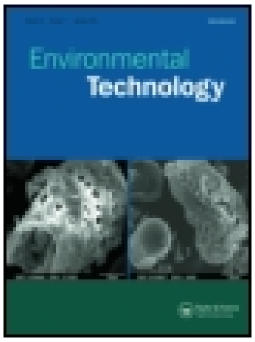

Environmental Technology

ISSN: 0959-3330 (Print) 1479-487X (Online) Journal homepage: http://www.tandfonline.com/loi/tent20

In search for a robust design of environmental sensor networks

Setia Budi, Ferry Susanto, Paulo de Souza, Greg Timms, Vishv Malhotra & Paul Turner

ENVIRONMENTAL TECHNOLOGY, 2017
http://dx.doi.org/10.1080/09593330.2017.1310303

In search for a robust design of environmental sensor networks

Setia Budi [a,b,c], Ferry Susanto [b,d], Paulo de Souza [b], Greg Timms [b], Vishv Malhotra [a] and Paul Turner [a]

[a]School of Engineering and ICT, University of Tasmania, Hobart, Australia; [b]Data61, Commonwealth Scientific and Industrial Research Organisation, Sandy Bay, Australia; [c]Faculty of Information Technology, Maranatha Christian University, Bandung, Indonesia; [d]College of Engineering and Science, Victoria University, Footscray, Australia

ABSTRACT
This paper presents an approach to the design of environmental sensor networks (ESN) which aims at providing a robust, fit-for-purpose network with minimum redundancy. A set of near optimum ESN designs is sought using an evolutionary algorithm, which incorporates redundancy and robustness as fitness functions. This work can assist the decision-making process when determining the number of sensor nodes and how the nodes are going to be deployed in the region of interest.

ARTICLE HISTORY
Received 5 April 2016
Accepted 18 March 2017

KEYWORDS
Sensor networks design; multi-objective optimisation; evolutionary algorithm; inverse distance weighting; spatial interpolation

1. Introduction

Determining changes in environmental parameters over an extensive region is vital for a number of human activities including farming [1,2], water usage [3–6], logistics [7–9], tourism and recreation [10,11], urban development [12–14] or emergency responses [15,16]. Over the past decades, there have been a number of technological advances in sensing instrumentation [17–20], data transmission [21,22], data format [23,24], management and storage [25], as well as in data processing and analytics [26,27]. Those developments have provided us with more accurate weather forecasts and better decision-making when environmental parameters are involved. Fundamental to any forecast modelling is the quality of the environmental sensing. The purpose of the sensing exercise should be a guiding principle for the deployment of environmental sensor networks (ESN).

In order to have a fit-for-purpose ESN, design is crucial prior to the deployment [28–30]. An ESN design will inform the choice of sensor quality that is required for the purpose, the number of sensor nodes for data collections which suit the user's needs; it should also determine the frequency of data reading and data transmission, data storage strategies, key metrics to quality assurance and quality control, operational period, and priorities in maintenance under budget constraints [30–33]. Some practical aspects are also important source of constraints to the design of ESN. These include considerations on the availability of power supply, costs associated to the experiment (e.g. hardware, communication services, deployment, maintenance, decommissioning), safety, interoperability, data management, meta-data capture, and storage.

One of the major foci in current ESN design practice is to satisfy a specific application requirement within a given budget. On the other hand, when sensors fail, the usefulness of the network degrades and no longer produces the data needed. Robustness of ESN is essential and should be considered in the design process. In principle, a robust ESN can be achieved by over sampling, at a potentially prohibitive cost. Nevertheless, redundancy in sensor nodes deployment would also introduce an undesirable increase in both deployment and maintenance costs. For this reason, it is important to find a good compromise of ensuring maximum robustness (fit-for-purpose) with minimum redundancy (cost-effective).

In the current work, we propose a methodology to search for the best placement for a given number of sensor nodes with redundancy and robustness as the primary factors to be optimised. A model output of temperature distribution over a large region is used to demonstrate the proposed method. The quality of the network is assessed with the purpose of representing a temperature distribution within the region under study.

CONTACT Setia Budi setia.budi@utas.edu.au School of Engineering and ICT, University of Tasmania, Private Bag 87, Hobart, TAS 7001, Australia; Data61, Commonwealth Scientific and Industrial Research Organisation, 15 College Road, Sandy Bay, TAS 7005, Australia; Faculty of Information Technology, Maranatha Christian University, Bandung 40164, Indonesia

2. Problem formulation

Design of an ESN is about deciding how many sensor nodes are required to best represent a given region, as well as where those nodes will be deployed, how frequently they should collect and communicate the data, and for how long they will operate. Design of ESN could also inform priorities for maintenance, the impact of sensor node failure, and even the quality of the sensors and their supporting hardware. Designing a sensor network is intrinsically an optimisation problem.

In the present study, we design a network of temperature sensors that best represent a region of interest (RoI). An ESN design is formulated as a set of ESN deployment plans which cover a number of sensor nodes and the placement of each node across the RoI. Representativeness is considered by comparing how close the measured values are from the model output generated by the ESN over the entire region. Suppose $Y = \{y_1, y_2, \ldots, y_N\}$ is a set of N sensor nodes deployed in the RoI; and $y_{n,t}$ is the averaged daily temperature measured by node y_n on the tth day. The RoI is mapped as a two dimension space X (described in Section 3.1) ; and each node in Y could be deployed anywhere in X (illustrated in Figure 1).

There is a significant growth in search space with a larger area of deployment. Assuming there is one sensor node going to be deployed in 8×8 space, it means, there are 64 possible sensor deployment schema. The number of possible locations will also grow significantly with the increase in the number of sensor nodes. As an example, the deployment of four sensor nodes in 8×8 space would lead to ${}^{64}C_4 = 64!/(4! \cdot (64-4)!)$ or 635,376 possible deployment schema. In this type of deployment, the search space is very large with each position yielding different levels of representativeness.

A robust ESN would be able to perform appropriately for its purpose even in a condition where loss or disruption of some of its sensors occurred. Measures, such as choosing excellent hardware, having a preventive and quick corrective maintenance schedule, adding more sensor nodes (if budget allows), and/or adopting a computational procedure to fill gaps in the data could promote the robustness of the network. Data gap filling can be achieved by two distinct approaches. The first approach is a temporal interpolation, when some knowledge about the dynamics of the phenomena of interest (e.g. temperature or solar radiation) is well known [34–36]. In the present study, we used a linear temporal interpolation to estimate missing values (i.e. gaps in the data) which were commonly found in the case of sensor failure or data communication issues. More accurate methods would include time-series modelling such as time-series analysis [37,38] or pattern matching [39,40]. Another approach is to use spatial interpolation, when data from neighbouring nodes can be applied to estimate the value that is missing [41–43]. More sophisticated computational solutions for this would also include temporal variability of the phenomena [44]. Robustness can also be achieved by

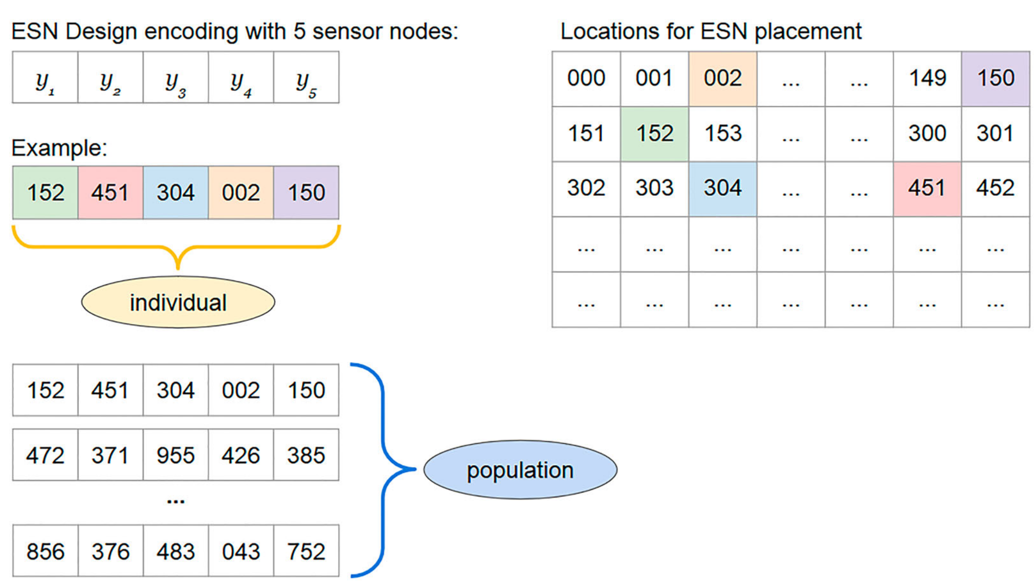

Figure 1. An example of how an ESN design consisting of five sensor nodes is encoded to form an individual (i.e. a possible solution) and a population (i.e. a set of possible solutions). The RoI in this study is gridded as a two-dimensional space (151 × 101) where each cell is indexed. The placement for each sensor node within the RoI is identified by the cell index. In this figure, the fifth sensor node (y_5) is placed in index 150 which is located on the top-right corner of the RoI.

simply adding nodes to the network, if budget is not an issue. However, adding sensors will lead to unnecessary redundancies. Redundancy within the network occurs when there are one or more sensor nodes generated data which most of the time are able to be estimated by their neighbouring nodes.

The ideal strategy to achieve robustness with minimum redundancy will require good design, good hardware choice, maintenance schedule planning, and some computational approaches to overcome inevitable data gaps. In the present work, a methodology to design a sensor network is proposed by employing two objective functions: one to improve robustness with temporal interpolation and another one to, simultaneously, reduce redundancies by using spatial interpolation.

3. Materials and methods

In this work, there are two key important questions to be addressed in designing ESN: (1) How many sensor nodes are needed to best represent the RoI ? and (2) Where those nodes should be deployed? The number of nodes within an ESN will impact the deployment exercise and maintenance costs. In addition, careful node placement could support the effectiveness of an ESN and the efficiency of its operation [45,46].

In order to answer the questions, a computational method is proposed. The method consists of three main procedures: ESN design optimisation, selecting the number of nodes, and node placement. Each of the procedures is described in detail in Sections 3.2–3.4, respectively. The ESN design optimisation incorporated both redundancy and robustness as its objective functions. The representativeness of each generated ESN design, given a number of sensor nodes, is quantified and is utilised as a guide to decide how many nodes to deploy. Further, the balance between redundancy and robustness, including the feasibility for the deployment of each optimised design (given the previously chosen number of nodes), is assessed in order to select the final ESN design (node placement). A year of temperature data from an atmospheric model output was used to demonstrate the performance of the proposed method; background information on the data is given in Section 3.1. Figure 2 provides an overview of the proposed method.

3.1. Dataset

In order to investigate the proposed method, we used data from an atmospheric model over a region of 15,000 km^2 in the northeast of Tasmania. This model was implemented by Katzfey and Thatcher [47] and it has been calibrated using scientific class weather stations. The dataset is stored in netCDF format [23] as a two-dimensional data grid with a size of 151×101 (15,251 cells) and it is publicly accessible [48]. Calibrated model outputs of air temperature data throughout the year 2013 were selected for the present work.

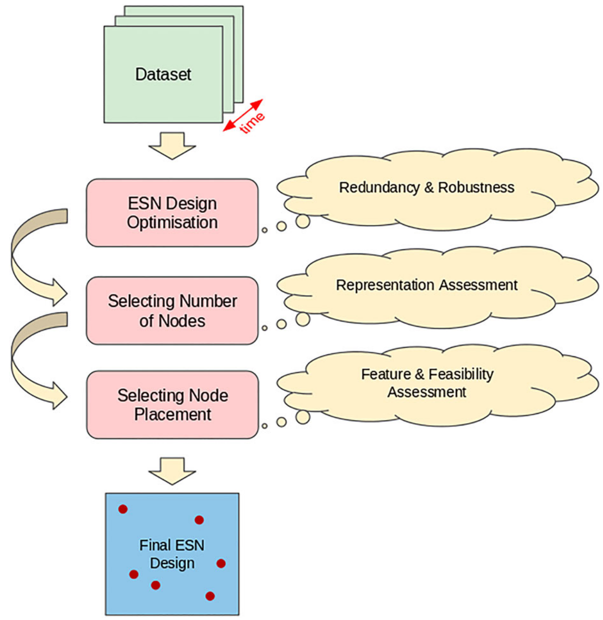

Figure 2. An overview of the proposed method in order to answer two key questions in ESN design (i.e. the number of sensor nodes and the placement of the nodes).

For clarity purposes, the dataset in this work is formalised in mathematical notation as follows: suppose $X = \{x_1, x_2, \ldots, x_N\}$ is a set of N locations (cells) on two-dimensional space; and $X_t = \{x_{1,t}, x_{2,t}, \ldots, x_{N,t}\}$ is a set of averaged daily temperature in two-dimensional space X on the tth day. In this case $x_{n,t}$ represents an averaged temperature at location (cell) x_n on the tth day.

3.2. ESN design optimisation

Since we aim to construct an ESN design based on a number of criteria (redundancy and robustness), we have a multi-objective optimisation problem. Several approaches are available to solve multi-objective optimisation problems [49]. Considering the large size of the search space (described in Section 2), we employed an evolutionary algorithm (EA) in our work as the optimisation technique.

EA [50] is a computational algorithm that mimics the biological evolutionary process. The algorithm is widely used for solving both single and multi-objective optimisation problems within a relatively large search space. It starts with an initial population, which consists of a number of individuals. Each individual represents a single possible solution, and the quality of each individual is calculated by the so-called fitness function. Further, three genetic-like operations (i.e. selection, recombination, and mutation) are applied to the current population to produce a new generation of the population. The selection operation preserves the fittest individuals for the next generation while the recombination and mutation operations take part in preserving variation within the population. The process runs iteratively for several generations until user-specified stopping criteria are met. Two stopping criteria are employed for the purpose of this work:

(1) when the maximum number of generations is achieved or
(2) when the Pareto Front (PF) [51,52] has not changed over a pre-specified number of generations. In other words, the 'saturation point' threshold is met.

Figure 3 illustrates the overall work-flow in EA. At the end of the iteration, a set of near-optimum individuals can be found within the population of the last generation [49]. This set of near-optimum individuals is known as the PF in the study of multi-objective optimisation.

In the present study, the optimisation procedure explored several possible sensor node placements within the search space (i.e. region covered by the ESN) for a given number of nodes. Each potential solution (i.e. position of sensor nodes) is encoded as an individual (illustrated in Figure 1) where each one is ranked

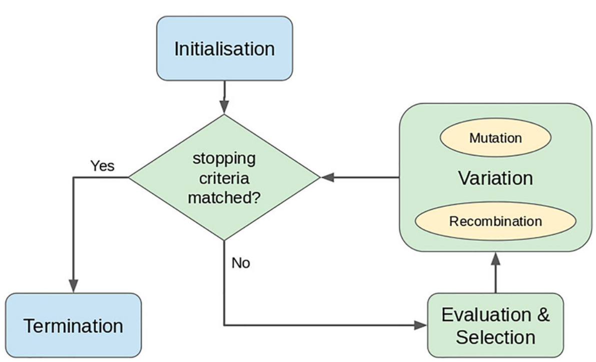

Figure 3. A general work-flow in an EA. The initialisation process generates a number of random individuals to form an initial population. The rest of the routines run iteratively until the predefined stopping criteria are met. The evaluation routine assesses each individual within the current population in respect to all predefined fitness functions (i.e. objective functions). The selection routine forms a new population for the next generation based on the fitness values assigned to each individual. Mutation and recombination maintain the variation within the population which is essential in order to explore the search space and to avoid local optima [50].

against its fitness (i.e. optimum robustness and redundancy). Figure 4 provides a visual description regarding the ESN design optimisation implemented in this study.

Both redundancy and robustness of an ESN design are translated as a set of fitness functions, which will be used to quantify and to evaluate the fitness of each individual produced in every generation. The formulation of these two functions is described in the following sections (Sections 3.2.1 and 3.2.2, respectively).

3.2.1. Redundancy

Redundancy in this work is defined as an unnecessary deployment of sensor nodes in which a node's role can be handled by its neighbouring nodes. Leave-one-out cross-validation (LOOCV) is utilised in conjunction with a spatial data interpolation technique to obtain the 'least redundant node placement' [53], where each node would cover the values which are unlikely to be estimated by its neighbouring nodes. For the purpose of clarity, Figure 5 provides a graphical illustration of the LOOCV. The function is formulated as follows:

$$\mathrm{LOOCV}(\hat{f}) = \sqrt{\frac{1}{N \cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - \hat{f}^{(-i)}(y_i, t))^2}, \quad (1)$$

where $\hat{f}$ is the spatial data interpolation function (Equation (2)), $y_{i,t}$ is the value measured by node y_i on tth day, $\hat{f}^{(-i)}(y_i, t)$ is the estimated value produced by $\hat{f}$ with the absence of ith node ($Y \backslash \{y_i\}$), N is the number of sensor nodes deployed in the RoI, T is the total number of days in a year.

Inverse distance weighting (IDW) proposed in [42] is chosen as the spatial data interpolation in the present work. Suppose y_o represents the node located within the RoI where a value is required to be estimated. The spatial data interpolation function is formulated as follows:

$$\hat{f}(y_o, t) = \sum_{i=1}^{N} \frac{d_i^{-1}}{\sum_{i=1}^{N} d_i^{-1}} \cdot y_{i,t}, \quad (2)$$

where $\hat{f}$ is the spatial data interpolation function (IDW), $\hat{f}(y_o, t)$ is the estimated value in node y_o on the tth day, N is number of known nodes (i.e. neighbouring nodes), d_i is the distance in space between y_o and y_i, $y_{i,t}$ is the value measured in node y_i on the tth day.

The function presented in Equation (1) is aiming to find a set of node locations in a way that each node is important and necessary to be deployed within the network. This allows the interpolator to estimate a good representation of the RoI. The absence of any node will greatly degrade the representativeness of the area under study, since each node is unlikely to be

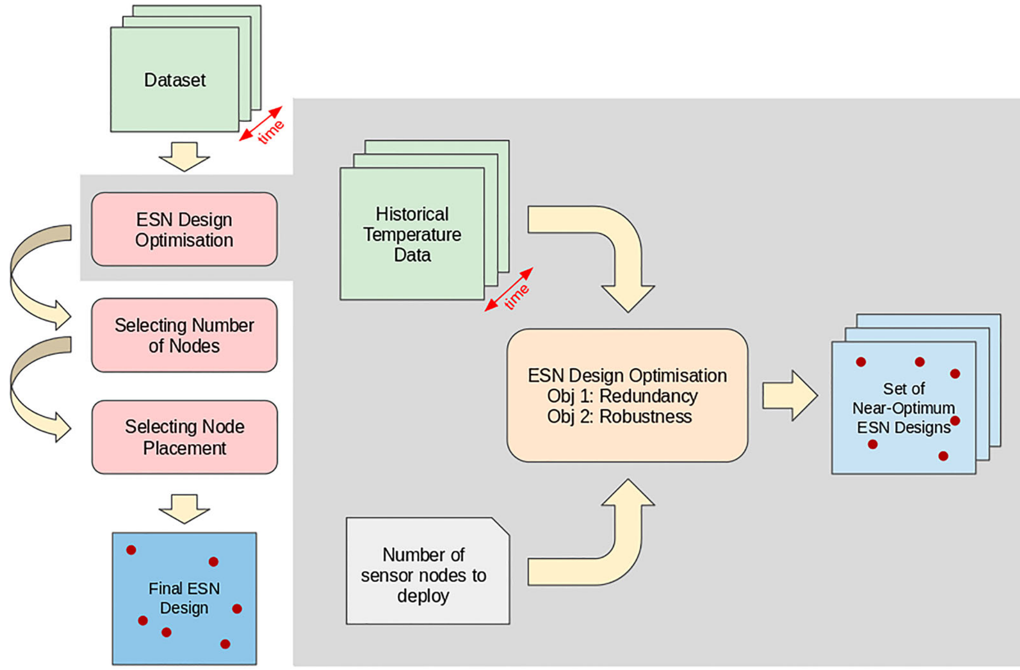

Figure 4. An overview of the ESN design optimisation procedure in this work.

estimated by its neighbouring nodes. In this case, a higher LOOCV($\hat{f}$) implies a lower degree of redundancy and an individual (ESN design) which yields a higher LOOCV($\hat{f}$) is comparatively better than the one which produces a lower value. The fitness function is calculated by maximising Equation (1).

3.2.2. Robustness

A robust ESN design, as defined in this work, is defined as a design that maintains the network's performance while dealing with the loss or disruption of a node within the network. The robustness is quantified by going through the *N* nodes, and uses linear temporal interpolation to

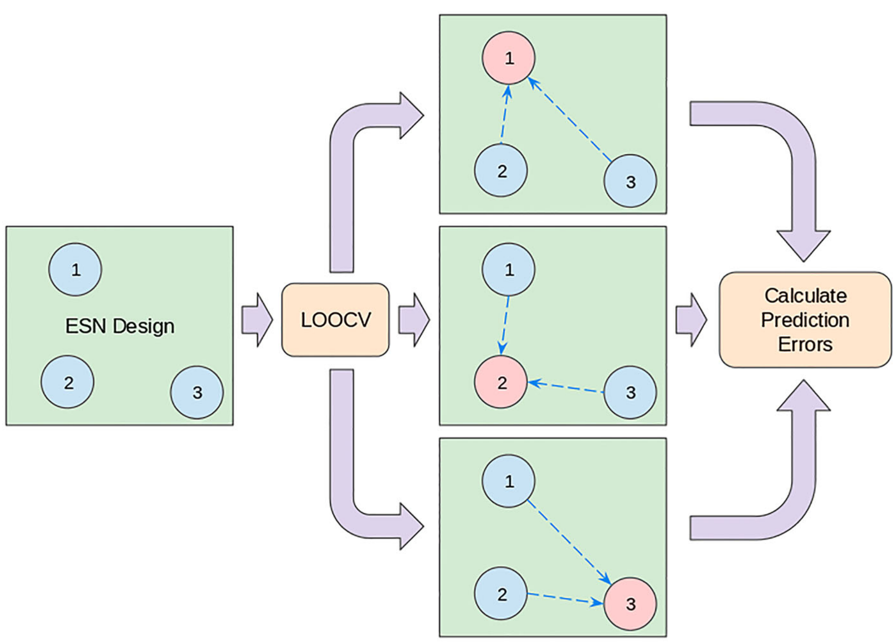

Figure 5. LOOCV applied in this study. For simplicity, the figure presents an example of an ESN formed by three sensor nodes. The LOOCV is conducted by omitting one sensor node in turn while the rest of the nodes are used to predict the omitted node. A spatial interpolation technique (IDW) is employed as a method to predict the node. In the end, the prediction errors are calculated as the output of the LOOCV.

estimate the averaged daily temperature at the location where each node is located, throughout the year (T days). The function is formulated as follows:

$$\mathrm{RMSE}(\hat{g}) = \sqrt{\frac{1}{N \cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - \hat{g}(y_i, t))^2}, \tag{3}$$

where $\mathrm{RMSE}(\hat{g})$ is the root-mean-squared error between the estimated value and the observed value, $\hat{g}$ is the temporal data interpolation function (Equation (4)), N is the number of sensor nodes deployed in the RoI, T is the total number of days in a year.

Linear interpolation is applied as the temporal data interpolation technique used in this work and it is formulated as follows:

$$\hat{g}(y_o, t) = 1/2 \cdot (y_{o,t-1} + y_{o,t+1}), \tag{4}$$

where $\hat{g}$ is the temporal data interpolation function, $\hat{g}(y_o, t)$ is the estimated value in node y_o on the tth day based on the past (t−1) and future (t+1) measurements within the same node (i.e. at the same location in space).

The function presented in Equation (3) measures the degree of error for each node in an ESN design to recover itself (i.e. by utilising a temporal data interpolation technique) in the case where missing values occur. In this case, a lower $\mathrm{RMSE}(\hat{g})$ implies a higher degree of robustness and an individual (i.e. ESN design) with lower $\mathrm{RMSE}(\hat{g})$ is relatively better compared to the one with higher value. The fitness function is calculated by minimising Equation (3).

3.3. Selecting number of nodes

Decision-making on how many sensor nodes should be deployed has never been an easy task in ESN design. Any decision will eventually have an impact on both the deployment and the maintenance cost of the networks. In order to assist such a decision-making process, we simulate several different numbers of nodes and apply the previously described optimisation procedure (Section 3.2). For each near-optimum design produced, we assessed its representativeness with respect to the RoI. We formulate the representativeness of an ESN design as the capability of the networks to interpolate the averaged daily temperatures across the entire space within the RoI. The performance of each design is measured based on the prediction error over a period of one year. In this case, a lower prediction error indicates a better representativeness. The prediction is calculated based on the spatial interpolation technique (i.e. IDW) as previously described in Equation (2). The formula adopted in this work to quantify the prediction error is as follows:

$$\mathrm{RMSE}(\hat{f}) = \sqrt{\frac{1}{N \cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{i,t} - \hat{f}(x_i, t))^2}, \tag{5}$$

where $\hat{f}$ is the spatial data interpolation function (Equation (2)), N is the total number of cells in two-dimensional space X, T is the total number of days in a year, $x_{i,t}$ is the average temperature data measured at location x_i on the tth day, $\hat{f}(x_i, t)$ is the estimated value for average temperature data at location x_i on the tth day.

In general, ESN designs which incorporate higher numbers of sensor nodes should better represent the RoI compared to those with fewer nodes. The procedure proposed within this work would be able to quantify the degree of representativeness which could be gained by incorporating certain numbers of sensor nodes into the design. The final decision in determining the number of sensors to be included would certainly be restricted by the budget constraints; and the proposed procedure could benefit in assisting such decision-making process. Figure 6 illustrates the procedure for selecting the number of sensor nodes in this work.

3.4. Selecting node placement

Once a number of nodes has been decided, the next decision to be made is selecting an ESN design from a set of near-optimum designs. Due to the inherent conflicting nature of the objectives for multi-objective optimisation problems, where optimising one objective may sacrifice the other objective, there rarely exists a single solution that simultaneously optimises all the objectives. Therefore, instead of having a single optimum solution, a PF is formed to capture a set of near-optimum solutions (i.e. non-dominated solutions) [50]. In this work, a given number of sensor nodes would yield a set of different possible placements of the nodes across the RoI. Each placement will produce a particular composition among two fitness values (redundancy and robustness). The selection of the final solution (the final ESN design) within the PF will involve the users' domain knowledge of the problem on hand. In our work, the process of selecting a final ESN design is divided into two stages: feature assessment and feasibility assessment. Figure 7 presents the procedure for selecting a final ESN design in this work.

3.4.1. Feature assessment

The proposed method uses two fitness functions while searching in the optimum space. One function measures robustness and the other one redundancy. Running both

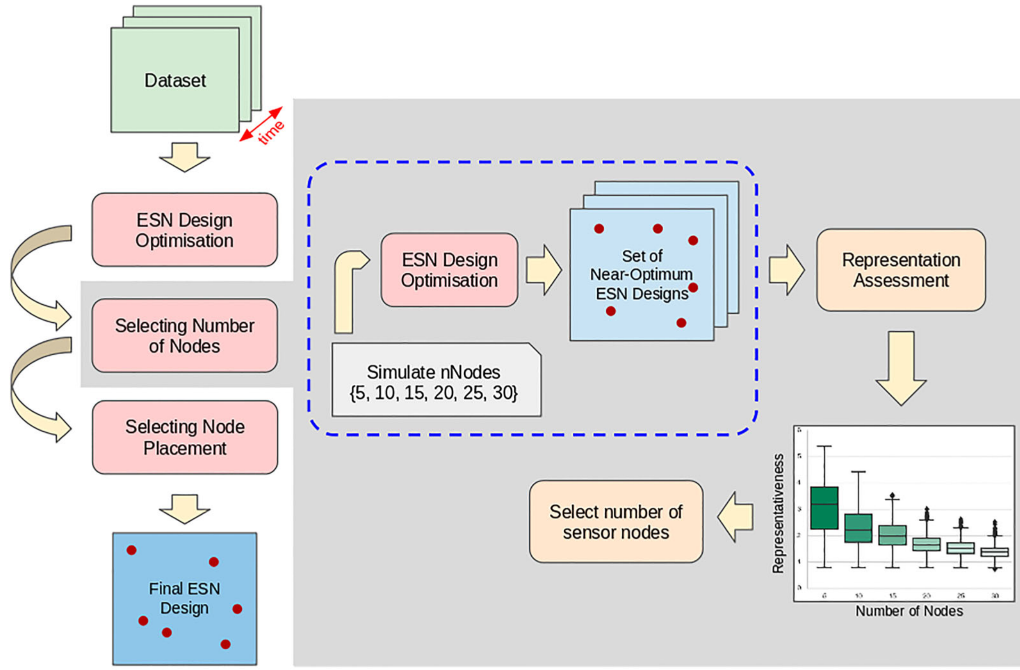

Figure 6. An overview of the procedure for selecting number of sensor nodes with respect to the representativeness yielded by all the optimised ESN designs for a given set of different number of nodes.

functions will generate several possible solutions where redundancy and robustness vary in value for an optimum design of the network. Feature assessment involves the decision maker's preferences while dealing with the trade-off between redundancy and the robustness. The advantage of this method is that the users can decide the level of redundancy and robustness they choose to have.

3.4.2. Feasibility assessment

A final phase of the decision on where sensor nodes should be deployed can be made by considering

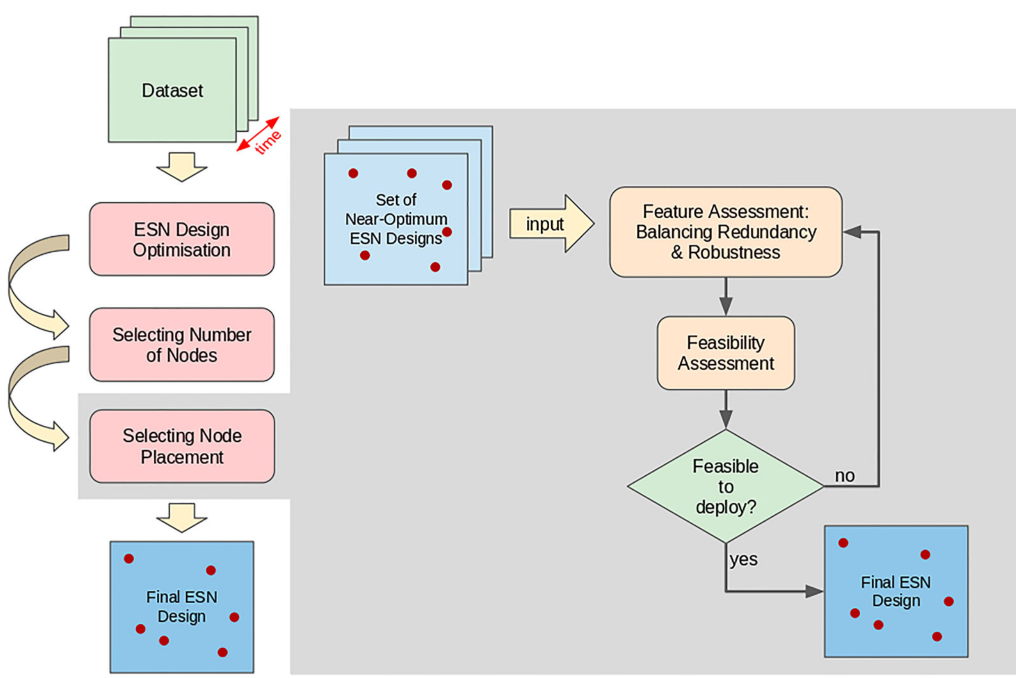

Figure 7. An overview of the procedure in selecting an ESN design.

Table 1. Parameters which are implemented in our work to run the EA.

Parameter	Value
Population size	30
Crossover probability	0.9
Mutation probability	0.01
Crossover operation	One point
Mutation operation	Uniform integer
Selection operation	NSGA2 [56]
Max. number of generations	1000
Saturation point threshold	30 generations

accessibility of the site, costs, safety, among other aspects associated with field work. Many of these attributes lack numerical values. This limitation is overcome by human judgement. The feasibility assessment would incorporate the decision makers' domain knowledge of the RoI:

- whether the design has 'captured' critical locations within the landscape;
- the feasibility of the design to be deployed in respect to the constraints in the physical landscape;
- the level of sparsity of the node locations to avoid clustering.

3.5. Experimental setup

The optimisation procedure in this work was implemented in Python and an open source Python module: distributed evolutionary algorithms in Python [54] was utilised to assist in the optimisation process. Grefenstette's parameter setting [55] was adopted as the EA's parameters (Table 1) in this work.

Replication is applied to evaluate our work. We utilised the method proposed in [57] to determine the required number of replications. The method is formulated as follows:

$$n\_rep = \left(\frac{z \times \sigma}{\mu \times 0.05}\right)^2, \tag{6}$$

where n_rep is the prescribed number of replications, z refers to z-score of 1.96 which leads to a 95% confidence interval; μ and σ are the mean and standard deviation obtained from a preliminary simulation; and 0.05 refers to the expected deviation percentage (5%) of μ. We ran our preliminary simulation over 10 different runs, each with a different random seed, for each set of sensor nodes. The prescribed number of replications resulting from our preliminary simulation is presented in Table 2.

We adopted 10 as the minimum number of replications in the case where $n\_rep < 10$. As shown by Table 2, there is no n_rep greater than 10, therefore we adopt

Table 2. Prescribed number of replications (n_rep) from our preliminary simulation.

N	5	10	15	20	25	30
Fitnessfunction 1	7	9	7	9	3	4
Fitness function 2	1	1	1	1	1	1

Note: The calculation is based on Equation (6) for each different number of sensor nodes with respect to the two fitness functions (Equation (1) and (3)).

10 as the replication number for all simulations in this work. In our experiment, we simulate six different numbers of sensor nodes to be deployed (5, 10, 15, 20, 25, and 30 nodes). For each number of nodes, we run the optimisation procedure (described in Section 3.2) with 10 replications. A unique random seed number is assigned for each replication.

4. Results and discussion

The impact of the increase in the number of sensor nodes on the representativeness of the ESN is presented in Figure 8. The figure suggests that an increase in the number of nodes will improve the representativeness of the networks, which can be seen by the decline in prediction error and also by the reduction in data deviation. This is not surprising, considering there are more data samples being collected from the RoI. A relatively more significant improvement can be achieved by increasing the number of nodes from 5, 10, and 15, up to 20 nodes. Once the number of nodes reached 20, adding more nodes resulted in less significant improvement.

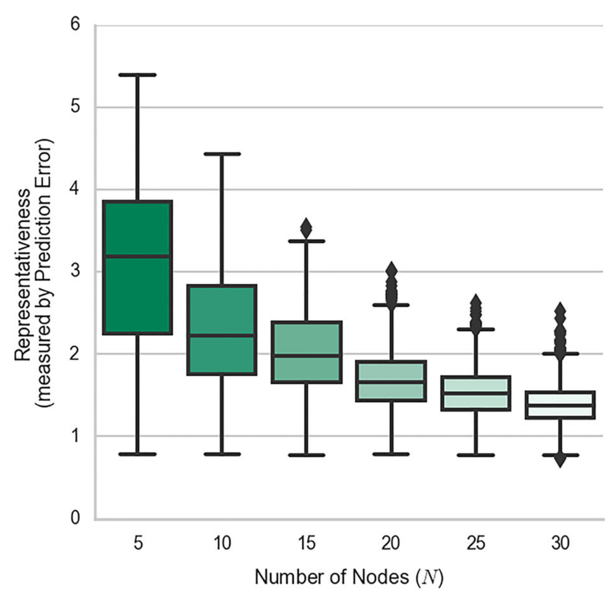

Figure 8. The impact of increasing the number of sensor nodes on the capability of the ESN to interpolate the entire space of the RoI over a one-year period of time. The plotting is based on 10 simulation runs with a different random seed in each run.

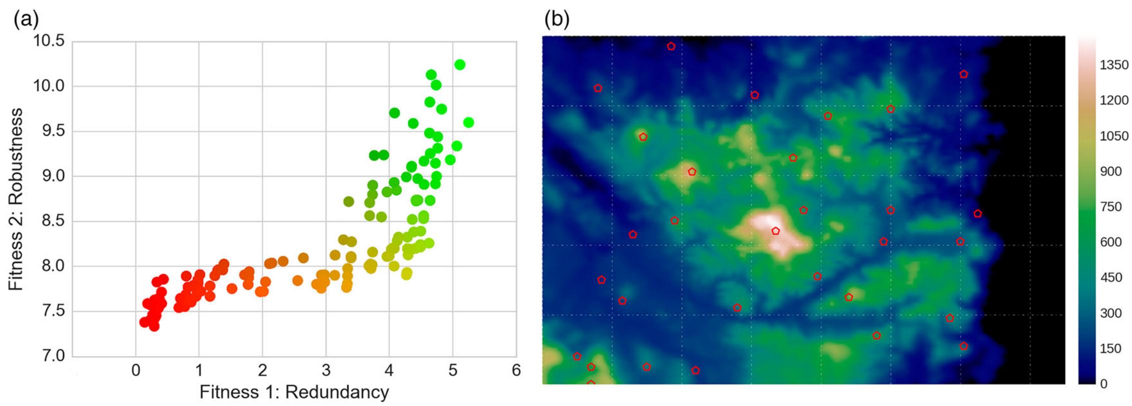

Figure 9. The figures are generated for 30 sensor nodes with the aim of assisting a decision maker in selecting a final ESN design. The markers plotted in (a) represent all the near-optimum ESN designs optimised for 30 sensor nodes (with respect to both redundancy and robustness as the features to be assessed). The green markers favour more redundancy (maximum LOOCV($\hat{f}$)) whereas the red ones favour more robustness (minimum RMSE($\hat{g}$)). (b) The placement for sensor nodes in the RoI chosen from one of the markers plotted in (a). The background colour represents the elevation in the RoI (in metres).

The decision to incorporate a certain number of sensor nodes would be driven by the purpose of the ESN and also the current budget situation. If the budget only allows for the deployment of a limited number of sensor nodes, then the decision maker could analyse whether the resulting representativeness is sufficient or not for the purpose of operating the network. The information generated from this simulation would benefit such a decision-making process.

Figure 9(a) presents all the near-optimum solutions obtained for the design with 30 sensor nodes. Each point of this plot represents a single near-optimum ESN design of a given number of nodes. All the designs are generated by our method (over 10 different runs) with respect to the two fitness functions (i.e. redundancy and robustness). Two different colours are employed in the figure to represent the degree of priority among the two fitness values (i.e. green and red to represent optimum redundancy and optimum robustness, respectively). Having two objective functions has the advantage of leaving it to the decision maker of the network to decide what property he wants the network to have: more or less robustness, or more or less redundancy, or a balance of both. The visualisation presented in Figure 9(a) would assist the decision maker in understanding the trade-off which they are dealing with while selecting a certain ESN design.

Decision makers will be further assisted by another data visualisation presented in Figure 9(b). The colour bar on the right side of the figure represents the elevation (i.e. altitude). This kind of visualisation assists a decision maker to examine the feasibility of a particular ESN design for deployment in the RoI. In the case where the feasibility criteria were not satisfied, decision makers could select some other alternative of ESN designs (Figure 9(a)) and iterate the procedure until the most desirable ESN design is found.

5. Conclusions

Design is an essential process prior to the deployment of an ESN. The decision in determining the number of sensor nodes and the placement of the nodes in the RoI would impact the effectiveness of the network and the efficiency of its operation. This paper presented an ESN design technique which is able to promote robustness (to maintain a fit-for-purpose ESN) while incorporating minimum redundancy. This technique will benefit a decision maker to find a balance between redundancy and robustness in the design of ESN. An EA is utilised in this work to locate a set of near-optimum ESN designs. The technique can support a decision maker, particularly in determining the number of sensor nodes and how the nodes will be deployed in the RoI.

Acknowledgements

The experiment was conducted using the SouthEsk data model owned by the Commonwealth Scientific Research and Industrial Organisation (CSIRO). The authors wish to thank Ray Williams (Data61, CSIRO) and Tom Gillard for proofreading the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

SB would like to thank Sense-T for the PhD scholarship and CSIRO for the top-up scholarship.

ORCID

Setia Budi http://orcid.org/0000-0003-1408-7808
Ferry Susanto http://orcid.org/0000-0002-7484-8253
Paulo de Souza http://orcid.org/0000-0002-0091-8925
Greg Timms http://orcid.org/0000-0001-5289-2619
Vishv Malhotra http://orcid.org/0000-0002-5732-8467
Paul Turner http://orcid.org/0000-0003-4504-2338

References

[1] Javier Ferrandez-Pastor F, Manuel Garcia-Chamizo J, Nieto-Hidalgo M, et al. Developing ubiquitous sensor network platform using internet of things: application in precision agriculture. Sensors. 2016;16:1141–1161.

[2] Pajares G, Peruzzi A, Gonzalez-de-Santos P. Sensors in agriculture and forestry. Sensors. 2013;13:12132–12139. Available from: http://www.mdpi.com/1424-8220/13/9/12132

[3] Heinz E, Kraft P, Buchen C, et al. Set up of an automatic water quality sampling system in irrigation agriculture. Sensors. 2014;14:212–228. Available from: http://www.mdpi.com/1424-8220/14/1/212

[4] Kim K, Myung H. Sensor node for remote monitoring of waterborne disease-causing bacteria. Sensors. 2015;15:10569–10579. Available from: http://www.mdpi.com/1424-8220/15/5/10569

[5] Tomperi J, Koivuranta E, Kuokkanen A, et al. Real-time optical monitoring of the wastewater treatment process. Environ Technol. 2016;37:344–351. Available from: http://dx.doi.org/10.1080/09593330.2015.1069898

[6] Robles A, Durán F, Ruano MV, et al. Instrumentation, control, and automation for submerged anaerobic membrane bioreactors. Environ Technol. 2015;36:1795–1806. Available from: http://dx.doi.org/10.1080/09593330.2015.1012180

[7] Shih CW, Wang CH. Integrating wireless sensor networks with statistical quality control to develop a cold chain system in food industries. Comput Standards Interfaces. 2016;45:62–78. Available from: http://www.sciencedirect.com/science/article/pii/S0920548915001452

[8] Badia-Melis R, Ruiz-Garcia L, Garcia-Hierro J, et al. Refrigerated fruit storage monitoring combining two different wireless sensing technologies: RFID and WSN. Sensors. 2015;15:4781–4795. Available from: http://www.mdpi.com/1424-8220/15/3/4781

[9] Mejjaouli S, Babiceanu RF. RFID-wireless sensor networks integration: decision models and optimization of logistics systems operations. J Manuf Syst. 2015;35:234–245. Available from: http://www.sciencedirect.com/science/article/pii/S0278612515 000114

[10] Gavalas D, Kenteris M. A web-based pervasive recommendation system for mobile tourist guides. Pers Ubiquit Comput. 2011;15:759–770. Available from: http://link.springer.com/article/10.1007/s00779-011-0389-x

[11] Joo J, Yim J, Lee CK. Protecting cultural heritage tourism sites with the ubiquitous sensor network. J Sustain Tour. 2009;17:397–406. Available from: http://dx.doi.org/10.1080/0966958080 2582498

[12] Li D, Cao J, Yao Y. Big data in smart cities. Sci China Inform Sci. 2015;58:1–12. Available from: http://link.springer.com/article/10.1007/s11432-015-5396-5

[13] Faye S, Chaudet C. Characterizing the topology of an urban wireless sensor network for road traffic management. IEEE Trans Veh Technol. 2016;65:5720–5725.

[14] Vera-Gómez JA, Quesada-Arencibia A, García CR, et al. An intelligent parking management system for urban areas. Sensors. 2016;16:931. Available from: http://www.mdpi.com/1424-8220/16/6/931

[15] Dutta R, Das A, Aryal J. Big data integration shows Australian bush-fire frequency is increasing significantly. R Soc Open Sci. 2016;3:150241. Available from: http://rsos.royalsociety publishing.org/content/3/2/150241

[16] Furquim G, Pessin G, Faiçal BS, et al. Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory. Neural Comput Appl. 2016;27:1129–1141. Available from: http://link.springer.com/article/10.1007/s00521-015-1930-z

[17] Bums DK, Kember PN, Taylor S, et al. Air pollution monitoring and the role of semiconductor sensors. Environ Technol Lett. 1980;1:259–272. Available from: http://dx.doi.org/10.1080/09593338009383975

[18] Hu B, Chen W, Zhou J. High performance flexible sensor based on inorganic nanomaterials. Sensors Actuators B: Chem. 2013;176:522–533. Available from: http://www.sciencedirect.com/science/article/pii/S0925400512009483

[19] Felemban E, Shaikh FK, Qureshi UM, et al. Underwater sensor network applications: a comprehensive survey. Int J Distrib Sensor Netw. 2015;11:896832. Available from: http://dsn.sagepub.com/content/11/11/896832

[20] Silva LMC, Salgado AM, Coelho MAZ. Development of an amperometric biosensor for phenol detection. Environ Technol. 2011;32:493–497. Available from: http://dx.doi.org/10.1080/09593330.2010.504234

[21] Bhandary V, Malik A, Kumar S. Routing in wireless multimedia sensor networks: a survey of existing protocols and open research issues. J Eng. 2016;2016:e9608757. Available from: http://www.hindawi.com/journals/je/2016/9608757/abs/

[22] Wu Y, Cardei M. Multi-channel and cognitive radio approaches for wireless sensor networks. Comput Commun. 2016;94:30–45. Available from: http://www.sciencedirect.com/science/article/pii/S014036641630295X

[23] Rew R, Davis G. NetCDF: an interface for scientific data access. IEEE Comput Graph Appl. 1990;10:76–82.

[24] Reed C, Buehler K, McKee L. OGC consensus: how successful standards are made. ISPRS Int J Geo-Inform. 2015;4:1693–1706. Available from: http://www.mdpi.com/2220-9964/4/3/1693

[25] Li W, Wu S, Song M, et al. A scalable cyberinfrastructure solution to support big data management and multivariate visualization of time-series sensor observation data. Earth Sci Inform. 2016;9:449–464. Available from: http://link.springer.com/article/10.1007/s12145-016-0267-1

[26] Liao Q, Shi L, He Y, et al. Visualizing anomalies in sensor networks. Proceedings of the ACM SIGCOMM 2011 conference (SIGCOMM'11). New York (NY): ACM; 2011. p. 460–461. Available from: http://doi.acm.org/10.1145/2018436.2018521

[27] Liao Y, Mollineaux M, Hsu R, et al. SnowFort: an open source wireless sensor network for data analytics in infrastructure and environmental monitoring. IEEE Sensors J. 2014;14:4253–4263.

[28] Younis M, Akkaya K. Strategies and techniques for node placement in wireless sensor networks: a survey. Ad Hoc Netw. 2008;6:621–655. Available from: http://www.sciencedirect.com/science/article/pii/S1570870507000984
[29] Budi S, P de Souza, Timms G, et al. Optimisation in the design of environmental sensor networks with robustness consideration. Sensors. 2015;15:29765–29781. Available from: http://www.mdpi.com/1424-8220/15/12/29765
[30] D'Este C, de Souza P, Sharman C, et al. Relocatable, automated cost-benefit analysis for marine sensor network design. Sensors. 2012;12:2874–2898. Available from: http://www.mdpi.com/1424-8220/12/3/2874
[31] Smith D, Timms G, de Souza P, et al. A Bayesian framework for the automated online assessment of sensor data quality. Sensors. 2012;12:9476–9501. Available from: http://www.mdpi.com/1424-8220/12/7/9476
[32] Timms GP, de Souza PA, Reznik L, et al. Automated data quality assessment of marine sensors. Sensors. 2011;11:9589–9602. Available from: http://www.mdpi.com/1424-8220/11/10/9589
[33] Campbell JL, Rustad LE, Porter JH, et al. Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data. BioScience. 2013;63:574–585.
[34] Sun H, Koch M. Case study: analysis and forecasting of salinity in apalachicola bay, Florida, using Box-Jenkins ARIMA models. J Hydraul Eng. 2001;127:718–727. Available from: http://dx.doi.org/10.1061/(ASCE)0733-9429(2001)127:9(718)
[35] Chatterjea S, Havinga P. Improving temporal coverage of an energy-efficient data extraction algorithm for environmental monitoring using wireless sensor networks. Sensors. 2009;9:4941–4954. Available from: http://www.mdpi.com/1424-8220/9/6/4941
[36] Kumar U, Jain VK. ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). Stoch Environ Res Risk Assess. 2010;24:751–760. Available from: http://link.springer.com/article/10.1007/s00477-009-0361-8
[37] Ho SL, Xie M, Goh TN. A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. Comput Ind Eng. 2002;42:371–375. Available from: http://www.sciencedirect.com/science/article/pii/S0360835202000360
[38] Valipour M, Banihabib ME, Behbahani SMR. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. J Hydrol. 2013;476:433–441. Available from: http://www.sciencedirect.com/science/article/pii/S002216941200981X
[39] He J, Zhang Y, Huang G, et al. CIRCE: correcting imprecise readings and compressing excrescent points for querying common patterns in uncertain sensor streams. Inform Syst. 2013;38:1234–1251. Available from: http://www.sciencedirect.com/science/article/pii/S0306437912000063
[40] Shahriar MS, de Souza P, Timms G. On retrieving patterns in environmental sensor data. Earth Sci Inform. 2012;5:43–59. Available from: http://link.springer.com/article/10.1007/s12145-012-0095-x
[41] Li J, Heap AD. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecol Inform. 2011;6:228–241. Available from: http://www.sciencedirect.com/science/article/pii/S1574954110001147
[42] Shepard D. A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM National Conference; 1968 Jan 1. New York (NY): ACM. p. 517–524.
[43] Cressie N. Spatial prediction and ordinary kriging. Math Geol. 1988;20:405–421. Available from: http://link.springer.com/article/10.1007/BF00892986
[44] Susanto F, de Souza P, He J. Spatiotemporal interpolation for environmental modelling. Sensors. 2016;16:1245. Available from: http://www.mdpi.com/1424-8220/16/8/1245
[45] Boukerche A, editor. Algorithms and protocols for wireless sensor networks. Hoboken (NJ): Wiley; 2009 (Wiley series on parallel and distributed computing).
[46] Franceschini F, editor. Distributed large-scale dimensional metrology: new insights. London: Springer; 2011.
[47] Katzfey J, Thatcher M. Ensemble one-kilometre forecasts for the South Esk Hydrological Sensor Web. 19th International Congress on Modelling and Simulation (modsim2011); 2011; Perth, Australia. p. 3511–3517. WOS:0003149893 03069
[48] Terhors A, Peters C, Katzfey J, Tasmanian Partnership for Advanced Computing. TAS sensor web – conformal cubic atmospheric model (CCAM) simulations – 1 km resolution – South Esk; 2011 [cited 2016 Nov 13]. Available from: https://researchdata.ands.org.au/tas-sensor-web-south-esk.
[49] Deb K. Multi-objective optimization using evolutionary algorithms. 1st ed. Chichester: Wiley; 2009.
[50] Coello CC, Lamont GB, Veldhuizen DAv. Evolutionary algorithms for solving multi-objective problems. 2nd ed. New York (NY): Springer; 2007.
[51] Ngatchou P, Zarei A, El-Sharkawi A. Pareto multi objective optimization. Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems; 2005; Washington, DC, USA. p. 84–91.
[52] Huang B, Fery P, Xue L, et al. Seeking the Pareto front for multiobjective spatial optimization problems. Int J Geograph Inform Sci. 2008;22:507–526. Available from: http://dx.doi.org/10.1080/13658810701492365
[53] Susanto F, Budi S, de Souza P, et al. Design of environmental sensor networks using evolutionary algorithms. IEEE Geosci Remote Sens Lett. 2016;13:575–579.
[54] Fortin FA, De Rainville FM, Gardner MAG, et al. DEAP: evolutionary algorithms made easy. J Mach Learn Res. 2012;13:2171–2175. Available from: http://dl.acm.org/citation.cfm?id=2503308.2503311
[55] Yu X, Gen M. Introduction to evolutionary algorithms. 2010th ed. London: Springer; 2010.
[56] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput. 2002;6:182–197. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=996017
[57] Jain R. The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling. New York (NY): Wiley; 1991.