International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

# Statistics-based outlier detection for wireless sensor networks

Y. Zhang , N.A.S. Hamm , N. Meratnia , A. Stein , M. van de Voort & P.J.M. Havinga

**To cite this article:** Y. Zhang , N.A.S. Hamm , N. Meratnia , A. Stein , M. van de Voort & P.J.M. Havinga (2012) Statistics-based outlier detection for wireless sensor networks, International Journal of Geographical Information Science, 26:8, 1373-1392, DOI: 10.1080/13658816.2012.654493

**To link to this article:** https://doi.org/10.1080/13658816.2012.654493

Published online: 27 Feb 2012.

*International Journal of Geographical Information Science*
Vol. 26, No. 8, August 2012, 1373–1392

Taylor & Francis
Taylor & Francis Group

# Statistics-based outlier detection for wireless sensor networks

Y. Zhang[a]*, N.A.S. Hamm[b], N. Meratnia[a], A. Stein[b], M. van de Voort[a] and P.J.M. Havinga[a]

[a]*Pervasive System Group, Department of Computer Science (EWI), University of Twente, Enschede, The Netherlands;* [b]*Department of Earth Observation Science, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands*

(*Received 13 December 2010; final version received 18 December 2011*)

Wireless sensor network (WSN) applications require efficient, accurate and timely data analysis in order to facilitate (near) real-time critical decision-making and situation awareness. Accurate analysis and decision-making relies on the quality of WSN data as well as on the additional information and context. Raw observations collected from sensor nodes, however, may have low data quality and reliability due to limited WSN resources and harsh deployment environments. This article addresses the quality of WSN data focusing on outlier detection. These are defined as observations that do not conform to the expected behaviour of the data. The developed methodology is based on time-series analysis and geostatistics. Experiments with a real data set from the Swiss Alps showed that the developed methodology accurately detected outliers in WSN data taking advantage of their spatial and temporal correlations. It is concluded that the incorporation of tools for outlier detection in WSNs can be based on current statistical methodology. This provides a usable and important tool in a novel scientific field.

**Keywords:** outlier detection; wireless sensor networks; spatial correlation; temporal correlation; time-series analysis; geostatistics



## 1. Introduction

Data acquisition is an issue of ongoing attention for geographical information science. Modern sensors may be mounted on satellites, aircraft, marine or terrestrial platforms. The quality of the acquired data is of central concern. This is addressed, in part, by increasing the sampling frequency in both space and time. A recent development in data acquisition concerns the use of wireless sensors, consisting of nodes that measure environmental variables such as temperature, humidity, sound, pressure, light, vibration or motion (Arampatzis *et al*. 2005). A collection of these devices forms a wireless sensor network (WSN) (Akyildiz *et al.* 2002). These nodes are equipped with sensing, processing, wireless communication and, recently, actuation capabilities (Liu *et al.* 2003). They are able to perform limited local data processing and transmit data via a single-or multi-hop routing to a base station.

WSN applications often require efficient, accurate, real-time analysis in order to facilitate situation awareness and critical decision-making (Roman *et al.* 2008). In this context, situation awareness refers to awareness about the environment. Accurate analysis and

*Corresponding author. Email: zhangy@cs.utwente.nl

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2012.654493
http://www.tandfonline.com

decision-making rely on the quality of sensor data as well as on additional information and context (Klein and Lehner 2009). Raw sensor observations, however, often have low accuracy, due to the limited WSN resources and harsh deployment environments (Zhang *et al.* 2010b). This affects the utility of WSNs for reliable, real-time decision-making and for situation awareness. This often results in outlying observations. In order to make effective use of WSN data, it is necessary to identify the outliers.

In the context of WSNs, outliers are defined as those observations that do not conform to the defined (expected) normal behaviour of the data (Subramaniam *et al.* 2006, Chandola *et al.* 2009). Based on this definition, outliers occurring in WSNs are classified into two different types (Zhang *et al.* 2007b):

- *Errors* refer to observations that deviate significantly from the true state of the measured phenomenon. These inaccurate observations may result from sensor malfunction and need to be corrected or removed.
- *Events* refer to observations that indicate a change in the state of the environment, relative to the predefined 'normal behaviour' (Claramunt and Thriault 1995). They may arise due to a gradual or sudden change in the real world, for example, a change in temperature due to rainfall. Events are interesting to the user and need to be investigated further.

Outlier detection for WSNs aims to identify outliers and to distinguish between errors and events with a high accuracy and with low false positive rates (FPRs). Outlier detection also has to satisfy the WSN resource constraints of communication as well as computational and memory complexity. The challenge is to develop an accurate outlier detection methodology that meets the resource constraints. A highly accurate technique that does not meet the resource constraints would simply be unusable.

The methodology developed in this article exploits spatial and temporal correlations existing in WSN data to define the normal behaviour. It then identifies outliers and distinguishes between errors and events in a distributed and online manner. The methodology is developed on the basis of time-series analysis (Chatfield 2004) and geostatistics (Cressie 1991). It was tested on a publicly available WSN data set collected at the Grand St. Bernard, Switzerland (Ingelrest *et al.* 2010), by using cross-validation as well as by reference to a data set where the observations had been labelled as normal or outlier. The evaluation addressed outlier detection accuracy as well as communication, computational and memory complexity.

## 2. Related work

Outlier detection has attracted much attention in the field of WSNs, and in recent years many outlier detection techniques specifically developed for WSNs have emerged (Zhang *et al.* 2010a). In general, these techniques can be classified as (i) those that do not utilize spatial or temporal correlation in the data and (ii) those that are based on spatial or temporal correlation only or on both.

In the first case Sheng *et al.* (2007) proposed a histogram-based technique to detect distance-based outliers in WSNs. Histogram hints indicating the data distribution, rather than the full set of accumulated data, were transmitted to the base station for centralized processing. Branch *et al.* (2006) proposed a technique based on distance similarity to identify outliers that exchanged a set of representative data among neighbouring nodes.

Zhang *et al.* (2007a), adopting the structure of an aggregation tree to prevent the broadcasting of each node in the network, proposed a distance-based outlier detection technique. Rajasegarar *et al.* (2006) proposed an outlier detection technique based on clustering sensor observations at a node and merging clusters before communicating with other nodes. Although these techniques aimed at reducing WSN communication overhead, they all eventually identified global outliers offline at the base station and are thus unsuitable for local, real-time decision-making and situation awareness. They also ignored the time order of sensor data and failed to predict future values. Furthermore, their experiments used communication overhead as a performance metric to evaluate performance of the proposed techniques, without considering the detection accuracy and computational and memory complexity.

To date, limited research has been undertaken that makes explicit use of spatial and temporal correlation for the purpose of outlier detection in WSNs. Wu *et al.* (2007) proposed an outlier detection technique that employed spatial correlation of the observations existing among neighbouring nodes to distinguish between outliers and event boundaries. Subramaniam *et al.* (2006) proposed an outlier detection technique based on temporal correlation of streaming sensor data where each node identified local outliers if the observations deviated significantly from the temporal correlation model. The accuracy of the above two outlier detection techniques was low because they ignored the temporal and spatial correlation, respectively. Elnahrawy and Nath (2004) and Ni and Pottie (2009) proposed Bayesian-based space–time techniques for fault detection in WSNs. They did not calculate explicitly the spatial and temporal correlations and only assumed the existence of such correlations. Shuai *et al.* (2008) proposed a Kalman filter-based outlier detection technique, which utilized spatial and temporal correlations in the data. This technique achieved optimal estimation of the state of the system with white noise disturbance and did not require much computation and large storage. Their way of modelling spatial correlation, by inverse distance weighting (IDW), however, resulted in a low prediction accuracy because there was no explicit model of the spatial correlation. Moreover, collecting observations from neighbouring nodes at each time epoch caused a high communication overhead.

To authors' knowledge, the research presented in this article is the first attempt to capture efficiently temporal and spatial correlations using time-series analysis and geostatistics for distributed and online outlier detection in WSNs.

## 3. Study site and data description

The WSN deployment investigated in this article was located at the Grand St. Bernard pass, situated between Switzerland and Italy (Ingelrest *et al.* 2010), running northeast–southwest through the Valais Alps at a maximum elevation of 2469 m, with coordinate equal to 45° 52′ 08″ N, 7° 10′ 14″ E.

The set-up consisted of 23 sensor nodes measuring several meteorological parameters during a period of 2 months (September–October 2007) with a sampling frequency of 2 minutes. The nodes were deployed in two clusters separated by approximately 500 m: a small cluster consisting of 5 nodes and a big cluster consisting of 18 nodes. Each cluster had a base station and all nodes within a cluster could communicate directly with each other via radio transmission. Furthermore, each node knew its own location as well as the locations of its nearest neighbours. Figure 1a illustrates the coordinates of the Grand St. Bernard deployment according to the Swiss coordinate system (Swiss Grid).

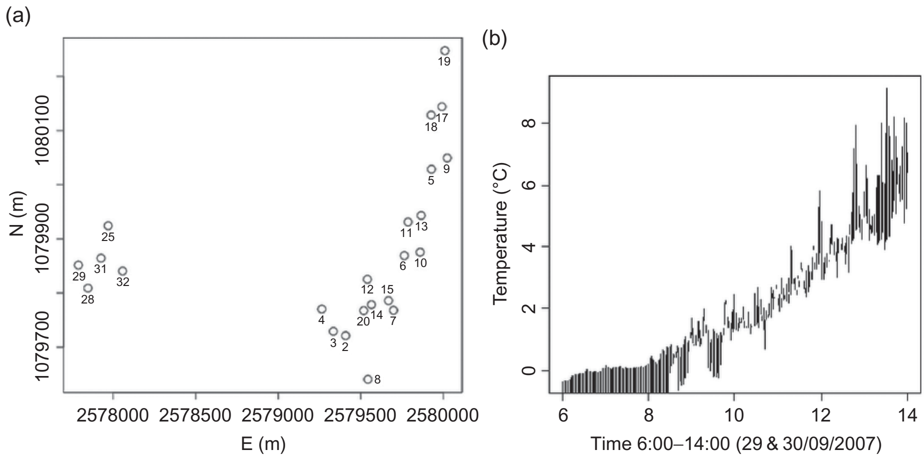


Figure 1. (a) The coordinates of the Grand St. Bernard deployment according to the Swiss coordinate system (Swiss Grid). (b) The similarity of the two data sets from 29 and 30 September at node 31.

The proposed methodology was developed and tested on the small cluster consisting of densely deployed sensor nodes 25, 28, 29, 31 and 32, in which observations were made at the same point in time, specifically for the period 06:00–14:00 on two consecutive days (29 and 30 September 2007) with one attribute, ambient temperature. The range of temperature measurements is –1°C to 10°C and the precision is ±0.3°C. Figure 1b shows the similarity of the two data sets from 29 to 30 September at node 31.

## 4. Methods

A WSN consists of $n$ densely deployed sensor nodes, in which observations are made at (nearly) equal times; all nodes can communicate directly with each other by radio transmission and each node knows its own location as well as the locations of its neighbours. Let $x(s, t)$ denote a sensor observation, where $s$ is the location of a node, and $t$ is the time epoch at which the observation is made, whereas $\hat{x}(s, t)$ denotes a predicted value of $x(s, t)$. The methodology is illustrated in Figure 2.

Section 4.1 presents models for spatial and temporal correlations based on time-series analysis and geostatistics. Section 4.2 proposes distributed and online outlier detection methodologies based on these models. Section 4.3 describes the experimental data set and accuracy assessment techniques.

### *4.1. Modelling spatial and temporal correlations for WSNs*

The usual process of time-series and geostatistical analysis requires expert knowledge as well as a high level of user interaction. The burden of computational and communication complexity is potentially high, especially in the context of WSNs. Resource-efficient solutions are required. These are described in this section.

#### *4.1.1. Modelling temporal correlation*

Time-series analysis involves three major steps (Chatfield 2004): (i) removing the trend and seasonality in order to achieve a stationary time series, (ii) fitting an auto-regressive

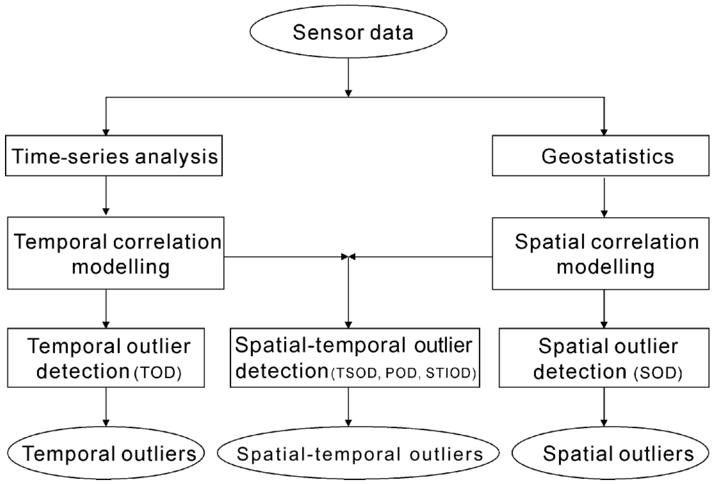


Figure 2. The fundamentals of the developed methodology.



moving average (ARMA) model to the stationary time series and (iii) predicting future values using the ARMA model.

This research was undertaken using data from two consecutive mornings (one to fit the model and one to test the model). A simple technique, *first differencing*, was used to eliminate the trend, resulting in a new time series $\{x'(s,t) = x(s,t) - x(s,t-1)\}$. The diurnal fluctuation was not modelled. Modelling diurnal and seasonal patterns is potentially resource-consuming for WSNs and is left as a topic for future research.

For the second step, the ARMA model was simplified to an AR($p$) model, where $p$ is the number of previous observations. The AR($p$) model implies that the current observation is only correlated with the previous $p$ observations. Moreover, $p$ was kept to the minimum possible to limit the model complexity. The AR($p$) model is formulated as follows:

$$x'(s,t) = \varepsilon(s,t) + \sum_{i=1}^{p} \alpha_i x'(s,t-i) \tag{1}$$

where $\alpha_i = \{\alpha_i : i = 1, \ldots, p\}$ are parameters and $\epsilon(s, t)$ is white noise.

For the third step, the AR model estimated at each node was used together with the previous observations to predict the next observation (based on Equation (1)). A confidence interval of prediction is $[\hat{x}(s,t) - \beta\hat{\sigma}, \hat{x}(s,t) + \beta\hat{\sigma}]$ (Chatfield 2004), where $\hat{\sigma}$ is the prediction standard error and $\beta$ is the coefficient for a given confidence level.

Time-series analysis requires an uninterrupted series; however, WSN data often contain obvious errors or show missing values. *Median smoothing* (Basu and Meckesheimer 2007) was used to replace obvious errors and missing data within a smoothing window, prior to undertaking the above analysis.

### *4.1.2. Modelling spatial correlation*

Geostatistical analysis involves two main steps: (i) modelling spatial correlation by calculating the sample variogram and fitting a model to it and (ii) using the model to predict at unsampled locations. Geostatisticians commonly refer to this prediction as 'kriging', after its inventor (Webster and Oliver 2007).

Variogram modelling typically requires a sample size of at least 100 nodes. The WSN used in this study, however, contained only 23 nodes (see Section 3), therefore the method of Sterk and Stein (1997) was adopted to alleviate this constraint. This method combines observations at different time periods from the limited number of locations to estimate the sample variogram. Its justification is based on the assumption that observations collected at different time periods can be characterized by the same spatial correlation structure. The formula for variogram was modified to

$$\hat{\gamma}(h) = \frac{1}{2n_t(h)} \sum_{t=1}^{m} \sum_{s=1}^{n_t(h)} (x(s,t) - x(s+h,t))^2 \tag{2}$$

where $m$ is the number of different time periods, $h$ is the lag distance and $n_t(h)$ denotes the number of point pairs for each $h$ at time period $t$. In studies where the WSN is large (>100 nodes) the usual method for variogram estimation (Webster and Oliver 2007) can be adopted.

For the second step, a predicted (kriged) value for any location is derived from the weights of its spatial neighbours and their observations, formulated as

$$\hat{x}(s_n, t) = \lambda_1 x(s_1, t) + \cdots + \lambda_{n-1} x(s_{n-1}, t) \tag{3}$$

where $\{x(s_1, t), \ldots, x(s_{n-1}, t)\}$ are the observations at the adjacent locations of $s_n$, and $\{\lambda_1, \ldots, \lambda_{n-1}\}$ are the weights based on the variogram values between $s_n$ and its adjacent locations such that $\sum_{s=1}^{n-1} \lambda_s = 1$.

In the usual ordinary kriging framework, predictions at measurement locations default to the measured value (kriging honours the data). As discussed below, outlier identification requires that it is known whether the kriged value differed significantly from the measured value. Hence, at each measurement location the value was predicted as if the measurement had not been taken. The confidence interval of prediction from Section 4.1.1 was then used for that purpose, where $\hat{\sigma}$ equals the prediction (kriging) standard error. This assumes a Gauss-distributed error. For each node, the weights from its corresponding neighbours need to be calculated. Based on the assumption that the nature of the spatial correlation does not change, this has only to be performed once. This keeps the computation and communication complexity low.

### *4.2. Statistics-based outlier detection techniques*

In this section, the distributed and online outlier detection techniques for WSNs are presented. They are classified into temporal, spatial and spatial-temporal outlier detection depending on the use of temporal and spatial correlations.

#### *4.2.1. Temporal outlier detection*

Temporal outlier detection (TOD) identifies outliers in an online manner using the time-series model. The AR model is used to predict the value at a given point in time, $\hat{x}(s, t)$, together with its confidence interval. A *temporal outlier* is recorded when an observation $x(s, t)$ falls outside the confidence interval of its associated predicted value $\hat{x}(s, t)$. Three additional questions were addressed:

(1) How should $x(s, t)$ be dealt with after identifying it as a normal observation or as an outlier?
(2) How can errors and events be distinguished?
(3) When and how should the time-series model be updated?

For the first question, measurements identified as normal are used directly to predict the next observation in time. Otherwise, after detecting an outlier $x(s, t)$, TOD uses measurements at the previous time instances $\{x(s, t-p), \ldots, x(s, t-1)\}$ and the AR model to predict the next observations in the sequence, as normal or as an outlier. Each new measurement in the sequence is identified as an outlier or as normal upon arrival. Here two possibilities exist:

- If all measurements in the sequence are identified as outliers, the sequence is classified as an event and indicates a change in the normal behaviour of the WSN data. Consequently, the actual measurements, including $x(s, t)$, are used for prediction.
- If only a few measurements in the sequence are detected as outliers, these are labelled as errors and are not used for the next prediction. Instead, the predicted values are used to predict the next observation.

For the second question, the length of the outlier sequence is used to distinguish between errors and events. This is partly a practical problem and depends on the application requirements and sampling rates. The duration of an event is not known beforehand and some events last longer than others. These characteristics complicate the identification of the entire event. Therefore, the aim is to identify changes in the normal behaviour that lead to an event and not to identify the entire event. Furthermore, TOD should not cause a considerable delay in the identification of the type of outlier. This means that the length of the sequence should be small and is set at 5 for defining an event in this article.

TOD is a modification to two typical ARMA prediction approaches. The first, denoted $S_1$, predicts into the future using only the current observations, for example, predict at $t + n$ using data only up to time $t$. Clearly, the further into the future the prediction is made, the lower the confidence in the prediction becomes. The second, denoted $S_2$, predicts at $t + 1$ using the data measured up to and including $t$.

The third question arises because, for many types of observation, the normal behaviour may change over time. For example, the value of meteorological observations may change because the weather changes. It may be necessary to update the time-series model to reflect this. This poses a challenge, because such an update may be memory and processor-intensive. This is not addressed in this article and is left for future research.

#### 4.2.2. *Spatial outlier detection*

Spatial real-data-based outlier detection (SOD) enables each node to identify outliers using only the spatial model. A *spatial outlier* is a measurement $x(s, t)$ that lies outside of the confidence interval of its predicted value $\hat{x}(s, t)$ in the spatial domain.

SOD uses real measurements from spatial neighbours for prediction and for outlier detection. Each node transmits its own observation to all its neighbours at each time instant. Once a node identifies an entire outlier sequence, it sends a notification message to all its neighbours. Upon receipt of a positive confirmation about the occurrence of this event from its neighbours, it confirms the occurrence of an event. Otherwise, it treats the outliers as errors and then uses the predicted values to replace them for the next prediction.

Note that such a frequent data transmission results in a large communication overhead and bandwidth occupation. Moreover, sending and receiving data from all the nodes to all their neighbours could lead to a considerable detection delay.

#### *4.2.3. Spatial-temporal outlier detection (POD, TSOD and STIOD)*

The methodologies described in Sections 4.2.1 and 4.2.2 are incomplete. TOD has insufficient information to distinguish well between errors and events because it identifies outliers in time at a single point in space. Conversely, SOD ignores the temporal context by identifying outliers in space at a single moment in time. This section introduces three spatial-temporal correlation-based outlier detection methodologies.

Temporal and spatial real-data-based outlier detection (TSOD) integrates TOD and SOD. Each node identifies temporal outliers and then checks whether these are also spatial outliers by obtaining neighbours' observations at corresponding time instants. This separately takes advantage of temporal and spatial correlations in the data for outlier detection. Unlike POD and STIOD (described below), actual measurements are used for both the spatial and temporal predictions.

It was expected that TSOD would have a reduced communication overhead as compared to TOD or SOD, although this could still be substantial. For this reason, two alternative approaches were developed and tested, as described below.

Spatial predicted-data-based outlier detection (POD) predicts neighbours' observations for spatial outlier detection without any actual data transmission. First, each node transmits the parameters $\{\alpha_i : i = 1, ..., p\}$ of its own AR model (based on Equation (1)) to its neighbours. Once each node receives these parameters from its neighbours, it first uses them together with its own previous observations (based on Equation (1)) to predict the current values for its neighbours. Afterwards, each node uses the newly predicted value of each of its neighbours together with their corresponding weights (based on Equation (3)) to predict its own current value. Accordingly, those actual measurements from each node that lie outside the confidence interval of the predicted value were considered as outliers.

Spatial and temporal integrated outlier detection (STIOD) integrates temporal and spatial correlations for outlier detection. The following are the main steps of STIOD:

(1) The spatial correlation is assessed and the weight for each node is calculated and sent to the nodes. As a result, each sensor node $s_j$ obtains the corresponding weights of its $n - 1$ spatial neighbours $\{\lambda_{js} : s = 1, ..., n - 1\}$.
(2) Each sensor node $s_j$ models its temporal correlation using the time-series analysis and obtained parameters $\{\alpha_{ji} : i = 1, ..., p\}$ of the temporal model and then sends these parameters to its neighbours.
(3) Each node combines the temporal correlation parameters of its neighbours with their weights. The integrated parameters are denoted $\left\{\sum_{s=1}^{n-1} \alpha_{si}\lambda_{js} : i = 1, ..., p\right\}$, representing the spatial integration.
(4) The parameters derived from Step (3) are further integrated with each node $s_j$ itself. The complete integrated parameters are denoted as $\left\{\frac{\alpha_{ji}+\sum_{s=1}^{n-1} \alpha_{si}\lambda_{js}}{2} : i = 1, ..., p\right\}$, integrating both spatial and temporal correlations.
(5) Each node uses the integrated parameters derived from Step (4) together with its own previous observations to predict its next observation (based on Equation (1)). It then compares the predicted value with its actual observation and identifies the actual observation as outlier or normal.

Afterwards, STIOD identifies outliers and distinguishes between errors and events in real time using the same strategy as TOD.

### 4.3. Evaluation methods

The data and study site are described in Section 3. Two consecutive mornings (06:00–14:00), 29 and 30 September, were selected. The data from 29 September were used to estimate the parameters of the time-series and of the geostatistical models. The time-series and geostatistical predictions were then evaluated using cross validation, as described in Section 4.3.1. The outlier detection methodologies were evaluated against the data from 30 September as described in Section 4.3.2.

#### 4.3.1. Cross validation

Cross validation is the simplest and most widely used method for estimating prediction errors (Webster and Oliver 2007) compared to other methods, for example, bootstrap (Efron 1979), and was used in this study. For leave-one-out cross-validation (LOOCV) (Webster and Oliver 2007), a node's observation is left out from the data set, and all other nodes' observations are used to estimate the variogram. This estimated variogram model is then used to predict the observations at the left-out node. This procedure is repeated until each node is left out once. For the time-series model, each observation was predicted using its AR model, defined in Equation (1). For both the spatial and temporal models, the difference between the measured and predicted value is the error, $e_i = \hat{x}_i - x_i$, where $i$ denotes a specific observation.

The mean prediction error ($\text{MPE} = \sum e_i/n$) and root mean squared error ($\text{RMSE} = \sum e_i^2/n$) are the two main metrics of cross validation. Ideally the MPE should be 0, which means that the prediction is unbiased. The RMSE is a measure of accuracy, hence the lower, the better.

#### 4.3.2. Outlier detection accuracy

The detection rate (DR) and false positive rate (FPR) were used as metrics of detection accuracy. DR indicates the percentage correctly detected as a proportion of the total number of true outliers. The FPR, also known as false alarm rate, is the percentage of normal data that are incorrectly detected as outliers. An effective outlier detection technique should achieve a high DR and low FPR.

In order to calculate the DR and FPR, a reference data set was necessary. To obtain this, every observation in the data set needed to be labelled as either normal or an outlier. No general purpose labelling technique exists, so the data were labelled based on the running average, Mahalanobis distance and density.

- The *running average-based labelling technique* uses a smoothing window and calculates the mean value for a fixed sample size. An outlier is defined by taking the absolute value of the difference between the measurements and the values calculated by applying the running average. Each measurement above the critical threshold is considered an outlier. The median instead of the mean is used to calculate the threshold, in order to minimize the influence of the outliers.
- The *Mahalanobis distance-based labelling technique* identifies outliers based on the measure of full dimensional distance between a point and its nearest neighbour in the

data set. Using the Mahalanobis distance to label the data, an outlier is considered to be a measurement whose Mahalanobis distance is larger than a certain threshold. This threshold is defined as the average value of the Mahalanobis distance values.
- The *density-based labelling technique* uses the local density to search for outliers and identifies local outliers in data sets with diverse clusters. A measurement is an outlier if it resides in an area of the grid whose density is lower than a fixed percentage of the density values.

A detailed description and discussion of the labelling techniques for WSNs can be found in Zhang (2010).

### *4.4. Software*

The R software environment for statistical computing (version 2.8.1) (R Development Core Team 2010) was used for the analysis. In particular, gstat (Pebesma 2004) was used for geostatistics and stats (specifically the ts function) was used for AR modelling (Jones *et al.* 2009).

## 5. Results

Section 5.1 presents the cross-validation evaluation of the predictions from the time-series and geostatistical models. Section 5.2 then evaluates the outlier detection methodologies.

### *5.1. Prediction accuracy*

LOOCV yielded an MPE of 0.015 and an RMSE of 0.8. For the time-series model, the MPE was 0.05 and RMSE was 0.4. The low MPE indicates that the models were unbiased and the low RMSE indicates that the models were accurate.

### *5.2. Outlier detection accuracy*

#### *5.2.1. Temporal correlation-based outliers*

The effects of several important parameters for TOD were examined . These parameters included the size of the smoothing window, the order $p$ of the AR($p$) model and the value of the confidence level. In the experiments, the size of the smoothing window was assigned values from {15, 30, 48, 60}, the order of the AR($p$) model varied between {1, 2, 3, 4} and the confidence level ranged from {90%, 95%, 99%, 99.7%}.

Table 1 shows the DR and FPR for temporal outliers using the three labelling techniques for different width smoothing windows. A wider smoothing window resulted in a lower DR whereas the FPR reduced slightly for all the three labelled data sets. The size of the smoothing window influenced the original data structure, resulting in a less accurate outlier detection. To ensure reliable outlier detection results for all the three labelling techniques, the size of the smoothing window was set to 15 (30 minutes) in the remaining experiments.

Table 2 shows the DR and FPR for temporal outliers using the three labelling techniques for different orders of the AR($p$) model. Increasing $p$ resulted in increased accuracy. The greatest increase in accuracy was observed when $p$ was increased from 1 to 2, whereas the increase in accuracy for larger values of $p$ was low. A larger value of $p$ resulted in more

Table 1. DR (%) and FPR (%) of TOD for different SW sizes.

| Labelling technique | TOD | SW = 15 | SW = 30 | SW = 48 | SW = 60 |
|---|---|---|---|---|---|
| Running average | DR (%) | 74.2 | 73.0 | 67.4 | 57.3 |
| | FPR (%) | 10.6 | 10.9 | 9.5 | 8.3 |
| Mahalanobis distance | DR (%) | 82.9 | 82.9 | 82.9 | 71.4 |
| | FPR (%) | 13.3 | 13.5 | 11.7 | 10.1 |
| Density | DR (%) | 100 | 100 | 100 | 100 |
| | FPR (%) | 14.9 | 15.0 | 13.2 | 11.4 |

Notes: DR, detection rate; FPR, false postive rate; TOD, temporal outlier detection; SW, smoothing window.
The 95% confidence interval was used for this evaluation.

Table 2. DR (%) and FPR (%) of TOD for different orders of the AR($p$) model.

| Labelling technique | TOD | $p = 1$ | $p = 2$ | $p = 3$ | $p = 4$ |
|---|---|---|---|---|---|
| Running average | DR (%) | 67.4 | 73.0 | 73.0 | 74.2 |
| | FPR (%) | 9.8 | 11.0 | 11.0 | 10.6 |
| Mahalanobis distance | DR (%) | 82.9 | 82.9 | 82.9 | 82.9 |
| | FPR (%) | 12.0 | 13.6 | 13.6 | 13.3 |
| Density | DR (%) | 100 | 100 | 100 | 100 |
| | FPR (%) | 13.6 | 15.1 | 15.1 | 14.8 |

Notes: DR, detection rate; FPR, false postive rate; TOD, temporal outlier detection.
The 95% confidence interval was used for this evaluation.

previous observations that were used in the AR model. A value of $p = 2$ was selected as a trade-off between increased accuracy and increased complexity.

Table 3 shows the DR and FPR of TOD using the three labelling techniques for different values of the confidence level. A relatively low confidence level resulted in a high DR and high FPR based on all three labelled data sets. High confidence levels, however, led to a low FPR and also to a low DR. The reason was that more outliers were included in the confidence interval if a high confidence level was used and these were then identified as being normal. Clearly, the opposite was true for the low confidence level. Hence, in the subsequent experiments, the confidence level was set to 95% (two standard errors).

The performance of prediction strategy of TOD was compared with two usual ARMA prediction strategies $S_1$ and $S_2$ (see Section 4.2.1). Figure 3 illustrates the detected temporal outliers by applying TOD, $S_1$ and $S_2$ on the running average-based labelling technique. Table 4 shows the DR and FPR for temporal outliers detected by TOD, $S_1$ and $S_2$. It can be seen that $S_1$ and $S_2$ had a low accuracy, because predictions at each step resulted in a high

Table 3. DR (%) and FPR (%) of TOD for different confidence levels.

| Labelling technique | TOD | CL = 90% | CL = 95% | CL = 99% | CL = 99.7% |
|---|---|---|---|---|---|
| Running average | DR (%) | 78.7 | 74.2 | 51.7 | 42.7 |
| | FPR (%) | 17.9 | 10.6 | 5.7 | 2.5 |
| Mahalanobis distance | DR (%) | 91.4 | 82.9 | 74.3 | 62.9 |
| | FPR (%) | 20.3 | 13.3 | 7.1 | 3.8 |
| Density | DR (%) | 100 | 100 | 100 | 100 |
| | FPR (%) | 22.0 | 14.8 | 8.5 | 5.0 |

Note: DR, detection rate; FPR, false postive rate; TOD, temporal outlier detection.

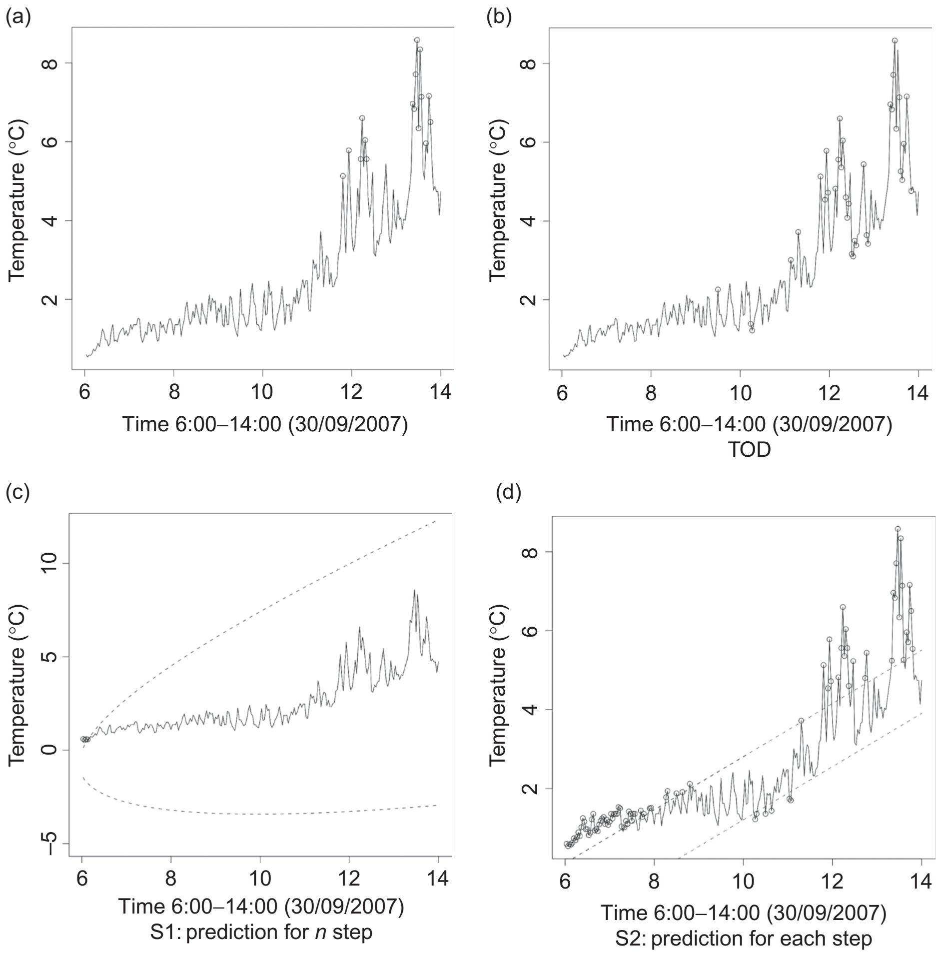


Figure 3. (a) Labelled data using running average-based labelling technique at node 29. (b) Temporal outliers detected at node 29 by TOD. (c) Temporal outliers detected at node 29 by $S_1$. (d) Temporal outliers detected at node 29 by $S_2$. Dashed lines illustrate the upper and lower bounds of the predicted values.
Note: TOD, temporal outlier detection.

FPR. Both $S_1$ and $S_2$ ignored the classification of new observations as normal or outlier. TOD achieved a lower FPR for the three labelling techniques.

Table 5 shows the number of outliers and events detected at nodes using TOD. As described in Sections 1 and 4.2.1, an event is a particular type of outlier. There were apparently fewer events than the total number of outliers, whereas the other outliers were classified as errors. Notice that each node detected a similar number of events relative to the nearby nodes.

### 5.2.2. *Spatial correlation-based outliers*

Equation (2) was used to calculate the sample variogram using data from all 23 nodes for each hour from 06:00 to 14:00 on 29 September. The sample variogram together with the fitted exponential model is shown in Figure 4. The weights, $\lambda$, for prediction were then

Table 4. DR (%) and FPR (%) for temporal outliers using three labelling techniques for prediction strategies.

| Labelling technique | | TOD | $S_1$ | $S_2$ |
|---|---|---|---|---|
| Running average | DR (%) | 72.3 | 1.1 | 95.5 |
| | FPR (%) | 10.5 | 5.2 | 41.6 |
| Mahalanobis distance | DR (%) | 100 | 2.9 | 100 |
| | FPR (%) | 15.0 | 5.0 | 43.9 |
| Density | DR (%) | 100 | 14.3 | 100 |
| | FPR (%) | 15.1 | 4.9 | 45.3 |

Notes: DR, detection rate; FPR, false postive rate.
The 95% confidence interval was used for this evaluation.



Table 5. Number of outliers and events detected at different nodes using TOD.

| Nodes | Number of outliers | Number of events |
|---|---|---|
| Node 25 | 42 | 5 |
| Node 28 | 47 | 5 |
| Node 29 | 35 | 4 |
| Node 31 | 36 | 5 |
| Node 32 | 24 | 2 |

Note: TOD, temporal outlier detection.

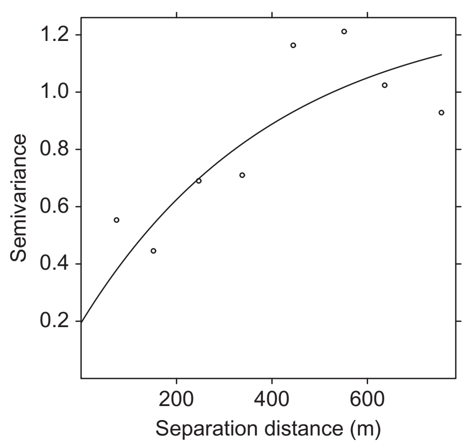


Figure 4. Sample variogram for all 23 nodes for 06:00–14:00, 29 September calculated according to Equation (2). The line shows the fitted exponential model, and partial sill = 1, range = 550, nugget = 0.2.

calculated and used in SOD, which was applied to 30 September. Figure 5a illustrates the spatial outliers detected at node 32, that is, using SOD in the small cluster. These outliers were identified because they lay outside the confidence interval of their predicted values in the spatial domain. Table 6 shows the DR and FPR for SOD. As for TOD, it shows a 100% DR for the Mahalabonis distance and density-based labelling technique, although the FPR was much lower (<5% compared to 15% for TOD). When assessed using the running average labelling technique, the accuracy was low for both metrics.

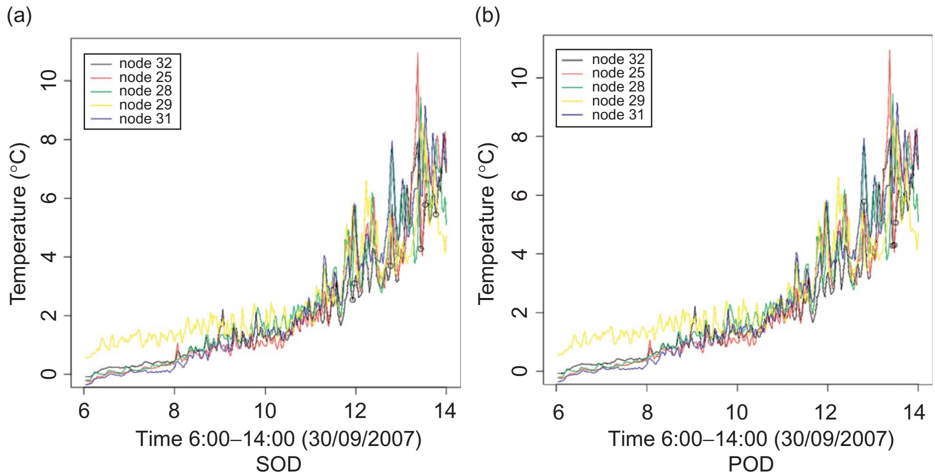


Figure 5. (a) Spatial outliers detected at node 32 by SOD in the small cluster. (b) Spatial-temporal outliers detected at node 32 by POD in the small cluster.
Note: SOD, spatial outlier detection; POD, spatial predicted outlier detection.

Table 6. DR (%) and FPR (%) for outliers using three labelling techniques for all five proposed techniques.

| Techniques | | Running average | Mahalanobis distance | Density |
|---|---|---|---|---|
| TOD | DR (%) | 72.3 | 100 | 100 |
| | FPR (%) | 10.5 | 15.0 | 15.1 |
| SOD | DR (%) | 24.5 | 100 | 100 |
| | FPR (%) | 3.3 | 4.6 | 4.7 |
| TSOD | DR (%) | 23.4 | 100 | 100 |
| | FPR (%) | 1.7 | 3.0 | 3.1 |
| POD | DR (%) | 29.8 | 80 | 75 |
| | FPR (%) | 1.8 | 3.7 | 3.8 |
| STIOD | DR (%) | 71.3 | 100 | 100 |
| | FPR (%) | 10.8 | 15.2 | 15.3 |

Note: DR, detection rate; FPR, false positive rate; TOD, temporal outlier detection; SOD, spatial outlier detection; TSOD, temporal and spatial outlier detection; POD, spatial predicted outlier detection; STIOD, spatial and temporal integrated outlier detection.

### *5.2.3. Spatial-temporal correlation-based outliers (TSOD, POD and STIOD)*

Table 6 shows the DR and FPR for TSOD, evaluated against all the three labelling techniques. As with TOD and SOD, it showed a 100% DR for the Mahalabonis distance and density-based labelling technique, although the FPR was the lowest (~3%). When assessed using the running average labelling technique, the accuracy was low for both metrics.

Figure 5b illustrates the outliers detected by SOD and POD at node 32. Table 6 shows that, in all cases, the accuracy of POD was low.

Finally, Figure 6 illustrates the performance of STIOD by evaluating the results of TSOD and STIOD against a data set labelled by the Mahalanobis distance-based labelling technique at node 28. From Table 6, it can be seen that STIOD achieved a comparable accuracy to TOD but, with exception of the running average labelling technique, lower accuracy than TSOD.

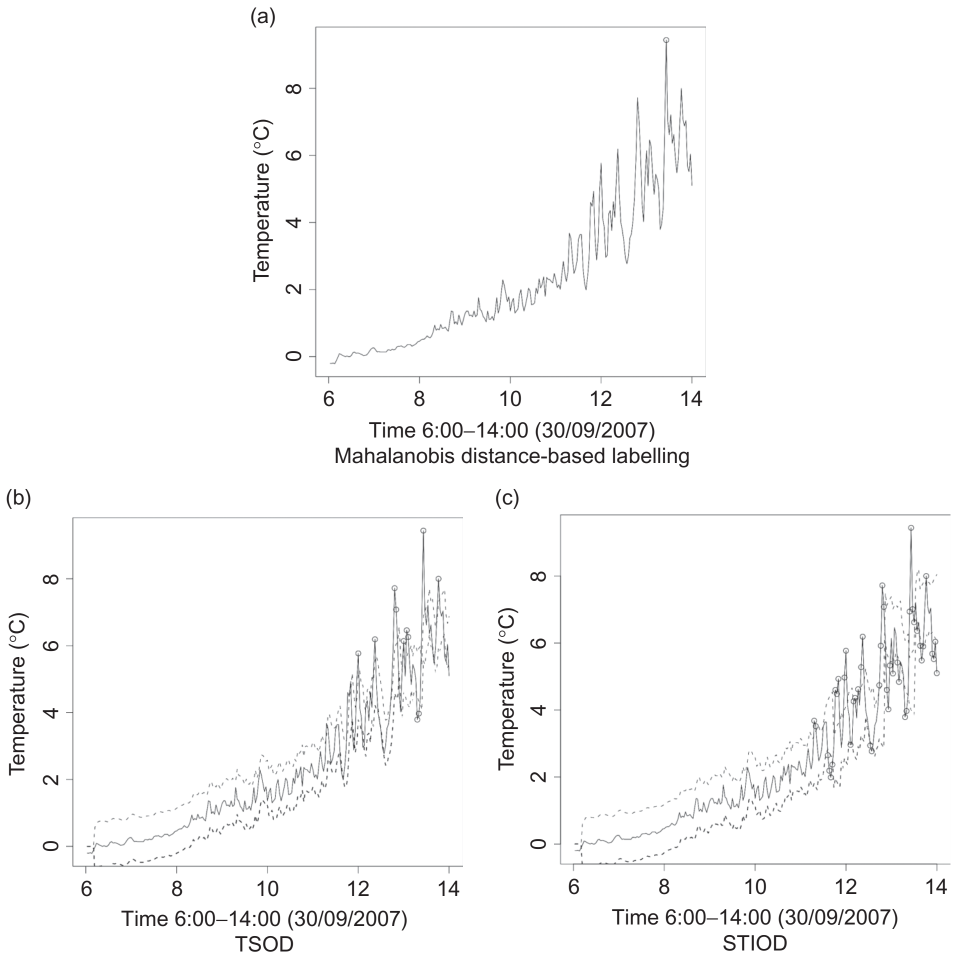


Figure 6. (a) Labelled data using Mahalanobis distance-based labelling technique at node 28. (b) Spatial-temporal outliers detected at node 28 by TSOD. (c) Spatial-temporal outliers detected at node 28 by STIOD.

## 6. Discussion

Section 6.1 discusses accuracy assessment based on the experimental results presented in Section 4. Section 6.2 then discusses the model complexity. This is important because both accuracy and complexity need to be considered. Finally Section 6.3 discusses other open issues.

### *6.1. Accuracy assessment*

Cross validation was used to evaluate the time-series and geostatistical models (Sections 4.3.1 and 5.1). This shows the models to be unbiased with high accuracy. Therefore, these models were appropriate for use in subsequent outlier detection methodologies.

Accuracy assessment is an important component of the analysis of geographic data (Fisher 1999). In order to assess the accuracy of outlier detection, a reference data set is required. No separate reference data set was available for this research so the reference data were generated using an *a posteriori* labelling of the data using the three techniques.

The running average-based technique labels whether an observation is an outlier depending on the surrounding values on the time axis, rather than the full range of values in the data set. TOD and STIOD, which use previous observations together with the temporal correlation model to identify outliers, achieved the highest DR, although FPR was relatively high. In contrast, SOD and POD, which are designed to identify spatial outliers occurring at each time instant, had a very low DR and a low FPR. TSOD also had a low DR because those temporal outliers that were detected by TOD were not also detected when SOD was applied; however, its FPR was low. The low DR for SOD, TSOD and POD may be explained by the fact that these methodologies operated in the spatial domain, whereas the labelling technique worked in the temporal domain. The conflict between DR and FPR made it difficult to judge which model was most accurate, although clearly none had a high accuracy.

The Mahalanobis distance-based and density-based techniques labelled outliers purely depending on the range of the values in the data set and ignored the temporal order in the data. Both the labelling techniques led to almost identical results for the DR and FPR for all the five outlier detection methodologies. With the exception of POD, all five models achieved a DR of 100% and could only be distinguished by the FPR. TSOD achieved the lowest FPR (~3%), although SOD was only slightly higher (~4.5%). Clearly TSOD uses most information (both spatial and temporal) to identify outliers and this led to the highest accuracy. The reason for the much higher FPR for TOD than SOD is unclear, but may be due to the fact that SOD used the five nodes (four neighbours + itself) to identify outliers whereas TOD used only a single node. As such TOD uses less information than SOD. POD and STIOD attempt to further reduce node-to-node communication by using predicted values rather than actual measurements. Hence they use less information and were less accurate than the other methodologies.

According to the running average-based labelling, all the models had low accuracy. According to the Mahalabonis distance-based and density-based labelling, TSOD had the highest accuracy, with SOD having a slightly lower accuracy. Apparently, the choice of which outlier detection techniques to use is dependent on which labelling technique is used. For this research, the Mahalanobis distance-based and density-based labelling are preferred over the running average-based technique. This is because they used all the data and label outliers based on *a posteriori* data analysis. Furthermore, since the running-average-based technique works only in the time domain, it is not well suited to the assessment of spatial outliers.

### *6.2. Model complexity*

The analysis of the WSN data is usually undertaken under high resource constraints. This is relevant when assessing the quality of an outlier detection model. A highly accurate, but resource-hungry, method is of little practical use. The complexity of the model includes communication overhead and computation and memory complexity. This is a key issue that distinguishes WSN data analysis from other scenarios, where analysis is performed offline and not in real time. In those latter situations, model complexity is less important, whereas for WSNs it is critical.

The communication complexity of the five models depends on the local transmission of model parameters and actual observations, required for spatial prediction (kriging). TOD requires no communication overhead because the analysis was performed locally, at each node. For SOD, each node sends its own observation at each time interval. The maximum communication overhead for each node is $O(m \cdot d)$, where $m$ is the number of new observations to be classified and $d$ is the number of variables. For POD and STIOD, each node transmits the parameters of the temporal correlation model once only, thus the maximum communication overhead for each node is $O(n \cdot d)$, with $n$ the number of adjacent nodes. For TSOD, each node needs to send its own observation when an observation is identified as a temporal outlier, hence the maximum communication overhead for each node is $O(m' \cdot d)$, where $m'$ with $(m' < m)$ is the number of detected temporal outliers at each node.

The computational complexity of TOD depended mainly on fitting the AR model and is represented as $O(p)$. Hence, the maximum computational complexity at each node in TOD is $O(c \cdot d \cdot p)$, where $c$ is the number of original observations to be modelled. The computational complexity in SOD and POD depended mainly on fitting the variogram and the computation of weights for spatial neighbours, represented as $O(q)$. Hence, the maximum computational complexity of each node is $O(c \cdot d \cdot q)$. The maximum computational complexity of each node in TSOD and STIOD is $O(c \cdot d \cdot (p + q))$.

The memory complexity of the five models arose mainly due to keeping observations in memory and is represented as $O(m \cdot d)$. The overhead of storing the parameters of temporal and spatial correlation was negligible. Hence, the maximum memory complexity of each node for each model is the same.

Table 7 allows the assessment of the complexity of each model. It has become clear that the key differentiating factor was the communication complexity. TOD carries no communication overhead but also it did not allow incorporation of spatial information into the analysis and yielded inaccurate results. SOD has an extremely high communication overhead that would be unsustainable for most WSNs. Hence, despite giving accurate results the model was not of a high quality. TSOD provides a compromise in this respect, since communication was limited to instances when a temporal outlier was detected. POD and STIOD aim to further reduce the communication complexity by replacing actual observations with predicted values, computed on the node. Their low level of accuracy, however, shows that these models are of low quality.

The outcome of the above discussion is that incorporating spatial data into outlier detection increased the overall complexity, specifically arising from communication. Incorporating temporal correlation helped to reduce this. Thus, it is concluded that TSOD performed best, since it provided a compromise between accuracy and complexity.

Table 7. Complexity analysis of our outlier detection techniques for each sensor node.

| Techniques | Communication complexity | Computational complexity | Memory complexity |
|---|---|---|---|
| TOD | – | $O(cdp)$ | $O(md)$ |
| SOD | $O(md)$ | $O(cdq)$ | $O(md)$ |
| TSOD | $O((m' d))$ | $O(cd(p + q))$ | $O(md)$ |
| POD | $O(nd)$ | $O(cdq)$ | $O(md)$ |
| STIOD | $O(nd)$ | $O(cd(p + q))$ | $O(md)$ |

Notes: TOD, temporal outlier detection; SOD, spatial outlier detection; TSOD, temporal and spatial outlier detection; POD, spatial predicted outlier detection; STIOD, spatial and temporal integrated outlier detection.

### *6.3. Other open issues*

The models presented in this article enabled each node to identify outliers in an online real-time manner. Furthermore, they allowed different types of outliers to be distinguished as errors and events. This is illustrated for TOD in Table 5. However, the evaluation of the five methodologies was performed only on the basis of outlier detection, and not on the ability to distinguish between errors and events. Developing a labelling technique that can distinguish between different types of outliers is an open topic for future research. One solution would be to label outliers that occur in a consecutive time sequence as events, whilst isolated outliers detected by a node would be labelled as errors.

For this article, the reference data were computed using three labelling techniques. Other labelling techniques exist as well (Elnahrawy and Nath 2004; Muthukrishnan *et al.* 2004). The choice for a particular labelling technique may lead to different conclusions about the accuracy of the analytical modelling technique. Choosing an appropriate labelling technique for a particular data set and modelling objective is important and requires further work.

The research presented in this article further advances the WSN analysis of temperature data. Previous work (Rajasegarar *et al.* 2007, 2008) required knowledge of all time series and performed outlier detection offline. Furthermore, these papers failed in detecting changes between two consecutive time series (Pokrajac *et al.* 2007) in real time. In contrast, the techniques proposed in this article detected outliers upon arrival of a new observation and also solved the problem of occurrence of missing values that could be replaced by predicted values. The proposed outlier detection techniques for WSNs could in principle be extended to other numerical data, for example, humidity, soil moisture, depending on the application constraints and accuracy requirements.

Finally, the research in this article identified outliers and classified them into errors and events, on a per-node basis. It may be useful to monitor the development of an event in space and time. For example, the user might be interested in the development of a fire or rainfall event. Monitoring the spatial-temporal evolution of an event is an area for future research. Granularity, recognized as spatial and temporal resolution, is an important issue that relates to this. In this study there was a fine temporal sampling interval (2 minutes) and a sparse sampling interval in space (23 measurements in a 500 m $\times$ 2000 m area). The variogram showed that the spatial sampling interval was dense enough to allow use of the spatial autocorrelation, since the observations in the small cluster were separated by less than 100 m and the variogram range was 550 m. This might not be the case in every study, and the user should consider the sample density next to the empirically derived correlations as well as their definition of outliers and events. In particular, the spatial and temporal resolution affects the ability to monitor the spatial-temporal evolution of an event.

## 7. Conclusion

In this article, five different distributed, online statistics-based outlier detection techniques for WSNs were proposed. These were assessed in terms of their accuracy and complexity. The first (TOD) identified temporal outliers at each node, whilst two others (SOD and POD) identified spatial outliers. Finally, spatial and temporal modelling were combined (TSOD and STIOD). Experimental results showed that TOD had the lowest communication complexity but yielded inaccurate results. SOD gave accurate results but had an extremely high communication overhead. POD and STIOD were specifically designed to further reduce communication, but were less accurate.

The final preferred technique, TSOD, enabled each node to accurately identify outliers, to detect changes in the normal behaviour of the data and to forecast observations whilst appropriately handling detected outliers. TSOD still carried a communication overhead, but this is unavoidable if incorporation of the spatial dimension is required.

The analysis highlighted the importance of using an appropriate labelling technique to generate the reference data set. The identification of generic guidance for labelling, given a specific data set and modelling objective, remains an open area for research. Further work should also focus on accuracy assessment for distinguishing between errors and events.

## Acknowledgements

This work was supported by the EU's Seventh Framework Programme in the context of the SENSEI project. This spatial and temporal analysis was developed when Y. Zhang was a PhD-intern in the Department of Earth Observation Science at the International Institute for Geo-Information Science and Earth Observation (ITC).



## References

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., 2002. A survey on sensor networks. *IEEE Communications Magazine*, 40 (8), 102–114.

Arampatzis, T., Lygeros, J., and Manesis, S., 2005. A survey of applications of wireless sensors and wireless sensor networks. *Proceedings of the 13rd Mediterranean conference on control and automation*, 27–29 June 2005. Limassol: Cyprus, 719–724.

Basu, S. and Meckesheimer, M., 2007. Automatic outlier detection for time series: an application to sensor data. *Journal of Knowledge and Information Systems*, 11 (2), 137–154.

Branch, J., Szymanski, B., Giannella, C., and Wolff, R., 2006. In-network outlier detection in wireless sensor networks. *Proceedings of the 26th IEEE international conference on distributed computing systems*.

Chandola, V., Banerjee, A., and Kumar, V., 2009. Anomaly detection: a survey. *ACM Computing Surveys*, 41 (3), 1–58.

Chatfield, C., 2004. *The analysis of time series: an introduction*. London, Chapman and Hall/CRC.

Claramunt, C. and Thriault, M., 1995. Managing time in GIS: an event-oriented approach. *In*: J. Clifford and A. Tuzhilin, eds. *Recent advances on temporal databases*. Zurich: Springer-Verlag, 23–42.

Cressie, N.A.C., 1991. *Statistics for spatial data*. New York, John Wiley & Sons.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7 (1), 1–26.

Elnahrawy, E. and Nath, B., 2004. Context-aware sensors. *In*: H. Karl, A. Willig and A. Wolisz, eds. *Wireless sensor networks: first European workshop, EWSN 2004*. Berlin: Springer, 77–93.

Fisher, P.F., 1999. Models of uncertainty in spatial data. *In*: P.A. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind, eds. *Geographical information systems: principles, techniques, management*. Chichester: John Wiley & Sons, 190–206.

Ingelrest, F., Barrenetxea, G., Schaefer, G., Vetterli, M., Couach, O., and Parlange, M., 2010. SensorScope: application-specific sensor network for environmental monitoring. *ACM Transactions on Sensor Networks*, 6 (2), 1–32.

Jones, O., Maillardet, R., and Robinson, A., 2009. *Introduction to scientific programming and simulation using R*. Boca Raton: Chapman and Hall/CRC.

Klein, A. and Lehner, W., 2009. Representing data quality in sensor data streaming environments. *Journal on Data and Information Quality*, 1 (2), 1–28.

Liu, J., Chu, P., Liu, J., Reich, J., and Zhao, F., 2003. State-Centric programming for sensor-actuator network systems. *IEEE Pervasive Computing*, 2 (4), 50–62.

Muthukrishnan, S., Shah, J., and Vitter, S., 2004. Mining deviants in time series data streams. *IEEE proceedings of the 16th international conference on scientific and statistical database management (SSDBM'04)*, IEEE Computer Society, 21–23 June 2004. Greece: Santorini, 41–50.

Ni, K. and Pottie, G., 2009. Sensor network data fault detection using hierarchical Bayesian space-time modeling. *Technical report*, TR-69, University of California.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*, 30, 683–691.

Pokrajac, D., Lazarevic, A., and Latecki, L.J., 2007. Incremental local outlier detection for data streams. *Proceedings of the IEEE symposium on computational intelligence and data mining*, IEEE Computer Society, 1–5 April 2007. Hawaii: 504–515.

Rajasegarar, S., Leckie, C., Palaniswami, M., and Bezdek, J.C., 2006. Distributed anomaly detection in wireless sensor networks. *Proceedings of IEEE international conference on communications*, IEEE Computer Society, 30 Oct–1 Nov 2006. Singapore: 1–5.

Rajasegarar, S., Leckie, C., Palaniswami, M., and Bezdek, J.C., 2007. Quarter sphere based distributed anomaly detection in wireless sensor networks. *Proceedings of IEEE international conference on communications*, IEEE Computer Society, 24–28 June 2007. Glasgow: 3864–3869.

Rajasegarar, S., Leckie, C., and Palaniswami, M., 2008. CESVM: centered hyperellipsoidal support vector machine based anomaly detection. *Proceedings of IEEE international conference on communications*. IEEE Computer Society, Beijing, China: 1610–1614.

R Development Core Team, 2010. *R: A language and environment for statistical computing* [online]. Vienna, Austria, R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available from: http://www.R-project.org. 26 June 2011.

Roman, R., Lopez, J., and Gritzalis, S., 2008. Situation awareness mechanisms for wireless sensor networks. *IEEE Communications Magazine*, 46 (4), 102–107.

Sheng, B., Li, Q., Mao, W., and Jin, W., 2007. Outlier detection in sensor networks. *Proceedings of the 8th ACM international symposium on mobile ad hoc networking and computing*, ACM Press, 9–14 September 2007. Montreal, Canada:

Shuai, M., Xie, K., Chen, G., Ma, X., and Song, G., 2008. A Kalman filter based approach for outlier detection in sensor networks. *Proceedings of international conference on computer science and software engineering*, IEEE Computer Society, 12–14 December 2008. Wuhan, China: 154–157.

Sterk, G. and Stein, A., 1997. Mapping wind-blown mass transport by modeling variability in space and time. *Soil Science Society of America Journal*, 61, 232–239.

Subramaniam, S., Palpanas, T., Papadopoulos, D., Kalogerakiand, V., and Gunopulos, D., 2006. Online outlier detection in sensor data using non-parametric models. *Proceedings of the 32nd international conference on very large data bases*, ACM Press, 12–15 September 2006. Seoul, Korea: 187–198.

Webster, R. and Oliver, M.A., 2007. *Geostatistics for environmental scientists*. Chichester: Springer.

Wu, W., Cheng, X., Ding, M., Xing, K., Liu, F., and Deng, P., 2007. Localized outlying and boundary data detection in sensor networks. *IEEE Transactions on Knowledge and Data Engineering*, 19 (8), 1145–1157.

Zhang, K., Shi, S., Gao, H., and Li, J., 2007a. Unsupervised outlier detection in sensor networks using aggregation tree. *Proceedings of the 3rd international conference on advanced data mining and applications*, 158–169.

Zhang, Y., Meratnia, N., and Havinga, P.J.M., 2007b. A taxonomy framework for unsupervised outlier detection techniques for multi-type data sets. *Technical Report*, TR-CTIT-07-79, The Netherlands: University of Twente.

Zhang, Y., Meratnia, N., and Havinga, P.J.M., 2010a. Outlier detection techniques for wireless sensor network: A survey. *IEEE Communications Surveys & Tutorials*, 12 (2), 159–170.

Zhang, Y., Meratnia, N., and Havinga, P.J.M., 2010b. Ensuring high sensor data quality through use of online outlier detection techniques. *International Journal of Sensor Networks*, 7 (3), 141–151.

Zhang, Y., 2010. *Observing the unobservable – distributed online outlier detection in wireless sensor networks*. Thesis. (PhD) The Netherlands: University of Twente.