Sensor Selection and Random Field Reconstruction for Robust and Cost-effective Heterogeneous Weather Sensor Networks for the Developing World

Pengfei Zhang  
University of Oxford  
Oxford, UK

Ido Nevat  
TUM CREATE  
Singapore

Gareth W. Peters  
Heriot-Watt University  
Scotland, UK

Wolfgang Fruehwirt  
University of Oxford  
Oxford, UK

Yongchao Huang  
University of Oxford  
Oxford, UK

Ivonne Anders  
ZAMG  
Vienna, Austria

Michael Osborne  
University of Oxford  
Oxford, UK

Abstract

We address the two fundamental problems of spatial field reconstruction and sensor selection in heterogeneous sensor networks: (i) how to efficiently perform spatial field reconstruction based on measurements obtained simultaneously from networks with both high and low quality sensors; and (ii) how to perform query based sensor set selection with predictive MSE performance guarantee. For the first problem, we developed a low complexity algorithm based on the spatial best linear unbiased estimator (S-BLUE). Next, building on the S-BLUE, we address the second problem, and develop an efficient algorithm for query based sensor set selection with performance guarantee. Our algorithm is based on the Cross Entropy method which solves the combinatorial optimization problem in an efficient manner.

1 Introduction

We consider the case where two types of sensors are deployed: the first consists of expensive, high quality sensors; and the second, of cheap low quality sensors, which are activated only if the intensity of the spatial field exceeds a pre-defined activation threshold (e.g. wind sensors). This type of heterogeneous sensor networks approach has gained attention in the last few years due to the vision of the Internet of Things (IoT) where networks may share their data over the internet [Gubbi et al. [2013], Vermesan et al. [2011]].

Two practical scenarios that are of importance are: firstly, high-quality sensors may be deployed by government agencies (e.g. weather stations). These are sparsely deployed

1In particular, developing world countries are constrained by their budget when purchasing equipment. Meanwhile, these countries are heavily effected by climate change. This combination makes the robust and cost-effective sensor selection in weather sensor networks a major concern for the developing world.
due to their high costs, limited space constraints, high power consumption etc. To improve the coverage of the WSN, low-quality cheap sensors can be deployed to augment the high-quality sensor network [Rajasegarar et al. 2014]. Secondly, High-quality sensors cannot be easily deployed in remote locations, for example in oceans, lakes, mountains and volcanoes. In these cases, battery operated low-quality cheap sensors can be deployed [Werner-Allen et al. 2006].

More specifically, the following two fundamental problems are the focus of this paper: Firstly, **Spatial field reconstruction**: the task is to accurately estimate and predict the intensity of a spatial random field, not only at the locations of the sensors, but at all locations [Peters et al. 2015], [Nevat et al. 2015, 2013], given heterogeneous observations from both sensor networks; Secondly, **Query based sensor set selection with performance guarantee**: the task is to perform on-line sensor set selection which meets the QoS criterion imposed by the user, as well as minimises the costs of activating the sensors of these networks [Calvo-Fullana et al. 2016], [Joshi and Boyd 2009], [Chepuri and Leus 2015].

## 2 System model

We now present the system model for the physical phenomenon observed by two types of networks.

**A1** Consider a random spatial phenomenon (e.g. wind) to be monitored defined over a 2-dimensional space \( \mathcal{X} \in \mathbb{R}^2 \). The mean response of the physical process is a smooth continuous spatial function \( f(\cdot): \mathcal{X} \rightarrow \mathbb{R} \), and is modelled as a Gaussian Process (GP) according to:

\[
f(x) \sim \mathcal{GP}(\mu_f(x; \theta_f), C_f(x_1, x_2; \Psi_f)),
\]

where the mean and covariance functions \( \mu_f(x; \theta_f), C_f(x_1, x_2; \Psi_f) \) are assumed to be known.

**A2** Let \( N \) be the total number of sensors that are deployed over a 2-D region \( \mathcal{X} \subseteq \mathbb{R}^2 \), with \( x_n \in \mathcal{X}, n = \{1, \ldots, N\} \) being the physical location of the \( n \)-th sensor, assumed known by the FC. The number of sensors deployed by Network 1 and Network 2 are \( N_{1H} \) and \( N_{L} \) respectively, so that \( N = N_{1H} + N_{L} \).

**A3** **Sensor network 1 includes high quality sensors.** The sensors have a 0-threshold activation and each of the sensors collects a noisy observation of the spatial phenomenon \( f(\cdot) \). At the \( n \)-th sensor, located at \( x_n \), the observation is given by:

\[
Y^H(x_n) = f(x_n) + \mathcal{W}(x_n), \ n = \{1, \ldots, N_{1H}\}
\]

where \( \mathcal{W}(x_n) \) is i.i.d Gaussian noise \( \mathcal{W}(x_n) \sim N(0, \sigma^2_W) \). **Sensor network 2 includes low quality sensors.** The sensors have a \( T \)-threshold activation and each of the sensors collects a noisy observation of the spatial phenomenon \( f(\cdot) \), only if the intensity of the field at that location exceeds the pre-defined threshold \( T \), (e.g. anemometer sensors for wind monitoring [Ada 2015, And 2015]). At the \( n \)-th sensor, located at \( x_n \), the observation is given by:

\[
Y^L(x_n) = \begin{cases} 
  f(x_n) + \mathcal{V}(x_n) & , f(x_n) \geq T \\
  \mathcal{V}(x_n) & , f(x_n) < T 
\end{cases}
\]

where \( \mathcal{V}(x_n) \) is i.i.d Gaussian noise \( \mathcal{V}(x_n) \sim N(0, \sigma^2_V) \).

## 3 Field Reconstruction via Spatial Best Linear Unbiased Estimator (S-BLUE)

To perform inference in our Bayesian framework, one would typically be interested in computing the predictive posterior density at any location in space, \( x_s \in \mathcal{X} \), denoted \( p(f_s|Y_N) \). Based on this quantity, a point estimator, like the Minimum Mean Squared Error (MMSE) estimator can be derived:

\[
\hat{f}_s^{\text{MMSE}} = \int_{-\infty}^{\infty} p(f_s|Y_N, x_1:N, x_s) f_s \, df_s
\]
We develop the spatial field reconstruction via Best Linear Unbiased Estimator (S-BLUE), which enjoys a low computational complexity \cite{Kay1998}. The S-BLUE does not require calculating the predictive posterior density, but only the first two cross moments of the model. The S-BLUE is the optimal (in terms of minimizing Mean Squared Error (MSE)) of all linear estimators and is given by the solution to the following optimization problem:

\[
\hat{f}_s := \hat{a} + \hat{B}Y_N = \arg\min_{\hat{a}, \hat{B}} \mathbb{E} \left[ (f_s - (a + B Y_N))^2 \right],
\]

where \(\hat{a} \in \mathbb{R}\) and \(\hat{B} \in \mathbb{R}^{1 \times N}\).

The optimal linear estimator that solves \((4)\) is given by

\[
\hat{f}_s = \mathbb{E}_{f_s, Y_N} [f_s Y_N] \mathbb{E}_{Y_N} [Y_N Y_N]^{-1} (Y_N - \mathbb{E} [Y_N]),
\]

and the Mean Squared Error (MSE) is given by

\[
\sigma^2 = k (x_s, x_s) - \mathbb{E}_{f_s, Y_N} [f_s Y_N] \mathbb{E}_{Y_N} [Y_N Y_N]^{-1} \mathbb{E}_{Y_N, f_s} [Y_N f_s].
\]

### 4 Query Based Sensor Set Selection with Performance Guarantee

In this Section we develop an algorithm to perform on-line sensor set selection in order to meet the requirements of a query made by users of the system. In this scenario users can prompt the system and request the system to provide an estimated value of the spatial random field at a location of interest \(x_s\). We defined the activation sets of the sensors in both networks by \(S_1 \in \{0, 1\}^{N_1}\), \(S_2 \in \{0, 1\}^{N_2}\). Then the sensor selection problem can be formulated as follows:

\[
S = \arg\min_{\begin{subarray}{l} s_1 \in \{0, 1\}^{N_1} \\ s_2 \in \{0, 1\}^{N_2} \end{subarray}} \frac{w_h |S_1| + w_l |S_2|}{\sigma^2_s},
\]

s.t. \(\sigma^2_s < \sigma^2_q\),

where \(\sigma^2_q\) is the maximal allowed uncertainty at the query location \(x_s\), and \(w_h\) and \(w_l\) are the known costs of activating a sensor from Network 1 and Network 2, respectively.

Suppose we wish to maximize a function \(U(x)\) over some set \(\mathcal{X}\). Let us denote the maximum by \(\gamma^*\); thus,

\[
\gamma^* = \max_{x \in \mathcal{X}} U(x).
\]

The Cross Entropy Method (CEM) solves this optimization problem by casting the original problem \((5)\) into an estimation problem of rare-event probabilities. By doing so, the CEM aims to locate an optimal parametric sampling distribution, that is, a probability distribution on \(\mathcal{X}\), rather than locating the optimal solution directly. To apply the CEM to solve our optimization problem in \((7)\), we need to choose a parametric distribution. Since the activation of the sensors is a binary variable (eg. 0 → don’t activate, 1 → activate), we choose an independent Bernoulli variable as our parametric distribution, with a single parameter \(p\) (ie. \(\mathbf{V} = p\)). The Bernoulli distribution is a member of the NEF of distributions, hence, an analytical solution of the stochastic program is available in closed form as follows:

\[
p_{t,j} = \sum_{i=1}^{K} \mathbb{I}(H_{i,j}^1 = 1) \mathbb{I}(U(k) \geq \beta_i)/\sum_{i=1}^{K} \mathbb{I}(U(k) \geq \beta_i).
\]

Since the optimization problem in Eq. \((7)\) is a constrained optimization problem, we introduce an Accept\-\-Reject step which rejects samples which do not meet the QoS criterion \(\sigma^2_s < \sigma^2_q\), as follows

\[
U(k) = \begin{cases} - (w_h |S^H| + w_l |S^L|), & \sigma^2_s(k) < \epsilon \\ -\infty, & \text{Otherwise} \end{cases}
\]
5 Experimental Results and Discussion

In order to test our algorithm on real data sets, we use fine grained datasets available from Hans-Ertel-Centre for Weather Research (HErZ) project\footnote{https://www.herz-tb4.uni-bonn.de/index.php/hans-ertel-centre-for-weather-research/funding}. Particularly, we choose Feb 23, 2015 as testing data since it has the highest wind speed across the whole year. The left panel of Fig. 1 shows the region of interest on the map. Both high and low quality sensors are selected randomly within the region. In this figure we randomly deployed 50 high quality and 50 low quality sensors. The right panel of Fig. 1 shows daily wind speed intensity across the region. In left panel of Fig. 2 we present a quantitative comparison of the MSE for various values of high and low quality sensors. The result shows a clear trend of MSE with the increasing of high and low quality sensors. We also illustrate how our sensor selection algorithm performs. For comparison, we use an optimal selection method which only selects the sensor set collections that minimize the U values and ensures that the QoS criterion is being met. The simulation parameters we have are: \( \{N_h = 5, N_l = 10, T = 0, w_h = 150, w_l = 30, \sigma_w = 0.001, \sigma_g = 0.003, k_f(x_*, x_*) = 5.8, x_* = 10, y_* = 50, \epsilon = \{3.4, 3.6, 3.8, 4.2, 4.4\}\) \footnote{ANEMO 4403 RF WINDSPEED METER (ANEMOMETER) WITH WM44 P RF DISPLAY UNIT WIRELESS WIND SPEED METER. Technical report, 2015. URL} . We fix the \( N_h = 5, N_l = 10 \). The comparison is shown in right panel of Fig. 2. We also increase the number of iterations in CE method from 1 to 10. It shows CE method converges quickly to the optimal selection algorithm within 10 iterations for all the \( \epsilon \) values.

References


