IEEE ICBEST 2015
31 August – 01 September 2015, Singapore

Outlier Detection and Data Filtering for Wireless Sensor and Actuator Networks in Building Environment

Manish Gupta, Krishnanand K.R., Hoang Duc Chinh, Sanjib Kumar Panda*
Department of Electrical & Computer Engineering,
National University of Singapore, Singapore
*eleskp@nus.edu.sg

Abstract — **The management systems of smart buildings aim to provide a comfort environment while reducing energy consumption. Efficient energy management requires changes not only in the way the energy is supplied but also in the way the devices are controlled. Such system requires various sensing data, data analysis and optimization algorithms in decision making. Wireless sensor networks have been employed widely for data collection. However, these sensors due to different factors can cause variations in raw sensor measurements, inaccurate information transferred to the base station may result in misguidance to the control system. The main objective in outlier detection is to find the data that is deviating from the other data based on the algorithm techniques. In this paper, Hodrick Prescott filters are adopted for reducing the noise and errors from real sensor data. Experiments and analysis conducted show that output data is more stable when compared with the results obtained by other methods.**

Keywords— wireless sensor network, moving average filter, hodrick-prescott filter, outlier detection

I. INTRODUCTION

Residential and office building consume around 40% of the total energy consumption in the world [1]. It is essential to cut the annual consumption of primary energy by 20% by year 2020 to reduce the consumption of buildings [1]. This can be obtained by integrating information and communication technologies in the design and operations of buildings to support continuous improvements in the energy performances of the buildings [2].

Advance of communication technology enables the popular use of wireless monitoring and control in Building Automation Systems (BAS) [1]. The exploitation of Wireless Sensor and Actuator Networks (WSANs) facilitates intelligent management of energy in buildings further increasing occupant comfort and reducing energy demand. In addition of achieving environmental and economic gains, the introduction of monitoring, control, analysis systems for building management system will allow easily handling for building owners, managers, operators, crews and other users in the buildings. With increase in internet power, the various protocols has allowed to access building information and control systems through smart handheld devices [1].

Compared to traditional BAS communication systems which used wired technologies and cabling installation, WSANs have an advantage of easy deployment, especially for building retrofits. Low cost and low power features allow WSANs to be used for various applications like environmental monitoring, health monitoring and industrial monitoring as well as distributed lighting and air conditioning control systems in a large scale with high density distribution. Therefore, a comprehensive understanding and control of the built environment can be achieved to reduce energy consumption and CO_2 emission, meanwhile human comfort is maintained [1].

Wireless sensors and actuators networks consists of group of sensor and actuator nodes which are connected to each other through wireless medium to perform distributive sensing and actuation tasks [2]. A sensor node senses environmental parameters, processes information, and transfers data to base stations or other nodes via on-board radio modules. An actuator is a device which converts an electrical control signal to a physical action in order to adjust environmental conditions. WSANs are highly flexible and this enables the deployment of sensors network in required infrastructure to perform measurements for predefined time. There are number of protocols and communication standards to support them such as ZigBee, 6LoWPAN, Wi-Fi, Wireless HART, and ISA100.11a [4-8].

In the WSANs, essential physical environmental parameters collected and pre-processed by sensor networks are transferred to control centers for analysis, modelling and prediction as well as feedback of the control system. The data obtained from these sensors are affected by environmental noise, inaccurate sensor devices, faulty pre-processing or temporal mal-functioning nodes. The issue will lead to imprecise modelling and prediction or failures of control actions. Thus, it is essential to have outlier or error detection of sensing values perceived by the sensors [2].

According to G. H. John [9], an outlier might also be surprising vertical data. The author states that a point which belongs to one class and residing in another class may results in veridical classification which is outlier to the user. An outlier can be considered as the noise points which are lying outside the pre-defined clusters and even separated from the noise itself. The outliers normally behave different from the norm [10]. Therefore, outlier detection is an important thing to be done in application processing since these points can lead to the abnormal behavior during processing of an application [11].

E. Elnahrawy et. al. introduced structure for cleaning and querying noisy sensors in [12]. The authors presented a probability approach called Bayesian approach for reducing the uncertainty related to the data that arise due to the random noise [12]. A survey of other techniques for outlier detection was also presented in [13]. Christoph Heinz and Bernhard Seeger suggested the detection of outliers with statistical method and investigated the augmentation of sensor network by meaningful statistical models in [14]. S. Subramanian et.al. proposed a framework operating in a distributed fashion. An approximation of multi-dimensional data distributions was given in order to enable complex applications in resource-constrained sensor networks [15].

This paper aims at the techniques which are involved in removing the outliers from the raw data which are distorted by the environmental noise. Outlier's detection is used to filter the noisy data, remove unwanted spikes, and detect faulty nodes and the events happening in the nodes. The target of removing the noise from the sensor data is to make the control action smooth which depends on the environmental parameters. This will ensure that there is less chattering in the control action from the main control station to the end point actuators.

II. PROBLEM STATEMENT

Wireless sensors networks (WSNs) normally consists of hundreds or thousands of tiny, low cost sensor nodes using multiple wireless communication standards, protocols operating in harsh environmental conditions. These sensors have constraints in terms of bandwidth, memory, energy consumption and many other factors which make them vulnerable to faults and malicious attacks. These issues can cause deviations from the raw sensor signals and the inaccurate readings are being transmitted from the sensor nodes to the base station.

The objective in outlier detection is to find all those data elements which are most deviated from the remaining data based on some measuring algorithm. There can be multiple sources of outlier such as mechanical faults, instrument error (faulty sensor node), human error, natural deviations, etc. Broadly classifying there can be two sources of outlier – errors and events. Errors refer to noise-related measurements or due to faulty sensors. Events refer to natural deviations or the particular happening around in the real-time world such as fire, dust (pollution) etc. Errors causing outliers are normally frequent and main source of faulty readings. On the other hand, outliers due to the events based have less probability of occurrence [16].

In this research, the outlier's readings in the readings obtained from the sensor nodes. Different types of smoothing filters have been applied on the sensors reading to remove the noise and distortions. These techniques are applied on the real-time sensor data points obtained from the designed wireless sensor networks.

III. SYSTEM DESCRIPTION

A. Building Automation System

A BAS is centralized and interlinked networks of hardware and software, which monitor and control the environment in commercial, industrial, and institutional facilities. While managing various building systems, the automation system ensures the operational performance of the facility as well as the comfort and safety of building occupants.

Typically, such control systems are installed in new buildings or as part of a renovation where they replace an outdated control system. Generally, building automation begins with control of mechanical, electrical, and plumbing (MEP) systems. For instance, the heating, ventilation, and air-conditioning (HVAC) system is almost always controlled, including control of its various pieces of equipment such as:

Chillers Boilers Air Handling Units (AHUs) Roof-top Units (RTUs) Fan Coil Units (FCUs) Heat Pump Units (HPUs) Variable Air Volume boxes (VAVs) Lighting control is, likewise, low-hanging fruit for optimizing building performance. Other systems that are often controlled and/or brought under a complete automation system include: Power monitoring Security Close circuit video (CCTV) Card and keypad access Fire alarm system Elevators/escalators plumbing and water monitors [17].

B. Wireless Sensors and Actuators Network (WSAN)

In building systems, a WSN consists of autonomous sensors that are distributed to monitor physical and environmental conditions such as light, temperature and humidity, etc. These collected data are transmitted to the main base system for further processing and consequent usage. In more complicated networks, the processed data can be used as feedbacks to control of different actuators so that desired environmental conditions are maintained. The introduction of actuators into WSNs forms the WSANs. Similar to WSNs, topology of WSANs can vary from star to multi-hop wireless mesh networks [18]. Fig. 1 show an example of a WSAN architecture in which data packets are transferred bi-directionally from nodes to server and vice versa to exchange sensing information and control commands.

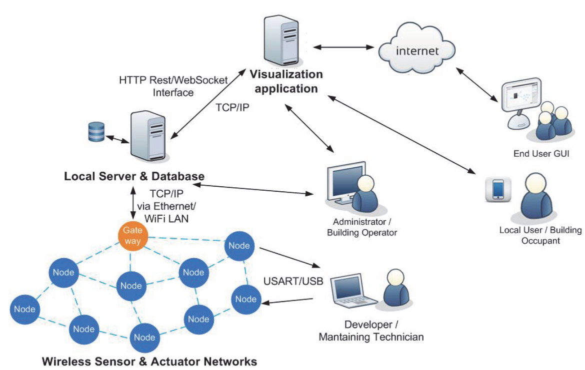

Figure 1 *Wireless Sensor and Actuator Network architecture*

In a BAS, the actuators can be air conditioning or lighting control board. Fig. 2 shows an actuator node we are using to control LED Lamp in Building Lighting system. The node consists of XBee module, an Arduino board and a driver circuit.

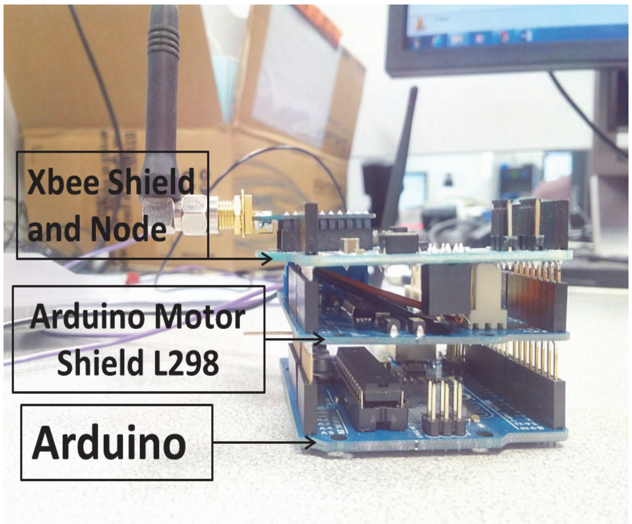

Figure 2 Arduino and Xbee assembly for controlling actuator node

Control command is sent towards the actuator nodes from a control centre and the status of the LED is updated once the action is performed. Communication between the control centre and the actuators is carried out via XBee module, which is a ZigBee-based wireless device as shown in Fig. 3. The set point included in control command can be achieved from occupant feedback or calculated based on sensor values. As the control centre is connected to Internet via Wi-Fi/Ethernet, occupants can use a mobile application to control the actuator by providing the preferred environment conditions and sending these values to control centre.

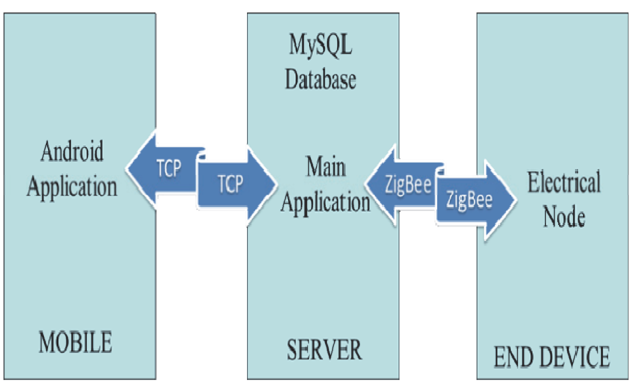

Figure 3 Data flow diagram for controlling electrical node

In order to obtain a fully automatic control system, observation of building environment is required. It is enabled by the wireless sensor network with the capabilities of sensing, networking and processing. In this work, the IRIS sensor nodes equipped with MTS400 sensor boards are used to monitor three parameters: illuminance, humidity and temperature of indoor environment. An external Telaire 7001 CO2 Sensor is also used to measure CO2 concentration. The CO2 sensor data is stored in PI database[19] and then the data is extracted from the database for further analysis.

25 nodes are deployed in a laboratory to collect data and forward to a data center facilitated by a MySQL Server. Fig. 4 show a graphical monitoring system displaying sensor deployment and sensing measurements of a selected node.

Figure 4 Graphical User interface for monitoring environmental parameters

These controlling & monitoring networks can be integrated together to control the different parameters as per user comfort in both manual and automatic modes with the support of motion detectors like PIR sensor or other localization systems.

Data mining in sensor networks is the process in which application oriented models and patterns with acceptable accuracy are extracted from a continuous, rapid and possibly non-ended flow of data streams from sensor networks. The whole data needs to be processed immediately and algorithm for the data mining should be fast enough to process high frequency data received from the sensor nodes [20],[21].

So, when there is large data set coming from the sensor nodes, the data will be consisting of outliers and noises, which can further affect your control decisions. So there is need of techniques which can remove unwanted noises and outliers from the dataset. This paper further focuses some of the techniques which are used as smoothening filter for removing unwanted noise and errors.

IV. METHODOLOGY

A. Hodrick-Prescott Filter (HP Filter)

The Hodrick–Prescott filter (also known as Hodrick–Prescott decomposition) is a mathematical tool used to remove the cyclical component of a time series from raw data.

It is a flexible de-trending method that is widely used in empirical macro research. Let's suppose that the original series x_t is composed of a trend component g_t and a cyclical component c_t[22].

$$x_t = g_t + c_t \quad (1)$$

The HP-Filter isolates the cycle component by following minimization problem.

$$\sum_{t=1}^{T}(x_t - g_t)^2 + \lambda \sum_{t=2}^{T-1}[(g_{t+1} - g_t) - (g_t - g_{t-1})]^2 \quad (2)$$

The first term is a measure of the fitness of the time series while the second term is a measure of the smoothness. There is a conflict between "goodness of fit" and "smoothness". To keep track of this problem there is a "trade-off"-parameter λ. Note that λ is 0, the trend component becomes equivalent to the original series while λ diverges to infinity, the trend component approaches a linear trend.

B. Moving Average Filters

In statistics, a moving average (rolling average or running average) is a calculation to analyze data points by creating a series of averages of different subsets of the full data set. It is also called a moving mean or rolling mean and is a type of finite impulse response filter [23].

Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward"; that is, excluding the first number of the series and including the next number following the original subset in the series. This creates a new subset of numbers, which is averaged. This process is repeated over the entire data series. The plot line connecting all the (fixed) averages is the moving average. A moving average is a set of numbers, each of which is the average of the corresponding subset of a larger set of datum points. A moving average may also use unequal weights for each datum value in the subset to emphasize particular values in the subset.

$$y[i] = \frac{1}{M}\sum_{j=0}^{M-1} x[i+j] \tag{3}$$

A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. The threshold between short-term and long-term depends on the application, and the parameters of the moving average will be set accordingly. Mathematically, a moving average is a type of convolution and so it can be viewed as an example of a low-pass filter used in signal processing. Viewed simplistically it can be regarded as smoothing the data. The greatest advantage of MA is that it fixes the irregular fluctuation in a Time series. MA is easy to understand and does not require any complex mathematical calculation.

V. Implementation

The above two described filters are being implemented to remove the noise from the sensor data consisting of temperature, humidity and illuminance. The data logic is being implemented in MATLAB. Further a data connection is being established with the MYSQL database using database toolbox in MATLAB. The program is written to get the data array for the stored parameters using MYSQL query defined in the program.

The fetched array is then applied through the filters described above and then filtered data is compared with the original data.

A. Results and Discussions

Different plots were being obtained for both moving average and HP filter. The results obtained are shown in the figures 5 - 7 below.

Fig 5 depicts the filtered data that has been obtained after applying filters on Humidity parameter. Fig 6 displays the filtered data that has been obtained after applying filters on Temperature parameter. Fig 7 displays the filtered data that has been obtained after applying filters on CO_2 parameter.

Figure below contains 2 subplots in which the first plot represents the raw data and second plot represents the processed data after applying filters.

For application of these filters, MATLAB connects to the MYSQL database and queries are being used to fetch the raw data for 7 days along with the time stamp for the data. This time stamp column is used to plot the graph for desired parameter with time stamp. In case of Moving average, 3rd order and 5th order, the resultant value is calculated with the average of subset containing 3 and 5 past values respectively. The benefit is to minimize the fluctuations that are occurring in the temperature/humidity parameters.

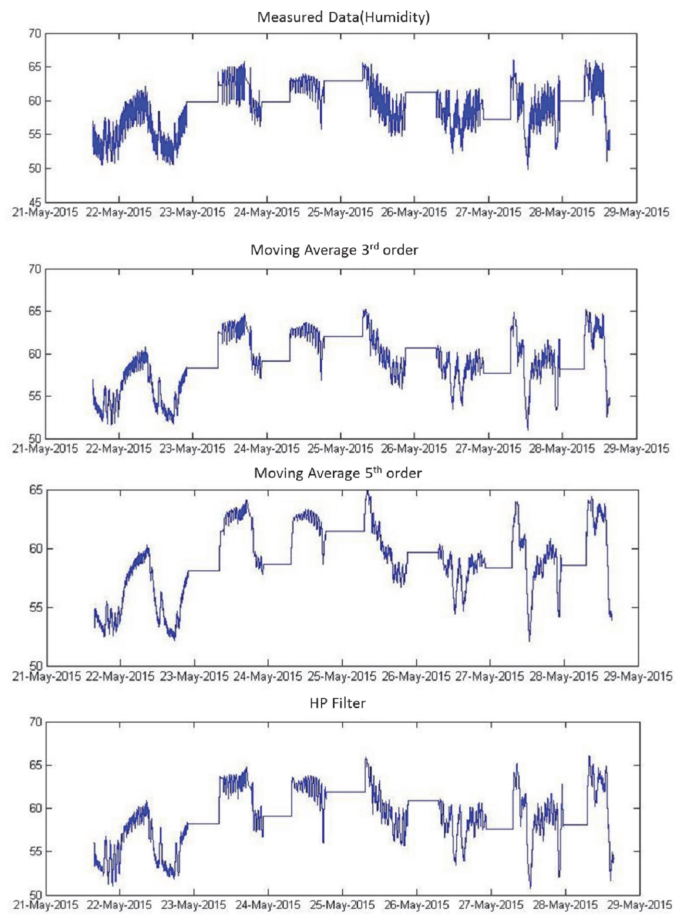

Figure 5 Raw (Measured) and Filtered Data using moving average filter 3th order, 5th order and HP filter for Humidity

The above Figure 5 represents the raw and filtered data plotted for humidity parameter. The graph contains 4 subparts in which 2nd, 3rd and 4th subpart represents filtered data for moving average 3rd order, moving average filter 5th order and HP filter respectively.

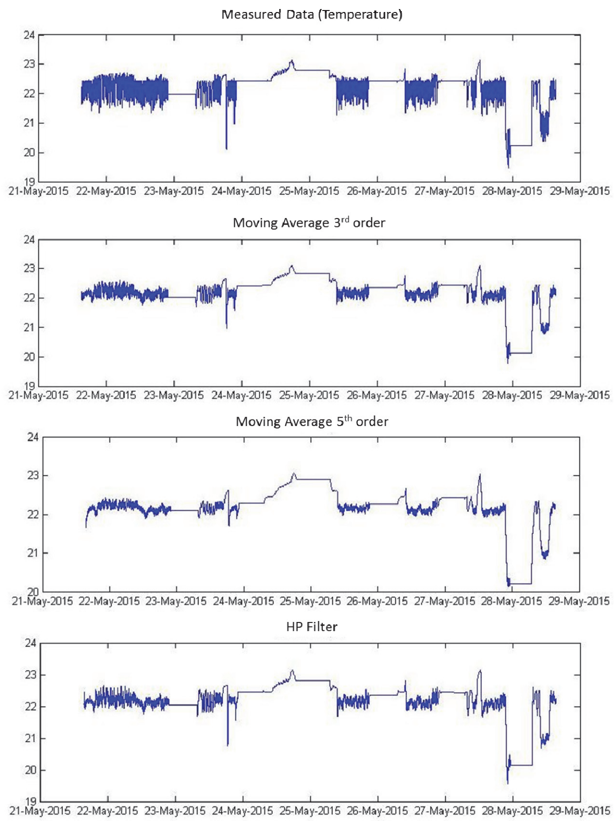

Figure 6 Raw (Measured) and Filtered Data using moving average filter 3[th] order, 5th order and HP filter for Temperature

The above Figure 6 represents the raw and filtered data plotted for temperature parameter. The graph contains 4 subparts in which 2[nd], 3[rd] and 4[th] subpart represents filtered data for moving average 3[rd] order, moving average filter 5[th] order and HP filter respectively.

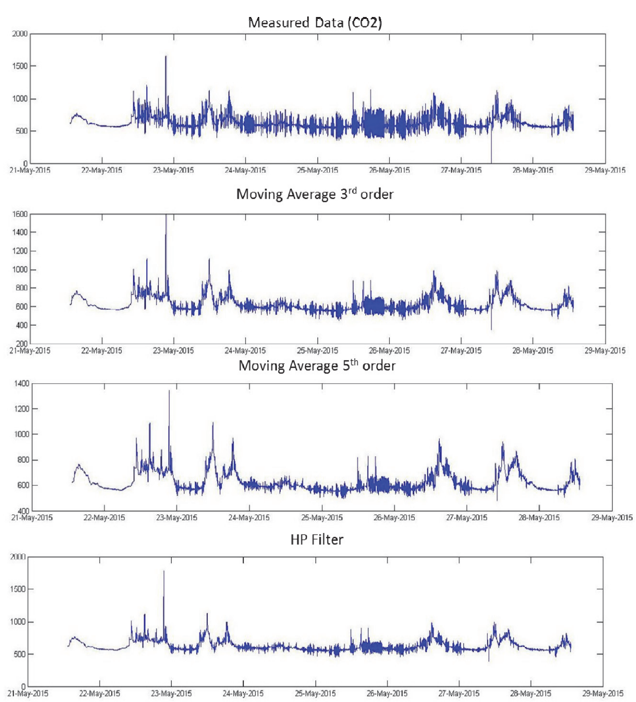

Figure 7 Raw (Measured) and Filtered Data using moving average filter 3[th] order, 5th order and HP filter for CO2 sensor

The above Figure 7 represents the raw and filtered data plotted for humidity parameter. The graph contains 4 subparts in which 2[nd], 3[rd] and 4[th] subpart represents filtered data for moving average 3[rd] order, moving average filter 5[th] order and HP filter respectively.

B. Discussions

Two different of filters have been analysed and are used on real-time database of environmental sensors. In the case of moving average filter, increment of filter order increases the smoothness of the data. Smoothness level can be enhanced further with higher order but at the expense of data loss. Thus, the filter order needs to be designed to compromise between data smoothness and data loss.

Computational time for all the techniques is being considered and the table below shows the time consumed for the filtering of the data.

Table I. Computational time of different filter in millisecond

Parameter	Moving Average (5[th] order)	Moving Average (3[rd] Order)	HP filter
Humidity	11.414499	11.432827	12.045537
Temperature	11.362975	11.691809	11.884996

From the above results, it can be seen that the processing time for the filters are approximately same with moving average filter taking less time than HP filter. Thus, it can be concluded that moving average is better than HP filter since in case of HP as due to intensive computational, the resources are wasted leading to increase in computational time. Also, choosing the factor λ can be tedious task which normally ensures the trade between the goodness of fit and the smoothness of the output data.

CONCLUSIONS

In this paper, a wireless sensor and actuator network in building applications is described. Two algorithms are investigated to analyze the data collected by this WSAN. The results show that the clean outcome can be used to serve the control actions. Future woks may include implementation of these techniques for online filtering, advanced data analysis to understand the behaviors of the buildings and their occupant activities as a part of the building operating systems. Therefore it is expected that further improvement of maintaining human comfort and energy savings can be achieved.

ACKNOWLEDGMENT

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

REFERENCES

[1] Hanne Grindvol, Ovidiu Vermesan, Tracey Crosbie, Roy Bahr, Nashwan Dawood, Gian Marco Revel in "A Wireless Sensor Network For Intelligent Building Energy Management Based On Multi Communication Standards – A Case Study" , ITcon Vol. 17,pages 43-62, 2012

[2]Anuja.P, T.Murugeswari, in "A novel approach towards Building Automation through DALI-WSN Integration" , International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2013

[3] Feng Shu, Malka N. Halgamuge, Wei Chen in "Building Automation Systems Using Wireless Sensor Networks: Radio Characteristics and Energy Efficient Communication Protocols", EJSE Special Issue: Sensor Network on Building Monitoring: from theory to real application, pages 66–73, 2009

[4] The ZigBee Alliance, Control your World, 2015, Available: http://www.zigbee.org/

[5] Texas Instruments, 2015, 6LoWPAN – Overview, Available: http://www.ti.com/lsds/ti /wireless_connectivity/6lowpan/overview.page

[6] Wi- Fi Alliance, 2015, Available: http://www.wi-fi.org/

[7] Hart Communication Protocol and Foundation, Available: http://en.hartcomm.org/

[8] ISA100.11a Technology Standards, Available: http://www.nivis.com/technology/ISA100.11a.php

[9] John, G. H. (1995), Robust linear discriminant trees, in "Fifth International Workshop on Artificial Intelligence and Statistics", Ft. Lauderdale, FL, pp. 285-291

[10] Charu C. Aggarwal , Philip S, Yu, Outlier detection for high dimensional data, ACM SIGMOD Record Volume 30 Issue 2, June 2001 Pages37-46

[11] Suneet Shukla, Jyotika Saxena in "An Efficient Approach for Outlier Detection in Wireless Sensor Network" in IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 14, Issue 5 (Sep. - Oct. 2013), PP 99-103

[12] E. Elnahrawy and B. Nath. 2003. Cleaning and Querying Noisy Sensors. In Proc. of the 2nd ACM International conference on WSNA'03, pages 78-87

[13] Victoria J. Hodge and Jim Austin 2004, a survey of techniques for outlier detection , Artificial Intelligence Review Volume 22, Issue 2 , pp 85-126

[14] Christoph Heinz and Bernhard Seeger. 2006. Statistical Modeling of Sensor Data and its application to Outlier Detection. Technical Report 2006/07, University of Stuttgart; 5. GI/ITG KuVS Fachgespräch "Drahtlose Sensornetze", Stuttgart.

[15] S. Subramaniam, T. Palpanas, D. Papadopoulos, V.Kalogeraki and D. Gunopulos. 2006. Online Outlier Detection in Sensor Data Using Non•Parametric Models. VLDB'06, September 12- 15, Seoul, Korea, pages 187-198.

[16] M. Syed Mohamed, T. Kavitha in "Outlier Detection Using Support Vector Machine in Wireless Sensor Network Real Time Data" in International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-2, May 2011

[17] Understanding Building Automation and Control Systems: http://www.kmccontrols.com/productsUnderstanding_Building_Automation_and_Control_Systems.aspx

[18] Anuja.P, T.Murugeswari, "A novel approach towards Building Automation through DALI-WSN Integration", in International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2013.

[19] PI Data Access,2015, Available: http://www.osisoft.com/software-support/products/PI_Data_Access.aspx

[20] Zuriana Abu Bakar, Rosmayati Mohemad, Akbar Ahmad, Mustafa Mat Deris, "A Comparative Study for Outlier Detection Techniques in Data Mining" in IEEE CIS 2006

[21] Azhar Mahmood, Ke Shi,Shaheen Khatoon, and Mi Xiao "Data Mining Techniques for Wireless Sensor Networks: A Survey" in Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013, Article ID 406316, 24 pages

[22] Morten O. Ravn and Harald Uhlig "On Adjusting the Hodrick-Prescott Filter for the frequency of OBSERVATIONS" The Review of Economics and Statistics, May 2002, 84(2): 371–380

[23] Digital Signal Processing: A Practical Guide for Engineers and Scientists, California Technical Publishing, Steven W. Smith, San Diego, 2002, pp. 277-284