{\lambda^*} I_2)$ prior has been taken and for λ^* a gamma (1.5,3) prior is considered. Finally, for $\boldsymbol{\theta}_k$, a bivariate normal $(0, D)$ prior has been considered where D is diagonal matrix with diagonal elements 2 and 3.

The parameters are generated from their respective priors and then using the generated parameter values, the state values are simulated for different sensors from three different clusters. This process is repeated 100 times and thus 100 different datasets are generated from the same model.

First, three different models derived from equation (3.17) are considered by changing the function $f_k(t)$ and the error components for different clusters and the model parameters for each cluster has been estimated separately using the simulated data. This approach is referred as Method 1. Next a single model has been considered for all the three clusters. In other words, the model given in equation (3.17) has been fitted without the subscript k for all 135 sensor nodes and estimated the model parameters using the combined data from three clusters. This approach is referred as Method II.

Then the proposed approach, referred as Method III is considered. Here also different model are fitted for different clusters but MSBP priors has been considered on the model parameters as proposed in Section 3.3. Such priors will allow information exchange on the model parameters across the clusters.

Finally, the following dynamic state-space model has been considered:

$$Y_{ik}(t) = X_{ik}(t) + d_{ik}(t), \quad X_{ik}(t) = aX_{ik}(t-1) + e_{ik}(t) \tag{3.18}$$

where $Y_{ik}(t)$ denotes the observed measurement from the i-th sensor of the k-th cluster at time t and $X_{ik}(t)$ denotes the unobserved state value (which is to be estimated). The measurement error $d_{ik}(t)$ and the system error $e_{ik}(t)$ are identically and independently distributed (iid) as Gaussian $(0,\nu^2)$ and Gaussian $(0,\kappa^2)$, respectively. In addition, it is assumed $X_{ik}(0)$ are iid Gaussian $(0,\tau^2)$. Also, $|a| < 1$ has been considered for the convergence of the parameters. The model parameters are estimated by KF [43] using sequential importance sampling. This approach has been referred as Method IV.

For each dataset, the model parameters for each method are estimated and then the mean squared error (MSE) is computed defined as:

$$\text{MSE} = \frac{1}{\sum\limits_{k=1}^{3} n_k} \sum_{k=1}^{3} \sum_{i=1}^{n_k} \left(X_{ik}(t_{ij}) - \hat{X}_{ik}(t_{ij}) \right)^2$$

where $\hat{X}_{ik}(t_{ij})$ is the estimated state value from the model under consideration.
For the Method IV,

$$\text{MSE} = \frac{1}{\sum\limits_{k=1}^{3} n_k} \sum_{k=1}^{3} \sum_{i=1}^{n_k} \left(Y_{ik}(t_{ij}) - \hat{X}_{ik}(t_{ij}) \right)^2$$

The average MSE is then computed by averaging three MSE values for three clusters. This process has been for all the 100 simulated datasets and finally average the MSE values across datasets.

Figure 3.3, shows the plot of average MSE for the four competing methods at 10 different time points as per simulation model. It may be noted that the average MSE for Method III (the proposed method) is consistently lower than the other methods over time. Method I gives the worst performance in terms of average MSE and Method II works better than Method I and Method IV but worse than Method III. Method IV performs better than Method I but worse than the other two methods. This illustrates better predictive power of the proposed approach for cluster based WSNs.

3.4.3 Application of the Proposed Method in Detecting Anomalous Node

Here, an application of the proposed approach has been illustrated in detecting an anomalous node in a cluster based WSN. An anomalous node is the node, which functions differently than the other sensor nodes, and thus can be a selfish and/or a malicious node. A malicious node is an internal intruder having detrimental effect on the surrounding sensor nodes. Detection of such nodes has become a major research topic in recent years [48, 51]. The cluster containing an anomalous node can be detected as the "target cluster" and the anomalous node can be detected by further splitting and merging of the clusters. Below, a numerical example has been provided for illustrating this application.

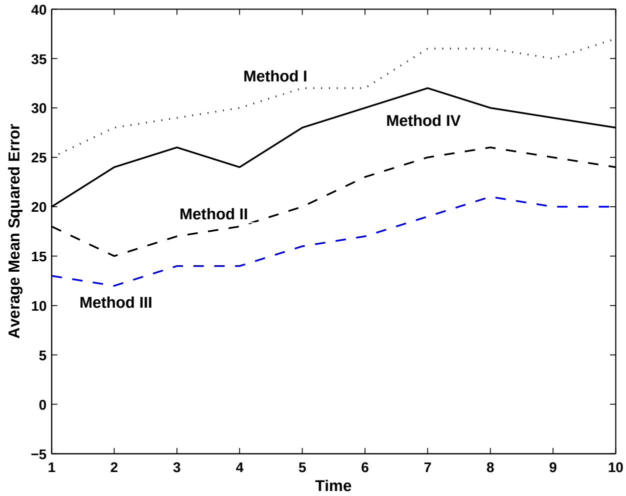


Figure 3.3: Model comparison in terms of average mean squared error (MSE) for Simulation 2. (Method III is the proposed method)

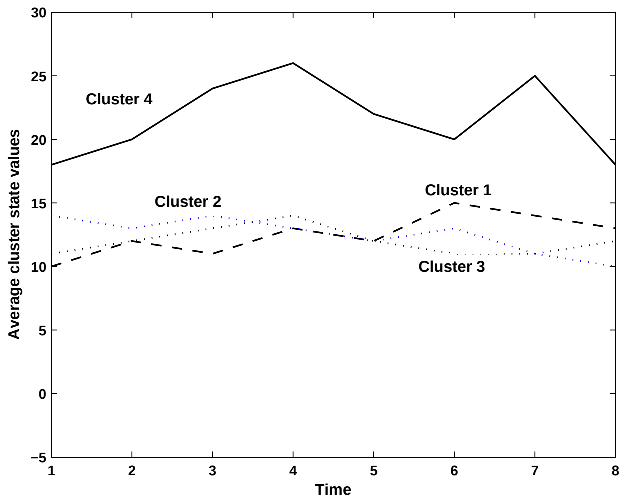


Figure 3.4: Average state values for different clusters in detecting anomalous node: before splitting

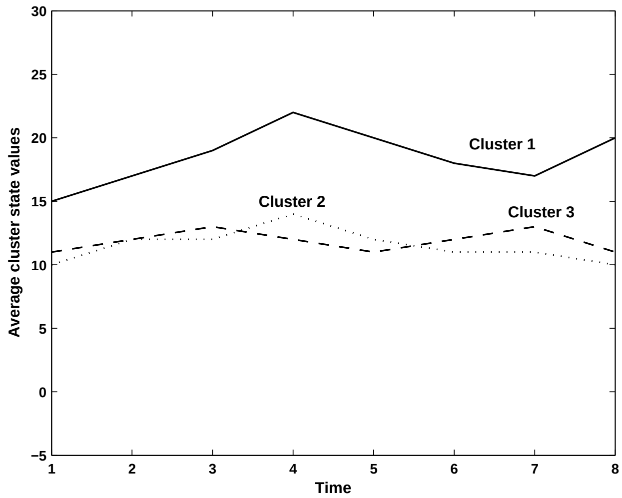


Figure 3.5: Average state values for different clusters in detecting anomalous node: after the first splitting

A network has been considered with 4 clusters containing 10 sensors each. Sensors are measured at 8 different evenly spaced time points. The state values of the i-th sensor belonging to the k-th cluster are simulated from equation (3.17) with the same specification of the function $f_k(t)$, with $k = 1, 2, 3, 4$. The residual errors are generated from Gaussian (0,1.5). Note that a Gaussian (0,1.5) distribution is centred at zero, so this specification considers errors to be small (close to zero). The following values of $\mathbf{b}_k$, $\mathbf{c}_k$ and $\boldsymbol{\theta}_k$, k=1,2,3,4 are taken;

$\mathbf{b}_1 = (1.3, 2.15, 1.07)$, $\mathbf{b}_2 = (1.26, 2.18, 1.35)$, $\mathbf{b}_3 = (1.31, 2.16, 1.08)$, $\mathbf{b}_4 = (1.4, 2.11, 1.03)$.

$\mathbf{c}_1 = (1.5, 0.96)$, $\mathbf{c}_2 = (1.6, 1.12)$, $\mathbf{c}_3 = (1.46, 1.01)$, $\mathbf{c}_4 = (1.8, 1.13)$.

$\boldsymbol{\theta}_1 = (1.2, 2.4)$, $\boldsymbol{\theta}_2 = (1.8, 3.0)$, $\boldsymbol{\theta}_3 = (1.3, 2.3)$, $\boldsymbol{\theta}_4 = (1.5, 2.9)$.

As in Section 3.4.1, one can consider a different set of parameters but the inference will remain unchanged. Here, one set of parameters has been considered for the illustration purpose. Assume that the third sensor in cluster 4 is an anomalous node, and hence, is the object of interest which needs to be identified. The state values are generated for this

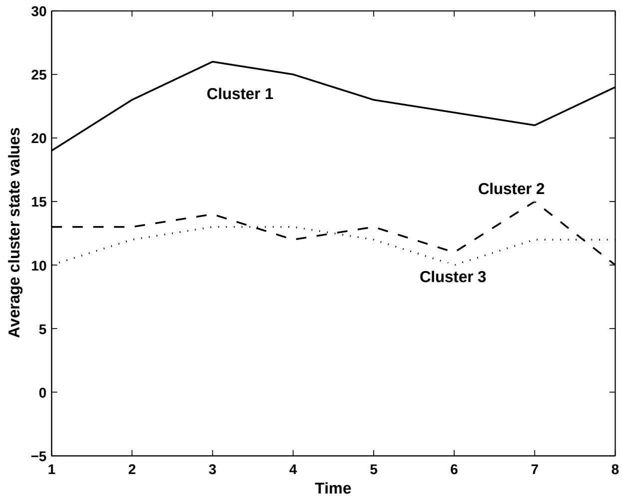


Figure 3.6: Average state values for different clusters in detecting anomalous node: after second splitting

particular sensor node from the following model:

$$X_i(t) = a_0 + a_1 t + a_2 t^2 + 1.76 X_i(t-1) + e_i(t) \tag{3.19}$$

where $a_0 = 2.87, a_1 = 2.25, a_2 = 3.3$ and the residuals are generated from Gaussian (0,10). Note that this specification is arbitrary and one can consider a different set of parameter values and a different Gaussian distribution (with mean=0).

Once the dataset is generated, the proposed model has been fitted as given in equation (3.1) and the proposed prior has been considered in section 3.3. The posterior estimates of the model parameters are obtained by MCMC and then the average state values are computed for each cluster at 8 different time points by simply averaging the estimated state values of all the sensors belonging to a particular cluster at a fixed time point. The author then plots the estimated average trajectory for each cluster as shown in Figure 3.4. It is noted that the average curve of cluster 4 is far away from the average curves of three other clusters. This reflects that cluster 4 contains the possible anomalous node, since the mean value is highly affected by the outliers.

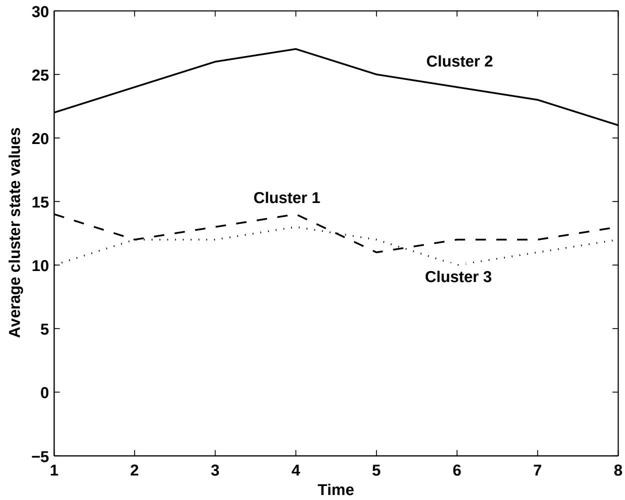


Figure 3.7: Average state values for different clusters in detecting anomalous node: after third and final splitting

Then cluster 4 is split into two sub-clusters; the first sub-cluster contains the first 5 nodes and the second one contains the remaining 5 nodes. Clusters 1,2 and 3 are merged into a single cluster. Thus, now three clusters are left containing 5,5 and 30 sensor nodes, respectively. The estimated model parameters are used and the average mean curves are estimated for these 3 clusters, which is shown in Figure 3.5. It is noticed that the estimated mean curve for cluster 1 (new) behaves differently from the other two mean curves, thus indicating that cluster 1 as the "target" cluster.

Next, cluster 1 is further split into two sub-clusters, the first one containing first three nodes and second one containing the remaining two. These are the newly formed clusters 1 and 2; the new cluster 3 is formed by merging clusters 2 and 3 from the previous step containing 35 nodes in total. Figure 3.6 shows the estimated mean curves and indicates cluster 1 as the outlier.

Finally, cluster 1 is divided into two sub-clusters, the first one contains nodes 1 and 2; the second one contains node 3. New cluster 3, containing 37 nodes, is again obtained by merging clusters 2 and 3 from the previous step. Figure 3.7 shows the average curves and reflects that the cluster 2 is the outlier. Since cluster 2 contains a single node, this particular node is correctly detected as an anomalous node, which is possibly a malicious node.

The False Positive Rate (FPR) [55] is computed (defined in Section 2.4.3) for the proposed method of detecting the anomalous node. 100 such datasets are generated from the above model and the above method is used to detect the anomalous node. The anomalous node is detected correctly for 96 datasets, thus estimated FPR=0.04. Also the average time to detect the outlier node is 186 seconds, which is pretty fast. All computations are performed in the software R which is freely available and efficient for statistical computations.

3.5 Summary

There is a rich literature on the state estimation of WSNs. Most of these works are based on the dynamic state-space model via KF or similar Markov models. In any network, information exchange is inevitable and in this chapter, a Bayesian non-parametric approach has been proposed for addressing this issue in the context of state estimation of WSNs.

- A cluster-based WSN has been considered and a discrete-time linear Markov model is proposed for estimating the state values of the sensor nodes over time. The proposed linear model considers effect of nearest neighbours on the current state value for each sensor node and thus allows information exchange within a cluster.

- Non-parametric MSBP priors has been considered for the cluster-specific model parameters. Such priors allow information exchange among the sensor nodes belonging to different clusters through the model parameters.

- The usefulness of the proposed model in locating an immobile anomalous node in the network is demonstrated.

Simulation studies are performed to assess the performance of the proposed model. The proposed approach will be useful in emergency monitoring, medical genetics, geosciences and many other disciplines where WSNs are frequently used for state estimation.

Chapter 4

TIME SYNCHRONIZATION IN WIRELESS SENSOR NETWORKS

4.1 Preamble

Applications of Wireless Sensor Networks (WSNs) are found in various disciplines including environmental studies, medical genetics, emergency monitoring etc. [66-69]. All these applications assume that the sensor nodes are synchronized to a common clock. Hence, time synchronization of the nodes in the network plays a key role for accurate state estimation and/or predictions of the surrounding environment. If different sensors in the network run according to their own clocks, then coordination among the sensors are affected and consequently the inference becomes inconsistent [70]. Time synchronization is also very important in randomly deployed sensor networks (e.g. hierarchical topology) because data transmission from leaf nodes to the sink in this type of network takes place in a multihop manner [71].

In recent years, many useful protocols have been proposed for time synchronization in WSNs; e.g. Reference-Broadcast Synchronization (RBS), Time-synchronization Protocol for Sensor Networks (TPSN), Flooding Time Synchronization Protocol (FTSP), Receiver-only Synchronization (ROS) etc. Consider a pair of sensor nodes within a network, say node A and node B, let us denote the time of the clocks in the sensor nodes A and B by

$C_A(t)$ and $C_B(t)$, respectively as the functions of time (t). Relative clock offset between the node A and node B at time t is defined as $C_A(t) - C_B(t)$. The derivative (the rate of variation) of the relative clock offset gives the relative clock skew. Thus the relative clock skew between node A and node B is given by $C'_A(t) - C'_B(t)$. A pair of nodes are synchronized by estimating the relative clock-offset and relative skew between the nodes. Statistical models have been developed where the offset and skew parameters are estimated by Least Squares (LS) method using a regression model as proposed in [30,31]. When a pair of sensor nodes exchange timing message, a group of neighbouring sensors overhear those messages and synchronize themselves accordingly [40]. Methods have been proposed for recursive clock skew estimation for WSNs using reference broadcasts [72], time synchronization in WSNs using max and average consensus protocol [73]. A Bayesian approach of time synchronization for a complex network model has been proposed in [74]. Protocols have been proposed for synchronization in more complex network structures [32,75]. Some recent approaches for time synchronization under different protocols are reviewed in [76,77]. Recent advancements in time synchronization protocol lead to design of different mechanisms to overcome the security issues in FTSP [32] .

All the above approaches assume that the set of time readings between a pair of nodes are uncorrelated, which might not be true in real applications. Since the readings are taken from the same pair of nodes at different time points, it is expected that these time readings will be correlated over time. In this chapter, the author considers an auto-regressive dependence among the time readings and synchronize the pair of nodes by estimating their relative clock offset and relative clock skew using Generalized Least Squares (GLS) method. The advantage of the proposed approach is not only in its accuracy and precision but also in its robustness in estimating the model parameters for correlated time readings.

When the successive time readings are correlated, GLS approach is an iterative method of parameter estimation, which is computationally expensive. The authors proposes an alternative Bayesian approach where some prior distributions are assumed for the model parameters and then the joint posterior distribution is computed. From this joint posterior, the author derives the full conditional density for each of the model parameters and use Markov Chain Monte Carlo (MCMC) to estimate the model parameters. Posterior estimates are indeed the weighted average of prior mean and the Maximum Likelihood Estimates (MLE) based on the data. Also, MCMC is a much faster algorithm than GLS and the author compares two methods in terms of their respective CPU times.

This chapter has two main contributions:

- The author proposes a new approach of estimating model parameters using GLS method for time synchronization in WSNs when the successive time readings are correlated.
- Further, a Bayesian model is developed where the parameters are estimated by MCMC. This approach is computationally fast and provides efficient estimates.

[1] The findings of this chapter has been published in a research paper.

4.2 Receiver-Only Synchronization (ROS) Method

The proposed approach is discussed in the context of Receiver-Only Synchronization [30]. However, it may be noted that this approach can also be implemented in the context of other synchronization approaches e.g. TPSN, RBS, FTSP etc.

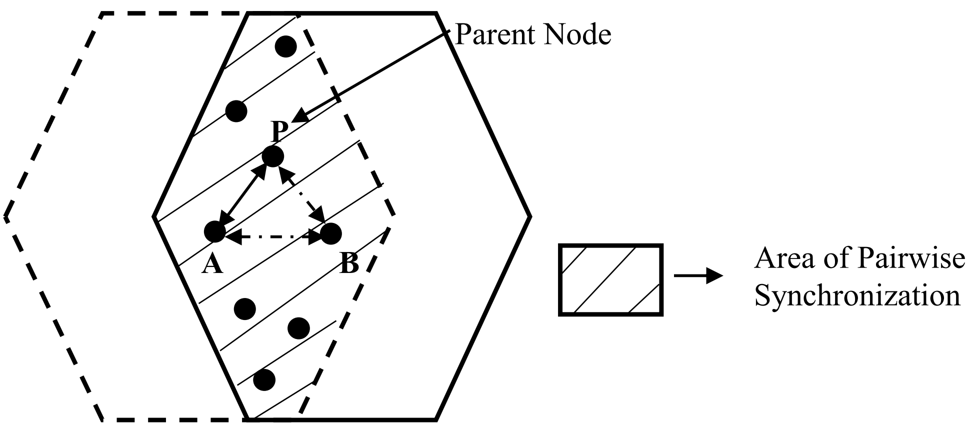


Figure 4.1: A pictorial representation of Receiver-Only Synchronization

In Figure 4.1, we consider two nodes P and A, where node P is the parent node.

[1]A. Chatterjee, and P. Venkateswaran, "An efficient statistical approach for time synchronization in wireless sensor networks", International Journal of Communication Systems, Vol 29, Issue 4, March 2016.

Communication range of each node is limited to a pre-specified area (represented by solid and dotted area respectively) and if a node belongs to the checked area, it can receive message from both A and P. A pairwise synchronization is performed between node P and node A and all other nodes belonging to the checked area receive a series of synchronization messages. Since all the neighbouring sensors are synchronized by simply overhearing the timing messages, ROS becomes an energy efficient synchronization protocol.

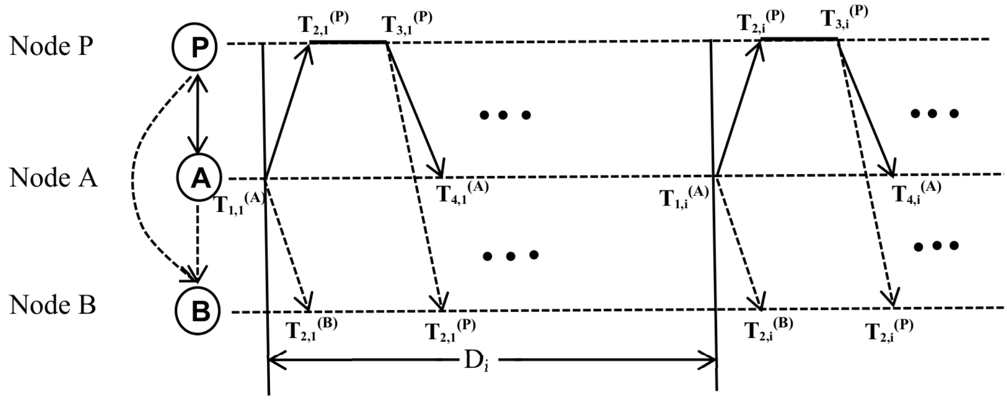


Figure 4.2: Model for timing message exchange in ROS

As shown in Figure 4.2 [30], node B which belongs to the checked area, receives a set of N time readings $\left(\left\{T_{2,i}^{(B)}\right\}_{i=1}^{N}\right)$ from node A and additionally receives another set of N readings $\left(\left\{T_{2,i}^{(P)}\right\}_{i=1}^{N}\right)$ from node P. Considering the relative clock offset and relative clock skew between A and P, $T_{2,i}^{(P)}$ can expressed as the following:

$$T_{2,i}^{(P)} = T_{1,i}^{(A)} + \theta_{offset}^{(AP)} + \theta_{skew}^{(AP)}(T_{1,i}^{(A)} - T_{1,1}^{(A)}) + d^{(AP)} + X_i^{(AP)} \tag{4.1}$$

where $\theta_{offset}^{(AP)}$ and $\theta_{skew}^{(AP)}$ denote the relative clock offset and relative clock skew, respectively between A and P. The fixed and random delays in the readings are denoted by $d^{(AP)}$ and $X_i^{(AP)}$. Similarly,

$$T_{2,i}^{(B)} = T_{1,i}^{(A)} + \theta_{offset}^{(AB)} + \theta_{skew}^{(AB)}(T_{1,i}^{(A)} - T_{1,1}^{(A)}) + d^{(AB)} + X_i^{(AB)} \tag{4.2}$$

Using the above two equations, one can write,

$$T_{2,i}^{(P)} - T_{2,i}^{(B)} = \theta_{offset}^{(BP)} + \theta_{skew}^{(BP)}(T_{1,i}^{(A)} - T_{1,1}^{(A)}) + d^{(AP)} - d^{(AB)} + X_i^{(AP)} - X_i^{(AB)} \tag{4.3}$$

It is assumed that the random error component in equation (4.3), $X_i^{(AP)} - X_i^{(AB)}$ follows Gaussian distribution with mean=0 and variance=σ^2. Note that in this assumption, the author consider the possible correlation between $X_i^{(AP)}$ and $X_i^{(AB)}$.

Let $w_i = X_i^{(AP)} - X_i^{(AB)}$, $\mu = d^{(AP)} - d^{(AB)}$, $x_i = T_{2,i}^{(P)} - T_{2,i}^{(B)} - \mu$.

Then equation (4.3) can be expressed as:

$$x_i = \theta_{offset}^{(BP)} + \theta_{skew}^{(BP)}(T_{1,i}^{(A)} - T_{1,1}^{(A)}) + w_i \tag{4.4}$$

Noh et al. [30] used equation (4.4) to estimate $\theta_{offset}^{(BP)}$ and $\theta_{skew}^{(BP)}$ using least squares method in a classical regression framework. This approach assumes that w_i's are identically and independently distributed random variables distributed as Gaussian $(0, \sigma^2)$ and consequently x_i's are uncorrelated. In reality, for WSNs one gets different state values with time stamps, which are not uncorrelated. Thus x_i's are correlated over different time points and this correlation needs to be modelled for an efficient and consistent estimation of the clock offset and clock skew. The author models this correlation by considering autoregressive model for w_i's as proposed:

$$w_i = \rho w_{i-1} + \nu_i \tag{4.5}$$

where ν_i's are Gaussian $(0, \tau^2)$.

Thus, from equation (4.4),

$$x_i - \rho x_{i-1} = \theta_{offset}^{(BP)}(1-\rho) + \theta_{skew}^{(BP)}\left[(T_{1,i}^{(A)} - T_{1,1}^{(A)}) - \rho(T_{1,i-1}^{(A)} - T_{1,1}^{(A)})\right] + (w_i - \rho w_{i-1}) \tag{4.6}$$

Substituting equation (4.5) in equation (4.4),

$$x_i^* = \theta_{offset}^* + \theta_{skew}^{(BP)}(T_{1,i}^{(A*)} - T_{1,1}^{(A*)}) + \nu_i \tag{4.7}$$

where $x_i^* = x_i - \rho x_{i-1}$, $\theta_{offset}^* = \theta_{offset}^{(BP)}(1-\rho)$, $T_{1,i}^{(A*)} = T_{1,i}^{(A)} - \rho T_{1,i-1}^{(A)}$, and $T_{1,1}^{(A*)} = T_{1,1}^{(A)}(1-\rho)$.

It is noted that equation (4.7) is a standard regression model where the residuals ν_i's are Gaussian $(0, \tau^2)$. Hence, the ordinary least squares method can be applied for estimating θ_{offset}^* and $\theta_{skew}^{(BP)}$. One needs to know ρ to carry out the analysis, but typically ρ is unknown. Hence one has to get the initial estimate of ρ from the OLS and use an iterative approach to estimate it. Here, the algorithm suggested by Cochrane-Orcutt [78] is followed for estimating ρ as outlined below:

1. Estimate the parameters in equation (4.4) by OLS and get the residual estimates $\hat{w}_i$.
2. Regress the residual $\hat{w}_i$ on the lagged residual $\hat{w}_{i-1}$ using equation (4.5) and get the estimate of ρ as $\hat{\rho}$.
3. Use $\hat{\rho}$ to get x_i^*, $T_{1,i}^{(A*)}$ and $T_{1,1}^{(A*)}$.
4. Regress x_i^* on $T_{1,i}^{(A*)} - T_{1,1}^{(A*)}$ using equation (4.7) and generate new residuals from here.
5. Use equation (4.5) again to get new estimate of ρ, say, $\hat{\rho}_{new}$.
6. Go to step 3 and repeat until the difference between the estimates from two consecutive iterations is arbitrarily small.

4.3 Simulation Results

4.3.1 Performance of the proposed Approach for Correlated Data

The performance of the proposed approach of estimating the relative clock-skew and clock-offset is compared to the one proposed in [30] through simulation studies. It is noted that one can consider any set of fixed parameter values for the simulation study and can illustrate the usefulness of the proposed model and estimation method. However, the author only shows the results from one set of prefixed parameter values. The author simulates w_i's from equation (4.5) with ρ=0.65 and τ^2=0.2. Then using equation (4.4), x_i's are simulated with $\theta_{offset}^{(BP)}$=1.45, $\theta_{skew}^{(BP)}$=0.98 and σ^2=0.1. Simulations are performed for many different values of time readings N. For each fixed N, 100 such datasets are replicated. For each simulated dataset, the model parameters are estimated by the proposed approach (using equation (4.7)).

For performance comparison, the model parameters are estimated using equation (4.4) as proposed by Noh et al. [30]. For both the approaches, the author calculates the Mean Squared Error (MSE) using the following formula:

$$\text{MSE}= \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x_i})^2$$

where $\hat{x_i}$ denotes the estimated value of x_i. Then the average MSE is calculated by averaging the MSE values for 100 replicated datasets for each N.

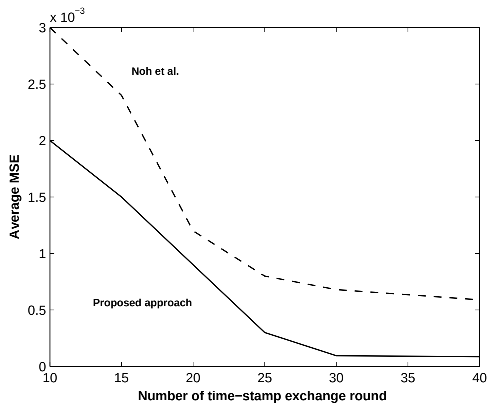


Figure 4.3: Performance of the proposed approach compared to Noh et al. [30] in terms of Average MSE for correlated data

In Figure 4.3, the author shows the simulation results for different values of N. It is noted that the proposed approach provides uniformly lower average MSE than the approach proposed by Noh et al. [30]. Thus the effectiveness of the proposed approach is assessed and verified for the case when the random components in equation (4.3) are correlated over different time points. The estimated value of ρ is 0.631 with Standard Error (SE)=0.24.

In Table 4.1, the author shows the model parameter estimates with the SE for both the approaches. Note that the proposed approach provides the estimates closer to the true parameter values with smaller SE than the other approach. Thus, accuracy (less bias) and precision (less variation) are gained in parameter estimates by the proposed algorithm compared to the one proposed in [30] for the correlated measurements.

Table 4.1: Parameter estimates and Standard Errors (SE) for the proposed approach and Noh et al. [30] for correlated data

Parameter	True value	Proposed approach Estimates (SE)	Noh et al. [30] Estimates (SE)
$\theta^{(BP)}_{offset}$	1.45	1.43(0.89)	1.26(1.17)
$\theta^{(BP)}_{skew}$	0.98	0.94(1.02)	1.13(1.25)
σ^2	0.1	0.13(0.27)	0.18(0.43)

4.3.2 Performance of the Proposed Approach for Uncorrelated Data

The author performs another simulation study for investigating the effectiveness of the proposed approach compared to the other approach [30] when the random components in equation (4.3) are not correlated over time. Again, the results are shown only for one set of fixed parameter values, but the result can be demonstrated with any fixed set of model parameter values. The author simulates w_i's which are identically and independently distributed as Gaussian (0,0.25) and then using equation (4.4), simulates the x_i's for $\theta^{(BP)}_{offset}$=1.28, $\theta^{(BP)}_{skew}$=0.85. Similar to the earlier simulation study, here also the author simulates data for many different values of N time readings. Here 100 datasets are replicated for each N and for each dataset the model parameters are estimated by the proposed approach using equation (4.7) and also for the approach suggested in [30]. For each N, average MSE is calculated by averaging over 100 replicated datasets.

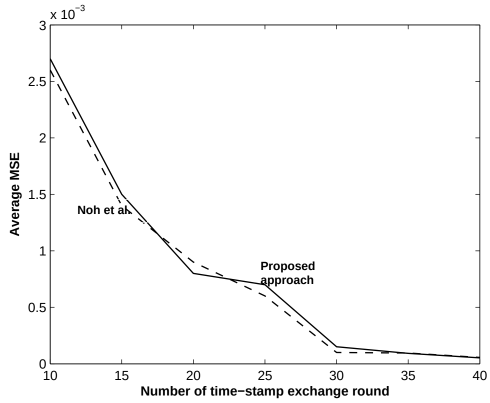


Figure 4.4: Performance of the proposed approach compared to Noh et al. [30] in terms of Average MSE for the uncorrelated data

In Figure 4.4, the author shows the results of the simulation study. Note that solid curve indicates the average MSE as per the proposed approach and the other one as per the approach proposed in [30]. It is observed that the average MSE from the proposed approach for almost all values of time reading N closely follows the average MSE from the approach proposed in [30]. Thus this simulation study demonstrates the effectiveness of the proposed estimation approach for uncorrelated data as well. From the simulation study, the estimated value of ρ is 0.0021 with SE 0.17, thus indicating the successive measurements are almost uncorrelated.

Table 4.2: Parameter estimates and standard errors for the proposed approach and the approach by Noh et al. [30] for uncorrelated data

Parameter	True value	Proposed approach Estimates (SE)	Noh et al. Estimates (SE)
$\theta^{(BP)}_{offset}$	1.28	1.32(1.19)	1.31(1.15)
$\theta^{(BP)}_{skew}$	0.85	0.84(1.08)	0.83(1.03)
σ^2	0.25	0.28(0.48)	0.23(0.43)

In Table 4.2, the author shows the estimated model parameters and the corresponding SE for both the approaches. Here both the approaches are quite comparable although the approach by Noh et al. [30] provides slight improvement in estimation. However, the improvement in estimation as per the approach in Noh et al. [30] is not much significant as noted in Table 4.2. Thus, even for the uncorrelated data, proposed algorithm works quite well.

The results from the first simulation study shows that for the correlated time readings, the proposed approach performs extremely well in comparison to the approach proposed by Noh et al. [30]. The results from the second simulation study shows that for the uncorrelated time readings the proposed approach performs almost similar to the approach proposed by Noh et al. [30]. Thus, the robustness of the proposed approach is assessed through simulation studies and hence in practice, the proposed approach would be a better method for the parameter estimation.

4.4 Parameter Estimation using Bayesian Approach and MCMC

4.4.1 Model and Priors

Now an alternative Bayesian approach is proposed to estimate the parameters in equation (4.7) using MCMC. The proposed approach is quite similar to the one in [79], and [80] who proposed Bayesian inference in standard regression models.

Note that the author rewrites equation (4.4) as the following:

$$x_i = Z_i\boldsymbol{\beta} + w_i \tag{4.8}$$

where $Z_i = \left[1, T_{1,i}^{(A)} - T_{1,1}^{(A)}\right]$ and $\boldsymbol{\beta} = \left[\theta_{offset}^{(BP)}, \theta_{skew}^{(BP)}\right]^T$.

Note that w_i's follow equation (4.5) and ν_i's are Gaussian $(0, \tau^2)$.

The unconditional density of x_i corresponding to the very first time reading (i=1) can be written as:

$$f(x_i|\boldsymbol{\beta}, \rho, \tau^2) = \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{(1-\rho^2)}{2\tau^2}(x_i - Z_i\boldsymbol{\beta})^2\right] \tag{4.9}$$

Also one can write the conditional density of x_i given all the past reading $x_1, \ldots, x_{i-1}$ as

$$f(x_i|x_1, x_2, \ldots, x_{i-1}, \boldsymbol{\beta}, \rho, \tau^2) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2\tau^2}\left\{(x_i - \rho x_{i-1}) - (Z_i - \rho Z_{i-1})\boldsymbol{\beta}\right\}^2\right] \tag{4.10}$$

Thus, the joint likelihood function of $x_1, \ldots, x_N$ can be expressed as the following:

$$f(x_1, x_2, \ldots, x_N|\boldsymbol{\beta}, \rho, \tau^2) = (2\pi\tau^2)^{-\frac{N}{2}}(1-\rho^2)^{\frac{1}{2}} \exp\left[-\frac{1}{2\tau^2}\sum_{i=1}^{N}(x_i^* - Z_i^*\boldsymbol{\beta})^2\right] \tag{4.11}$$

where x_i^* and Z_i^* are defined as the following:

$$x_i^* = \begin{cases} \sqrt{1-\rho^2}x_i, & i = 1 \\ x_i - \rho x_{i-1}, & i = 2, \ldots, N \end{cases}$$

and

$$Z_i^* = \begin{cases} \sqrt{1-\rho^2} Z_i, & i = 1 \\ Z_i - \rho Z_{i-1}, & i = 2, \ldots, N \end{cases}$$

Note that here one has to estimate $\boldsymbol{\beta}, \rho$ and τ^2 for which one needs to assume some prior densities. The author considers a non-informative prior for $\boldsymbol{\beta}$, Uniform[-1,1] prior for ρ and Inverse Gamma (a, b) prior for τ^2. One can also assume an informative prior (for example, normal prior) for $\boldsymbol{\beta}$ when such prior knowledge is available.

Here the prior structure is as the following:

$$\pi(\boldsymbol{\beta}) \propto \text{constant},$$
$$\pi(\rho) = \text{Uniform}(-1, 1),$$
$$\pi(\tau^2) \propto (\tau^2)^{-a-1} \exp\left(-\tfrac{b}{\tau^2}\right)$$

4.4.2 Posterior and Full Conditionals

Combining the above priors and the joint likelihood function given in equation (4.11), one gets the following joint posterior density:

$$\pi(\boldsymbol{\beta}, \rho, \tau^2 | X) \propto (\tau^2)^{-\frac{N}{2}-a-1} (1-\rho^2)^{\frac{1}{2}} \exp\left[-\frac{1}{2\tau^2} \sum_{i=1}^{N} (x_i^* - Z_i^* \boldsymbol{\beta})^2\right] \exp\left(-\frac{b}{\tau^2}\right) I_{[-1,1]}(\rho) \tag{4.12}$$

where $X = (x_1, x_2, \ldots, x_N)$.

From the above joint posterior distribution, the author derives the full conditional distribution for each of the model parameters following Robert and Casella [25], Carlin and Louis [81], Das et al. [56]. The full conditional densities are given as follows:

$$\pi(\boldsymbol{\beta}|\rho, \tau^2) \sim N\left(\boldsymbol{\beta}^*, \left(\tfrac{1}{\tau^2} \sum_{i=1}^{N} Z_i^{*T} Z_i^*\right)^{-1}\right); \text{ where } \boldsymbol{\beta}^* = \left(\sum_{i=1}^{N} Z_i^{*T} Z_i^*\right)^{-1} \left(\sum_{i=1}^{N} Z_i^{*T} x_i^*\right)$$

$$\pi(\rho|\boldsymbol{\beta}, \tau^2) \propto (1-\rho^2)^{\frac{1}{2}} \exp\left[-\tfrac{1}{2\tau^2} \sum_{i=1}^{N} (x_i^* - Z_i^* \boldsymbol{\beta})^2\right] I_{[-1,1]}(\rho)$$

$$\pi(\tau^2|\boldsymbol{\beta}, \rho) \sim IG\left(\tfrac{N}{2} + a, b + \tfrac{1}{2} \sum_{i=1}^{N} (x_i^* - Z_i^* \boldsymbol{\beta})^2\right)$$

It is noted that for $\boldsymbol{\beta}$ and τ^2, the full conditionals are known densities and hence Gibbs sampler can be used for posterior sampling. However for ρ, one has to use Metropolis-Hastings algorithm for sampling [80].

4.4.3 Performance Evaluation

In this section, the author evaluates the performance of the proposed approach using Bayesian method through simulation studies for the correlated data. Simulation has been carried out using equation (4.4) with ρ=0.65, τ^2=0.2, $\theta^{(BP)}_{offset} = 1.45$ and $\theta^{(BP)}_{skew} = 0.98$. Note that any fixed set of model parameters will provide the results, but the author considers only one set of parameters for the current presentation. The author considers N=20 and uses the simulated data to fit into the proposed Bayesian model as given in equation (4.8). Then, the model parameters ρ, τ^2, and $\theta^{(BP)}_{offset}$ are estimated by a hybrid combination of Gibbs sampler and Metropolis-Hastings (MH) algorithm [80] as outlined below to run MCMC.

(1) Start with some initial values of the model parameters, $\boldsymbol{\beta}_0, \rho_o$ and τ_0^2.

(2) At the j-th iteration, (i.e. for j=1,2,3,...) simulate the parameters from their respective posterior densities as follows:

(a) Simulate $\boldsymbol{\beta}_j$ from $\pi(\boldsymbol{\beta}_j|\rho_{j-1}, \tau^2_{j-1})$ which is the full conditional distribution of $\boldsymbol{\beta}_j$ given all the other parameters.

(b) Simulate ρ_j from $\pi(\rho_j|\boldsymbol{\beta}_j, \tau^2_{j-1})$. MH algorithm is used for this simulation. Generate ρ^* from Uniform[-1,1].

Let us define 'acceptance probability'=$\omega(\rho_{j-1}, \rho^*)$ as the following:

$\omega(\rho_{j-1}, \rho^*) = min\left[1, \frac{\pi(\rho^*|\boldsymbol{\beta}_j, \tau^2_{j-1}, X)}{\pi(\rho_{j-1}|\boldsymbol{\beta}_j, \tau^2_{j-1}, X)}\right]$.

Set

$$\rho_j = \begin{cases} \rho^*, & \text{with probability} = \omega(\rho_{j-1}, \rho^*) \\ \rho_{j-1}, & \text{otherwise.} \end{cases}$$

(c) Next, simulate τ_j^2 from $\pi(\tau_j^2|\boldsymbol{\beta}_j, \rho_j)$. The full conditional distribution of τ_j^2 is an Inverse

Gamma distribution.

(3) Discard the first 2,000 'burn-in' iterations to remove the effect of the initial values of the parameters.

(4) Assess the convergence of the chains graphically as suggested by Brooks and Gelman [36].

(5) Use posterior means as the respective parameter estimates.

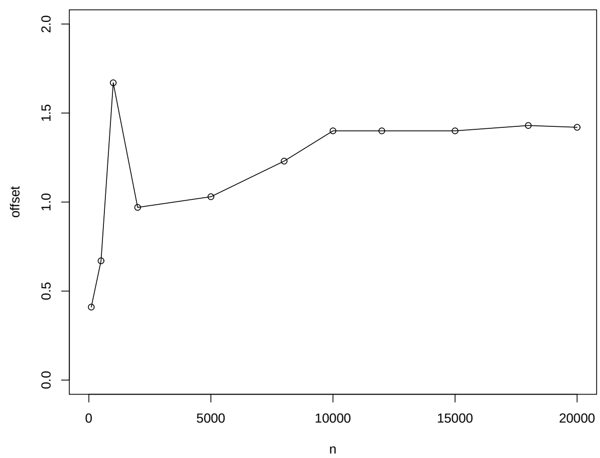


Figure 4.5: MCMC estimates of the offset parameter for different number of iterations (n)

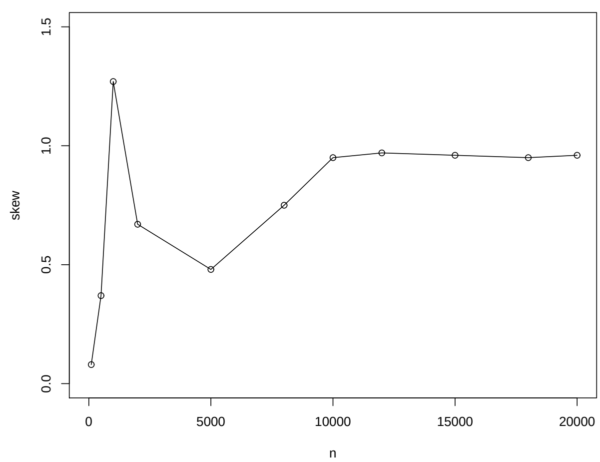


Figure 4.6: MCMC estimates of the skew parameter for different number of iterations (n)

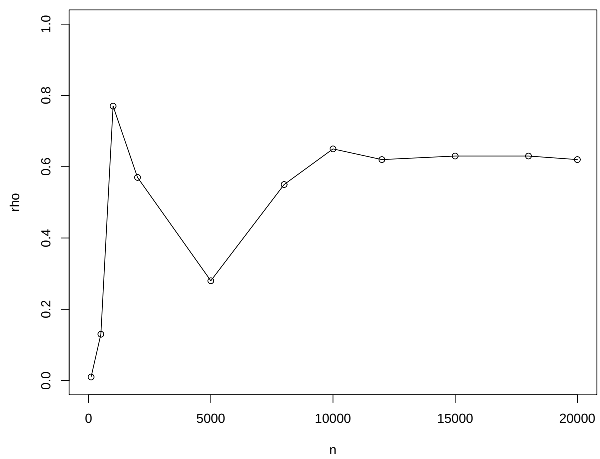


Figure 4.7: MCMC estimates of rho for different number of iterations (n)

20,000 chains are run after removing the 'burn-in' iterations. The estimates of the parameters for different number of the iteration (n) are shown in Figures 4.5 to 4.8, and it is noticed that for n=10,000 onward, the estimates are quite stable. In Table 4.3, the author shows the parameter estimates with SEs for the correlated data obtained from Simulation 1 as described in Section 4.3.1, for Bayesian and GLS approach. It can be observed that in terms of the accuracy and precision of the estimates, both the approaches are equally effective. Thus the Bayesian approach provides very similar estimates to those obtained from GLS technique.

Table 4.3: Parameter estimates and standard errors for the Bayesian approach and GLS approach for Simulation 1

Parameter	True value	GLS approach Estimates (SE)	Bayesian approach Estimates (MCSE)
$\theta^{(BP)}_{offset}$	1.45	1.43(0.89)	1.42(0.77)
$\theta^{(BP)}_{skew}$	0.98	0.94(1.02)	0.96(1.14)
τ^2	0.2	0.24(0.78)	0.17(0.86)
ρ	0.65	0.631(0.24)	0.628(0.29)

In Table 4.4, the author compares the performance of both the proposed Bayesian approach and the GLS approach for correlated data in terms of their respective CPU times. For different values of N, it is observed that the CPU times for the Bayesian approach

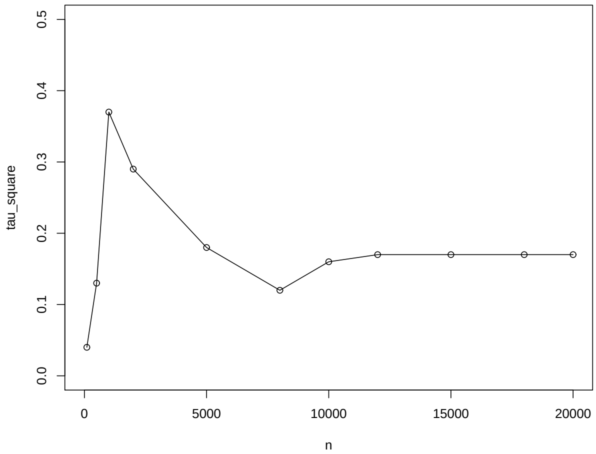


Figure 4.8: MCMC estimates of tau-square for different number of iterations (n)

Table 4.4: Comparison of the Bayesian approach and GLS approach in terms of CPU times

	CPU times (seconds)	
N	Bayesian approach	GLS
5	53	72
10	54	78
20	62	103
25	62	108
30	65	118
40	68	125

is much lower than the GLS approach. Although both the approaches are comparable in terms of accuracy and precision of the estimates, Bayesian approach is much faster than GLS method. Note that in Bayesian approach, all the parameters are updated simultaneously in a single chain, but the other approach is a two-step method in which ρ has to be updated before updating all other parameters. All computations are performed using R (in WINDOWS 8 Intel Core i7 Processor), because it is free software, and there are in-built functions for OLS, GLS, and MCMC.

4.4.4 Model Comparison

Finally, the author compares the performance of the proposed MCMC based Bayesian approach with the one proposed in Noh et al. [30] through simulation study. Similar to the Simulation 1 in Section 4.3, the author first simulates w_i's using equation (4.5) with ρ=0.45 and τ^2=0.34. It is again note that the choice of the parameter values does not matter for the final results, and the author only considers one set of model parameters for the illustration purpose. Using equation (4.4) with $\theta^{(BP)}_{offset}$=1.70, $\theta^{(BP)}_{skew}$=1.04 and σ^2=0.40, the author simulates x_i's. Consider N=30 for this simulation study.

First, the parameters are estimated using the simple linear regression as suggested in Noh et al. [30]. OLS method is used for such estimation. Once the parameters are estimated, the author calculates the estimated bias as the difference between the actual and estimated parameter value, i.e. bias = (Estimated value - Actual value). Then the proposed MCMC based approach is used for parameter estimation and calculate the estimated bias.

100 replicated datasets are considered, each with N=30, and calculate the bias for each dataset. Finally, the author calculates the Average Bias for the parameters in all 100 replications and also computes the SE of the parameter estimates. In Table 4.5, the author shows the average bias and the estimated SE for the model parameters. It can be observed that the proposed approach provides estimates with lower average bias and smaller SE in comparison to the method given in Noh et al. [30]. Thus even for moderately correlated time readings (ρ <0.5), this approach is much better than the other method. Hence for highly correlated time readings, proposed method will perform better than the one proposed in Noh et al. [30].

Table 4.5: Average Bias and Standard Error for the model parameters from MCMC and the method by Noh et al.[30]

	MCMC		Noh et al.[30]	
Parameter	Average Bias	Standard Error	Average Bias	Standard Error
$\theta^{(BP)}_{offset}$	0.23	1.05	0.67	1.39
$\theta^{(BP)}_{skew}$	0.17	0.97	0.58	1.26
σ^2	0.24	1.12	0.72	1.54

4.5 Summary

In this chapter, the author has proposed a powerful method for estimating the model parameters for time synchronization in WSNs. Joint estimation of clock-offset and clock-skew has been proposed in the literature using the simple regression framework. Here it has been shown that simple regression poorly estimates the parameters due to the inherent correlation among successive time readings between two sensor nodes.

- An alternative autoregressive model is proposed and GLS method is used for estimating the relative offset and skew parameters.

- An alternative Bayesian approach is also proposed for the parameter estimation considering correlated readings between two sensors. The effectiveness of the proposed approach in comparison to the existing approach has been investigated.

Chapter 5

CONCLUSION

Nowadays Wireless Sensor Networks (WSNs) are used in a variety of applications. For a long time, WSNs have been used for intrusion detection in military surveillance application [82-86], for parameter monitoring and information gathering in health care applications [87-93] and agricultural studies [94-100]. In military applications, sensor nodes sense the environment and alert the forces accordingly. In SensorScope Project reported in [99], WSNs are used for gathering massive data from environment monitoring. The sensor nodes measure air temperature, wind speed, humidity, wind direction, soil water content etc. over time and send it to the base station where the collected raw data are processed and used for future prediction of environment. In health care system [93], sensor nodes collect data on different parameters e.g. heart rate, blood pressure etc. and send to the base station. At the base station, statistical models are used for data analysis and prediction of the future condition of the patient. In agricultural studies [100], WSNs are used for monitoring soil conditions including water content, mineral content, salinity, soil temperature etc. over time.

In all the applications mentioned above, the general principle is to use appropriate statistical or mathematical models for data analysis and prediction of the future 'states'. In this thesis work, the author has addressed this generic problem, where data are collected from a cluster-based WSN and state estimation is performed using powerful statistical models. Since the sensor nodes collect data dynamically, dynamic statistical models for state estimation has been used. Traditionally, Kalman-Filter (KF) based state-space models are used for such estimation. However, in this thesis, the author has proposed a linear statistical model and estimated the model parameters by Maximum Likelihood (ML) estimation method. Additionally, a Bayesian model where Markov Chain Monte Carlo (MCMC) iterations are used for parameter estimation, has been proposed. Although the application of the proposed approach is demonstrated only in anomaly detection, it may be noted that the proposed method can be effectively used in many other applications as well.

The entire research work carried out in this thesis and presented in different chapters has been summarized below:

In **Chapter 1**, the issues associated with state estimation in a cluster-based discrete-time WSN has been introduced. The basic structure of WSNs, and the state estimation problems are presented through appropriate literature review. KF based state-space models, which are traditionally used for state estimation, are presented, and the limitations of such models are also presented. Further, the basic principles of Bayesian approach which is being followed and developed for different applications has been introduced in this chapter.

In **Chapter 2**, a powerful linear statistical model has been proposed for estimating the state values of the sensor nodes longitudinally and the estimated state values are used for detecting the anomalous nodes in WSNs. Detection of anomalous node in distributed WSNs is extremely important for powerful inference and network reliability.

- The proposed approach is powerful since it considers the effect of the nearest neighbours on the current state values and then detects the anomalous nodes based on the estimated state values.
- Proposed method is energy efficient, since it can also estimate the missing state values of the sensor nodes within the cluster, which are kept in sleep mode for energy conservation.
- Alternative Bayesian model is also proposed which is computationally faster for state

estimation and anomaly detection.

- Performance of all the three models (i) Proposed ML based regression model (ii) Proposed MCMC based Bayesian model and (iii) Traditional KF based state-space model are compared.

- It is found that the proposed MCMC based Bayesian method is more appealing since it is computationally faster and provides the smallest Average Mean Squared Error (AMSE) compared to the other two methods.

Besides anomalous node detection, the proposed model can be effectively used in other applications as well such as security surveillance, pattern recognition, habitat monitoring, etc.

In **Chapter 3**, a Bayesian non-parametric approach for simultaneous state estimation and anomaly detection in WSNs has been proposed. Although, there is a rich literature on the state estimation of WSNs but most of these works are based on the dynamic state-space model via KF or similar Markov models.

- A cluster-based WSN and a discrete-time linear Markov model is considered for estimating the state values simultaneously of all the sensor nodes across the clusters over time. The proposed linear model considers the effect of the nearest neighbours on the current state value for each sensor node and allows information exchange among different sensor nodes within a cluster.

- For allowing the information exchange (through the model parameters) across different clusters in the network, non-parametric Matrix Stick-Breaking Process (MSBP) priors are considered for the cluster-specific model parameters. Further, the similarity of different parameters across the clusters is assessed through simulation studies.

- The author demonstrated the usefulness of the proposed model in locating an immobile anomalous node in the network and computed the time to locate the anomalous object. The False Positive Rate (FPR) of the proposed approach is computed and presented.

- Through simulation studies, the usefulness of the proposed approach has been compared and assessed with the traditional approach i.e. state-space model via KF.

The proposed approach will be useful in emergency monitoring, medical genetics, geosciences and many other disciplines where WSNs are frequently used for decision making based on the estimated state.

In **Chapter 4**, a powerful method for estimating the model parameters for time synchronization in WSNs has been proposed. Joint estimation of clock-offset and clock-skew has been proposed in the literature using the standard regression framework. Here, the author has shown that simple regression using Ordinary Least Square (OLS) poorly estimates the parameters due to the inherent correlation among successive time readings between two sensor nodes.

- An alternative autoregressive model has been proposed where the relative offset and skew parameters are estimated by Generalized Least Squares (GLS) method.
- A computationally efficient Bayesian approach is also proposed for parameter estimation considering correlated readings between two sensors.
- The effectiveness of the proposed approach in comparison to the existing approach is investigated in terms of Standard Error (SE) and CPU times through extensive simulation studies.
- Three methods of estimation: (i) exiting OLS method in Noh et al. [30], (ii) GLS method, and (iii) Proposed Bayesian method, are compared in terms of computational time (CPU time) and AMSE through computer simulations.
- It is observed that Bayesian method outperforms the other two methods since it provides the smallest AMSE, and works much faster.

The advantage of the proposed approach is not only in its accuracy and precision but also in its robustness in estimating the model parameters for correlated time readings.

In view of the current research trends in the state estimation and time synchronization of discrete-time WSNs, the author believes that this thesis is a collection of original research papers that are well documented and finally published in the Engineering journals of international repute.

BIBLIOGRAPHY

[1] M. Winkler, K. D. Tuchs, K. Hughes, and G. Barclay, “Theoretical and practical aspects of military wireless sensor networks”, *Journal of Telecommunications and Information Technology*, Vol. 2, pp. 37–45, 2008.

[2] H. Qian, P. Sun, and Y. Rong, “ Design proposal of self-powered WSN node for battle field surveillance”, *Energy Procedia*, Vol. 16, pp. 753–757, 2012.

[3] M. Nagarajan, and S. Karthikeyan,“ Health Condition Observation with High Level Efficiency in Wireless Sensor Network”, *Wireless Communication*, Vol. 1, No. 5, pp. 183–189, 2009.

[4] J. Shi-yong, W. Fe, B. Jing-fui, and Y. Ting, “Geological disaster monitoring system based on WSN and GSM dual-network integration technology”, *in Proc. of 14th International Conference on Communication Technology (ICCT)*, IEEE, 2012.

[5] Y. Gong, F. Ruan, Z. Fan, T. Li, and J. Wang, “Study on Security Issues in Wireless Sensor Network”, *International journal of Security and Its Applications* Vol. 134, pp. 146–151, 2016.

[6] J. Valverde, V. Rosello, G. Mujica, J. Portilla, A. Uriarte, and T. Riesgo,“ Wireless sensor network for environmental monitoring: application in a coffee factory”, *International Journal of Distributed Sensor Networks*, Vol. 8, No. 1, 18 pages, 2012.

[7] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “ Wireless sensor networks: a survey”, *Computer networks*, Vol. 38, No. 4, pp.393–422, 2002.

[8] G. Anastasi, M. Conti, M.D. Francesco, and A. Passarella, “Energy conservation in wireless sensor networks: A survey”, *Ad hoc networks*, Vol. 7, No. 3, pp. 537–568,

2009.

[9] A. Chatterjee, and D. Mukherjee, "Variety event detection in Wireless Sensor Networks through single hop cluster topology", *Tenth International Conference on Wireless and Optical Communications Networks (WOCN)*, pp. 1–5, IEEE, 2013.

[10] M. Michalik, "Base station for Wireless sensor network", *Diss. Masarykova univerzita*, Fakulta informatiky, 2013.

[11] A. Chatterjee, and D. Mukherjee, "Reliable multipath wireless sensor network routing protocol scheme for network lifetime maximization", *2nd International Conference on Business and Information Management (ICBIM)*, pp. 64–68, IEEE, 2014.

[12] A. Hac, "Wireless sensor network designs", *John Wiley & Sons Ltd*, 2003.

[13] C. Fischione, P. Park, P. D. Marco, and K. H. Johansson,"Design principles of wireless sensor networks protocols for control applications", *Wireless Networking Based Control*, Springer New York, pp. 203–238, 2011.

[14] A. G. Lazaropoulos "Wireless sensor network design for transmission line monitoring, metering, and controlling: introducing broadband over power lines-enhanced network model (BPLeNM)", *ISRN Power Engineering*, Hindawi Publishing Corporation, Vol. 2014, Article ID 894628, 22 pages, 2014.

[15] K. Gupta, and V. Sikka, "Design Issues and Challenges in Wireless Sensor Networks", *International Journal of Computer Applications*, Vol. 112, No. 4, 2015.

[16] J. Zhu, L. Tan, H. Xi, and Z. Zhang, "Reliability analysis of wireless sensor networks using Markovian model", *Journal of Applied Mathematics*, 2012.

[17] M. Mahmood, and K. Winston, "Event reliability in wireless sensor networks", *Seventh International Conference onIntelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, IEEE, 2011.

[18] I. Silva, L. Guedes, P. Portugal, and F. Vasques, "Reliability and availability evaluation of wireless sensor networks for industrial applications", *Sensors*, Vol. 12, No. 1, pp. 806–838, 2012.

[19] A. Dmaso, N. Rosa, and P. Maciel, "Reliability of wireless sensor networks", *Sensors*

Vol. 14, No. 9, pp. 15760–15785, 2014.

[20] X. Zhu, Y. Lu, J. Han, and L. Shi, "Transmission reliability evaluation for wireless sensor networks", *International Journal of Distributed Sensor Networks*, Vol. 12, No. 2, 10 pages, 2016.

[21] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False Data Injection Attacks against State Estimation in Wireless Sensor Networks", *49th IEEE Conference on Decision and Control*, pp. 5967–5972, 2010.

[22] Q. Li, R. G. Mark, and G.D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter", *Physiological measurement*, Vol. 29, No. 1, pp. 15, 2007.

[23] B. Ristic, S. Arulampalam, and N. Gordon, "Beyond the Kalman filter: Particle filters for tracking applications", *Artech house*, 2003.

[24] E.A. Wan, and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation", *In Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC*, pp. 153–158, IEEE, 2000.

[25] C.P. Robert, and G. Casella, "Monte Carlo Statistical Methods", *Springer-Verlag*, 1999.

[26] D. Blackwell, and J.B. MacQueen, "Ferguson distributions via Polya urn schemes", *Annals of Statistics*, Vol.1, no.2, pp. 353–355, March, 1973.

[27] T.S. Ferguson, "A Bayesian analysis of some nonparametric problems", *Annals of Statistics*, Vol.1, no.2, pp. 209–230, March, 1973.

[28] J. Sethuraman, "A constructive definition of Dirichlet priors", *Statistica Sinica*, Vol.4, no.3, pp. 639–650, 1994.

[29] D.B. Dunson, Y. Xue, and L. Carin, "The matrix stick-breaking process: Flexible Bayes meta-analysis", *Journal of American Statistical Association*, Vol.103, no.481, pp. 317–327, 2008.

[30] K.L. Noh, E. Serpedin, and K. Qaraqu, "A new approach for time synchronization in wireless sensor networks: pairwise broadcast synchronization", *IEEE Transactions on Wireless Communications*, Vol.7, no.9, pp. 3318-3322, September, 2008.

[31] K.L. Noh, Q. Chaudhari, E. Serpedin, and B.W. Suter, "Novel clock phase offset and skew estimation using two-way timing message exchanges for wireless sensor networks", *IEEE Transactions on Communication*, Vol.55, no.4, pp. 766–776, April, 2007.

[32] D. Huang, W. Teng, and K. Yang, "Secured flooding time synchronization protocol with moderator", *International Journal of Communication Systems*, Vol. 26, No. 9, pp. 1092–1115, 2013.

[33] M.F. Othman, and K. Shazali, "Wireless sensor network applications: A study in environment monitoring system", *Procedia Engineering* Vol. 41, pp. 1204–1210, 2012.

[34] C.Y. Lin, W.C. Peng, and Y.C. Tseng, "Efficient in-network moving object tracking in wireless sensor networks", *Mobile Computing, IEEE Transaction* Vol. 5, No. 8, pp. 1044–1056, 2006.

[35] T. Banerjee, B. Xie, and D.P. Agrawal, "Fault tolerant multiple event detection in a wireless sensor network", *Journal of parallel and distributed computing* Vol. 68, No. 9, pp. 1222–1234, 2008.

[36] Y.Y. Li, and L.E. Parker, "Intruder detection using a wireless sensor network with an intelligent mobile robot response", *Southeastcon IEEE*, pp. 37–42, 2008.

[37] D. Malan, T. Fulford-Jones, M. Welsh, and S. Moulton, "Codeblue: An ad hoc sensor network infrastructure for emergency medical care", *In International workshop on wearable and implantable body sensor networks*, Vol. 5, 2004.

[38] S. Jin, W. Yue, and Q. Sun, " Performance analysis of the sleep/wakeup protocol in a wireless sensor network", *International Journal of Innovative Computing Information and Control*, Vol. 8, No. 5, pp. 3833–3844, 2012.

[39] A.I. Khan, and P. Mihailescu, "Parallel pattern recognition computations within a wireless sensor network", *In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference*, Vol.1, pp. 777–780, IEEE, August, 2004.

[40] J. Liang, Z. Wang, and X. Liu, "Distributed State Estimation for Discrete-Time Sensor Networks with Randomly Varying Nonlinearities and Missing Measurements", *IEEE Transactions on neural networks* Vol. 22, pp. 486–496, 2011.

[41] J. Xu, and J. Li, "State Estimation with Quantized Measurements in Wireless Sensor Networks", *Proceedings of the 29th Chinese Control Conference*, pp. 4857-4862, 2010.

[42] C. Sun, P. Zhang, and Q. Tang, "Detection and States Estimation of Multiple Mobile Targets in Wireless Sensor Network", *International Conference on Communications and Intelligence Information Security*, pp. 147-151, 2010.

[43] D. Quevedo, A. Ahlen, and K. Johansson, "State Estimation Over Sensor Networks With Correlated Wireless Fading Channels", *IEEE Transactions on Automatic Control*, Vol. 58, pp. 581–593, 2013.

[44] M. Xie, S. Han, B. Tian, and S. Parvin, "Anomaly detection in wireless sensor networks: A survey", *Journal of Network and Computer Applications*, Vol. 34, No. 4, pp. 1302–1325, 2011.

[45] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks", *IEEE Wireless Communications*, Vol. 15, No. 4, pp. 34–40, 2008.

[46] C. O'Reilly, A. Gluhak, M.A. Imran, and S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-stationary environment", *Communications Surveys & Tutorials, IEEE*, Vol. 16, No. 3, pp. 1413–1432, 2014.

[47] O. Salem, A. Guerassimov, A. Mehaoua, A. Marcus, and B. Furht, "Anomaly Detection in Medical Wireless Sensor Networks using SVM and Linear Regression Models", *E-Health and Telemedicine: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, Vol. 23, 2015.

[48] B. Sun, X. Shan, K. Wu, and Y. Xia", "Anomaly detection based secure in-network aggregation for wireless sensor networks", *Systems Journal, IEEE*, Vol. 7, No. 1, pp. 13–25, 2013.

[49] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, "Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks", *International Conference on Communications, ICC'07*, pp. 3864–3869, IEEE, 2007.

[50] X. Wang, M. Fu, and H. Zhang, "Target Tracking in Wireless Sensor Networks Based on the Combination of KF and MLE Using Distance Measurements", *IEEE Transactions*

on mobile computing, Vol. 11, pp. 567-576, 2012.

[51] A. Roy, P. Kar, S. Misra, and M.S. Obaidat, "D3: distributed approach for the detection of dumb nodes in wireless sensor networks", *International Journal of Communication Systems*, Vol. 30, No. 1, 2015

[52] L. Wu, P. Shi, and H. Gao, "State Estimation and Sliding-Mode Control of Markovian Jump Singular Systems", *IEEE Transactions on Automatic Control*, Vol. 55, pp. 1041–1046, 2010.

[53] M. Di, E. M. Joo, and L.H. Beng, "A comprehensive study of Kalman filter and extended Kalman filter for target tracking in Wireless Sensor Networks", *IEEE International Conference on Systems, Man and Cybernetics*, pp. 2792–2797, 2008.

[54] A. Chatterjee, and P. Venkateswaran, "An efficient statistical approach for time synchronization in wireless sensor networks", *International Journal of Communication Systems*, Vol. 29, No. 4, pp. 722–733, 2016.

[55] P. Bose, H. Guo, E. Kranakis, A. Maheshwari, P. Morin, J. Morrison, M. Smid, and Y. Tang, "On the false-positive rate of Bloom filters", *Information Processing Letters*, Vol. 108, No. 4, pp. 210-213, 2008.

[56] K. Das, R. Li, S. Sengupta, and R. Wu, " A Bayesian Semi-parametric Model for Bivariate Sparse Longitudinal Data", *Statistics in Medicine*, Vol. 32, No. 22, pp. 3899–3910, 2013.

[57] K. Das, and M. J. Daniels, "A semiparametric approach to simultaneous covariance estimation for bivariate sparse longitudinal data", *Biometrics*, Vol. 70, No. 1, pp. 33–43, 2014.

[58] I. Nevat, G.W. Peters, F. Septier, and T. Matsui, "Estimation of Spatially Correlated Random Fields in Heterogeneous Wireless Sensor Networks", *IEEE Transactions on Signal Processing*, Vol.63, no.10, pp. 2597–2609, March, 2015.

[59] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering using consensus strategies", *46th IEEE Conference on Decision and Control*, 2007.

[60] H. Zhang, J, Moura , and B. Krogh, "Dynamic Field Estimation Using Wireless Sen-

sor Networks: Tradeoffs Between Estimation Error and Communication Cost", *IEEE Transactions on Signal Processing*, Vol. 57, no. 6, pp. 2383–2395, June, 2009.

[61] J.T. Gaskins,and M.J. Daniels, "A nonparametric prior for simultaneous covariance estimation", *Biometrika*, Vol.100, no.1, pp. 125–138, March, 2013.

[62] Y. Li, X. Lin, and P. Muller, "Bayesian inference in semiparametric mixed models for longitudinal data", *Biometrics*, Vol.66, no.1, pp. 70–78, March, 2010.

[63] D. Ruppert, M.P. Wand, and R.J. Carroll, "Semiparametric Regression", Cambridge University Press, New York, 2003.

[64] R. Tibshirani, " Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, Vol.58, no.1, pp. 267–288, 1996.

[65] S.P. Brooks, and A. Gelman, " General methods for monitoring convergence of iterative simulations", *Journal of Computational and Graphical Statistics*, Vol.7, no.4, pp. 434–455, 1998.

[66] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks", *Proc. IEEE Int. Symp. Intell., Limassol, Cyprus*, pp. 719–724, 2005.

[67] E. Msechu, and G. Giannakis, "Sensor-centric data reduction for estimation with ESNs via censoring and quantization", *IEEE Trans. Signal Process*, Vol. 60, No. 1, pp. 400–414, 2012.

[68] M. Erol-Kantarci, and H.T. Mouftah, "Wireless sensor networks for smart grid applications", *Proc. Int. Conf. on Elec. Commun and Photonic*, pp. 1–6, 2011.

[69] Y.C. Wu, Q. Chaudhari, and E. Serpedin, "Clock-synchronization of wireless sensor networks", *IEEE Signal Process. Mag.*, Vol. 28, No. 1, pp. 124–138, 2011.

[70] D. Adami, R. Garroppo, S. Giordano, and S. Lucetti, "On synchronization techniques: performance and impact on time metrics monitoring", *International Journal of Communication Systems*, Vol. 16, No. 4, pp. 273–290, 2003.

[71] H. Chen, J. Feng, and C.K. Tse, "Performancelifetime tradeoff for source extraction in multihop sensor networks", *International Journal of Communication Systems*, Vol. 23,

No. 12, pp. 1613–1631, 2010.

[72] H. Yigitler, A. Mahmood, R. Virrankoski, and R. Jantti, "Recursive clock skew estimation for wireless sensor networks using reference broadcasts", *IET Wireless Sensor Systems*, Vol. 12, No. 4, pp. 338–350, 2012.

[73] D. Zhao, Z. An, and Y. Xu, "Time Synchronization in Wireless Sensor Networks Using Max and Average Consensus Protocol", *International Journal of Distributed Sensor Networks*, Article ID: 192128, 2013.

[74] J. Du, and Y-C Wu, "Fully distributed clock skew and offset estimation in wireless sensor networks", *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4499–4502, IEEE, 2013.

[75] P. Briff, A. Lutenberg, L. Rey Vega, and F. Vargas, "On Energy Efficient Time Synchronization for Wireless Networks under Large-Scale and Small-Scale Fading Effects", *Wireless Sensor Network*, Vol. 5, pp. 181–193, 2013.

[76] I-K Rhee, J. Lee, J. Kim, E. Serpdin, and Y-C Wu, "Clock synchronization in wireless sensor networks: an overview", *Sensors* Vol. 9, pp. 57–85, 2009.

[77] B. Kaur, and A. Kaur, "A Survey of Time Synchronization Protocols for Wireless Sensor Networks", *International Journal of Computer Science and Mobile Computing*, Vol.2, no.9, pp. 100–106, September, 2013.

[78] D. Cochrane, and G.H. Orcutt, "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms, *Journal of the American Statistical Association*, Vol. 44, pp. 32–61, 1949.

[79] S. Chib, "Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach", *Journal of Econometrics*, Vol. 58, pp. 275–294, 1993.

[80] S. Chib, and E. Greenberg, "Bayes Inference in Regression Models with ARMA(p,q) Errors", *Journal of Econometrics*, Vol. 64, pp. 184–206, 1994.

[81] B.P. Carlin, and T.A. Louis, "Bayes and Empirical Bayes Methods for data Analysis" *Chapman and Hall*, 1996.

[82] M.P. urii, Z. Tafa, G. Dimi, and V. Milutinovi, "A survey of military applications of wire-

less sensor networks", *Mediterranean Conference on Embedded Computing (MECO)*, pp. 196–199, IEEE, 2012.

[83] T. He, S. Krishnamurthy, J.A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, "Energy-efficient surveillance system using wireless sensor networks", *In Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pp. 270–283, ACM, 2004.

[84] M. Winkler, K.D. Tuchs, K. Hughes, and G. Barclay,"Theoretical and practical aspects of military wireless sensor networks", *Journal of Telecommunications and Information Technology*, pp.37–45, 2008.

[85] K. Chakrabarty, S.S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks", *IEEE transactions on computers*, Vol. 51, No. 12, pp.1448–1453, 2002.

[86] C. Gui, and P. Mohapatra, "Power conservation and quality of surveillance in target tracking sensor networks", *In Proceedings of the 10th annual international conference on Mobile computing and networking*, pp. 129–143, ACM, 2004.

[87] H. Alemdar, and C. Ersoy, "Wireless sensor networks for healthcare: A survey", *Computer networks*, Vol. 54, No. 15, pp. 2688–2710, 2010.

[88] G.Z. Yang, and G. Yang, "Body sensor networks", *London: Springer*, Vol. 1, 2006.

[89] B.P. Lo, S. Thiemjarus, R. King, and G.Z. Yang, "Body sensor networka wireless sensor platform for pervasive healthcare monitoring", pp. 77–80, 2005.

[90] J. Ko, C. Lu, M.B. Srivastava, J.A. Stankovic, A. Terzis, and M. Welsh, "Wireless sensor networks for healthcare", *Proceedings of the IEEE*, Vol. 98, No. 11, pp.1947–1960, 2010.

[91] M. Al Ameen, J. Liu, and K. Kwak, "Security and privacy issues in wireless sensor networks for healthcare applications", *Journal of medical systems*, Vol. 36, No. 1, pp.93–101, 2012.

[92] J.A. Stankovic, Q. Cao, T. Doan, L. Fang, Z. He, R. Kiran, S. Lin, S. Son, R. Stoleru, and A. Wood, "Wireless sensor networks for in-home healthcare: Potential and chal-

lenges", *In High confidence medical device software and systems (HCMDSS) workshop*, Vol. 2005, 2005.

[93] R.S. Dilmaghani, H. Bobarshad, M. Ghavami, S. Choobkar, and C. Wolfe, "Wireless Sensor Networks for Monitoring Physiological Signals of Multiple Patients", *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 5, No. 3, pp. 347-356, 2011.

[94] F.J. Pierce, and T.V. Elliott, "Regional and on-farm wireless sensor networks for agricultural systems in Eastern Washington", *Computers and electronics in agriculture*, Vol. 61, No. 1, pp. 32–43, 2008.

[95] T. Wark, P. Corke, P. Sikka, L. Klingbeil, Y. Guo, C. Crossman, P. Valencia, D. Swain, and G. Bishop-Hurley, "Transforming agriculture through pervasive wireless sensor networks", *IEEE Pervasive Computing*, Vol. 6, No. 2, 2007.

[96] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard computing: Sensor networks in agricultural production", *IEEE Pervasive computing*, Vol. 3, No. 1, pp. 38–45, 2004.

[97] D. Culler, D. Estrin, and M. Srivastava, "Guest editors' introduction: Overview of sensor networks", *Computer*, Vol. 37, No. 8, pp. 41–49, 2004.

[98] A. Baggio, "Wireless sensor networks in precision agriculture", *In ACM Workshop on Real-World Wireless Sensor Networks (REALWSN 2005)*, pp. 1567–1576, 2005.

[99] G. Barrenetxea, F. Ingelrest, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange, "Sensorscope: Out-of-the-box environmental monitoring", *In Proceedings of the 7th international conference on Information processing in sensor networks*, pp. 332–343, IEEE Computer Society, 2008.

[100] X. Yu, P. Wu, W. Han, and Z. Zhang, "Overview of wireless underground sensor networks for agriculture", *African Journal of Biotechnology*, Vol. 11, No. 17, pp. 3942–3948, 2012.

REPRINTS OF PUBLISHED JOURNAL PAPERS (FIRST PAGES ONLY)

Received: 25 December 2015 | Revised: 2 June 2016 | Accepted: 6 August 2016

DOI 10.1002/dac.3191

WILEY

RESEARCH ARTICLE

A unified approach of simultaneous state estimation and anomalous node detection in distributed wireless sensor networks

Aditi Chatterjee[1,2 *] | Palaniandavar Venkateswaran[1] | Debaprasad Mukherjee[3]

[1]Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India
[2]Department of Electronics and Communication Engineering, Dr B.C. Roy Engineering College, Durgapur, India
[3]Department of Computer Science and Information Technology, Dr B.C. Roy Engineering College, Durgapur, India

Correspondence
Aditi Chatterjee, Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India.
Email: eice.aditi@gmail.com

SUMMARY

Detection of anomalous node in distributed wireless sensor networks is extremely important for powerful inference and network reliability. In this paper, we propose a powerful linear statistical model for estimating the state values of the sensor nodes longitudinally, and the estimated state values are used for detecting the anomalous nodes. Our proposed approach is powerful because it considers the effect of the nearest neighbors on the current state values and then detects the anomalous nodes based on the estimated state values. Our method can estimate the missing state values of the sensor nodes, which are kept in sleep mode for energy conservation. We also propose an alternative Bayesian model that is computationally faster for state estimation and anomaly detection. The effectiveness of the proposed model is investigated through extensive simulation studies, and the usefulness of our algorithm is numerically assessed. The performance of the proposed approach is compared to that of the traditional approaches through simulation studies. The proposed model can be effectively used in security surveillance, pattern recognition, habitat monitoring, etc.

KEYWORDS

anomalous nodes, Gibbs sampling, information exchange, linear regression model, wireless sensor networks.

1 | INTRODUCTION

Wireless sensor networks (WSNs) consist of a mass of sensor nodes distributed over a physical space for monitoring the environmental conditions such as temperature, pressure, humidity, acoustics, and resonance. Recent advancements in the state estimation of WSNs have drawn attention of researchers mainly because of the extensive applications of WSNs in habitat monitoring,[1] object tracking,[2] event detection,[3] intruder locating[4] etc. Thus WSNs find applications in medical science,[5] security surveillance,[6] pattern recognition,[7] and many other disciplines. Models and methods have been proposed for the state estimation of both discrete and continuous-time sensor networks. Note that the term "state" in the context of WSNs is quite subjective in the sense that it depends on the objective of the study and the quantity being measured from the sensor nodes over time. Specifically, the state of a sensor node at time t is estimated based on the available measurements till time $t-1$. State of a sensor node thus can be continuous (e.g., air pressure and humidity) as well as binary (e.g., occurrence of an event). However, we consider WSNs with continuous state values in this paper.

In a purely probabilistic framework Liang et al.[8] proposed models for distributed state estimation of discrete-time WSNs. Similar models have been proposed by Xu and Li.[9] Sun et al.[10] proposed a very powerful model for state estimation of multiple mobile targets. Quevedo et al.[11] proposed models that can efficiently estimate the states of WSNs with correlated wireless fading channels. Mo et al.[12] proposed model that can handle the false data injection attacks in state estimation of sensor networks. Because there are several resource constraints in WSNs, anomaly detection using state estimation becomes very essential for reliable networks.

In this paper, we propose a new approach of estimating the state values of the sensor nodes and then detecting the possible anomalous node(s) in the network based on the estimated state values. Detection of anomalous node is important because such nodes might have detrimental effects on the surrounding sensors and thus may affect the performance of the entire network over time.[13,14] Analysis of sensor measurements is very important for anomaly detection in WSNs. However, sensor measurements contain spatio-temporal correlations. Because the sensor nodes are densely deployed, spatial correlation exists among the neighboring sensor nodes. Temporal correlation occurs because of the predictable relationship that exists in sequential measurements of the sensor nodes. Presence of anomaly is affirmed when one (or more) sensor node behaves differently from most the sensor nodes.[15]

Detection of anomalous node is widely applied in a variety of fields. Sensor nodes in wireless body area networks enable real-time global patient and health care monitoring.[16] Sun et al.[17] used extended Kalman filter for detecting the false injected data from the anomalous behavior of the sensor nodes. Rajasegarar et al.[18] used distributed one-class quarter-sphere support vector machines to distinguish anomalous measurements from the obtained data.

There is a rich literature on object tracking using WSNs[2,19] but relatively few papers on anomalous node detection.[18–20] Almost all these papers detect an anomalous node based on its relative performance compared to the neighboring sensor nodes. We propose a novel anomaly detection method based on the estimated state values of the sensor nodes. While the state estimation issue has been addressed from various perspectives, mostly the suggested methods rely on state-space models based on Kalman filter (KF) and stochastic differential equations. Examples include Wu et al.,[21] Quevedo et al.,[11] and Di et al.[22] Our proposed model is fundamentally different from the previous works and statistically powerful for the state estimation of discrete-time WSNs. We propose a dynamic regression model that allows the "exchange of the relevant information" among the spatially close nodes. We note that information exchange is inevitable in any network and hence the models for state estimation should consider this appropriately for reliable state estimation. Our approach is based on the maximum likelihood (ML) estimation and related inferential properties. In recent years there is a growing interest in Bayesian models and estimating the model parameters by Monte Carlo Markov chain (MCMC). Bayesian models can combine the prior information on the network (specifically the parameter values) with the observed likelihood and compute the posterior distribution for inference. For complex models, implementation of the traditional methods like ML, KF is really challenging and computationally expensive. However, MCMC can easily handle complex models and estimate the model parameters iteratively in a relatively less time.[23] Hence we also propose an alternative Bayesian model for state estimation and anomaly detection.

The current paper has two major contributions. First we propose a dynamic statistical model for state estimation, which allows the appropriate information exchange among the spatially close sensors. Such models have not been used yet in WSNs literature for state estimation. We develop an algorithm for locating the possible anomalous node(s) in the network. Second, we propose an alternative Bayesian model and estimate the state values using MCMC in a shorter time. We compare the performance of all the three methods and conclude that state-space model based on KF should not be considered as the "best alternative" in the context of state estimation of WSNs.

The rest of the paper is organized as follows. In Section 2, we explain the proposed model and parameter estimation using ML method. In Section 3, we propose an alternative Bayesian model and MCMC algorithm. Simulation results for assessing the performance of our model are discussed in Section 4. Finally, Section 5 concludes.

2 | METHOD OF LONGITUDINAL STATE ESTIMATION

Throughout this paper, we use the following notations. Consider a particular WSN consisting of N sensor nodes with one or more anomalous nodes. For each sensor node, the state values are estimated at T different discrete time points $1, 2, \ldots, T$. At each time point t, the state value of the i-th sensor is denoted by $X_i(t)$, which is communicated to the "sink" of the network. Let (θ_i, δ_i) be the coordinates of the i-th sensor node, $i = 1, 2, \ldots N$. The Euclidean distance between the sensor i and the sensor j is denoted by D_{ij}. Based on the D_{ij} values, we first "group" the sensor nodes into several clusters so that the sensors belonging to a common cluster are spatially close to each other. In Figure 1, we show a cluster-based distributed WSN.

2.1 | Linear statistical model for state estimation

Traditionally in the state estimation of WSNs, state-space models are used.[11,21,22] In state-space models, the state values y_i's are modeled as a function of the unknown states x_i's and then auto-regressive models are used for the unknown states. Mathematically one can summarize a state-space model as the following: $y_i|x_i = f(x_i) + d_i$; $x_i|x_{i-1} = g(x_{i-1}) + e_i$, where d_i and e_i respectively denote the observation error and system error. The functions f and g are assumed to be linear, and the errors are assumed to be Gaussian for estimating the parameters via KF. However, such models do not consider the information exchange among the sensor nodes, which are spatially close to each other. In any network, information exchange is inevitable and must be taken into account. Considering this, we propose the following linear model for estimating the state values of the sensor nodes at different discrete time points:

$$X_i(t) = f(t) + \alpha X_i(t-1) + \beta Z_i(t-1) + \epsilon_i(t), \tag{1}$$

where $X_i(t)$ denotes the state value of the i-th sensor node at time t belonging to a particular cluster. The smooth function

Simultaneous State Estimation of Cluster-Based Wireless Sensor Networks

Aditi Chatterjee, Palaniandavar Venkateswaran, *Senior Member, IEEE*, and Kiranmoy Das

***Abstract*— There is rich literature on the state estimation of wireless sensor networks (WSNs). Most of these works are based on the dynamic state-space model via Kalman filter or similar Markov models. In any network, information exchange is inevitable, and in this paper, we propose a Bayesian non-parametric approach for addressing this issue in the context of state estimation of WSNs. We consider a cluster-based WSN and consider a discrete-time linear Markov model for estimating the state values of the sensor nodes over time. For measuring the amount of information shared by the model parameters across different clusters, we consider non-parametric matrix stick-breaking priors for the cluster-specific model parameters. We demonstrate the usefulness of our proposed model in locating an immobile anomalous node in the network. We compute the time to locate the anomalous object and the false positive rate of our proposed approach. Simulation studies are performed to assess the operating characteristics of the proposed model. The proposed approach will be useful in emergency monitoring, medical genetics, geosciences, and many other disciplines where WSNs are frequently used for decision making.**

***Index Terms*— Cluster, Dirichlet process, linear Markov models, matrix stick-breaking process, wireless sensor networks.**

I. Introduction

IN RECENT years we have witnessed a revolution in the scientific research related to wireless sensor networks (WSNs) and their applications in various disciplines. Sensor nodes are battery-powered tiny devices which are able to sense, process, store and exchange information. WSNs typically consist of ten to thousands of sensor nodes which are deployed (randomly or systematically) over certain physical places of interest. In recent times, WSNs are used for monitoring environment [1], tracking object of interest [2], detecting events [3], locating intruders [4] etc. Thus WSNs find applications in medical science [5], security surveillance [6], pattern recognition [7] and many other disciplines.

Most of the real applications of WSNs typically depend on the accuracy and precision related to the state estimation of the network over different time points. In reality, the state values of each sensor node are estimated longitudinally and thus the state values of the entire network is estimated by combining the individual state values of the sensors. Efficient state estimation of the sensor nodes belonging to a network depends on two things: (1) time synchronization among the sensor nodes and (2) statistical models for estimating the sensor specific state values. There is a rich literature on time synchronization of the sensor nodes, examples include Noh et al. ([8], [9]), Kaur and Kaur [10], Chatterjee and Venkateswaran [11] and many more. These authors mainly used linear models to estimate the relative clock-offset and clock-skew for the network synchronization. There is indeed a richer literature on state estimation and most of these works are based on state-space models via Kalman-Filter ([12], [13]). Liang et al. [14] proposed probabilistic models for distributed state estimation of a discrete-time WSN. Models proposed by Sun et al. [15] can estimate the state values of multiple mobile targets. Mo et al. [16] proposed models for state estimation when false data are injected in the network. Under various complex scenarios, the state estimation of the sensor nodes has become the most important task in our time. Thus, powerful statistical models have arguably become more important in the recent research works related to WSNs.

Information exchange among the sensor nodes at different time points is fundamentally important in WSNs and hence the statistical models for state estimation must be different from the traditional regression models where subjects are assumed to be independent [17]. In the dynamic state-space models, the observations y_i's are modeled as a function of the unknown state values x_i's and then Markov (auto-regressive) models are used for the unknown states. Mathematically we can write, $y_i|x_i = f(x_i) + d_i$; $x_i|x_{i-1} = g(x_{i-1}) + e_i$, where d_i and e_i respectively denote the measurement error and system error. The functions f and g are assumed to be linear and the errors are assumed to be Gaussian for estimating the parameters via Kalman-Filter (K-F). Although there is a rich literature on state estimation using Kalman-Filter which enable information exchange among the sensor nodes [18], [19], relatively little attention has been given for information exchange among different clusters of a network. Our proposed linear Markov model (in Section 2.1) is different from the existing works since (i) we specify a single regression model with only one error component and (ii) we allow the information exchange among the sensors belonging to the same cluster by considering the "neighbourhood effect" in the model and also among different clusters by appropriately sharing the model parameters in a non-parametric Bayesian approach.

Manuscript received November 18, 2015; revised May 9, 2016, August 15, 2016; accepted September 1, 2016. Date of publication September 16, 2016; date of current version December 8, 2016. The associate editor coordinating the review of this paper and approving it for publication was M. Vu.

A. Chatterjee and P. Venkateswaran are with the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata 700032, India.

K. Das is with the Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata 700108, India (e-mail: kmd@isical.ac.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2016.2610429

Dirichlet Process (DP) priors, originally proposed by Blackwell and MacQueen [20], Ferguson [21], have been used in non-parametric Bayesian literature for classification and information sharing. The popularity of DP prior is mainly due to its computational ease and powerful inferential properties which come as the consequence of the stick breaking formulation of Dirichlet Process due to Sethuraman [22]. More recently Dunson et al. [23] formulated matrix stick breaking priors which can handle the information sharing across the model parameters for the datasets coming from different related "groups". Gaskins and Daniels [24], Das and Daniels [25] extended the matrix stick breaking priors for sharing the large covariance parameters for different related groups. We build our work on the seminal paper by Dunson et al. [23] and use the matrix stick breaking priors for sharing the model parameters across different groups.

In this paper, we consider the cluster topology for a general WSN. The sensor nodes are "grouped" into "clusters" based on their relative Euclidean distances. Thus the nodes which are spatially close to each other are kept in the same cluster. The linear model we propose considers the effect of the nearest neighbours on the current state value for each sensor node and thus allows within cluster information exchange. We also propose non-parametric matrix stick breaking priors for the cluster specific model parameters and thus consider the possibility of information exchange between the clusters in the network under consideration.

We illustrate an application of the proposed approach in detecting immobile anomalous node. An anomalous node may be a foreign object, a selfish node or a malicious node which might have detrimental effect on the surrounding sensor nodes over time. Our approach can accurately detect such nodes. Removal of such nodes makes the network more effective.

The rest of the article is organized as follows. In section 2, we discuss the proposed linear Markov model and penalized splines for estimating the general effect of time. The proposed prior structure and parameter estimation are discussed in section 3. Results from the simulation studies are discussed in section 4. We also numerically illustrate a practical application of our proposed approach in this section. Finally section 5 concludes.

II. Proposed Model

A. Linear Markov Model

Consider a sensor network with K clusters and the k-th cluster consists of n_k sensors with a total of $n = \sum_{k=1}^{K} n_k$ sensors. Note that clusters are formed by considering the Euclidean distance between the sensors, i.e. for some fixed δ, all the sensors which are in the δ-neighborhood of each other are kept in a single cluster. Figure 1 demonstrates the cluster formation. Note that δ should be chosen such that the number of clusters is neither too small nor too large (typically less than 10 in our setting). However, in reality, the number of clusters (and hence the value of δ) depends on the coverage area and the nature of the experiment.

We consider a discrete-time state estimation procedure in this article. For making an energy efficient protocol,

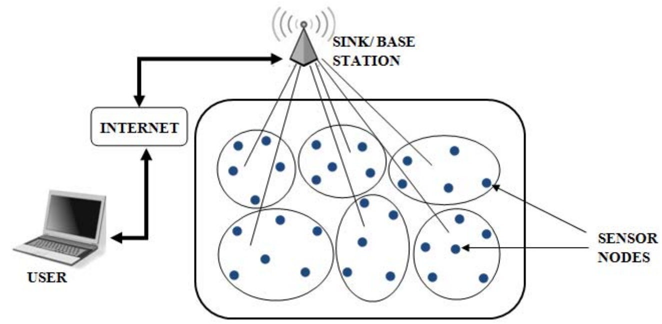


Fig. 1. Cluster-based wireless sensor network.

we assume that all the sensors within a cluster are not necessarily measured exactly at the same time points. Some sensors might be kept in the sleep mode for some time and since in the sleep mode the sensors consume very little battery power, the network becomes energy efficient. We assume that the i-th sensor belonging to the k-th cluster ($k = 1, 2, \ldots, K$) is measured at T_i^k different time points and $X_{ik}(t_{ij})$ denotes its state value at time t_{ij} ($j = 1, 2, \ldots, T_i^k$). However, we also assume that at each time point at least one (if not all) sensor from each cluster is measured to keep the cluster active. Our linear Markov model for estimating the state value of the i-th sensor at time t_{ij} based on the observed measurements till time $t_{i(j-1)}$ can be expressed as the following:

$$X_{ik}(t_{ij}) = f_k(t_{ij}) + \theta_{k1} X_{ik}(t_{i(j-1)}) I(|t_{ij} - t_{i(j-1)}| < p) + \theta_{k2} Z_{ik}(t_{i(j-1)}) + e_{ijk}, \quad (1)$$

where f_k is the cluster specific general effect of time which we model using Penalized splines. The effect of time on the state values is possibly different for different clusters and hence we include the subscript k in the function $f(.)$. Here the indicator function $I(|t_{ij} - t_{i(j-1)}| < p)$ takes value 1, if $|t_{ij} - t_{i(j-1)}| < p$ and 0, otherwise. Note that θ_{k1} is the cluster specific effect of the previous available measurement on the current state value and will be estimated based on the available data. The previous available measurement will influence the current state value only when the time difference is below a fixed (known) threshold p (typically p=3). This is based on the assumption that measurements corresponding to the closer time points are more related than those for the further time points. Such assumptions are quite natural for any longitudinal study.

In equation (1), $Z_{ik}(t_{i(j-1)})$ denotes the average measurement from all the sensors belonging to the k-th cluster (except the i-th sensor) which are measured at time $t_{i(j-1)}$. Since we assumed that at each time point, some sensors are measured from each cluster, we can easily get $Z_{ik}(t_{i(j-1)})$ based on the available data for the k-th cluster. Hence θ_{k2} basically denotes the "neighbourhood effect" on the state value and needs to be estimated from the available data. Note that by introducing Z variable, we essentially incorporate the information sharing among the sensors within a single cluster. The residual errors e_{ijk}'s are assumed to be independently normally distributed with mean=0 and unknown variance=σ^2. We have to estimate σ^2 from the data. The above model is Markovian as the

INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS
Int. J. Commun. Syst. 2016; **29**:722–733
Published online 21 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/dac.2944

An efficient statistical approach for time synchronization in wireless sensor networks

Aditi Chatterjee[1,*,†] and Palaniandavar Venkateswaran[2]

[1]*Department of Electronics and Communication Engineering, Dr. B.C. Roy Engineering College, Durgapur, India*
[2]*Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India*

SUMMARY

In this paper, we propose a powerful method of estimating the model parameters for time synchronization in wireless sensor networks (WSNs). Joint estimation of clock offset and clock skew has been proposed in the literature using the standard regression framework. Here, we claim that simple regression poorly estimates the parameters because of the inherent correlation among successive time readings between two sensors. We propose an alternative autoregressive model and use generalized least squares for estimating the relative offset and skew parameters. A computationally efficient Bayesian approach is also proposed for the parameter estimation considering correlated readings between two sensors. The effectiveness of the proposed approach compared with the earlier approach has been investigated through extensive simulation studies. Copyright © 2015 John Wiley & Sons, Ltd.

Received 3 July 2014; Revised 4 November 2014; Accepted 21 January 2015

KEY WORDS: autoregressive model; Bayes estimates; clock offset; clock skew; generalized least squares; time synchronization

1. INTRODUCTION

Wireless Sensor Networks (WSNs) are composed of ten to thousands of sensor nodes, which are multifunctional tiny devices with low power and bandwidth. Sensor nodes are used to sense the environment, process the crude data and communicate over a local area. WSNs have been a very promising research area in the last 10 years, and consequently there is a rich literature on state estimation, lifetime maximization, wireless communications, channel fadding, and so forth. Applications of WSNs can be found in various disciplines including but not limited to environmental studies, medical genetics, emergency monitoring, and so forth [1–4]. WSNs find applications in a variety of phenomena in the real world. Most of these applications assume that the sensor nodes are synchronized to a common clock. Thus, the time synchronization of the nodes in the network plays a key role for accurate state estimations and/or predictions of the surrounding environment. If different sensors in the network run according to their own clocks, then coordination among the sensors are affected and consequently the inference becomes inconsistent and hence less powerful [5]. Time synchronization is also very important in randomly deployed sensor networks because data transmission from leaf nodes to the sink in this type of network takes place in multihop manner [6].

In recent years, many useful protocols have been proposed for time synchronization in WSNs, for example, reference-broadcast synchronization (RBS), Time synchronization Protocol for Sensor Networks (TPSN), Flooding Time Synchronization Protocol (FTSP), and Receiver-only Synchronization (ROS). If we have a pair of sensor nodes, say node A and node B, let us denote the time of the clocks in the sensor nodes A and B by $C_A(t)$ and $C_B(t)$, respectively, as the functions of time (t).

*Correspondence to: Aditi Chatterjee, Dr. B.C. Roy Engineering College, Durgapur, India.
†E-mail: eice.aditi@gmail.com

Relative clock offset between the node A and node B at time t is defined as $C_{\mathrm{A}}(t) - C_{\mathrm{B}}(t)$. The derivative (the rate of variation) of the relative clock offset gives the relative clock skew. Thus, the relative clock skew between node A and node B is given by $C_{\mathrm{A}}^{'}(t) - C_{\mathrm{B}}^{'}(t)$. Practically speaking, a pair of nodes are synchronized by estimating the relative clock offset and relative clock skew between the nodes. Statistical models have been developed where the offset and skew parameters are estimated by least squares method using a regression model in [7, 8]. When a pair of sensor nodes exchange the timing message, a group of neighboring sensors overhear those messages and synchronize themselves accordingly [8]. Methods have been proposed for recursive clock skew estimation for WSNs using reference broadcasts [9], time synchronization in WSNs using max and average consensus protocol [10]. A Bayesian approach of time synchronization for a complex network model has been proposed in [11]. Protocols have been proposed for synchronization in more complex network structures [12, 13]. Nice reviews of the recent approaches for time synchronization under different protocols are given in [14, 15]. Recent advancements in time synchronization protocol lead to design of different mechanisms to overcome the security issues in FTSP [16].

All the above approaches assume that the set of time readings between a pair of nodes are uncorrelated, which is typically not the case in practice. Because the readings are taken from the same pair of nodes at different time points, there are some kind of dependence among these time readings. In this paper, we address this issue by considering autoregressive dependence among the time readings and synchronize the pair of node by estimating their relative offset and relative skew using generalized least squares method. The advantage of the proposed approach is not only in its accuracy and precision but also in its robustness in estimating the model parameters for correlated time readings.

When the successive time readings are correlated, the generalized least squares approach becomes an iterative method of parameter estimation, which is computationally expensive. We propose an alternative Bayesian approach where prior distributions are assumed for the model parameters and then the joint posterior distribution is computed. From this joint posterior, we derive the full conditional density for each of the model parameters and then use Markov chain Monte Carlo (MCMC) to estimate the parameters. The advantage of Bayesian approach is, one can incorporate a prior knowledge about the model parameters to get the updated posterior estimates. Posterior estimates are indeed the weighted average of prior mean and the maximum likelihood estimates based on the data. Also MCMC is a much faster algorithm than the generalized least squares, and we compare two methods in terms of their respective CPU times.

The current paper has two main contributions. First, we propose an alternative approach of estimating model parameters for time synchronization in WSNs when the successive time readings are correlated. Second, we propose an alternative Bayesian model where the parameters can be estimated by MCMC. The Bayesian approach is economical in terms of the computational cost.

The rest of the paper is organized as follows. In Section 2, we review the estimation approach proposed in [7] and propose our generalized least squares approach for efficient estimation. The higher accuracy of the proposed method is shown in Section 3 by simulation studies. We propose an alternative Bayesian approach of estimating the model parameters by MCMC in Section 4 and compare its performance to the generalized least squares approach. Some concluding remarks are given in Section 5.

2. RECEIVER-ONLY SYNCHRONIZATION

Here we discuss our proposed approach in the context of ROS [8]. However, note that for other synchronization approaches, for example, TPSN, RBS, and FTSP, we can use the proposed estimation approach for consistent and powerful estimation.

In Figure 1 [8], consider two nodes P and A, where node P is the parent node. Communication range of each node is limited to a circle of prefixed radius, and if any node belongs to the checked area, it can receive message from both A and P. A pairwise synchronization is performed between node P and node A and all other nodes belonging to the checked area, receive a series of synchronization messages. As all the neighboring sensors are getting synchronized by simply overhearing the timing messages, ROS becomes an energy efficient synchronization protocol.

DOI: 10.1002/dac

BIOGRAPHY OF THE AUTHOR

Aditi Chatterjee was born on August 13, 1985. She received her B.Tech and M.Tech (Electronics and Communication Engineering) degrees from the West Bengal University of Technology in 2008 and 2011 respectively. She started her career as an Assistant Professor in Bengal College of Engineering and Technology, Durgapur. Later, she taught at Dr. B.C.Roy Engineering College, Durgapur as an assistant professor for 6 years during 2012-2017. She has been a Guest Faculty at Jadavpur University, Kolkata, since January 2018.

Her research interests include state estimation, anomaly detection, time synchronization for discrete-time wireless sensor networks. At present, she is working on Internet of Things (IoT) and statistical estimation theory. She has published 3 journal articles, 4 conference paper and 1 book chapter from Springer.

Apart from research work, she loves teaching and interacting with students. She enjoys music, film, drawing, cooking and travelling. She is blessed with a baby boy who takes most of her current time.