

## Article

# Optimisation in the Design of Environmental Sensor Networks with Robustness Consideration

Setia Budi <sup>1,2,\*</sup>, Paulo de Souza <sup>2</sup>, Greg Timms <sup>2</sup>, Vishv Malhotra <sup>1</sup> and Paul Turner <sup>1</sup>

Received: 25 October 2015 / Accepted: 18 November 2015 / Published: 27 November 2015

Academic Editor: Leonhard M. Reindl

<sup>1</sup> School of Engineering and ICT, University of Tasmania, Private Bag 87, Hobart, TAS 7001, Australia; Vishv.Malhotra@utas.edu.au (V.M.); Paul.Turner@utas.edu.au (P.T.)

<sup>2</sup> Commonwealth Scientific and Industrial Research Organisation, 15 College Road, Sandy Bay, TAS 7005, Australia; Paulo.Desouza@csiro.au (P.S); Greg.Timms@csiro.au (G.T.)

\* Correspondence: Setia.Budi@utas.edu.au; Tel.: +61-3-6226-2189; Fax: +61-3-6226-2913

**Abstract:** This work proposes the design of Environmental Sensor Networks (ESN) through balancing robustness and redundancy. An Evolutionary Algorithm (EA) is employed to find the optimal placement of sensor nodes in the Region of Interest (RoI). Data quality issues are introduced to simulate their impact on the performance of the ESN. Spatial Regression Test (SRT) is also utilised to promote robustness in data quality of the designed ESN. The proposed method provides high network representativeness (fit for purpose) with minimum sensor redundancy (cost), and ensures robustness by enabling the network to continue to achieve its objectives when some sensors fail.

**Keywords:** environmental sensor networks; sensor networks design; sensor networks deployment; optimisation; evolutionary algorithm; spatial regression test; gap filling; noise detection; data quality

## 1. Introduction

### 1.1. Environmental Sensor Networks

Automated environmental monitoring started with simple automatic logging systems that recorded several environmental properties at predetermined intervals. These simple monitoring systems had no communication capability. They required field scientists to visit the site regularly and download the data manually. Technological advancements enabled these passive logging systems to evolve into intelligent sensor networks where each sensor node actively communicates its own observed data to nearby sensor nodes. Moreover, these interconnected sensor nodes also have a capability to process and communicate their data to a remote data centre without any operator intervention. These monitoring systems are known as Environmental Sensor Networks (ESNs) [1–4].

ESNs have a significant role to support the quality of our life on this planet. They play a part in many different areas such as agriculture, forestry, science, health and safety, insurance, mining, weather forecast, etc.

Production in agriculture and forestry are highly dependent on the changes in the environmental parameters (temperature, humidity, rain fall, solar radiation). Apart from the need to increase production, agricultural management should also be practiced with a degree of precision (known as Precision Agriculture) to provide an alternative and realistic means to reduce the use of potentially harmful compounds and promote sustainability. Precision Agriculture is an emerging area where ESNs play an important role [5–8].

In forestry, ESNs are also utilised for fire detection systems [9–11]. The networks can alarm on the origin of the fire before it is spread uncontrollably. A major forest fire can destroy thousands of hectares and incur social, environmental and economic costs.

As living beings, water and air are crucial to support our life. The quality of the water and the air which we consume and breath every day directly impact our health. The need to promote better healthcare also motivates the extensive use of ESNs to monitor the quality of both water [12–14] and air [15–17].

In the scientific field, ESNs enable us to have a better understanding of the planet on which we live. It helps us to answer many questions which could not be answered in the past and also to promote more questions which have never been asked before. The changes in climate across the earth would never be able to be identified without ESNs. Nowadays, scientists around the world have more data than before to unveil the climate change and deeply analyse its impact [18,19].

The advancements in ESNs also benefit our day to day life by providing more accurate weather information (weather report and forecast), which are crucial in certain areas like tourism and transportation. Such information is also used to support personal decisions as simple as deciding what kind of clothes to wear in the day to suit the weather.

## 1.2. ESN Design and Its Challenges

In order to have a fit for purpose ESN, design is a critical process prior to the deployment phase. There are two fundamental questions that need to be addressed: how many sensor nodes are required to fit the application purposes and where should the nodes be deployed in the Region of Interest (RoI) [20,21].

In current ESN design practice, one of the major focus for reducing costs is to minimise the total number of sensor nodes required to cover a specific RoI [22]. However, when sensors fail, the usefulness of the network degrades. The ESN no longer produces the data needed; it is not advisable, or even possible, to rely on data from such a network for decision-making. Improving robustness of ESNs is paramount.

Sensor nodes placement significantly impacts the effectiveness of an ESN and the efficiency of its operation [23]. Creating an optimised sensor node placement is not an easy problem, and it has been proven to be NP-hard (Non-deterministic Polynomial-time hard) for most formulations of sensor deployment [24]. Complexity is introduced especially when dealing with the requirement to have a fully operational ESN, which meets the application purposes, with the lowest possible number of sensor nodes. Moreover, the uncertainty in a sensor's ability to function properly, resulting from disruptions that may be caused by terrain or harsh operational conditions in outdoor environmental monitoring, introduces further complexity.

A number of studies have been carried out in the past few years with the aim of optimising the placement of sensor nodes. As an overview, Younis and Akkaya presented a comprehensive survey of strategies and techniques in sensor networks deployment prior to 2008 [24,25]. Cheng *et al.* compared six different ESN deployment strategies (single static sink, mobile data sink, multiple data sink/clustering, non-uniform energy, non-uniform placement, non-uniform traffic) to determine the maximum achievable networks lifetime with minimum deployment cost. They mainly considered communication cost as the main contributor in power consumption; excluded some other factors such as sensing and processing cost. Linear programming is employed in their work to find the maximum lifetime for a given scenario [26]. Griboaud *et al.* employed the modelling power of Interacting Markovian Agent to evaluate the performance of on-off strategies in a sensor network; where sensors are distributed in a continuous finite geographical area (based on a known spatial Poisson density) [27]. Bhondekar *et al.* constructed a hypothetical application involving deployment of three types of sensors, which measure three different environment properties for different applications, on a two dimensional field. A Genetic Algorithm (GA) is employed to optimise application specific parameters, connectivity parameters and energy parameters formulated into a single fitness function

(weighted sum approach) [28]. Aziz *et al.* formulated the coverage problem in the deployment of ESN as an optimisation issue and Particle Swarm Optimisation (PSO) is employed to discover the near optimum placement of the sensor nodes (within a two dimensional square area) which leads to the maximum possible coverage. The fitness for each discovered solution in every iteration is evaluated based on Voronoi Diagram (VD) [29]. Beccuti *et al.* utilised Petri Nets and Markov Decision Process to solve the problem of finding a good trade off between the power consumption and the sensor network reliability [30]. Akbarzadeh *et al.* proposed a method to find a near optimum sensor deployment with the main objective to maximise the coverage within three dimensional space. Distance, orientation, and visibility are included in their work as constraint factors. The study utilised Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and linked it to a Geographical Information System (GIS) to provide essential environmental data such as elevation of the RoI and obstacles in the area, to compute the fitness of individuals. A mountainous area in North Carolina was selected as the case study in their work [31]. Fan *et al.* worked on the deployment strategy that meets the coverage requirement of a sensor network by using a minimum number of sensor nodes, which are randomly and uniformly deployed in the monitored field (two spatial dimensions). Two deployment strategies are proposed in their work: Expected-area Coverage Deployment (ECD) and BOundary Assistant Deployment (BOAD) [22]. Senel *et al.* proposed an approach to increase the network connectivity and better spread the load among the relay nodes in sensor networks. The study is inspired by the behavior of a spider, which establishes a web for spanning gaps between objects [32]. Kulkarni and Venayagamoorthy presented a brief review for the application of Particle Swarm Optimisation (PSO) in sensor networks, including the networks deployment. For sensor networks deployment, the survey focused on optimising the coverage, connectivity, and energy consumption in three different kind of deployments: stationary sensor node deployment, mobile sensor node deployment, and base station deployment [33]. Another approach to deploy sensor nodes within three spatial dimensions is proposed by Unaldi *et al.* Bresenham's line of sight (LOS) algorithm is included in the fitness function while maximising the coverage of sensor networks using a GA [34]. Mamun provided detailed descriptions of existing topologies in wireless sensor networks, including a comparative discussion of the performance of different topologies [35]. D'Este *et al.* proposed an automated method for generating and combining cost and benefit values in the deployment of sensor networks in coastal regions. The work was focused on the marine environment of Australia and used the Tasmanian Marine Analysis Network (TasMAN) project [36,37] as a case study. The sensor placement solution described in their work is called Automated Cost-Benefit Analysis (ACBA) [21]. Rodger suggested a fuzzy multi-sensor data fusion Kalman model to support the Integrated Vehicle Health Maintenance System (IVHMS) based on fault detection and feedback [38]. Mansouri *et al.* worked on the deployment of an ESN that involved three types of sensors, to measure three different environmental properties for different applications, on a two dimensional space. GA is employed to optimise application specific parameters, connectivity parameters and energy consumption parameters; using a single fitness function (weighted sum approach) [39]. Banimelhem *et al.* proposed a GA-based approach to extend the network lifetime by finding the near optimum location of nodes which would reduce the required communication energy. This method acts as a refinement of the existing clustering protocol, Low Energy Adaptive Clustering Hierarchy (LEACH), in which the sensor nodes are organised into clusters in order to reduce the amount of energy consumed in the communication between the nodes. In their work, they also assumed that the sensor nodes were mobile and deployed on a two dimensional space [40]. Ayinde and Barnawi also proposed another strategy in ESN deployment which would prolong the network lifetime while still satisfying a cost budget and a minimum required connectivity. They named their technique Enhanced Lifetime Deployment with Cost Constraints (ELDwCC), which is developed based on Differential Evolution (DE). DE is employed to navigate around the search space to deploy nodes in three spatial dimensions [41]. Cerotti *et al.* presented the implementation of Markovian Agent models in the deployment of wireless sensor network for forest fire monitoring system [42]. Rodger and

George proposed an optimisation algorithm which provide set of solutions given certain probabilities and impacts of risks to sustainability [43]. Lanza-Gutierrez and Gomez-Pulido considered energy efficiency, coverage, and reliability in the deployment of sensor nodes within three dimensional space. Their objectives were to reduce the energy cost, increasing the network life time, maximising the amount and diversity of the information provided by the network. Five different optimisation techniques were employed and compared in their work: Multiobjective Evolutionary Algorithm Based on Decomposition (MOEA/D), Non-dominated Sorting Genetic Algorithm-II (NSGA-II), Strength Pareto Evolutionary Algorithm 2 (SPEA2), Multi-Objective Artificial Bee Colony (MO-ABC), and Multi-Objective Firefly (MO-FA) [44]. Rebai *et al.* focused on maximising the coverage of the sensing area and maintaining the connectivity between sensor nodes while reducing the number of sensor nodes deployed within a two dimensional space. Integer Linear Programming Model, Local Search (LS) and GA were employed in the study to find the optimal placement of sensor nodes [45].

The existing works on ESN design have focused on three objectives (or constraints): the coverage of sensing and measuring points, the network connectivity, and the energy consumption, which would have direct impact on the lifetime of the networks. From the current literature, we discovered that network representativeness and data quality have not been addressed sufficiently as crucial issue in the design of an ESN. It is true that adding more sensor nodes in the RoI will certainly boost the data quality and robustness of the ESN, yet redundancy in the placement of sensor nodes would also introduce an undesirable increase in deployment and maintenance costs. Further, the existing works only focused on the spatial dimension and none of the aforementioned approaches considered temporal data to optimise sensor node placement.

### 1.3. Our Work

We address the issues in the deployment of an ESN and considers data quality issues (to promote robustness) while designing the placement of sensor nodes in two spatial dimensions. We also introduce temporal data as a third dimension. In this work, we are focusing our experiment on temperature data measured on an hourly basis.



**Figure 1.** One of the weather stations operated by Commonwealth Scientific and Industrial Research Organisation (CSIRO), deployed in Geeveston (Tasmania, Australia). Similar weather stations are also deployed in other regions in Tasmania, including in South Esk. The instrument is manufactured by Campbell Scientific (model GRWS100) [46] and equipped with a set of sensors to measure several environmental properties: air temperature, relative humidity, barometric pressure, rain fall, solar radiation, wind speed and wind direction. The station also capable of sending the sensed data to a central data repository over 3G connection. Energy is provided by a solar panel.

Some fundamental assumptions underpin our work. This study is specifically dealing with the design of an ESN consisting of several weather stations acting as sensor nodes. Each node is stationary deployed and also acts as a base station, which is equipped with solar panel and telemetry (with 3G connection) to send the sensed data to a centralised data repository. Network connectivity and energy consumption are not considered as parameters to be optimised. Figure 1 shows one of our weather stations deployed in the field.

## 2. Experimental Approach

There are two main (and interrelated) components in this study: the first is optimisation of the ESN design, and the second is the data quality assessment. The optimisation mainly focused on finding the location to place each sensor node, given a certain number of sensor nodes, which will produce the best representativeness of the area. The data quality assessment focused on two common issues in ESNs: gaps and noise. Gaps in ESN data mainly occur due to sensor or communication failure, which introduce some missing values in the data. In the case of noise, the sensor still produces some data, however, the measured data does not accurately represent the actual condition. The experimental part (including simulations) in this work is written in Python programming with IPython Notebook as the platform.

### 2.1. Dataset

The experimental part of this study was conducted using the SouthEsk Hydrological model [47] as a dataset. The model is produced by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and covers a set of environmental parameters such as temperature, relative humidity, wind speed, wind direction, and solar radiation in the North East of Tasmania ( $-41.0^\circ$  to  $-42.0^\circ$  latitude and  $147.0^\circ$  to  $148.5^\circ$  longitude). The region under study is presented in Figure 2.



**Figure 2.** Map of Tasmania (Australia). The red colored rectangular area in the north east region indicates the Region of Interest (RoI) under study (South Esk).

The dataset itself is stored in Network Common Data Form (netCDF) format [48] as a multi-dimensional matrix (temporal and two spatial dimensions). The spatial area is mapped into a data grid with a size of  $151 \times 101$ . In mathematical notation, the dataset can also be expressed as a three dimensional matrix of  $D$ , where the first, second, and third dimensions are time, latitude, and longitude respectively. Therefore,  $D_{t,i,j}$  is the environmental data measured at time index  $t$ , within the spatial coordinate latitude index  $i$  and longitude index  $j$ . This study will focus specifically on temperature data over 730 h, recorded on an hourly basis.



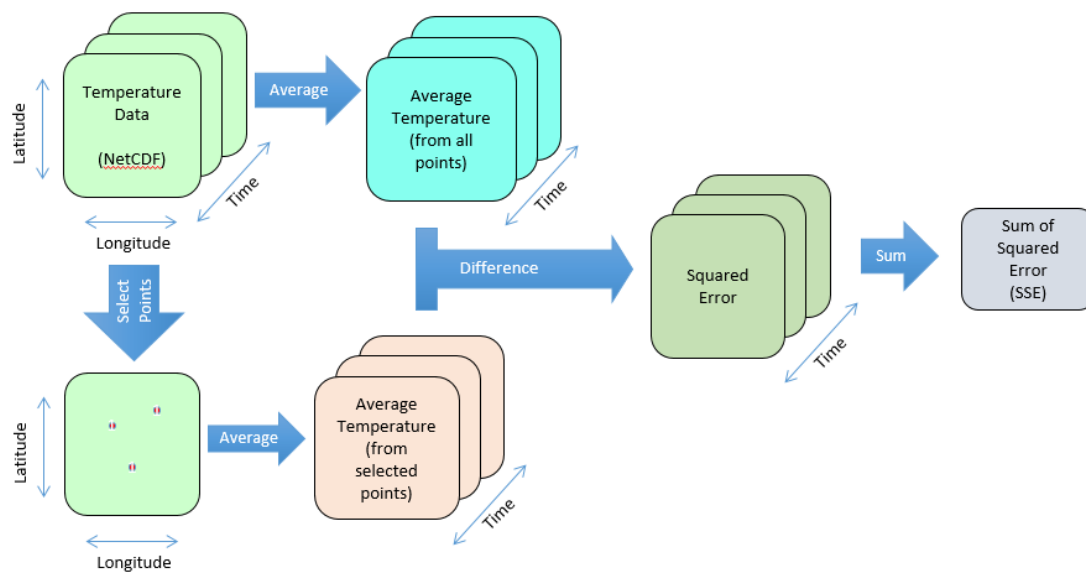
## 2.2. ESN Design Optimisation

The main focus of this component is to find the optimum placement for sensor nodes to produce the best representativeness of the RoI at minimal cost. This component covers two aspects: the representativeness measure for ESN design and the optimisation technique for placement of the sensor nodes.

### 2.2.1. Representativeness in ESN Design

In the designing process, the representativeness formulation should relate back to the purpose of the ESN deployment in the RoI. For the aim of this work, we utilised the average spatial temperature to quantify the representativeness of an ESN design as a case study. This decision is made to simplify the formulation of the representativeness. This approach follows these steps (also presented in Figure 3):

- (1) The average spatial temperature is calculated from all points across the two spatial dimensions for each time step.
- (2) Certain points are selected from the spatial dimensions to place the sensor nodes.
- (3) The average spatial temperature is calculated from the temperature data measured by the available sensor nodes for each time step.
- (4) The difference between these two average spatial temperatures is calculated for each time step and the Sum Squared Error (SSE) is produced as the result.



**Figure 3.** The figure shows the work flow to measure the representativeness of an ESN in respect to the Region of Interest (RoI) based on the average spatial temperature. The representativeness is calculated according to the difference between the actual average spatial temperature and the average spatial temperature measured by the deployed sensor nodes over periods of time.

The average spatial temperature at a particular time index could be expressed as in Equation (1) and the design of the ESN would be defined as in Equation (2).

$$\bar{D}_t = \frac{\sum_{i=1}^{n_2} \sum_{j=1}^{n_3} D_{t,i,j}}{n_2 n_3} \quad (1)$$

where:

- $t$  is the time index
- $i$  is the latitude index
- $j$  is the longitude index
- $n_2$  is the length of the second dimension (latitude)
- $n_3$  is the length of the third dimension (longitude)
- $D_{t,i,j}$  is the temperature data at time, latitude, and longitude index of  $t, i, j$  respectively
- $\bar{D}_t$  is the average spatial temperature at time index  $t$

$$\begin{aligned} sn &= (lat, lon) \\ \vec{SN} &= (sn_1, sn_2, \dots, sn_z) \end{aligned} \quad (2)$$

where:

- $sn$  is the location of a particular sensor node which is represented as a tuple of latitude and longitude
- $z$  is the number of sensor nodes included in the ESN design
- $\vec{SN}$  is the ESN design which is defined as a vector of sensor nodes ( $sn$ )

The temperatures and the average spatial temperature measured by the sensor nodes in an ESN design could be expressed in Equation (3). Furthermore, the representativeness of an ESN design would be defined as the Sum Squared Error (SSE) between the average spatial temperature and the average temperature produced by the ESN over time as described in Equation (4); where a lower SSE indicates a better representativeness to the RoI.

$$\begin{aligned} D_{t,\vec{SN}} &= D_{t,lat_1,lon_1}, D_{t,lat_2,lon_2}, \dots, D_{t,lat_z,lon_z} \\ \bar{D}_{t,\vec{SN}} &= \frac{\sum D_{t,\vec{SN}}}{z} \end{aligned} \quad (3)$$

$$SSE = \sum_{t=1}^{n_1} (\bar{D}_t - \bar{D}_{t,\vec{SN}})^2 \quad (4)$$

where:

- $D_{t,\vec{SN}}$  is a vector of temperature data measured by all sensor nodes at time index  $t$
- $\bar{D}_{t,\vec{SN}}$  is the average spatial temperature measured by all sensor nodes at time index  $t$
- $z$  is the number of sensor nodes included in the ESN design
- $n_1$  is the length of the first dimension (time)
- $SSE$  is the sum squared error of the average spatial temperature

### 2.2.2. Optimisation Problem in ESN Design

The problem of finding the sensor node locations which lead to the maximum representativeness with the minimum number of sensor nodes (and hence cost) is an optimisation problem. There is an exponential growth in search space with a larger area of deployment. We assume that we have a two dimensional space of  $8 \times 8$ , and we have one sensor node to be deployed. It means within the given two dimensional space there are  $2^8$  possible locations to deploy the sensor node. The number of possible locations will grow exponentially with the increase in the number of sensor nodes. With four sensor nodes, there are  $[2^8 \times (2^8 - 1) \times (2^8 - 2) \times (2^8 - 3)]$  possible sensor deployment schema. In this type of deployment, the search space is considerably large with each position yielding different levels of representativeness.

There are several approaches available to solve optimisation problems, namely: Dynamic Programming, Gradient Method, Evolutionary Algorithm, *etc.* Considering the large size of the search space, this study employed an Evolutionary Algorithm (EA) as an optimisation technique.

The algorithm is able to handle a large search space and is also capable of avoiding local optima during the search process. Each possible solution is treated as an individual, and a set of individuals forms a population [49–51]. In the present work, each individual represents a certain ESN design, which consists of a set of sensor nodes including their locations.

The optimisation process is commenced with an initial randomly generated population. Next, a set of offspring are produced from the population through the crossover operation. This process is then followed by a selection process, where every individual from the previous population and their offspring are evaluated and the best solutions chosen to form a new population (the next generation). The process runs iteratively for several generations and produces several new placement designs in each generation. At the end of the iteration process (once the algorithm has converged), the best sensor placement design will be presented as the near-optimum solution. In order to avoid premature convergence, a mutation operation is applied to escape from the local optima and to diversify the population in the next generation.

In this study, the selection process uses the representativeness of an ESN design as a fitness function in order to compare two solutions (ESN designs) and determine which one is dominant over the other. Solutions which are not dominated by any other solution, will be included in the next generation population. As a final result, the population in the last generation consists of a set of solutions (ESN design) which are better (not dominated by) any other possible solution within the search space.

The algorithm in this experiment is implemented using Distributed Evolutionary Algorithms in Python (DEAP) [52], an open source evolutionary computation framework developed at the Computer Vision and Systems Laboratory of Universite Lava, Canada. We referred to De Jong's parameters setting [53–55] to run EA in our experiment. The complete parameters set is presented in Table 1.

**Table 1.** Evolutionary Algorithm (EA) parameters.

Parameter	Value
Number of generation	1000
Population size	50
Crossover probability	0.6
Mutation probability	0.001
Selection operation	NSGA2 [51]
Crossover operation	one-point crossover
Mutation operation	uniform integer mutation
Seed number	0

### 2.3. ESN Data Quality Component

In order to analyse the impact of certain data quality issues on the representativeness of the ESN, artificial gaps and noise are randomly generated in a simulation.

#### 2.3.1. Gap Simulation

Artificial gaps were introduced in the simulation to analyse their impact on the representativeness of the previously discovered optimum ESN design. A gap in this simulation is formulated as a random time period where a random sensor node does not produce any data. It could also be expressed in mathematical notation as presented in Equation (5).

$$\begin{aligned}
 gap &= rand(t, sn) \\
 D_{gap} &= null \\
 \overrightarrow{GAP} &= gap_1, gap_2, \dots, gap_p \\
 p &= gap\_percentage \times n_1 \times z
 \end{aligned} \tag{5}$$



where:

$rand(t, sn)$  is a function to select randomly a time index and a sensor node  
 $p$  is the total number of gaps  
 $\overrightarrow{GAP}$  is the vector of gaps

The gap simulation starts by generating a certain number of random time slots which are used to represent the time points where the gap occur in the data. The number of random time slots is controlled by the gap percentage parameter. In this study, the gap percentage was in the range of 10%, 20%, 30%, and 40%. Therefore, in the case of 730 measurements, 10% of gaps mean that there will be 73 points in the data which have no value. The gaps were introduced randomly to every sensor node available in the ESN design in turn, and the representativeness was determined for every sensor node. The final representativeness value of an ESN design after the inclusion of data gaps is calculated as the average representativeness of all the nodes within the ESN.

### 2.3.2. Gap Filling

In order to overcome the loss of representativeness through gaps in data, a gap filling technique was applied. The Spatial Regression Test (SRT) [56,57] was selected as a gap filling technique due to its capability to consider both spatial and temporal aspects when predicting the missing value. Mathematically, the SRT can be expressed as in Equation (6).

$$\hat{y} = \frac{\sum_{i=1}^n \frac{y_i}{S_i^2}}{\sum_{i=1}^n \frac{1}{S_i^2}} \quad (6)$$

where:

$\hat{y}$  is the SRT predicted value  
 $n$  is the number of neighbouring sensor nodes  
 $y_i$  is the linear regression predicted value from neighbour sensor node  $i$   
 $S_i$  is the linear regression error from neighbour sensor node  $i$

Gap filling was paired with gap simulation in order to present an overview of both the impact of a certain proportion of gaps in the data to the representativeness of the ESN, and how far the gap filling technique could minimise the gap impact and promote the robustness of the ESN.

### 2.3.3. Noise Simulation

Artificial noise were also introduced in a simulation to analyse their impact on the representativeness of an ESN design. Noise in this simulation is formulated as random time points where a random sensor node produces defective data. The number of random time slots is controlled by the noise percentage parameter. In this study, the noise percentage was in the range of 10%, 20%, 30%, and 40%. The noise simulation in this study is presented in Equation (7).

$$\begin{aligned} noise &= rand(t, sn) \\ D_{noise} &= D_{noise} + rand(noise\_range) \\ \overrightarrow{NOISE} &= noise_1, noise_2, \dots, noise_q \\ q &= noise\_percentage \times n_1 \times z \end{aligned} \quad (7)$$

where:

$rand(t, sn)$  is a function to select randomly a time index and a sensor node  
 $q$  is the total number of noises  
 $\overrightarrow{NOISE}$  is the vector of noises

Unlike gap simulation which only has one parameter, in noise simulation there is a second parameter called noise range. The actual measurement drift itself was chosen randomly within a certain range which is determined by the noise range parameter. Here, the noise range is between  $-5\text{ }^{\circ}\text{C}$  and  $+5\text{ }^{\circ}\text{C}$ , excludes zero. Noises were introduced into every sensor node available in the ESN design in turn, and the representativeness was then calculated for every sensor node. The final representativeness value of an ESN design after the introduction of noises is calculated as the average representativeness of all the nodes within the ESN.

#### 2.3.4. Noise Detection

Unlike gaps, noises in the data are much harder to detect. In order to overcome the issue, a temperature threshold was used as an automated data quality control. The threshold was calculated based on the average and the standard deviation of the temperature data within a six hour window (three hours before and three hours after) of the data point. The threshold was generated for each data point to be assessed according to Equation (8). The constant value of two is applied in this study in order to create a 95% of confidence interval

$$threshold = \bar{x} \pm C_x \times S_x \quad (8)$$

where:

$\bar{x}$  is the average temperature data within the window period

$C_x$  is a constant value.

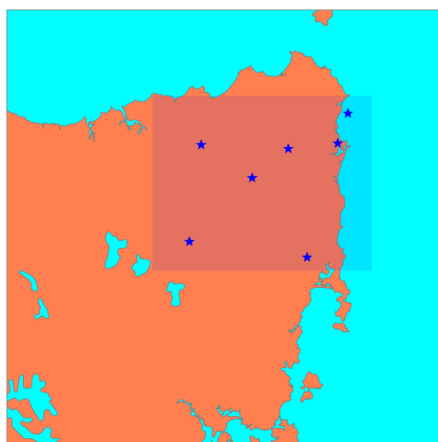
$S_x$  is the standard deviation of the temperature data within the window period

This noise detection technique was used in conjunction with noise simulation in order to produce an overview of the impact of a certain proportion of noises to the representativeness of the ESN, as well as the effectiveness of the automated data quality technique in identifying noise. If further data quality assessment was required, we could adopt the automated marine data quality assessment technique as presented in [58,59].

### 3. Results and Discussion

#### 3.1. Optimum ESN Design

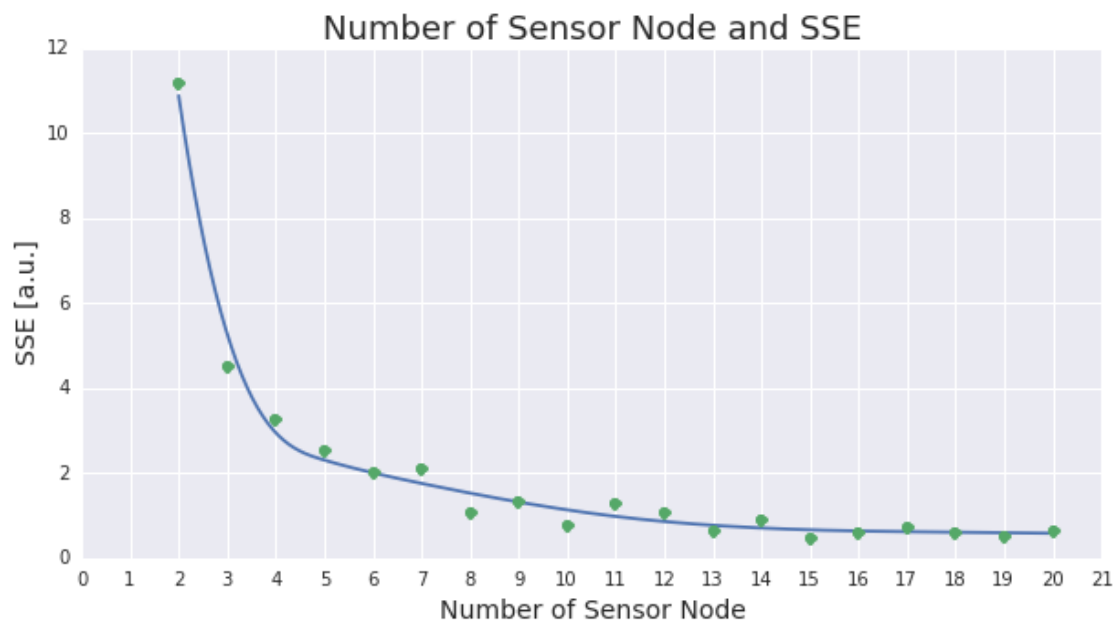
In our case study, the number of sensor nodes was set in the range from two to twenty nodes. Figure 4 presents an example result for the optimum configuration of a network consisting of seven nodes.



**Figure 4.** The stars indicate the optimum placement of seven sensor nodes suggested by the proposed method. The South Esk (Region of Interest in this study) is indicated by the rectangular region.

Figure 5 presents a comparison between the number of sensor nodes in each optimum ESN design and its fitness value (representativeness).

The figure indicates a significant improvement in representativeness by increasing the number of sensor nodes from two to five. Another remarkable improvement is found by adding nodes from five to ten nodes. Once the number of nodes reached thirteen, adding more nodes did not bring any significant improvement.

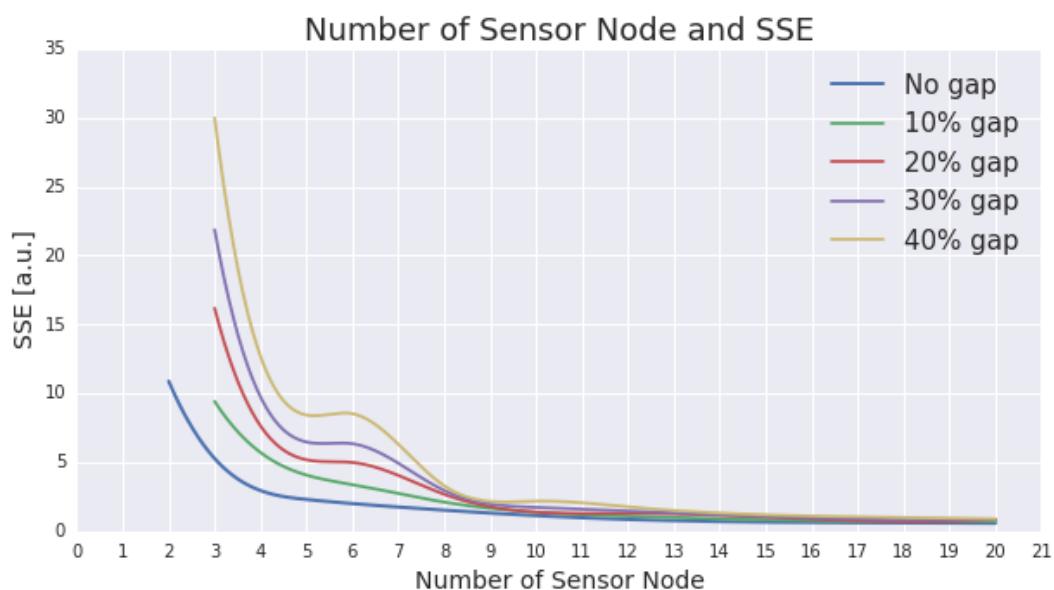


**Figure 5.** This figure presents the representativeness which could be gained from the increase in the number of deployed sensor nodes. The representativeness is measured by the difference (SSE) between the actual average spatial temperature and the average spatial temperature produced by the deployed sensor nodes over periods of time. The figure incorporates number of sensor nodes between two and twenty.

### 3.2. Impact of Data Gaps on ESN Design

Figure 6 presents the result from gap simulation which incorporates 10%, 20%, 30%, and 40% of gaps. The representativeness of the optimal ESN design is also presented in the figure as a comparison of how a certain degree of gaps would impact the representativeness. The gaps within an ESN design with only two nodes will significantly degrade the representativeness. For presentation purposes, in Figure 6, gaps are introduced started with the number of sensor nodes set to three.

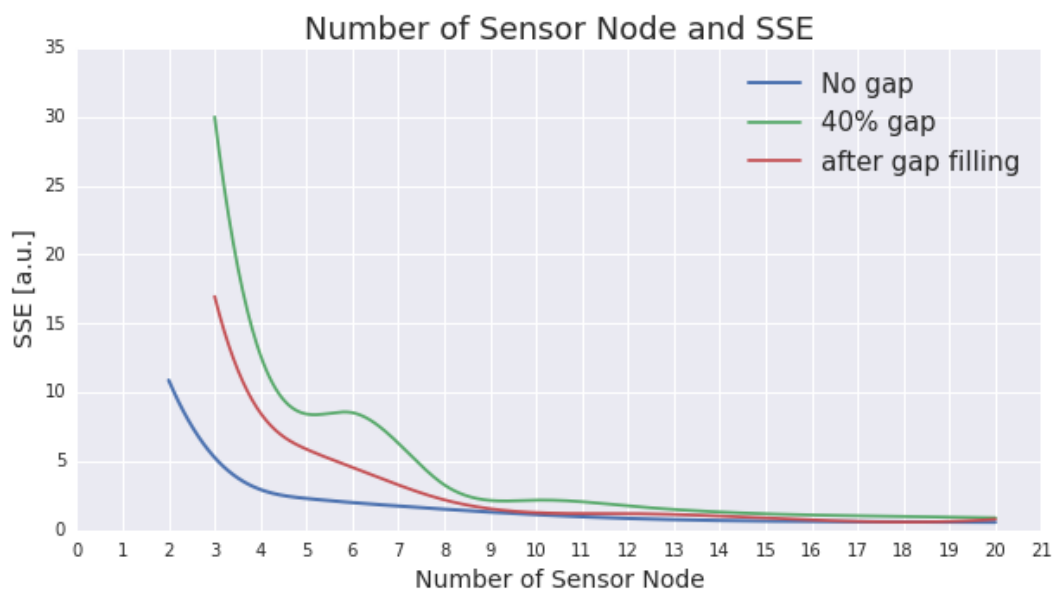
The figure indicates that, unsurprisingly, the more nodes that are incorporated in the design, the less the impact when gaps are introduced. Designs with less than five nodes will experience significant performance degradation once the gaps are introduced. Another notable impact also occurs in ESN with six to nine nodes. On the other hand, the gaps no longer produce any significant impact on the representativeness in an ESN design, which incorporates fifteen or more nodes.



**Figure 6.** The figure presents the impact of certain degree of gaps to the performance of the proposed ESN design (sensor nodes placement). The gaps are set within the range of 10%, 20%, 30%, and 40% (illustrated by green, red, purple, and yellow lines, respectively). For comparison purposes, the performance of the ESN design without gap is also presented (using a blue line).

### 3.3. Gap Filling Result

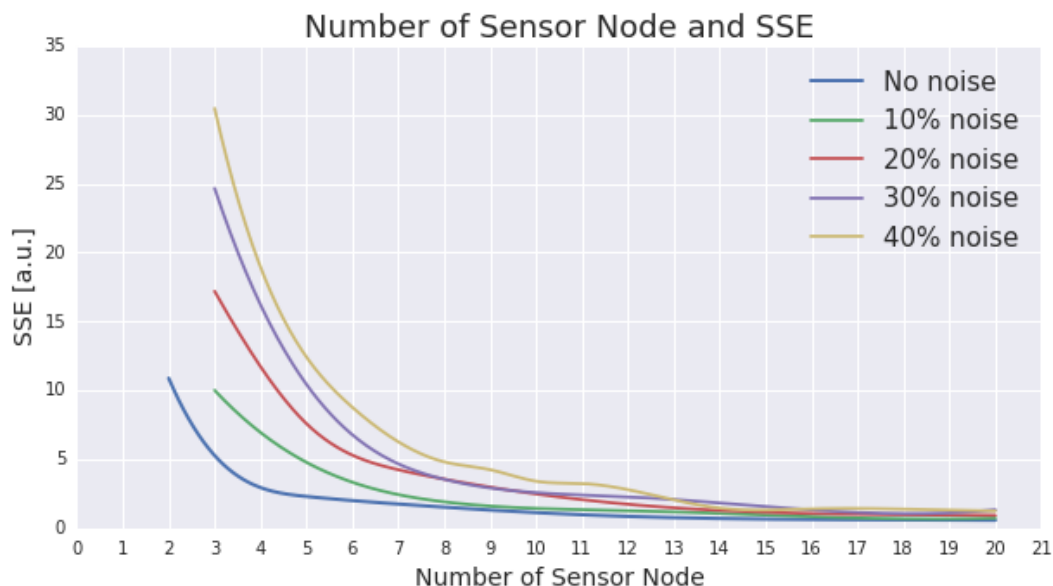
Figure 7 presents the result following gap filling using SRT with 40% of gap occurrences. A notable improvement resulting from the gap filling technique can be seen in the design with three to nine nodes. However, the improvement seems to be less significant for a design with more than thirteen nodes.



**Figure 7.** The figure presents the improvement promoted by a gap filling process (using Spatial Regression Test) to the ESN design with 40% of gaps. For comparison purposes, the performance of the ESN with no gap, with 40% of gap, and after the gap filling process are illustrated using blue, green, and red lines, respectively.

### 3.4. Impact of Noises on ESN Design

Figure 8 presents the results from the noise simulation. In a similar vein to the gap simulation, the noise simulation also shows that an ESN design that incorporates more sensor nodes experiences less impact from the outlying data. Significant impact to the representativeness can be found in the designs which incorporate eight or less sensor nodes. However, after reaching fifteen nodes, having more sensor nodes in the ESN does not bring significant performance improvement when dealing with noise.



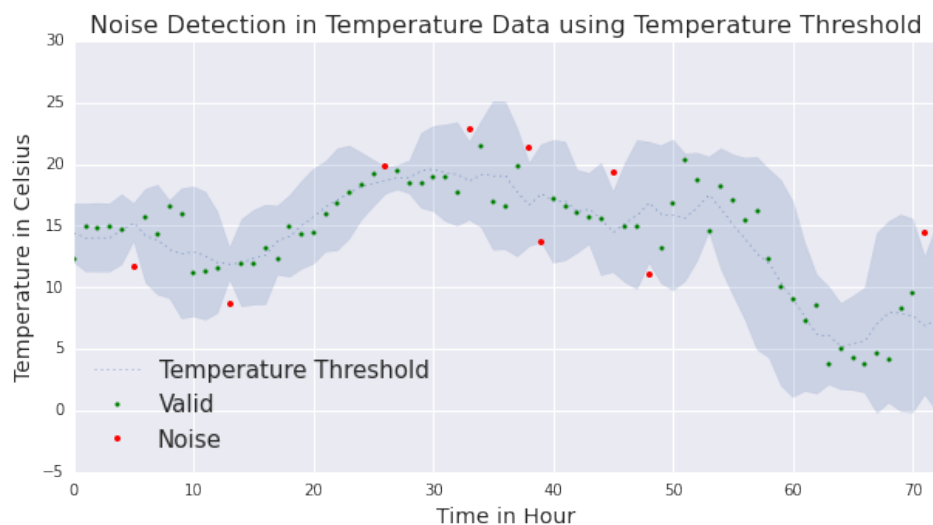
**Figure 8.** The figure presents the impact of certain degree of noise to the performance of the proposed ESN design (sensor nodes placement). The noise are set within the range of 10%, 20%, 30%, and 40% (illustrated by green, red, purple, and yellow lines, respectively). For comparison purposes, the performance of the ESN design without noise is also presented (using blue line).

### 3.5. Noise Detection Results

Figure 9 shows the result from noise detection using the temperature threshold technique. The dots in the figure represent the measured temperature and the grey area represents the temperature threshold calculated using Equation (8). The measured data within the threshold would be considered as valid data, and are represented as green dots in the figure. On the other hand, the measured data outside the threshold would be considered as noise or invalid data, and are represented as red dots.

### 3.6. Suggestion from the Results

As a recap from the experiment, without considering any data quality issues, the ideal number of sensor nodes to be deployed in the South Esk region would be nine. However, if there is a restriction in the deployment budget, the number could be reduced down to five nodes. The deployment with 10 or more nodes would not bring any notable improvement to the representativeness of the ESN toward the RoI, and could be considered as redundant nodes deployment. On the contrary, if data quality issues (gap and noise) were to be considered in ESN design, the ideal number of nodes would be between nine and fourteen. Deploying more than fourteen nodes in the South Esk would not produce any remarkable robustness or improvement against the occurrence of data quality issues (up to 40% of gaps or noises).



**Figure 9.** The figure indicates the detection of noise in the temperature data. Temperature Threshold (Equation (8)) is applied as a noise detection technique.

#### 4. Conclusions

This study proposes a design of ESNs, which is not only focused on minimising the redundancy of sensor nodes but also incorporates data quality requirements (to promote robustness). Temporal data is included during the optimisation process, in addition to two spatial dimensions. The impact of gap and noise in ESN data, including approaches to address the issues, are formalised in a simulation. The simulation presents how far the performance of an ESN, which consists of a certain number of sensor nodes, would degrade given a certain proportion of data gaps or noises. Furthermore, the simulation also demonstrates the significance of gap filling (using SRT) and noise detection (using temperature threshold) techniques in maintaining the robustness of the overall ESN performance. This information will help decision makers to determine the number and the placement of sensor nodes, which are crucial in the design of an ESN.

One challenge in our approach is that it requires an existing spatial temporal data model. Moreover, the representativeness calculations for the spatial temporal matrix are complex and the required computational time increases for larger space and longer time periods.

**Acknowledgments:** The experiment was conducted using the SouthEsk data model owned by Commonwealth Scientific Research and Industrial Organisation (CSIRO). Setia Budi acknowledges support from Sense-T for a Ph.D. scholarship and CSIRO for top up scholarship.

**Author Contributions:** In this manuscript, Setia Budi contributed with the experimental work, data analysis, and manuscript writing. Paulo de Souza contributed in the experiment design, data analysis, and manuscript writing. Greg Timms, Vishv Malhotra and Paul Turner contributed in data analysis and manuscript writing.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Abbreviations

<b>CSIRO</b>	Commonwealth Scientific and Industrial Research Organisation
<b>DEAP</b>	Distributed Evolutionary Algorithms in Python
<b>EA</b>	Evolutionary Algorithm
<b>ESN</b>	Environmental Sensor Networks
<b>GA</b>	Genetic Algorithm
<b>netCDF</b>	Network Common Data Form
<b>NP-hard</b>	Non-deterministic Polynomial-time hard
<b>NSGA2</b>	Non Sorting Genetic Algorithm 2
<b>RoI</b>	Region of Interest
<b>SRT</b>	Spatial Regression Test
<b>SSE</b>	Sum Squared Error



## References

1. Chong, C.Y.; Kumar, S. Sensor networks: Evolution, opportunities, and challenges. *IEEE Proc.* **2003**, *91*, 1247–1256.
2. Hart, J.K.; Martinez, K. Environmental Sensor Networks: A revolution in the earth system science? *Earth-Sci. Rev.* **2006**, *78*, 177–191.
3. Martinez, K.; Hart, J.; Ong, R. Environmental sensor networks. *Computer* **2004**, *37*, 50–56.
4. Porter, J.H.; Nagy, E.; Kratz, T.K.; Hanson, P.; Collins, S.L.; Arzberger, P. New Eyes on the World: Advanced Sensors for Ecology. *BioScience* **2009**, *59*, 385–397.
5. Pierce, F.J.; Elliott, T.V. Regional and on-farm wireless sensor networks for agricultural systems in Eastern Washington. *Comput. Electron. Agric.* **2008**, *61*, 32–43.
6. Heinz, E.; Kraft, P.; Buchen, C.; Frede, H.G.; Aquino, E.; Breuer, L. Set Up of an Automatic Water Quality Sampling System in Irrigation Agriculture. *Sensors* **2013**, *14*, 212–228.
7. Pajares, G.; Peruzzi, A.; Gonzalez-de Santos, P. Sensors in Agriculture and Forestry. *Sensors* **2013**, *13*, 12132–12139.
8. Bitella, G.; Rossi, R.; Bochicchio, R.; Perniola, M.; Amato, M. A Novel Low-Cost Open-Hardware Platform for Monitoring Soil Water Content and Multiple Soil-Air-Vegetation Parameters. *Sensors* **2014**, *14*, 19639–19659.
9. Lloret, J.; Garcia, M.; Bri, D.; Sendra, S. A Wireless Sensor Network Deployment for Rural and Forest Fire Detection and Verification. *Sensors* **2009**, *9*, 8722–8747.
10. Tsiourlis, G.; Andreadakis, S.; Konstantinidis, P. SITHON: A Wireless Network of in Situ Optical Cameras Applied to the Early Detection-Notification-Monitoring of Forest Fires. *Sensors* **2009**, *9*, 4465–4482.
11. Martinez-de Dios, J.R.; Merino, L.; Caballero, F.; Ollero, A. Automatic Forest-Fire Measuring Using Ground Stations and Unmanned Aerial Systems. *Sensors* **2011**, *11*, 6328–6353.
12. Storey, M.V.; van der Gaag, B.; Burns, B.P. Advances in on-line drinking water quality monitoring and early warning systems. *Water Res.* **2011**, *45*, 741–747.
13. Kim, K.; Myung, H. Sensor Node for Remote Monitoring of Waterborne Disease-Causing Bacteria. *Sensors* **2015**, *15*, 10569–10579.
14. Dong, J.; Wang, G.; Yan, H.; Xu, J.; Zhang, X. A survey of smart water quality monitoring system. *Environ. Sci. Pollut. Res.* **2015**, *22*, 4893–4906.
15. Shi, W.; Wong, M.S.; Wang, J.; Zhao, Y. Analysis of Airborne Particulate Matter (PM<sub>2.5</sub>) over Hong Kong Using Remote Sensing and GIS. *Sensors* **2012**, *12*, 6825–6836.
16. Malaver, A.; Motta, N.; Corke, P.; Gonzalez, F. Development and Integration of a Solar Powered Unmanned Aerial Vehicle and a Wireless Sensor Network to Monitor Greenhouse Gases. *Sensors* **2015**, *15*, 4072–4096.
17. Elen, B.; Peters, J.; Poppel, M.V.; Bleux, N.; Theunis, J.; Reggente, M.; Standaert, A. The Aeroflex: A Bicycle for Mobile Air Quality Measurements. *Sensors* **2012**, *13*, 221–240.
18. Isaak, D.J.; Wollrab, S.; Horan, D.; Chandler, G. Climate change effects on stream and river temperatures across the northwest U.S. from 1980–2009 and implications for salmonid fishes. *Clim. Chang.* **2011**, *113*, 499–524.
19. Ganguly, A.; Steinhäuser, K. Data Mining for Climate Change and Impacts. In Proceedings of the IEEE International Conference on Data Mining Workshops, ICDMW, Pisa, Italy, 15–19 December 2008; pp. 385–394.
20. Onur, E.; Ersoy, C.; Delic, H. How many sensors for an acceptable breach detection probability? *Comput. Commun.* **2006**, *29*, 173–182.
21. D’Este, C.; Souza, P.D.; Sharman, C.; Allen, S. Relocatable, Automated Cost-Benefit Analysis for Marine Sensor Network Design. *Sensors* **2012**, *12*, 2874–2898.
22. Fan, G.; Wang, R.; Huang, H.; Sun, L.; Sha, C. Coverage-Guaranteed Sensor Node Deployment Strategies for Wireless Sensor Networks. *Sensors* **2010**, *10*, 2064–2087.
23. McGrath, M.J.; Scanail, C.N. *Sensor Technologies: Healthcare, Wellness and Environmental Applications*, 1st ed.; Apress: Berkeley, CA, USA, 2013.
24. Younis, M.; Akkaya, K. Strategies and techniques for node placement in wireless sensor networks: A survey. *Ad Hoc Netw.* **2008**, *6*, 621–655.

25. Younis, M.; Akkaya, K. Node Positioning for Increased Dependability of Wireless Sensor Networks. In *Algorithms and Protocols for Wireless Sensor Networks*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2008; pp. 225–266.
26. Cheng, Z.; Perillo, M.; Heinzelman, W. General Network Lifetime and Cost Models for Evaluating Sensor Network Deployment Strategies. *IEEE Trans. Mob. Comput.* **2008**, *7*, 484–497.
27. Gribaudo, M.; Cerotti, D.; Bobbio, A. Analysis of On-off policies in Sensor Networks Using Interacting Markovian Agents. In Proceedings of the Annual IEEE International Conference on Pervasive Computing and Communications, Hong Kong, China, 17–21 March 2008; pp. 300–305.
28. Bhondekar, A.P.; Vig, R.; Singla, M.L.; Ghanshyam, C.; Kapur, P. Genetic Algorithm Based Node Placement Methodology For Wireless Sensor Networks. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, China, 18–20 March 2009.
29. Aziz, N.; Mohemmed, A.; Alias, M. A wireless sensor network coverage optimization algorithm based on particle swarm optimization and Voronoi diagram. In Proceedings of the International Conference on Networking, Sensing and Control, ICNSC'09, Okayama, Japan, 26–29 March 2009; pp. 602–607.
30. Beccuti, M.; Codetta-Raiteri, D.; Franceschinis, G. Multiple Abstraction Levels in Performance Analysis of WSN Monitoring Systems. In Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools; ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, Belgium, 20–22 October 2009; pp. 1–10.
31. Akbarzadeh, V.; Ko, A.H.R.; Gagne, C.; Parizeau, M. Topography-Aware Sensor Deployment Optimization with CMA-ES. In *Parallel Problem Solving from Nature, PPSN XI*; Schaefer, R., Cotta, C., Kolodziej, J., Rudolph, G., Eds.; Springer Berlin Heidelberg: Berlin, Germany, 2010; pp. 141–150.
32. Senel, F.; Younis, M.; Akkaya, K. Bio-Inspired Relay Node Placement Heuristics for Repairing Damaged Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2011**, *60*, 1835–1848.
33. Kulkarni, R.; Venayagamoorthy, G. Particle Swarm Optimization in Wireless-Sensor Networks: A Brief Survey. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **2011**, *41*, 262–267.
34. Unaldi, N.; Temel, S.; Asari, V.K. Method for Optimal Sensor Deployment on 3D Terrains Utilizing a Steady State Genetic Algorithm with a Guided Walk Mutation Operator based on the Wavelet Transform. *Sensors* **2012**, *12*, 5116–5133.
35. Mamun, Q. A Qualitative Comparison of Different Logical Topologies for Wireless Sensor Networks. *Sensors* **2012**, *12*, 14887–14913.
36. Timms, G.; McCulloch, J.; McCarthy, P.; Howell, B.; de Souza, P.; Dunbabin, M.; Hartmann, K. The Tasmanian Marine Analysis Network (TasMAN). In Proceedings of the OCEANS 2009—EUROPE, Bremen, Germany, 11–14 May 2009; pp. 1–6.
37. Hugo, D.; Howell, B.; D'Este, C.; Timms, G.; Sharman, C.; de Souza, P.; Allen, S. Low-cost marine monitoring: From sensors to information delivery. In Proceedings of the OCEANS, Waikoloa, HI, USA, 19–22 September 2011; pp. 1–7.
38. Rodger, J.A. Toward reducing failure risk in an integrated vehicle health maintenance system: A fuzzy multi-sensor data fusion Kalman filter approach for IVHMS. *Expert Syst. Appl.* **2012**, *39*, 9821–9836.
39. Mansouri, M.; Nounou, H.; Nounou, M. Genetic Algorithm-based Adaptive Optimization for Target Tracking in Wireless Sensor Networks. *J. Signal Process. Syst.* **2013**, *74*, 189–202.
40. Banimelhem, O.; Mowafi, M.; Taqieddin, E.; Awad, F.; Al Rawabdeh, M. An efficient clustering approach using genetic algorithm and node mobility in wireless sensor networks. In Proceedings of the 2014 11th International Symposium on Wireless Communications Systems (ISWCS), Barcelona, Spain, 26–29 August 2014; pp. 858–862.
41. Ayinde, B.; Barnawi, A. Differential evolution based deployment of wireless sensor networks. In Proceedings of the 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), Doha, Qatar, 10–13 November 2014; pp. 131–137.
42. Cerotti, D.; Gribaudo, M.; Bobbio, A. Markovian agents models for wireless sensor networks deployed in environmental protection. *Reliab. Eng. Syst. Saf.* **2014**, *130*, 149–158.
43. Rodger, J.; George, J. Environmental Energy Strategy: An Induced Linguistic Ordered Weighted Averaging Approach to a Sustainable Economy. *Int. J. Soc. Sustain. Econ. Soc. Cult. Context* **2015**, *10*, 1–18.

44. Lanza-Gutierrez, J.M.; Gomez-Pulido, J.A. Assuming multiobjective metaheuristics to solve a three-objective optimisation problem for Relay Node deployment in Wireless Sensor Networks. *Appl. Soft Comput.* **2015**, *30*, 675–687.
45. Rebai, M.; Le berre, M.; Snoussi, H.; Hnaien, F.; Khoukhi, L. Sensor deployment optimization methods to achieve both coverage and connectivity in wireless sensor networks. *Comput. Oper. Res.* **2015**, *59*, 11–21.
46. GRWS100: General Research-Grade Weather Station. Available online: <https://www.campbellsci.com.au/grws100> (accessed on 21 November 2015).
47. Katzfey, J.; Thatcher, M. Ensemble one-kilometre forecasts for the South Esk Hydrological Sensor Web. In Proceedings of the 19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011; pp. 3511–3517.
48. Rew, R.; Davis, G. NetCDF: An interface for scientific data access. *IEEE Comput. Graph. Appl.* **1990**, *10*, 76–82.
49. Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*, 1st ed.; Wiley: New York, NY, USA, 2009.
50. Coello, C.C.; Lamont, G.B.; Veldhuizen, D.A.V. *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed.; Springer: New York, NY, USA, 2007.
51. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* **2002**, *6*, 182–197.
52. Fortin, F.A.; de Rainville, F.M.; Gardner, M.A.; Parizeau, M.; Gagné, C. DEAP: Evolutionary Algorithms Made Easy. *J. Mach. Learn. Res.* **2012**, *13*, 2171–2175.
53. Jong, K.A.D.; Spears, W.M. An Analysis of the Interacting Roles of Population Size and Crossover in Genetic Algorithms. In *Parallel Problem Solving from Nature*; Schwefel, H.P., Manner, R., Eds.; Springer Berlin Heidelberg: Berlin, Germany, 1990; pp. 38–47.
54. Jong, K.D. Parameter Setting in EAs: A 30 Year Perspective. In *Parameter Setting in Evolutionary Algorithms*; Lobo, F.G., Lima, C.F., Michalewicz, Z., Eds.; Springer Berlin Heidelberg: Berlin, Germany, 2007; pp. 1–18.
55. Yu, X.; Gen, M. *Introduction to Evolutionary Algorithms*; Springer: New York, NY, USA, 2010.
56. Hubbard, K.G.; You, J. Sensitivity Analysis of Quality Assurance Using the Spatial Regression Approach—A case study of the Maximum/Minimum air temperature. *J. Atmos. Ocean. Technol.* **2005**, *22*, 1520–1530.
57. You, J.; Hubbard, K.G.; Goddard, S. Comparison of methods for spatially estimating station temperatures in a quality control system. *Int. J. Climatol.* **2008**, *28*, 777–787.
58. Timms, G.P.; de Souza, P.A.; Reznik, L.; Smith, D.V. Automated Data Quality Assessment of Marine. *Sensors* **2011**, *11*, 9589–9602.
59. Smith, D.; Timms, G.; de Souza, P.; D'Este, C. A Bayesian Framework for the Automated Online Assessment of Sensor Data Quality. *Sensors* **2012**, *12*, 9476–9501.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).