Optimising Sensor Layouts for Direct Measurement of Discrete Variables

X. Rosalind Wang
CSIRO ICT Centre
PO Box 76, Epping
NSW 1710, Australia
Rosalind.Wang@csiro.au

George Mathews
Australian Centre for Field Robotics
The University of Sydney
NSW 2006, Australia
g.mathews@acfr.usyd.edu.au

Don Price
CSIRO Materials Science and Engineering
PO Box 218, Lindfield
NSW 2070, Australia
Don.Price@csiro.au

Mikhail Prokopenko
CSIRO ICT Centre
PO Box 76, Epping
NSW 1710, Australia
Mikhail.Prokopenko@csiro.au

Abstract

An optimal sensor layout is attained when a limited number of sensors are placed in an area such that the cost of the placement is minimised while the value of the obtained information is maximised.In this paper, we discuss the optimal sensor layout design problem from first principles, show how an existing optimisation criterion (maximum entropy of the measured variables) can be derived, and compare the performance of this criterion with three others that have been reported in the literature for a specific situation for which we have detailed experimental data available. This is achieved by firstly learning a spatial model of the environment using a Bayesian Network, then predicting the expected sensor data in the rest of the space, and finally verifying the predicted results with the experimental measurements. The development of rigorous techniques for optimising sensor layouts is argued to be an essential requirement for reconfigurable and self-adaptive networks.

1. Introduction

Distributed sensing is becoming increasingly important in many areas of modern society. Sensing for environmental monitoring (on scales from personal spaces - rooms, buildings, vehicles - to catchments, ecosystems, continents and oceans), industrial process monitoring, monitoring for security and safety, health and well-being, and structural health monitoring are examples of application areas in which distributed sensing is becoming important.

Critical issues that must be addressed in the development of a distributed sensing system are what is to be sensed, where sensors should be located (and how many are required), and how should the data be communicated and processed. In practice, limitations (sometimes severe) are placed on the number of sensors that may be deployed by considerations of cost, weight, sensor size, and the ability to communicate and process the large volume of data effectively. Therefore, it is important to find methods to optimise the placement of a limited number of sensor nodes in the area of interest such that the *cost* is minimised while the *value* or *utility* of the obtained information is maximised. The ultimate aim of an optimisation process must be to enable a reliable and acceptably-accurate estimate to be made of the relevant state of the sensed object at minimal cost.

In this context, cost may include the installation cost, the energy cost of using the sensors, communication costs, the cost of processing and using the data, and any additional cost of operating the structure with the sensors deployed (e.g. increased operating costs of an aircraft due to additional weight of sensors). Value is understood as the expected benefit of using the sensor configuration (or layout) and is expressed in terms of the information gain and the usefulness of the information. The usefulness (utility) of the information must incorporate the notion of an acceptably-accurate estimate of the state of the object: the utility of information is high if it improves the accuracy of the estimate towards the required level, but is low if it doesn't, or if it leads to leads to an estimate of greater accuracy than is required.

Furthermore, as conditions, situations and/or system knowledge change, the optimum configuration of the network may change from the initial layout. In such dynamic situations there will be a need for reconfigurable or adaptive optimisation of the network. In the simplest case, sensors can be initially deployed over one area, some of which are later removed (after an optimal layout is determined), and deployed in new areas. Reconfiguration may also be performed autonomously (self-adaptation) using mobile sensors, but however it is carried out it must be based on a rigorous optimisation criterion. It is therefore clear that principled optimisation of sensor layouts is an essential requirement of self-adapting sensing systems.

A number of information-based criteria have been pro-

posed for optimising sensor placements, and some of these will be outlined briefly below. This paper presents some initial results of a larger program whose aim is to develop general principles and techniques for deriving and applying appropriate optimisation criteria for a wide range of sensing requirements, and to review, compare and evaluate a range of different criteria. Optimisation requires a model of the system, and we will study both data-driven (statistical) system models and physical models. The ultimate aim of this research is to develop self-adaptive (and self-organizing) sensing systems for a range of application domains: optimising the sensor layout is a necessary step in making the system reconfigurable and eventually self-adapting.

The present paper describes the first stage of this work, in which the performances of four existing criteria are compared for a specific situation for which we have detailed experimental data available to derive a statistical model (in this example a Bayesian network model is derived), but no reliable physical model. The data consists of direct measurements of the quantity of interest (in this case soil moisture).

Later work will investigate a range of other sensing situations, that can be modelled either physically or statistically, and which in some cases will require inferential (or indirect) sensing of important quantities rather than the direct sensing used in the present example.

The paper proceeds as follows. Section 2 discusses the optimal sensor layout design problem from first principles, which is essentially a decision-theoretic problem. However, while such problems may be readily formulated, defining the computational details (in this case information utility and cost) is often very difficult [1] and information-based approximations are used. Defining expected costs information-theoretically leads to representation of the sensor layout problem as an entropy optimisation problem.This is then contrasted with three other reported criteria, outlined in Section 3, some of which are decision-theoretic while others are also information-theoretic. Thus we compare the performance of four existing criteria for an optimal sensor layout (for direct sensing), which are:

1) maximum entropy of the measured variables (Equation 18)
2) minimum residual entropy of the system state variables, given the sensor measurements (Equation 20)
3) maximum mutual information between the state variables that haven't been measured, and the sensor measurements (Equation 21)
4) maximum information coverage of the sensor measurements (Equation 24)

These criteria are compared using data from an existing wireless sensor network that measures soil moisture on a dense, approximately rectangular grid. The experimental setup of the sensor network and the environment model (a graphical model based on a Bayesian Network) are presented in Section 4. To compare the criteria, we first compute optimal configurations, then predict the sensor measurements in the rest of the space and verify the results using the available measurements. This is a somewhat artificial problem, in the sense that data from a dense network is used to develop a statistical model of the sensed environment, but it is one that may occur in practice in the design of environmental monitoring networks. Small areas of a larger region to be eventually monitored may be provided with a dense coverage of sensors to allow an optimal network design to be developed, as outlined in the paper. The cost of sensors usually limits the number available, so after the optimal sensor distribution for one area has been deduced, the excess sensors are re-deployed to another area. Section 5 discusses the results of the experiment.

2. Optimal Design Problem

In this section, we present a general approach to optimising sensor layouts. Arguably, taking observations is aimed at improving the outcome of a future action. Consider the decision problem where the cost incurred by a future decision is described by the function $C : \mathcal{A} \times \mathcal{X} \rightarrow \Re$, where $\mathcal{A}$ is the set of possible decisions or actions, a, and $\mathcal{X}$ is the set of possible world states, x, that are relevant to the decision problem. There may be different costs for taking or failing to take a particular decision (to repair a defect, for example) when the environmental conditions require it.

If the true state $x \in \mathcal{X}$ is known, the optimal action is easily found by

$$a^* = \arg\min_{a \in \mathcal{A}} C(a, x) \tag{1}$$

However, for realistic systems, it is often not known what is the exact state x. For example, the true environmental conditions that require a repair action may not be known. Thus, it is necessary to consider the state to be a random variable $\mathbf{X}$ with a given probability distribution $\mathcal{P}_{\mathbf{X}} \triangleq [P(x_1), P(x_2), \ldots, P(x_{|\mathcal{X}|})]$, where $P(x) \triangleq \Pr(\mathbf{X} = x)$. The distribution $\mathcal{P}_{\mathbf{X}}$ is called the *prior belief*, and defines a model of the environment.

Now, since the state is not known, the actual cost of a particular action cannot be determined with certainty and the expected cost of an action should be considered. The expected cost of an action is given by:

$$J(a, \mathcal{P}_{\mathbf{X}}) = \sum_{x \in \mathcal{X}} P(x) C(a, x), \tag{2}$$

with the optimal expected cost given by

$$J^*(\mathcal{P}_{\mathbf{X}}) = \min_{a \in \mathcal{A}} J(a, \mathcal{P}_{\mathbf{X}}). \tag{3}$$

This is a function of the given distribution $\mathcal{P}_{\mathbf{X}}$. This generic characterisation does not include actual observations.

2.1. Including Observations

Now consider a scenario where a sensor can be deployed prior to taking the action to provide some information about the value of the state x. Hopefully this will lead to an improved final decision. Formally, the output of the sensor is modelled using the random variable $\mathbf{Z}$, in the discrete domain $\mathcal{Z}$. The dependency of the sensor readings on the state x is modelled using a conditional probability distribution function

$$\mathcal{P}_{\mathbf{Z}|\mathbf{X}} = \begin{bmatrix} P(z_1|x_1) & \cdots & P(z_1|x_{|\mathbf{X}|}) \\ \vdots & \ddots & \vdots \\ P(z_{|\mathbf{Z}|}|x_1) & \dots & P(z_{|\mathbf{Z}|}|x_{|\mathbf{X}|}) \end{bmatrix} \tag{4}$$

where $P(z|x) \triangleq \Pr(\mathbf{Z} = z|\mathbf{X} = x)$. The distribution $\mathcal{P}_{\mathbf{Z}|\mathbf{X}}$ is called the *sensor model* and completely specifies the characteristics of the sensor.

After a specific measurement z has been made, a new conditional distribution $\mathcal{P}_{\mathbf{X}|z} \triangleq [P(x_1|z), \dots, P(x_{|\mathcal{X}|}|z)]$ for the value of $\mathbf{X}$ can be generated using Bayes rule:

$$P(x|z) = \frac{P(z|x)\ P(x)}{\sum_{x\in\mathcal{X}} P(z|x)P(x)}, \tag{5}$$

The distribution $\mathcal{P}_{\mathbf{X}|z}$ will be referred to as the *posterior belief* and is defined on the same domain as $\mathcal{P}_{\mathbf{X}}$. The optimal action, for this belief, has an expected cost of

$$J^*\big(\mathcal{P}_{\mathbf{X}|z}\big) = \min_{a\in\mathcal{A}} \sum_{x\in\mathcal{X}} P(x|z)C(a,x). \tag{6}$$

This is a *posterior* measure, since it requires the value of the observation z. An *a priori* measure can be constructed by considering the expectation over all observations $z \in \mathcal{Z}$ of $J^*\big(\mathcal{P}_{\mathbf{X}|z}\big)$ and a distribution over the observations, $P(z)$:

$$\begin{aligned} G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) &= \sum_{z\in\mathcal{Z}} P(z) J^*\big(\mathcal{P}_{\mathbf{X}|z}\big) \\ &= \sum_{z\in\mathcal{Z}} P(z) \min_{a\in\mathcal{A}} \Big\{ \sum_{x\in\mathcal{X}} P(x|z)C(a,x) \Big\}. \end{aligned} \tag{7}$$

The function G^* is completely specified by the prior belief, $\mathcal{P}_{\mathbf{X}}$, and the sensor model, $\mathcal{P}_{\mathbf{Z}|\mathbf{X}}$.

Now, if the function J^* is concave (for example, if $J^*(\mathbf{Y}) = \sum_{y\in\mathbf{Y}} y \log \frac{1}{y}$, for some vector $\mathbf{Y}$), then Jensen's inequality gives us:

$$\begin{aligned} G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) &= \sum_{z\in\mathcal{Z}} P(z) J^*\big(\mathcal{P}_{\mathbf{X}|z}\big) \\ &\leq J^*\big(\sum_{z\in\mathcal{Z}} P(z)\mathcal{P}_{\mathbf{X}|z}\big) = J^*(\mathcal{P}_{\mathbf{X}}) \end{aligned} \tag{8}$$

This means that $G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}})$ is never greater than $J^*(\mathcal{P}_{\mathbf{X}})$, or put another way, on average utilising a sensor is never detrimental to the final decision problem, irrespective of the sensor model. We will prove later that the function, J^*, is concave. The difference between G^* and J^* can be used to determine the value or expected benefit of using the sensor modelled by $\mathcal{P}_{\mathbf{Z}|\mathbf{X}}$ for a given prior belief $P_{\mathbf{X}}$:

$$V(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) = J^*(\mathcal{P}_{\mathbf{X}}) - G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) \geq 0. \tag{9}$$

Here, the value of using a sensor with a given model $\mathcal{P}_{\mathbf{Z}|\mathbf{X}}$ is explicitly dependent on what information is already available, defined by $\mathcal{P}_{\mathbf{X}}$.

2.2. Optimal Sensor Layout

The question that this paper attempts to explore is if there are multiple sensors layouts available, each associated with a different sensor model, which is the best to use.

To begin, let the outputs of the available sensors be described by the random variables $\{Z^1, Z^2, \dots, Z^n\}$, where each has an associated sensor model of $\mathcal{P}_{Z^i|\mathbf{X}}$. The value of using a particular sensor $i \in \{1, \dots, n\}$ is given by its expected value, defined in (9). Then, if the ith sensor has a cost $K(i)$, the optimal sensor can be found by simply maximising the net benefit, for the given prior belief $\mathcal{P}_{\mathbf{X}}$:

$$i^* = \arg\max_i \big[V(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{Z^i|\mathbf{X}}) - K(i)\big] \tag{10}$$

This formulation explicitly captures the prior information $\mathcal{P}_{\mathbf{X}}$, the usage cost $K(i)$ and the characteristics of the sensor model $\mathcal{P}_{Z^i|\mathbf{X}}$.

A reason why this process is not often used for sensor layout design tasks (such as where to place temperature sensors in a building) is the difficulty caused in the formulation of the final decision problem (what is the temperature information going to be used for?). In other words, the function $C(a,x)$ is often not given in advance. This problem is typically overcome by ignoring the final decision task and simply reformulating the problem of optimal sensor layout as an inference problem that selects the sensor layout which maximises the information collected by the sensors about the variable $\mathbf{X}$. This will be described in the next section.

2.3. Information-theoretic Measures

Information theory provides the tools required to quantify what is meant by "information collected by the sensors". Entropy provides a measure of the uncertainty associated with a belief $\mathcal{P}_{\mathbf{X}}$:

$$H(\mathbf{X}) \triangleq \sum_{x\in\mathcal{X}} P(x) \log \frac{1}{P(x)} \tag{11}$$

This was first derived by Shannon in [2] from 3 basic axioms that a metric of uncertainty should satisfy. The base of the logarithm determines the unit of the measure, with base 2 corresponding to "bits".

Thus, to formulate the sensor selection problem as an information maximisation problem, the optimal expected

cost J^* of a distribution is replaced by the entropy H, that is $J^*(\cdot) = H(\cdot)$ and $J^*(\cdot|\cdot) = H(\cdot|\cdot)$. Further, the entropy is a concave function, therefore, the inequality in Equation (8) is satisfied. Intuitively, the information maximising criterion suggests to use the sensor that, on average, produces the least uncertain posterior belief. It is noted that these two different criteria (generic and information maximising) may select different optimal sensors.

Substituting $J^*(\cdot)$ with $H(\cdot)$ in (7) transforms the function $G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}})$ into the conditional entropy of $\mathbf{X}$ given $\mathbf{Z}$,

$$G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) = \sum_{z \in \mathcal{Z}} P(z) H(\mathbf{X}|z) \triangleq H(\mathbf{X}|\mathbf{Z}) \tag{12}$$

and the value of a configuration becomes the mutual information between $\mathbf{X}$ and $\mathbf{Z}$

$$V(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) = J^*(\mathcal{P}_{\mathbf{X}}) - G^*(\mathcal{P}_{\mathbf{X}}, \mathcal{P}_{\mathbf{Z}|\mathbf{X}}) \triangleq I(\mathbf{X}; \mathbf{Z}) \tag{13}$$

If each sensor has a different usage cost, the optimal selection problem becomes ill posed, since there is no direct method of trading off information with usage cost without explicitly considering what the information will be used for (an information theoretic approach was used to avoid this). To overcome this problem it will be assumed that the usage costs are constant and the optimal design problem can be formulated as the maximisation, over all sensors i, of the mutual information between $\mathbf{X}$ and Z^i

$$i^* = \arg\max_i I(\mathbf{X}; Z^i). \tag{14}$$

Alternatively, we can convert the information maximisation problem into an entropy minimisation problem when selecting the best sensor:

$$\begin{aligned} i^* &= \arg\max_i \left[H(\mathbf{X}) - H(\mathbf{X}|Z^i) \right] \\ &= \arg\min_i \left[H(\mathbf{X}|Z^i) \right] \end{aligned} \tag{15}$$

The equivalence between (14) and (15) does not scale to the selection of multiple sensors (see Section 3.1).

2.4. Maximum Entropy Criterion

At this stage, we apply this approach to direct measurements. In the simplest case of selecting one optimal sensor $\mathbf{Z}$, which is a deterministic function of the state $\mathbf{X}$, the conditional entropy of $\mathbf{Z}$ given $\mathbf{X}$ must be zero. This occurs because once the state $\mathbf{X}$ is known there is no uncertainty in $\mathbf{Z}$ (Figure 1).

It is clear then that the mutual information, $I(\mathbf{X}; \mathbf{Z})$, between the observation $\mathbf{Z}$ and the state $\mathbf{X}$ is equal to the entropy of the observation $H(\mathbf{Z})$. Using this in the sensor selection problem (14) yields the optimal sensor as follows:

$$i^* = \arg\max_i I(\mathbf{X}; Z^i) = \arg\max_i H(Z^i) \tag{16}$$

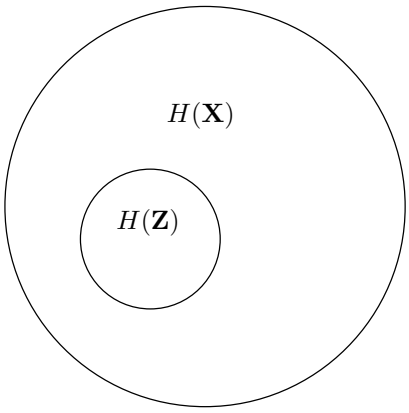


Figure 1. I diagram for a system where the sensor observations are a deterministic function of the variable $\mathbf{X}$.

We now consider a general case of problems where the state of interest $\mathbf{X}$ is defined as a set of variables $\mathbf{X} = \{X^1, X^2, \ldots, X^m\}$. It is assumed that a set of sensors $\mathbf{Z} = \{Z^1, Z^2, \ldots, Z^m\}$ exists, where each can measure the value of an associated variable X_i, this is the case of direct measurement as discussed earlier. If the sensors $\mathbf{Z}$ does not measure the associated variables $\mathbf{X}$, then we call this case an indirect measurement and this problem will not be explored further in this paper.

The design task becomes to select a subset of sensors $\nu \subseteq \{1, 2, \ldots, m\}$ to deploy. To avoid the case of selecting all variables $\nu = \{1, 2, \ldots, m\}$, a constraint is generally imposed on ν. For simplicity it is assumed this constraint imposes a maximum limit r on the number of elements in ν. To be able to refer to the elements of ν explicitly, this set will be denoted by $\nu = \{i_1, i_2, \ldots, i_r\}$.

Now consider an abstract compound random variable $\mathbf{Z}^\nu$, representing the combined output of all selected variables and defined as $\mathbf{Z}^\nu \triangleq \{X^i \colon \forall i \in \nu\} = \{X^{i_1}, \ldots, X^{i_r}\}$. Thus the optimal sensor selection task becomes

$$\nu^* = \arg\max_{|\nu| \leq r} I(\mathbf{X}; \mathbf{Z}^\nu) \tag{17}$$

$$= \arg\max_{|\nu| \leq r} H(\mathbf{Z}^\nu) \tag{18}$$

3. Current Methods

3.1. Minimum Residual Entropy Criterion

Krause and Guestrin [3] introduced a local reward function, which is defined on the marginal probability distribution of the variables in $\mathbf{X}$. The local reward is set for each variable X_i as the negative conditional entropy given the observation variable $\mathbf{Z}^\nu$, that is

$$R_i(P(X_i|\mathbf{Z}^\nu)) = -H(X^i|\mathbf{Z}^\nu). \tag{19}$$

The objective of the optimisation then becomes the minimisation of the sum of conditional entropies (residue entropy):

$$\nu_R^* = \arg\min_{|\nu|\leq r} \sum_{i=1}^{m} H(X^i|\mathbf{Z}^\nu) \tag{20}$$

It is noted that in general this is not the same criterion as the one developed in the previous section, since it does not take into account the dependencies between the variables X^i, as demonstrated in the Appendix.

3.2. Maximum Mutual Information Criterion

Guestrin *et al.* [4] proposed a criterion that gives an optimum subset of sensor locations that minimises the uncertainty about the estimates in the "rest of the space". The problem is formulated by searching for $\mathbf{Z}^\nu$ that reduces the entropy over the rest of the space $\mathbf{X}\backslash\mathbf{Z}^\nu = \mathbf{X}^{\bar{\nu}} = \{X^i : i \notin \nu\}$. Formally, the optimal subset is:

$$\begin{aligned}\nu_{MI}^* &= \arg\max_{|\nu|\leq r} \left[H(\mathbf{X}^{\bar{\nu}}) - H(\mathbf{X}^{\bar{\nu}}|\mathbf{Z}^\nu)\right] \\ &= \arg\max_{|\nu|\leq r} I(\mathbf{X}^{\bar{\nu}}; \mathbf{Z}^\nu),\end{aligned} \tag{21}$$

Thus this measure is equivalent to finding the maximum mutual information between $\mathbf{X}^{\bar{\nu}}$ and $\mathbf{Z}^\nu$. This only takes into account the mutual information between the observed and unobserved variables and not the remaining uncertainty of the unobserved variables, as demonstrated in the appendix.

It is noted that the criterion in Equation 21 is exactly the same as that introduced by Tononi et al. in [5], where the authors considered the mutual information between a subset $\mathbf{X}^k$ with k components, from a system $\mathbf{X}$ of n components, and its complement $\mathbf{X}^{\bar{k}}$. This concept of mutual information between two complementary subsets can be generalized to express the deviation from independence among the components in $\mathbf{X}$ by:

$$In(\mathbf{X}) = \left(\sum_{i=1}^{n} H(x_i)\right) - H(\mathbf{X}) \tag{22}$$

$$= In(\mathbf{X}^k) + In(\mathbf{X}^{\bar{k}}) + I(\mathbf{X}^k; \mathbf{X}^{\bar{k}}) \tag{23}$$

$In(\mathbf{X})$ is called *integration* by the authors and is related to their *complexity* criterion [5]. Now, the task of selecting a subset of sensors according to the maximum mutual information criterion is to find a split within $\mathbf{X}$, i.e. the k components, such that the third term in Equation 23 is maximized. Further, since $\mathbf{X}$ does not change when selecting a subset, then $In(\mathbf{X})$ reminds constant. Therefore, maximizing $I(\mathbf{X}^k; \mathbf{X}^{\bar{k}})$ is related to of minimizing $In(\mathbf{X}^k) + In(\mathbf{X}^{\bar{k}})$. Since mutual information is a measure of redundancy between two elements, then the MI criterion is a method of finding k, such that the redundancy between the two subsets is maximized while the redundancy within each subset is minimized.

3.3. Maximum Information Coverage Criterion

Olsson *et al.* [6] suggested a method that uses a combination of mutual information and information metric [7] to describe an *information coverage* (IC):

$$\begin{aligned}\nu_{IC}^* = \arg\max_{|\nu|\leq r} \sum_{i\in\nu} \sum_{\substack{j\in\nu \\ i\neq j}} [w_1 I(Z^i; Z^j) + \\ w_2(H(Z^i|Z^j) + H(Z^j|Z^i))],\end{aligned} \tag{24}$$

where the mutual information $I(Z^i; Z^j)$ is used as a measure for redundancy between measurements Z^i and Z^j, and the information metric (i.e. the information distance between two sensors) is used as a measure for novelty, w_1 and w_2 are the weights used to emphasis redundancy and novelty. Redundancy in this case is the ability of the sensors to be robust against the noise in the environment. Novelty is the measure that captures as much different information as possible from the environment. It is noted that neither model of environment, $\mathcal{P}_\mathbf{X}$, nor sensor models, $P(Z^i|\mathbf{X})$, are used in this approach.

If we set $w_1 = w_2 = 1$, then Equation (24) reduces to:

$$\nu_{IC}^* = \arg\max_{|\nu|\leq r} \sum_{i\in\nu} \sum_{\substack{j\in\nu \\ i\neq j}} H(Z^i, Z^j). \tag{25}$$

Thus, it is important to choose the ratio between w_1 and w_2, to capture redundancy or novelty in the system. Following Olsson [6], we used $w_1 = 1$ and $w_2 = 4$ to put more emphasis on the novelty.

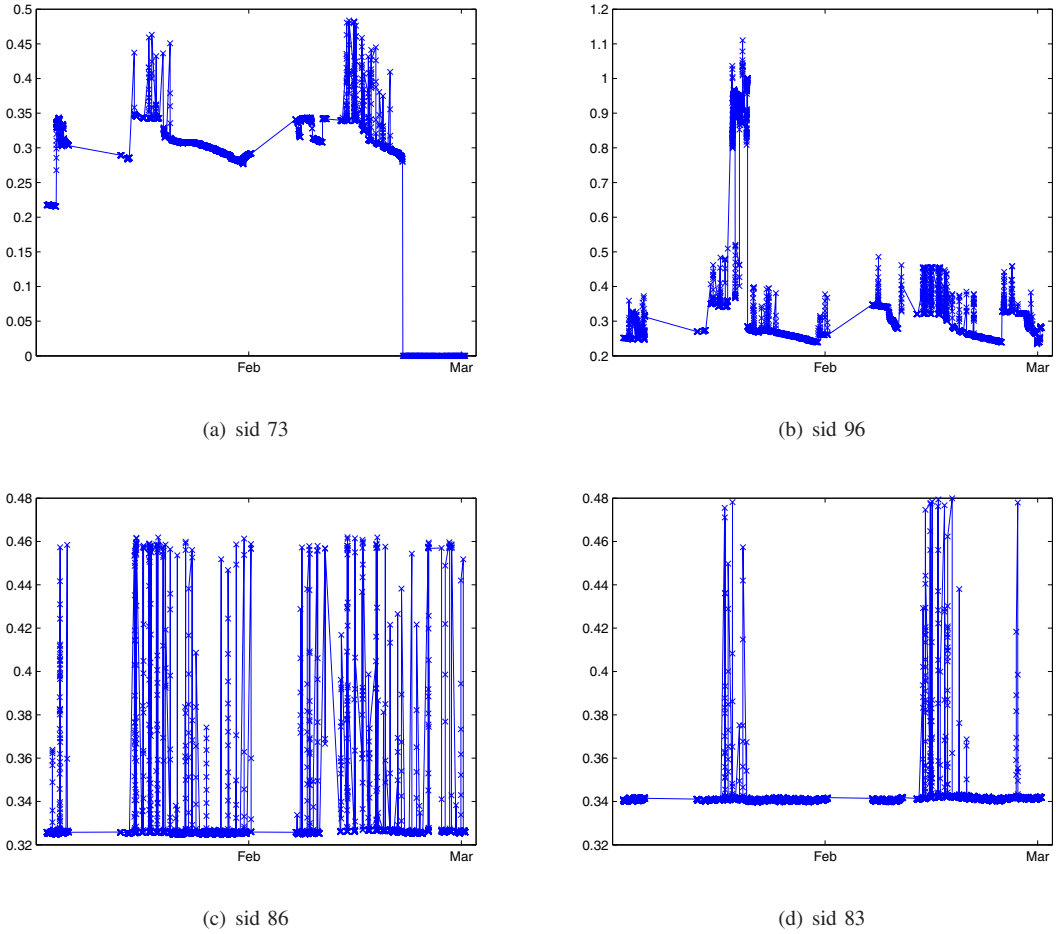
4. Experimental Setup

This section describes an experiment in which real network data from soil moisture measurements is used to derive optimal sensor placements using the four criteria outlined above. Data from sensors on an approximately rectangular 4x4 grid is used to find the optimal locations on the grid if only 2, 3 or 4 sensors were used. The resulting sensor layouts are tested using another set of data, obtained from a different time period.

4.1. Data

The data set used for this paper was obtained from a current wireless sensor network in Queensland, Australia, 670 km north of Brisbane. The network was set up as a test bed for environmental and animal behaviour monitoring at Belmont Research Station. The fixed sensor nodes that provided the data used for this paper are solar powered and have onboard sensors for monitoring soil moisture, battery voltage, and solar voltage.

Approximately two months of soil moisture data from January and February of 2008 was used. Each node takes a



(a) sid 73

(b) sid 96

(c) sid 86

(d) sid 83

Figure 2. The soil moisture data collected by the first four sensors in the network during the months of January and February 2008. 'x' mark actual readings, and the y-axis show raw data collected by the sensor.

sensor reading at roughly one minute intervals independent of its neighbours. Thus the data are preprocessed such that all the readings occur at the same time on 0 seconds of a minute. Further, due to various environmental and onboard issues, some nodes may not record any data for a period of time. The irregularity in the individual sensor's data time stamps combined with drop out in data recording of individual sensors means not all sensor nodes in the network will record a reading at a given time stamp t. In other words, the data set contains (roughly 40%) missing values. The data recorded by the first four nodes in the network is shown in Figure 2.

4.2. Bayesian Networks

Because a physical (hydrological) model of the sensed area at Belmont is not available, graphical models, specifically Bayesian Networks (BN) [8], were used to model the environment, $\mathcal{P}_{\mathbf{X}}$. For the case of direct sensing of a single quantity, in this case soil moisture, the BN can be mapped directly onto the physical network, since each physical node corresponds to a single sensed variable. Thus, each node of the BN represents a sensor node.

In this paper, three different network structures, corresponding to three different models for the environment, are considered. Figure 3 shows the first structure (Network One) of the Bayesian Network used, and Figure 4 shows the second structure (Network Two) where all dependencies from Network One are reversed. These two networks are constructed from the assumption that the neighbouring nodes in the sensor network are interdependent. Specifically, each node, X^i, in the Bayesian Network is a parent to two neighbouring nodes, and a child to two neighbouring nodes.

Bayesian Networks need to be *learned* before they can be used. Learning is the process of estimating the parameters as well as finding the structure of the network. The structures of the first two network are assumed to be known, thus only the parameters need to be learnt. The Expectation Maximisation

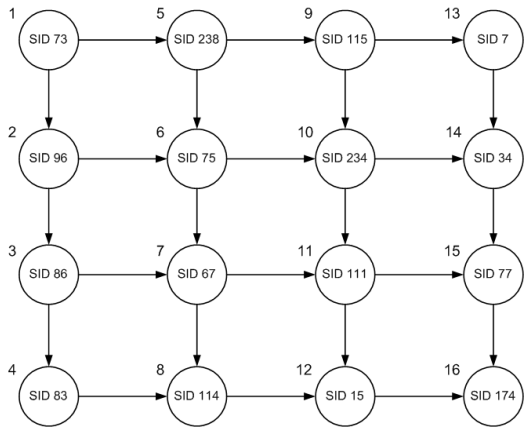


Figure 3. The Bayesian Network model referred to as Network One. The numbers on the top left side of each node denote the respective BN node numbers.

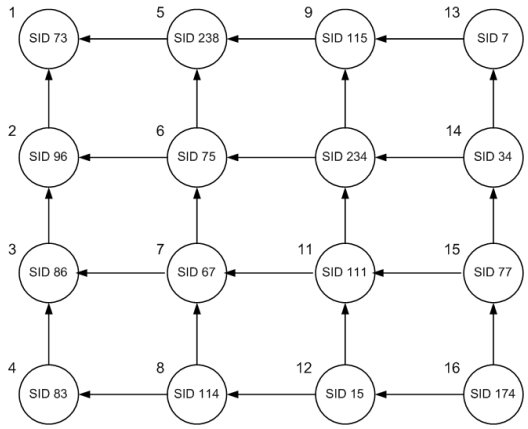


Figure 4. The Bayesian Network model used for Network Two.

(EM) algorithm [9], [10] is used to estimate the parameters since the data contains missing values, that is, each sensor node may not record the sensor readings at every time stamp t. The EM algorithm provides a general approach to maximum-likelihood parameter estimation when training data is incomplete.

Finally, we used a structure (Network Three) learnt from the Structural EM algorithm [11]. The basic topology of this network is the same as the first two, however, the dependencies between the nodes are now learnt from the data. The parameters of the network are also learnt during the Structural EM process.

4.3. Optimal Layout and Prediction

We use the learnt Bayesian Network to find the optimal layout of the sensors by finding probability distributions of $P(\mathbf{Z}^\nu)$ and $P(\mathbf{X}^{\bar{\nu}})$, through marginalisation of the joint distribution. Using these distributions we compute the relevant information-theoretic quantities and obtain the optimal layouts derived from all four criteria.

The optimal layouts can then be used to predict measurements from "the rest of space" using data only from the sensors of each layout and the learnt BN. That is, given the observed values of some of the nodes in the network, compute the probability distributions of the other nodes. Inference allows us to perform *prediction* on the data, that is, the posterior probability distribution of a node can be computed given the values of the parent and child nodes connected to it. The prediction results are then compared with the ground truth provided by the test data to compare the performances of the various sensor layout criteria.

The data set was divided into two, where the first half was used for training and the second half for testing. We further processed the data to give a discretisation of 3 values, {low, median, high}, to be used in the discrete nodes. No other pre-processing was carried out.

5. Results and Discussion

This section presents the results of the experiments for finding optimal layouts for two, three and four sensors using the criteria described earlier. The optimal layouts are presented, followed by the inference results for the other nodes in the network. A greedy search process was employed, to give the best theoretically possible results for the various criteria.

5.1. Network One

Table 1, second column summarises the results of the optimum layout for 2 sensors for the first network. It can be seen that the $H(\mathbf{Z})$ and the IC criteria gave the same sensor configuration at (1, 13). Figure 3 shows these two sensors are located at the top two corners of the Bayesian Network, which seems to confirm the proposition in [4] that the entropy-based method "pushes" the sensors to the edges of the network. For the other two criteria, $\sum_i H(X^i|\mathbf{Z}^\nu)$ and $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$, both gave a configuration with one node on the edge and the other node near the middle.

A similar search was performed for the optimal layout with 3 and 4 sensors, and the results are summarised in Table 1. In these cases, there are only two sets of configurations given by the 4 criteria: those from the $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$ criterion and those from the others. Moreover, all resulting layouts have at least one node near the middle of the network. This does not agree with the conclusion by Guestrin *et al.* [4] that using the entropy criterion will result in sensors placed far apart along the boundary of the space.

These sensor placements are assessed by performing inferences on the rest of the nodes, i.e. $\mathbf{X}\backslash\mathbf{Z}^\nu$, and comparing their prediction results with the actual measurements.

Table 1. Results of optimal layout for multiple sensors for Network One.

	2 sensors	3 sensors	4 sensors
$\arg\max H(\mathbf{Z})$	1, 13	1, 10, 13	1, 8, 10, 13
$\arg\min \sum_i H(X^i\|\mathbf{Z}^\nu)$	1, 10	1, 10, 13	1, 8, 10, 13
$\arg\max I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$	2, 6	1, 3, 10	2, 5, 10, 13
$\arg\max \mathrm{IC}$	1, 13	1, 10, 13	1, 8, 10, 13

Table 2. Inference results of optimal layout for multiple sensors for Network One. The results are shown here as harmonic weighted average of prediction accuracy in percentages.

	2 sensors	3 sensors	4 sensors
$\arg\max H(\mathbf{Z})$	99.11	99.15	99.22
$\arg\min \sum_i H(X^i\|\mathbf{Z}^\nu)$	96.95	99.15	99.22
$\arg\max I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$	91.75	97.78	93.36
$\arg\max \mathrm{IC}$	99.11	99.15	99.22

Table 3. Results of optimal layout for multiple sensors for Network Two.

	2 sensors	3 sensors	4 sensors
$\arg\max H(\mathbf{Z})$	1, 2	1, 2, 3	1, 2, 3, 4
$\arg\min \sum_i H(X^i\|\mathbf{Z}^\nu)$	3, 4	1, 3, 4	2, 3, 4, 5
$\arg\max I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$	1, 7	1, 7, 9	1, 7, 9, 15
$\arg\max \mathrm{IC}$	2, 5	1, 2, 3	1, 2, 3, 4

Table 4. Inference results of optimal layout for multiple sensors for Network Two. The results are shown here as harmonic weighted average of prediction accuracy in percentages.

	2 sensors	3 sensors	4 sensors
$\arg\max H(\mathbf{Z})$	67.73	72.82	63.35
$\arg\min \sum_i H(X^i\|\mathbf{Z}^\nu)$	72.63	74.85	61.02
$\arg\max I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$	82.06	76.02	75.26
$\arg\max \mathrm{IC}$	69.62	72.82	63.35

Table 2 shows the prediction results as harmonic weighted averages of the prediction accuracies. The harmonic average is used because it provides the truest average when the quantities involved are rates or ratios as the case here. From Table 2, it can be seen that the prediction results using the layout obtained from the $H(\mathbf{Z})$ and IC criteria gave the best accuracy in all three cases. This suggests that the sensors placed at the corners of the network do provide enough information to infer the possible sensor measurement at other locations, such as the case for 2 sensors.

We used two-tailed hypothesis tests to compare the pairwise prediction results statistically, taking the different sizes of the ground truth data sets into account. The null hypothesis H_0 is set to be the hypothesis that the results were from the same distribution and the α-value was set at 0.05. It was found that the resulting P-values[1] of all comparisons were all less than 1×10^{-20}, much smaller than the α value, which means the observed differences are significant, and thus the null hypothesis can be rejected. Therefore, the better sensor combination obtained by $H(\mathbf{Z})$ and the IC criteria have statistically significant performances.

5.2. Network Two

Table 3 summarises the results of the optimum placements for 2, 3, and 4 sensors for the Network Two. Comparing this table with the optimal layouts in Table 1, we can see that there are some differences.Firstly, all of the deduced optimal layouts are different from those for Network One. Secondly, the layouts obtained by the $H(\mathbf{Z})$ and the IC criteria are not always the same. Thirdly, the layouts obtained by $\sum_i H(X^i|\mathbf{Z}^\nu)$ are never the same as those from $H(\mathbf{Z})$. Therefore, we can conclude that different graphical models of the environment will result in different optimal layouts. However, the layouts obtained by the $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$ criterion are still always different from those resulting derived from $H(\mathbf{Z})$.

The prediction results from these layouts are shown in Table 4. It cab be seen that in this case the $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$ criterion gives the best prediction accuracies, while the others have similar results. We performed two-tailed hypothesis tests to compare these results, using the same H_0 and α-value. We found all but one (the results between $H(\mathbf{Z})$ and $\sum_i H(X^i|\mathbf{Z}^\nu)$ for 3 sensors) of the resulting P-values were less than the α-value, thus showing the differences between the prediction results are significant, and $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$ is the best criterion for this network structure.

5.3. Network Three

Network Three is the network whose structure is learnt using the Structural EM algorithm. Figure 5 shows the resultant network structure.

Table 5 shows the resulting optimal layouts using the network structure learnt and Table 6 shows the prediction results. The results here are very similar to those from Network One. The layouts obtained using $H(\mathbf{Z})$ and IC are the same, and are almost the same from $\sum_i H(X^i|\mathbf{Z}^\nu)$ except for 2 sensors, and those obtained from using $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$ are totally different. Further, the prediction results from the layout from $I(\mathbf{X}^{\bar{\nu}};\mathbf{Z}^\nu)$ are the worst among all layouts, and those from $H(\mathbf{Z})$ and IC are consistently the best.

1. The P-value, or significance value, is the probability of observing the test statistic if the null hypothesis is true.

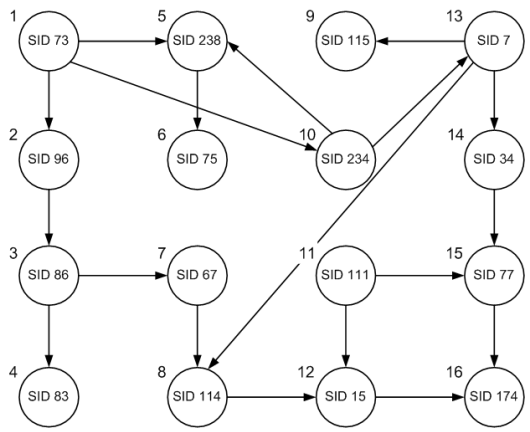


Figure 5. The Bayesian Network learnt using the Structural EM algorithm.

Table 5. Results of optimal layout for multiple sensors for Network Three.

	2 sensors	3 sensors	4 sensors
$\arg\max H(\mathbf{Z})$	8, 9	5, 8, 9	1, 5, 8, 9
$\arg\min \sum_i H(X^i \mid \mathbf{Z}^\nu)$	5, 9	5, 8, 9	1, 5, 8, 9
$\arg\max I(\mathbf{X}^{\bar{\nu}}; \mathbf{Z}^\nu)$	1, 5	1, 5, 13	1, 3, 5, 13
$\arg\max \mathrm{IC}$	8, 9	5, 8, 9	1, 5, 8, 9

Table 6. Inference results of optimal layout for multiple sensors for Network Three. The results are shown here as harmonic weighted average of prediction accuracy in percentages.

	2 sensors	3 sensors	4 sensors
$\arg\max H(\mathbf{Z})$	83.57	89.88	96.03
$\arg\min \sum_i H(X^i \mid \mathbf{Z}^\nu)$	83.38	89.88	96.03
$\arg\max I(\mathbf{X}^{\bar{\nu}}; \mathbf{Z}^\nu)$	77.01	77.90	77.70
$\arg\max \mathrm{IC}$	83.57	89.88	96.03

5.4. Discussion

From the tables of the resulting optimal layouts, it is interesting to note the incremental nature of these layouts: when a new sensor is added to the system, the previous configuration is often retained. For example, in Table 5, the layouts for Network Three, using the $H(\mathbf{Z})$ criterion for 2 sensors gives the optimal locations at (8, 9), which becomes (5, 8, 9) for 3 sensors, and (1, 5, 8, 9) for 4 sensors. Similar patterns can be observed in most cases. If this pattern proves to be more general, it could be an advantage when applying these optimal layout criteria in practical situations: should the cost budget or the accuracy requirement increase, new sensors can be added without re-deployment of existing sensors.

Comparing the prediction results in Tables 2, 4, and 6, it can be seen that no one criterion will consistently give the best prediction results for any of the environment models investigated. However, in general the prediction accuracies for Networks One and Three are better than for Network Two. Further, since the structure of Network Three has been learnt from the data vis the Structural EM algorithm, and because it has an overall structure similar to that of Network One (in general the dependencies "flow" from upper left to lower right), this suggests that these two structures are closer to the "true" topology of the environment. Since the goal is to find a layout that will provide good measurements not only at sensor locations but also at locations without the sensors, then for the case study presented here, the $H(\mathbf{Z})$ or IC criteria are the best options given the correct network structure.

To differentiate between the two criteria, we may compare their computational complexities. The IC criterion only requires pairwise entropies (entailing marginalisations over at most two sensor models), but the entropy criterion requires entropy computation for $|\nu|$ sensors. Thus, these two criteria have similar computational complexity for two sensors, but the IC criterion has less complexity as the number of sensors increases. Therefore, arguably, when using discrete variables and direct measurements, the optimal sensor layout is best found using the information coverage criterion given both prediction accuracy and computational complexity.

There is also another peculiar feature in the prediction results, which can be seen most prominently in Table 4, that is, the prediction accuracies from using 3 sensors outperforms those using 4 sensors. The reason for the discrepancy in the result is the number of ground truth data points available for test with each layout due to the missing data. For example, with sensor combination (1, 2, 3) in Table 3, 1241 sets of test data were available, where for (1, 2, 3, 4) only 575 sets were available. The quantity "set" here means the sensor data in the "rest of the space". A set of data is used only if none of the sensors in the layout have missing values. Although the prediction accuracy on the data common to both was similar, the prediction accuracy in the other data was a lot higher, causing this decrease in performance.

6. Conclusion

This paper has considered methods for selecting a limited number of sensor nodes in an area of interest such that the cost of the placement is minimised while the value of the obtained information is maximised — the optimal sensor layout problem. Specifically, we focused on direct measurements, and sensor layouts where sensors are placed over only a subset of possible locations, leaving the rest of the space without sensors. We compared and verified four criteria: maximum entropy of sensor measurements, minimum aggregated residue entropy (RE), maximum mutual

information (MI) between sensor and rest of the space, and maximum information coverage (IC).

We compared the four criteria mathematically, and found that both minimum RE and maximum MI criteria do not take into account terms that are contained within the maximum entropy criterion. The maximum IC criterion was shown to be the same as maximum entropy criterion under certain conditions. Some of these criteria are related to attempts to formalise complexity in self-organising systems. One may hypothesise that as the sensor layout approaches its optimal configuration, the sensing system's statistical complexity would increase. Investigation of the relationship between layout optimisation and system complexity will be the subject of future research [12].

Experimental verification was carried out using data from an existing wireless sensor network. Environment models were learnt as Bayesian Networks. Each of the criteria was applied independently producing in general a different optimal sensor layout. We found that for three or more sensors deployed in a layout, all four criteria can place some sensors on the edges and some near the middle of the area. Each layout was used to predict sensor measurements in the rest of the space. To verify the predictive performance of each of the layouts, we used the rest of the sensor data (i.e. that not utilised in learning) as the ground truth by comparing it with the predicted measurements. We found there is not one criterion that will give the best predictive results for all the environmental models tested. However, for the network structure with the better prediction accuracies, the maximum entropy and IC criteria gave the best results.

Future work will include similar comparisons for indirect measurements and extensions to different sensor networks.

Appendix

Comparison between Equation (20), the minimum residual entropy criterion, and Equation (17):

$$\begin{aligned}
&\arg\max_{|\nu|\le r} I(X^1,\ldots,X^m;\mathbf{Z}^\nu)\\
&=\arg\max_{|\nu|\le r}\Big[\underbrace{H(X^1,\ldots,X^m)}_{\text{Constant}}-H(X^1,\ldots,X^m|\mathbf{Z}^\nu)\Big]\\
&=\arg\min_{|\nu|\le r} H(X^1,\ldots,X^m|\mathbf{Z}^\nu)\\
&=\arg\min_{|\nu|\le r}\Big[\sum_{i=1}^{m} H(X^i|\mathbf{Z}^\nu)\\
&\qquad\underbrace{-I(X^1;X^{2:m}|\mathbf{Z}^\nu)-\cdots-I(X^{m-1};X^m|\mathbf{Z}^\nu)}_{\text{Terms not accounted for in (20)}}\Big]
\end{aligned}$$

Comparison between Equation (21), the maximum mutual information criterion, and Equation (17):

$$\begin{aligned}
&\arg\max_{|\nu|\le r} I(X^1,\ldots,X^m;\mathbf{Z}^\nu)\\
&=\arg\max_{|\nu|\le r}\Big[\underbrace{H(X^1,\ldots,X^m)}_{\text{Constant}}-H(X^1,\ldots,X^m|\mathbf{Z}^\nu)\Big]\\
&=\arg\max_{|\nu|\le r}\Big[-H(X^1,\ldots,X^m|\mathbf{Z}^\nu)\Big]\\
&=\arg\max_{|\nu|\le r}\Big[-H(\mathbf{X}^{\bar\nu}|\mathbf{Z}^\nu)\underbrace{-H(\mathbf{X}^\nu|\mathbf{Z}^\nu)}_{=0}\underbrace{+I(\mathbf{X}^{\bar\nu};\mathbf{X}^\nu|\mathbf{Z}^\nu)}_{=0}\Big]\\
&=\arg\max_{|\nu|\le r}\Big[H(\mathbf{X}^{\bar\nu})-H(\mathbf{X}^{\bar\nu}|\mathbf{Z}^\nu)-H(\mathbf{X}^{\bar\nu})\Big]\\
&=\arg\max_{|\nu|\le r}\Big[I(\mathbf{X}^{\bar\nu};\mathbf{Z}^\nu)\qquad\underbrace{-H(\mathbf{X}^{\bar\nu})}_{\text{Term not accounted for in (21)}}\Big]
\end{aligned}$$

Acknowledgment

We gratefully acknowledge the support of a grant from the US Air Force Research Laboratory, Asian Office of Aerospace Research and Development, Case Number AOARD-074045 [13].

We also like to thank the High Performance Computing and Communications Centre for the use of their supercomputer clusters in performing the experiments for this paper.

References

[1] D. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[2] C. Shannon, "A mathematical theory of communication," *Bell System Tech. Journal*, vol. 27, pp. 623–656, 1948.

[3] A. Krause and C. Guestrin, "Optimal nonmyopic value of information in graphical models: Efficient algorithms and theoretical limits," in *19th Int. Joint Conf. on Artificial Intelligence*, 2005, pp. 1339–1345.

[4] C. Guestrin, A. Krause, and A. Singh, "Near-optimal sensor placements in Gaussian processes," in *Int. Conf. on Machine Learning*, 2005, pp. 265–272.

[5] G. Tononi, O. Sporns, and G. M. Edelman, "A measure for brain complexity: Relating functional segregation and integration in the nervous system," *Proc. Natl. Acad. Sci. USA*, vol. 91, pp. 5033–5037, May 1994.

[6] L. Olsson, C. L. Nehaniv, and D. Polani, "Information tradeoffs and the evolution of sensory layout," in *Proceedings of the Ninth Int. Conf. on Artificial Life*, 2004, pp. 119–124.

[7] J. P. Crutchfield, "Information and its metric," in *Nonlinear Structures in Physical Systems Pattern Formation, Chaos and Waves*, L. Lam and H. C. Morris, Eds. Springer Verlag, 1990, pp. 119–130.

[8] F. V. Jensen, *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag, 2001.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–39, 1977.

[10] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*. New York: Springer-Verlag, 1999.

[11] N. Friedman, "The Bayesian structral EM algorithm," in *Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, USA, 1998.

[12] M. Prokopenko, F. Boschetti, and A. Ryan, "An information-theoretic primer on complexity, self-organisation and emergence," *Complexity*, vol. DOI: 10.1002/cplx.20249, 2008.

[13] X. R. Wang, G. Mathews, D. Price, and M. Prokopenko, "Sentient structures: Optimising sensor layouts for direct measurement of discrete variables," Tech. Rep. CMSE-2008-412, 2008, report to US Air Force Research Laboratory, Asian Office of Aerospace Research and Development (AOARD).