Estimation of Spatially Correlated Random Fields in Heterogeneous Wireless Sensor Networks

Ido Nevat, Gareth W. Peters, François Septier, and Tomoko Matsui, *Member, IEEE*

Abstract—We develop new algorithms for *spatial field reconstruction, exceedance level estimation* and *classification* in heterogeneous (mixed analog & digital sensors) Wireless Sensor Networks (WSNs). We consider spatial physical phenomena which are observed by a heterogeneous WSN, meaning that it consists partially of sparsely deployed high-quality sensors and partially of low-quality sensors. The high-quality sensors transmit their (continuous) noisy observations to the Fusion Centre (FC), while the low-quality sensors first perform a simple thresholding operation and then transmit their binary values over imperfect wireless channels to the FC. The resulting observations are mixed continuous and discrete (1-bit decisions) observations, and are combined in the FC to solve the inference problems. We first formulate the problem of *spatial field reconstruction, exceedance level estimation* and *classification* in such heterogeneous networks. We show that the resulting posterior predictive distribution, which is key in fusing such disparate observations, involves intractable integrals. To overcome this problem, we develop an algorithm that is based on a multivariate series expansion approach resulting in a Saddle-point type approximation. We then present comprehensive study of the performance gain that can be obtained by augmenting the high-quality sensors with low-quality sensors using real data of insurance storm surge database known as the Extreme Wind Storms Catalogue.

Index Terms—Wireless sensor networks, detection, Gaussian processes, Kernel methods, imperfect communication channels.

I. INTRODUCTION

WIRELESS SENSOR NETWORKS (WSN) have attracted considerable attention due to the large number of applications, such as environmental monitoring [1], weather forecasts [2]–[5], surveillance [6], health care [7], structural safety and building monitoring [8]and home automation [5], [9]. We consider a WSN which consists of a set of spatially distributed sensors that may have limited resources, such as energy and communication bandwidth. These sensors monitor a spatial physical phenomenon containing some desired attributes (e.g., pressure, temperature, concentrations of substance, sound intensity, radiation levels, pollution concentrations, seismic activity etc.) and regularly communicate their observations to a Fusion Centre (FC) in a wireless manner (for example, as in [10]–[14]). The FC collects these observations and fuses them in order to reconstruct the signal of interest, based on which effective actions are made [9].

The majority of recent research on WSN address problems relates to addressing estimation of a single point source, such as source localization [15]–[22], or source detection (i.e., hypothesis testing) [23]–[28]class of problems. In [18]–[21], location estimation algorithms of a scalar point source were developed, and in [22]the Posterior Cramér-Rao lower bound (PCRLB) for a single target tracking in WSN with quantization was approximated via particle filters. In [23], [24], decision fusion algorithms for a single source detection were developed, and in [25]a vector-valued quantity of a single source was estimated in WSN with censoring and quantization. In this paper we show how to utilize the entire set of sensor data to not just obtain estimation of a given point source localization but instead to reconstruct the entire spatial field under a statistical model. Hence, we move beyond the estimation of a single location parameter by developing models to reconstruct the entire spatial random field which exhibits spatial dependency structure that we capture via either a homogeneous or non-homogeneous spatial covariance function, depending on the statistical properties of the observed spatial field.

In many cases these WSN use a small set of high-quality and expensive sensors (such as weather stations) [29]. While these sensors are capable of reliably measuring the environmental physical phenomenon, the low spatial deployment resolution prohibits their use in spatial field reconstruction tasks. To overcome this problem, sparse high-quality sensor deployment can be augmented by the use of complementary low-quality sensors that can be deployed more densely due to their low costs [2], [3]. This exact same concept was further strengthened when the US Environmental Protection Agency (EPA) published its shift in the paradigm of data collection which promotes the notion of augmenting sparse deployments of high-quality sensors with dense deployment of low-quality and inaccurate sensors [30].

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

ployed 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 analog sensor network [3].

2) High-quality sensors cannot be easily deployed in remote locations, for example in oceans, lakes, mountains and volcanoes. In these cases, battery operated low-cost sensors can be deployed and use simple wireless transmission techniques for data aggregation to the FC [31].

The low-quality sensors we consider in this paper are capable of transmitting a single bit for every analog observation they obtain, making them very energy efficient. The FC then receives a vector of observations which are mixed continuous (high-quality) and discrete (low-quality 1-bit values). This makes the data fusion a very complex inference problem. As such, no previous attempts have been made to join (fuse) the observations from different types of sensors to solve the problem of *Spatial Random Filed Reconstruction*.

The main goal of this paper is to develop low complexity algorithms to solve the problems of *spatial field reconstruction*, *exceedance level* and *spatial classification* of spatial Gaussian random fields in WSN under practical scenarios of mixed analog and digital sensors. More specifically, the following three fundamental problems are the focus of this paper:

1) **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. An example is the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) network for *Ocean Colour Reconstruction from Remote Sensing Networks* [32], [33]. This WSN monitors the health and status of the world's oceans and simultaneously estimates both atmospheric and ocean color data, which has been widely used to study aerosol and ocean variations in applications such as climatic forcing, bio-geochemical cycling, and radiative effects in the global and regional scales [34], [35].

2) **Spatial exceedance level**: the task is to identify regions where the intensity of the spatial random field exceeds some pre-defined regulation limits. An example is air quality monitoring, where the concentration of air pollution levels, if exceeded, cause violation of air safety standards and therefore require a public health warning [36]. Other examples in environmental modelling include identifying areas where ozone concentrations exceed safety standards, meaning there is high risk for tornadoes or floods, or heavy-metal levels are dangerously high. Application of such problems are presented in [4], [37]and an application in wireless communications is presented in [38].

3) **Spatial classification**: the task is to perform binary classification of the spatial field. Spatial binary classification occurs frequently in environmental and ecological research, for instance when the data correspond to presence or absence of a certain invasive plant species at a location, or when the data happen to fall into one of two categories, say two soil types [39], [40].

We develop an algorithm to jointly solve these problems which involves obtaining the predictive distribution of the spatial random field. We show that calculating this quantity requires solving intractable multiple integrals. We present a novel method to approximate this quantity by developing a series expansion of the Saddle-point (Laplace) type via a Gaussian basis [41]. This transforms the intractable multiple integrals to produce simple closed form expressions. Based on these expressions we derive new algorithms and provide closed form solutions for both key problems of interest. To obtain low computational complexity algorithms, we couple the Laplace method with Iterated Conditioning on the Modes (ICM) method, which is an iterative optimization method [42], [43]. The ICM method decomposes the resulting multi-variate high-dimensional optimization problem into a sequential series of one-dimensional optimization problems which can be solved easily, thus obtaining efficient algorithms.

We demonstrate the usefulness of our algorithm by testing it on real data sets of wind speeds measurements collected in Europe. In addition we study the impact that different deployments configurations have on the performance and the trade-offs between analog and digital deployments.

II. WIRELESS SENSOR NETWORK SYSTEM MODEL AND DEFINITIONS

We consider a generic WSN where the sensors deployed in the field are composed of two types of sensors:

1) High-quality sensors : analog sensors which transmit their noisy observations over Additive White Gaussian Noise (AWGN) channels.

2) Low-quality sensors : digital sensors, which perform a thresholding operation by processing the noisy observations to obtain a 1-bit value, and then transmit this information over imperfect wireless channels to the FC.

The FC combines all received signals and calculates an estimate of the spatial phenomenon at any point of interest in space. We first provide a formal definition of the spatial random Gaussian field followed by detailed WSN assumptions.

A. Spatial Gaussian Random Fields Background

The observed phenomenon is a spatially dependent continuous process with a spatial correlation structure. The degree of the spatial correlation in the process increases with the decrease of the separation between two observing locations and can be accurately modelled as a Gaussian random field[1] [4], [13], [14], [37], [38], [44]–[47]. A Gaussian process (GP) defines a distribution over a space of functions and it is completely specified by the equivalent of sufficient statistics for such a process, and is formally defined as follows.

Definition 1: (Gaussian process [48], [49]): Let $\mathcal{X} \subset \mathbb{R}^D$ be some bounded domain of a d-dimensional real valued vector space. Denote by $f(\mathbf{x}) : \mathcal{X} \mapsto \mathbb{R}$ a stochastic process parametrized by $\mathbf{x} \in \mathcal{X}$. Then, the random function $f(\mathbf{x})$ is a Gaussian process if all its finite dimensional distributions are Gaussian, where for any $m \in \mathbb{N}$, the random variables $(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_m))$ are normally distributed.

---

[1]We use Gaussian Process and Gaussian random field interchangeably.

We can therefore interpret a GP as formally defined by the following class of random functions:

$$f(\cdot) \sim \mathcal{GP}(\mu(\cdot;\boldsymbol{\theta}), \mathcal{C}(\cdot,\cdot;\boldsymbol{\Psi})), \text{ with}$$

$$\mu(\mathbf{x};\boldsymbol{\theta}) := \mathbb{E}[f(\mathbf{x})],$$

$$\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\Psi}) := \mathbb{E}[(f(\mathbf{x}_i) - \mu(\mathbf{x}_i;\boldsymbol{\theta}))(f(\mathbf{x}_j) - \mu(\mathbf{x}_j;\boldsymbol{\theta}))],$$

where at each point the mean of the function is $\mu(\cdot;\boldsymbol{\theta})$, parameterized by $\boldsymbol{\theta}$, and the spatial dependence between any two points is given by the covariance function (Mercer kernel) $\mathcal{C}(\cdot,\cdot;\boldsymbol{\Psi})$, parameterized by $\boldsymbol{\Psi}$, see detailed discussion in [48].

It will be useful to make the following notational definitions for the cross correlation vector and auto-correlation matrix, respectively:

$$k(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}) := \mathbb{E}[f(\mathbf{x}_*) f(\mathbf{x}_{\mathcal{N}})] \in \mathbb{R}^{1 \times \mathcal{N}}$$

$$\mathbf{K}(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}) := \begin{bmatrix} \mathcal{C}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \mathcal{C}(\mathbf{x}_1, \mathbf{x}_{\mathcal{N}}) \\ \vdots & \ddots & \vdots \\ \mathcal{C}(\mathbf{x}_n, \mathbf{x}_1) & \cdots & \mathcal{C}(\mathbf{x}_n, \mathbf{x}_{\mathcal{N}}) \end{bmatrix} \in \mathcal{S}^+(\mathbb{R}^n),$$

with $\mathcal{S}^+(\mathbb{R}^n)$ is the manifold of symmetric positive definite matrices. Having formally specified the semi-parametric class of Gaussian process models, we proceed with presenting the system model.

B. Wireless Sensor Network System Model and Definitions

We present the heterogeneous WSN system with analog/digital sensors:

1) Consider a random spatial phenomenon 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 $f(\cdot) : \mathcal{X} \mapsto \mathbb{R}$, and is modelled as a Gaussian Process (GP) according to

$$f(\cdot) \sim \mathcal{GP}(\mu(\cdot;\boldsymbol{\theta}), \mathcal{C}(\cdot,\cdot;\boldsymbol{\Psi})).$$

2) Let N be the number of sensors that are deployed over a 2-D region $\mathcal{X} \subseteq \mathbb{R}^2$, with $\mathbf{x}_n \in \mathcal{X}, n = \{1, \ldots, N\}$, the physical location of the n-th sensor, assumed known by the FC. The number of analog and digital sensors is N_{A} and N_{D}, respectively, so that $N = N_{\mathrm{A}} + N_{\mathrm{D}}$.

3) **Sensors measurement model:** each sensor collects a noisy observation of the spatial phenomenon $f(\cdot)$. At the n-th sensor, the observation is expressed as:

$$Z(\mathbf{x}_n) = f(\mathbf{x}_n) + W_n, n = \{1, \ldots, N\}$$

where W_n is i.i.d Gaussian noise $W_n \sim N(0, \sigma_{\mathrm{W}}^2)$.

4) **Analog sensors processing:** each of the N_{A} analog sensors transmits its noisy observation to the FC over AWGN channels, as follows:

$$Y_n^A = Z(\mathbf{x}_n) + V_n, n = \{1, \ldots, N_{\mathrm{A}}\},$$

where V_n is i.i.d Gaussian noise $V_n \sim N(0, \sigma_{\mathrm{V}}^2)$.

5) **Digital Sensors processing:** each of the N_{D} digital sensors first performs a thresholding based decision based on its noisy observations and then transmits the 1-bit decision

over imperfect wireless channels [23]–[25]. These steps are summarized as follows:

a) **Thresholding:** at the n-th digital sensor, $n = \{1, \ldots, N_{\mathrm{D}}\}$, the binary decision rule is given by:

$$B(\mathbf{x}_n) = \begin{cases} 1, & Z(\mathbf{x}_n) \geq \lambda \\ 0, & \text{Otherwise} \end{cases}$$

where λ is a pre-defined threshold.

b) **Wireless Communications to Fusion Centre:** the 1-bit decision, $B(\mathbf{x}_n)$, is transmitted to the FC over imperfect binary wireless channels [19]. Under this model, the probability of a received sensor observation taking a specific value $Y_n^D = m$, conditional on the 1-bit decision, is given by $\mathbb{P}(Y_n^D = j|B(\mathbf{x}_n) = i)$. We denote the transition probabilities using a row-stochastic matrix $\mathbf{A}$ as follows:

$$\mathbf{A} = \begin{bmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{bmatrix},$$

where $p_{i,j} := \mathbb{P}(Y_n^D = j|B(\mathbf{x}_n) = i) \; \forall i,j \in \{0,1\}$, and $\sum_{l=0}^{L-1} p_{i,l} = 1 \; \forall i \in \{0,1\}$.

We now present the following definitions:

- $\mathbf{x}_{\mathcal{N}}$ is the physical location (in terms of $[x, y]$ coordinates) of the N sensors deployed in the field.
- $\mathbf{Y}_{\mathcal{N}} = \{Y_1, \ldots, Y_N\} \in \mathbb{R}^{N \times 1}$ is the collection of observations from all sensors (both analog and binary) at the fusion center.
- $\mathbf{Y}_{\mathcal{A}} \subseteq \mathbf{Y}_{\mathcal{N}}$ is the collection of observations from all analog sensors at the fusion center.
- $\mathbf{Y}_{\mathcal{D}} \subseteq \mathbf{Y}_{\mathcal{N}}$ is the collection of observations from all digital sensors, at the fusion center.
- $\mathbf{f}_{\mathcal{N}} = \{f_1, \ldots, f_N\} \in \mathbb{R}^{N \times 1}$ is the realization of $\mathbf{f}$ at the sensors located at $\mathbf{x}_{\mathcal{N}}$.
- $\mathbf{f}_{\mathcal{A}} \subseteq \mathbf{f}_{\mathcal{N}}$ is the realization of $\mathbf{f}$ at the analog sensors, located at $\mathbf{x}_{\mathcal{A}} \subseteq \mathbf{x}_{\mathcal{N}}$.
- $\mathbf{f}_{\mathcal{D}} \subseteq \mathbf{f}_{\mathcal{N}}$ is the realization of $\mathbf{f}$ at the digital sensors, located at $\mathbf{x}_{\mathcal{D}} \subseteq \mathbf{x}_{\mathcal{N}}$.
- $\mathbf{x}_{\mathcal{N}\backslash\mathrm{n}} := [\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_N]$.

C. Notations and Definitions

The following notation is used throughout: random variables are denoted by upper case letters and their realizations by lower case letters. In addition, bold will be used to denote a vector or matrix quantity, and lower subscripts refer to the element of a vector or matrix. We denote $\phi(x; \mu, \sigma^2) = N(x; \mu, \sigma^2)$ as the Probability Density Function (PDF) of a random Normal (Gaussian) variable with mean μ and variance σ^2. Its Cumulative Distribution Function (CDF) is denoted by $\Phi(\lambda, \mu, \sigma^2) = \int_{-\infty}^{\lambda} \phi(x; \mu, \sigma^2) dx$.

III. ESTIMATION CRITERIA FOR SPATIAL INTENSITY, EXCEEDANCE MAP & SPATIAL CLASSIFICATION

As discussed in Section I, the three problems we consider are *spatial field reconstruction, exceedance level estimation* and *spatial classification*. We now present formal definitions of these three problems.

1) **Objective I: MMSE spatial random field reconstruction**—Accurately reconstruct (i.e., estimate) the spatial

random field at un-monitored locations, $\mathbf{x}_* \in \mathcal{X}$, from samples collected by the N sensors. We denote the estimation of the spatial random process at any location $\mathbf{x}_* \in \mathcal{X}$ by $f_* = f(\mathbf{x}_*)$. A widely used criterion is the Minimum Mean Squared Error (MMSE), which utilizes the following distortion metric:

$$D(\widehat{f}_*, f_*) := \mathbb{E}[(f_* - \widehat{f}_*)^2].$$

The optimal solution in the sense of minimizing this distortion metric is the posterior predictive mean [50], given by the solution to the following integral:

$$\widehat{f}_* = \mathbb{E}[f_*|\mathbf{x}_\mathcal{N}, \mathbf{x}_*, \mathbf{Y}_\mathcal{N}]$$
$$= \int f_* p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) \mathrm{d}f_*. \quad (1)$$

2) **Objective II: spatial exeedance map**—Construct a spatial exceedance map estimation is quantified by the following metric: find a region $D_e \subset \mathcal{X}$ such that, with a certain given probability, $f(\mathbf{x}) \geq T$ for all $\mathbf{x} \in D_e$ for a given level T. This can be specified as a set of locations where the posterior predictive probability exceeds some given value:

$$D_e = \left\{ \mathbf{x} : \mathbb{P}(f_* \geq T|\mathbf{x}_\mathcal{N}, \mathbf{x}_*, \mathbf{Y}_\mathcal{N}) \geq 1 - \alpha \right\}$$
$$= \left\{ \mathbf{x} : \int_T^\infty p(f_*|\mathbf{x}_\mathcal{N}, \mathbf{x}_*, \mathbf{Y}_\mathcal{N}) \mathrm{d}f_* \geq 1 - \alpha \right\}, \quad (2)$$

where T is a pre-defined threshold and α is the confidence level and D_e is the domain or set of $\mathbf{x}$ values satisfying the exceedance of the spatial field.

3) **Objective III: Spatial Classification**—Given the observation vector $\mathbf{Y}_\mathcal{N}$ and class labels $B_* \in \{0, 1\}$ drawn according to the probability density function $p_{\mathbf{Y}_\mathcal{N}, B_*}(\mathbf{Y}_\mathcal{N}, B_*)$, the goal is to predict the confidence for each class at un-monitored locations, $\mathbf{x}_* \in \mathcal{X}$. That means that we find the classifier $\widehat{B}_* : \mathcal{X} \leftrightarrow \{0, 1\}$ that minimizes the error probability $\mathbb{P}(B_* \neq \widehat{B}_*)$, at an arbitrary location $\mathbf{x}_* \in \mathcal{X}$. The classifier is given by:

$$\widehat{B}_* = \begin{cases} 1, & \mathbb{P}(B_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) \geq T \\ 0, & \mathbb{P}(B_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) < T, \end{cases} \quad (3)$$

where T is the pre-defined threshold.
The common feature of Objectives 1–3 is the posterior predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$, which as we show in the next Section, involves solving intractable integrals. In the next Section we develop a novel approach to derive this quantity, and in Section V we present the algorithm to compute these goals.

IV. DERIVING THE POSTERIOR PREDICTIVE DISTRIBUTION $p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$

In this section we develop a novel approach to deriving the posterior predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$, which is common to all our estimation objectives in (1), (2) and (3).

To begin, we decompose the posterior predictive distribution as follows:

$$p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$$
$$= \int \cdots \int_{\mathbb{R}^N} p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N}) p(\mathbf{f}_\mathcal{N}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) \mathrm{d}\mathbf{f}_\mathcal{N}$$
$$= \int \cdots \int_{\mathbb{R}^N} p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N}) p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$$
$$\times p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) \mathrm{d}\mathbf{f}_\mathcal{N}, \quad (4)$$

where $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ represents the posterior distribution for the spatial phenomenon at the digital sensor locations given observations; $p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ is the posterior distribution for the spatial phenomenon at the analog sensor locations given observations and the values of the spatial phenomenon at the digital sensor locations; and $p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N})$ is the conditional predictive prior distribution. The steps we shall take to calculate $p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ are presented in a generic form in Algorithm 1.

Algorithm 1: Calculation of the Posterior Predictive Distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ in (4)

1: Calculate $p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N})$ (presented in Lemma 1).
2: Calculate $p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ (presented in Lemma 2).
3: Derive $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ (presented in Lemma 3 and Proposition 1).
4: Calculate $p(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ by solving the integral in (4) using the quantities derived in Steps 1–3. (presented in Theorem 1).

We begin with the calculation of Steps 1–2 in Algorithm 1, presented in Lemma 1 and Lemma 2.

Lemma 1: The conditional predictive prior distribution, $p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N})$, is given by:

$$p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N}) = N\left(f_*; \mu_{f_*|\mathbf{f}_\mathcal{N}}, \sigma^2_{f_*|\mathbf{f}_\mathcal{N}}\right),$$

where

$$\mu_{f_*|\mathbf{f}_\mathcal{N}} := \mu(\mathbf{x}_*) + k(\mathbf{x}_*, \mathbf{x}_\mathcal{N})\mathbf{K}^{-1}(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})(\mathbf{f}_\mathcal{N} - \mu(\mathbf{x}_\mathcal{N})),$$
$$\sigma^2_{f_*|\mathbf{f}_\mathcal{N}} := k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{x}_\mathcal{N})\mathbf{K}^{-1}(\mathbf{x}_\mathcal{N}, \mathbf{x}_\mathcal{N})k(\mathbf{x}_\mathcal{N}, \mathbf{x}_*).$$

Proof: This is a simple application of conditional Normal distribution properties which stems from the fact that we use a GP prior on $\mathbf{f}$ and therefore $p(f_*|\mathbf{f}_\mathcal{N}, \mathbf{x}_*, \mathbf{x}_\mathcal{N})$ is a multivariate Normal distribution. ∎

Lemma 2: The conditional distribution at the analog sensors, $p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$, is given by:

$$p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) = N(\mathbf{f}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}}, \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}}),$$

where

$$\boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}} := \left(\boldsymbol{\Sigma}^{-1}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}} + \sigma_W^{-2}\mathbf{I}\right)^{-1}$$
$$\times \left(\boldsymbol{\Sigma}^{-1}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}\boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}} + \sigma_W^{-2}\mathbf{Y}_\mathcal{A}\right)$$
$$\boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}} := \left(\boldsymbol{\Sigma}^{-1}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}} + \sigma_W^{-2}\mathbf{I}\right)^{-1}.$$

Proof: See Appendix A. ∎
We now proceed and calculate Step 3 in Algorithm 1.

A. Deriving Analytic Expression for $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ via Saddle-Point (Laplace) Method

We now derive analytic expression of the posterior distribution for the spatial phenomenon at the digital sensor locations given observations, $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ based on a series expansion of the Saddle-point (Laplace) type via a Gaussian basis.

Using Bayes' law, the posterior distribution for the spatial phenomenon at the digital sensor locations is given by

$$p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) = \frac{\mathbb{P}(\mathbf{Y}_\mathcal{N}|\mathbf{x}_\mathcal{N}, \mathbf{f}_\mathcal{D})p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N})}{\mathbb{P}(\mathbf{Y}_\mathcal{N}|\mathbf{x}_\mathcal{N})}$$
$$= \frac{\mathbb{P}(\mathbf{Y}_\mathcal{N}|\mathbf{x}_\mathcal{N}, \mathbf{f}_\mathcal{D})p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N})}{\int \cdots \int_{\mathbb{R}^N} \mathbb{P}(\mathbf{Y}_\mathcal{N}|\mathbf{x}_\mathcal{N}, \mathbf{f}_\mathcal{D})p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N})\mathrm{d}\mathbf{f}_\mathcal{D}}.$$

The multivariate posterior distribution is intractable due to the fact that the marginal likelihood, $\mathbb{P}(\mathbf{Y}_\mathcal{N}|\mathbf{x}_\mathcal{N})$, cannot be evaluated pointwise. We therefore utilize an approximation of the distribution $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$. This approximation has a specific form such that it will make the integral in (4) to become closed form. Our solution is based on a series expansion of the posterior distribution of the Saddle-point (Laplace) type via a Gaussian basis. This involves deriving a multi-variate Taylor series expansion for the log posterior, explained next.

We begin by presenting the definition of the multivariate Taylor series expansion. From this we will motivate the basic concept of an asymptotic expansion in terms of the 2-nd order type known as a Saddle-point approximation, which we will show is a powerful tool for obtaining accurate expressions for the posterior distribution.

Definition 2: (Multivariate Taylor Series Expansion [51]) Consider the function $g(\mathbf{x})$ (given by $g(\mathbf{x}) := \log(p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}))$ in this paper) which is a scalar function of vector $\mathbf{x}$, with all partial derivatives continuous in the neighborhood of $\widetilde{\boldsymbol{b}}$ denoted $n.e.(\widetilde{\mathbf{x}})$. Then in this neighborhood the k-th order multivariate Taylor series is well defined and given by

$$g(\mathbf{x}) = g(\widetilde{\boldsymbol{b}}) + g^{(1)}(\widetilde{\boldsymbol{b}})(\mathbf{x} - \widetilde{\boldsymbol{b}}) + \frac{1}{2}(\mathbf{x} - \widetilde{\boldsymbol{b}})^T g^{(2)}(\widetilde{\boldsymbol{b}})(\mathbf{x} - \widetilde{\boldsymbol{b}})$$
$$+ \frac{1}{3!}((\mathbf{x} - \widetilde{\boldsymbol{b}})^T \otimes (\mathbf{x} - \widetilde{\boldsymbol{b}})^T)g^{(3)}(\widetilde{\boldsymbol{b}})(\mathbf{x} - \widetilde{\boldsymbol{b}}) + \cdots$$
$$+ \frac{1}{k!}\begin{pmatrix} k-1 \\ \otimes (\mathbf{x} - \widetilde{\boldsymbol{b}})^T \end{pmatrix} g^{(k)}(\widetilde{\boldsymbol{b}})(\mathbf{x} - \widetilde{\boldsymbol{b}}) + R_{\boldsymbol{\beta}}(\mathbf{x}),$$

where the partial derivatives are given by

$$g^{(k)}(\widetilde{\boldsymbol{b}}) = \left.\frac{\partial \mathrm{vec}\left[g^{(k-1)}(\mathbf{x})\right]}{\partial \mathbf{x}^T}\right|_{\mathbf{x}=\widetilde{\boldsymbol{b}}}$$

The accuracy of such approximation has been studied in a range of places, see [44], [51]–[53] for further details. These depend on the regularity assumptions assumed for the function of interest g, see [52] for details. We apply this methodology to approximate the 2nd order series expansion for the posterior distribution at the digital sensors, $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ ([51], Section 4, Equation 4.3).

Lemma 3: The posterior distribution at the digital sensors, $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$, is given by:

$$\log p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) = \log q\left(\mathbf{f}_\mathcal{D}; \widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}, H\right) + R_3(\mathbf{f}_\mathcal{D}).$$

where

$$q\left(\mathbf{f}_\mathcal{D}; \widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}, H\right) = N\left(\mathbf{f}_\mathcal{D}; \widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}, H^{-1}\right),$$
$$\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}} = \arg\max_{\mathbf{f}_\mathcal{D}} p\left(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \in \mathbb{R}^{N\times 1},$$
$$[H]_{i,j} = -\frac{\partial^2}{\partial f_i \partial f_j}\log p\left(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right)|_{\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}},$$
$$R_3(\mathbf{f}_\mathcal{D}) = g\left(\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}\right) + \widetilde{R}_3(\mathbf{f}_\mathcal{D}) + \log((2\pi)^n|H|^{1/2})$$

Next, based on Lemma 3, we approximate $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ as follows.

Proposition 1: The posterior distribution $p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})$ is approximated as follows:

$$p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) \simeq \widehat{p}(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) = q\left(\mathbf{f}_\mathcal{D}; \widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}, H\right).$$

To obtain the conditional distribution in Proposition 1, we need to calculate the MAP estimate, $\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}$, presented in the next section.

B. Deriving the MAP Estimate $\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}$

We now present an optimal and computationally efficient algorithm to calculate the MAP estimate. First, in Lemma 4 we present the objective function. Then, in Lemma 5 we prove that the objective function is quasi-convex and can therefore be solved exactly using any gradient based approach. Finally, in Lemma 6 we present an efficient procedure to solve the resulting optimization problem via an application of the ICM method.

Lemma 4: The MAP estimate is given by the solution for the following N_D-dimensional optimization problem:

$$\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}$$
$$= \arg\max_{\mathbf{f}_\mathcal{D}}\left(\sum_{n=1}^{N_\mathrm{D}}\log\left(\sum_{l=0}^{1}\mathbb{P}\left(Y_n^D|B_n = l\right)\mathbb{P}(B_n = l|f_n)\right)\right.$$
$$+ \log N\left(\mathbf{Y}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}, \left(\sigma_\mathrm{V}^2 + \sigma_\mathrm{W}^2\right)\mathbf{I}_{N_\mathrm{A}} + \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}\right)$$
$$\left. + \log N\left(\mathbf{f}_\mathcal{D}; \boldsymbol{\mu}(\mathbf{x}_\mathcal{D}), \mathbf{K}(\mathbf{x}_\mathcal{D}, \mathbf{x}_\mathcal{D})\right)\right).$$

Proof: See Appendix B ∎

We now prove that the objective function in Lemma 4 is quasi-convex and then present an efficient algorithm to calculate $\widehat{\mathbf{f}}_\mathcal{D}^{\mathrm{MAP}}$.

Lemma 5: The objective function in Lemma 4 is quasi-convex (unimodal) in $\mathbf{f}_\mathcal{D}$.

Proof: See Appendix C. ∎

To calculate the MAP estimate in Lemma 4 efficiently, we utilize the Iterative Conditioning on the Modes (ICM) estimation method [42], [43]. The ICM approach decomposes the multivariate problem into a sequential series of N_D one-dimensional optimization problems which can be easily solved. The algorithm iterates each of the N_D univariate optimization problems recursively until convergence.

The solution to the optimization problem is presented in the following Lemma.

Lemma 6: The MAP estimate of the n-th component of $\mathbf{f}_{\mathcal{D}}$, $\widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}} = \mathrm{argmax}_{f_n}\, p(f_n|\mathbf{x}_{\mathcal{N}}, \widehat{\mathbf{f}}_{1:\mathrm{N}_{\mathcal{D}}\backslash n}, \mathbf{Y}_{\mathcal{N}})$, is given by the solution of the following one-dimensional non-linear equation:

$$\frac{\phi\left(\lambda, f(\mathbf{x}_n), \sigma_{\mathrm{W}}^2\right)\left(\mathbb{P}(Y_n|B_n=0) - \mathbb{P}(Y_n|B_n=1)\right)}{\mathbb{P}(Y_n|B_n=1) + \Phi\left(\lambda, f(\mathbf{x}_n), \sigma_{\mathrm{W}}^2\right)}$$
$$\times \frac{1}{\mathbb{P}(Y_n|B_n=0) - \mathbb{P}(Y_n|B_n=1)}$$
$$= \left(\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{A}}|\mathbf{f}_{\mathcal{D}}} - \mathbf{Y}_{\mathcal{A}}\right)^T$$
$$\times \left(\left(\sigma_{\mathrm{V}}^2 + \sigma_{\mathrm{W}}^2\right)\mathbf{I}_{N_{\mathcal{A}}} + \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{A}}|\mathbf{f}_{\mathcal{D}}}\right)^{-1} K\left(\mathbf{x}_{\mathcal{A}}, \mathbf{x}_{\mathcal{D}}\right)$$
$$\times K^{-1}\left(\mathbf{x}_{\mathcal{D}}, \mathbf{x}_{\mathcal{D}}\right)\mathbb{1}_n + \frac{\left(f(\mathbf{x}_n) - \mu_{\mathbf{x}_n|\mathbf{f}_{\mathcal{D}}\backslash n}\right)}{\sigma_{\mathbf{x}_n|\mathbf{f}_{\mathcal{D}}\backslash n}^2}.$$

Proof: See Appendix D. ∎

We now proceed and calculate Step 4 in Algorithm 1, which is then presented in Theorem 1.

C. Deriving the Posterior Predictive Distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$ in (4)

We now combine the results from Lemma 1, Lemma 2 and Proposition 1 to obtain the posterior predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$ in (4). We begin by deriving the posterior distribution $p(\mathbf{f}_{\mathcal{N}}|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$ and then combine that with the conditional predictive prior distribution to obtain the posterior predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$.

Lemma 7: The conditional distribution $p(\mathbf{f}_{\mathcal{N}}|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$ in (4) is given by:

$$p(\mathbf{f}_{\mathcal{N}}|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}) = p\left(\mathbf{f}_{\mathcal{A}}|\mathbf{f}_{\mathcal{D}}, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right) p\left(\mathbf{f}_{\mathcal{D}}|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right)$$
$$\simeq \widehat{p}\left(\mathbf{f}_{\mathcal{N}}|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right)$$
$$= p\left(\mathbf{f}_{\mathcal{A}}|\mathbf{f}_{\mathcal{D}}, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right) q\left(\mathbf{f}_{\mathcal{D}}; \widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}}, H\right)$$
$$= N\left(\mathbf{f}_{\mathcal{N}}; \boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}, \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}\right),$$

where

$$\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} = \begin{bmatrix} \widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}} \\ \boldsymbol{\mu}_{\mathbf{f}_{\mathcal{A}}|\widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}}, \mathbf{Y}_{\mathcal{N}}} \end{bmatrix}, \tag{5}$$

and:

$$\boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} = \begin{bmatrix} H & H\mathbf{A}^T \\ \mathbf{A}H & \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{A}}|\widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}}, \mathbf{Y}_{\mathcal{N}}} + \mathbf{A}H\mathbf{A}^T \end{bmatrix},$$

with

$$A := \left(\boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{A}}|\widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}}}^{-1} + N_{\mathcal{A}}\sigma_W^{-2}\mathbf{I}\right)^{-1}$$
$$\times \left(\boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{A}}|\widehat{\mathbf{f}}_{\mathcal{D}}^{\mathrm{MAP}}}^{-1} K\left(\mathbf{x}_{\mathcal{A}}, \mathbf{x}_{\mathcal{D}}\right) K^{-1}\left(\mathbf{x}_{\mathcal{D}}, \mathbf{x}_{\mathcal{D}}\right)\right).$$

Proof: This is the result of combining Lemma 2 and Proposition 1 using the Aitken Block-Diagonalization formula [50]. ∎

Next, we evaluate the integrands of the integral in (4):

Lemma 8: The joint distribution $p([\mathbf{f}_{\mathcal{N}}\ f_*]|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$ is given by

$$p\left([\mathbf{f}_{\mathcal{N}}\ f_*]^T|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right) = p\left(f_*|\mathbf{f}_{\mathcal{N}}, \mathbf{x}_*, \mathbf{x}_{\mathcal{N}}\right) p\left(\mathbf{f}_{\mathcal{N}}|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right)$$
$$= N\left([f_*\ \mathbf{f}_{\mathcal{N}}]^T; \boldsymbol{\mu}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}, \boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}\right),$$

where

$$\boldsymbol{\mu}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}$$
$$= \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} \\ \mu\left(\mathbf{x}_*\right) \\ +k(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}})K^{-1}\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}\right)\left(\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} - \mu(\mathbf{x}_{\mathcal{N}})\right) \end{bmatrix}$$

and the covariance is given by:

$$\boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} = \begin{bmatrix} \boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{11} & \boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{12} \\ \boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{21} & \boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{22} \end{bmatrix},$$

where

$$\boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{11} = \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}},$$
$$\boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{12} = \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} K^{-1}\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}\right) k(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_*),$$
$$\boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{21} = k\left(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}\right) K^{-1}\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}\right) \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}},$$
$$\boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{22} = \sigma_{f_*|\mathbf{f}_{\mathcal{N}}}^2 + k\left(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}\right) K^{-1}\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}\right) \boldsymbol{\Sigma}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}$$
$$\times K^{-1}\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}\right) k\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_*\right).$$

Proof: This is a simple application of the Aitken Block-Diagonalization formula [50]. ∎

Finally, the following Theorem provides the solution for the predictive distribution in (4).

Theorem 1: The posterior predictive distribution, $p(f_*|\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$, is given by:

$$\widehat{p}\left(f_*|\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right) = \int p\left([\mathbf{f}_{\mathcal{N}}\ f_*]^T|\mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}\right) d\mathbf{f}_{\mathcal{N}}$$
$$= N\left(f_*; \mu_{f_*|\mathbf{Y}_{\mathcal{N}}}, \sigma_{f_*|\mathbf{Y}_{\mathcal{N}}}^2\right), \tag{6}$$

where

$$\mu_{f_*|\mathbf{Y}_{\mathcal{N}}} = \mu(\mathbf{x}_*) + k\left(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}\right) K^{-1}\left(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}\right)$$
$$\times \left(\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}} - \mu\left(\mathbf{x}_{\mathcal{N}}\right)\right),$$
$$\sigma_{f_*|\mathbf{Y}_{\mathcal{N}}}^2 = \boldsymbol{\Sigma}_{f_*, \mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}^{22}.$$

Proof: This result comes from the property of marginalization of Normal distributions. ∎

Remark 1: We note that the mean of the posterior predictive distribution $\mu_{f_*|\mathbf{Y}_{\mathcal{N}}}$ has the same structure as Gaussian Process regression, where instead of using the observations directly, we substitute them with the estimate of the intensity of the field in all sensor locations, $\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}}|\mathbf{Y}_{\mathcal{N}}}$.

In the next section we shall utilize this result to derive the objectives 1–3 of Section III.

V. ALGORITHM DEVELOPMENT FOR THE THREE ESTIMATION CRITERIA

Now that we have derived the posterior predictive distribution $p(f_*|\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}})$, we present the solution to *spatial field reconstruction, exceedance level estimation* and *spatial classification* presented in Objectives 1–3 of Section III.

A. Objective I: Spatial Random Field Reconstruction

The MMSE estimate for the spatial random field reconstruction in (1) is given by

$$\widehat{f}_* = \mathbb{E}[f_* | \mathbf{x}_{\mathcal{N}}, \mathbf{x}_*, \mathbf{Y}_{\mathcal{N}}]$$
$$\simeq \int f_* \widehat{p}(f_* | \mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}) \mathrm{d} f_*$$
$$= \int f_* N\left(f_*; \mu_{f_* | \mathbf{Y}_{\mathcal{N}}}, \sigma^2_{f_* | \mathbf{Y}_{\mathcal{N}}}\right) \mathrm{d} f_*$$
$$= \mu(\mathbf{x}_*) + k\left(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}\right) K^{-1}(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}) \left(\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}} | \mathbf{Y}_{\mathcal{N}}} - \mu(\mathbf{x}_{\mathcal{N}})\right),$$

where $\boldsymbol{\mu}_{\mathbf{f}_{\mathcal{N}} | \mathbf{Y}_{\mathcal{N}}}$ is defined in (5).

B. Objective II: Spatial Exceedance Map

The exceedance map in (2) is based on solving for the following quantity,

$$p\left(f_* \geq \lambda | \mathbf{x}_{\mathcal{N}}, \mathbf{x}_*, \mathbf{Y}_{\mathcal{N}}\right) \simeq 1 - \Phi\left(\lambda, \mu_{f_* | \mathbf{Y}_{\mathcal{N}}}, \sigma^2_{f_* | \mathbf{Y}_{\mathcal{N}}}\right).$$

which can then be used to obtain the locations $\mathbf{x}$ required to obtain the region

$$D_e = \{\mathbf{x} : \mathbb{P}\left(f_* \geq T | \mathbf{x}_{\mathcal{N}}, \mathbf{x}_*, \mathbf{Y}_{\mathcal{N}}\right) \geq 1 - \alpha\} \qquad (7)$$

C. Objective III: Spatial Classification

In this section we present the algorithm to calculate the spatial classification of the spatial physical phenomenon.

Theorem 2: The binary predictive distribution is given by:

$$\mathbb{P}\left(B_* = 0 | \mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}, \lambda\right) = \Phi\left(\lambda, \mu_{f_* | \mathbf{Y}_{\mathcal{N}}}, \sigma^2_{\mathrm{W}} + \sigma^2_{f_* | \mathbf{Y}_{\mathcal{N}}}\right),$$
$$\mathbb{P}\left(B_* = 1 | \mathbf{x}_*, \mathbf{x}_{\mathcal{N}}, \mathbf{Y}_{\mathcal{N}}, \lambda\right) = 1 - \Phi\left(\lambda, \mu_{f_* | \mathbf{Y}_{\mathcal{N}}}, \sigma^2_{\mathrm{W}} + \sigma^2_{f_* | \mathbf{Y}_{\mathcal{N}}}\right).$$

where $\mu_{f_* | \mathbf{Y}_{\mathcal{N}}}$ and $\sigma^2_{f_* | \mathbf{Y}_{\mathcal{N}}}$ are defined in Theorem 1.
    Proof: See Appendix E ∎

VI. SIMULATION RESULTS

In this section we consider two studies, the first is based on synthetic data generated from a known model. We use this controlled scenario to demonstrate the properties of our estimation methods and illustrate how accurate they will be in different settings. Then we study a real data application which involves analysis of wind-speed data with the application in mind related to storm surge modelling in Europe, under the class of weather events known in insurance modelling as wind-storms or storm surge insurance storms. This type of application is of direct relevance for both safety assessment and insurance pricing purposes, see [54] and [55].

A. Synthetic Example

To evaluate the performance of the proposed algorithms and the improvement they provide we generated 2-D realizations from a Gaussian Process with the following attributes: the mean
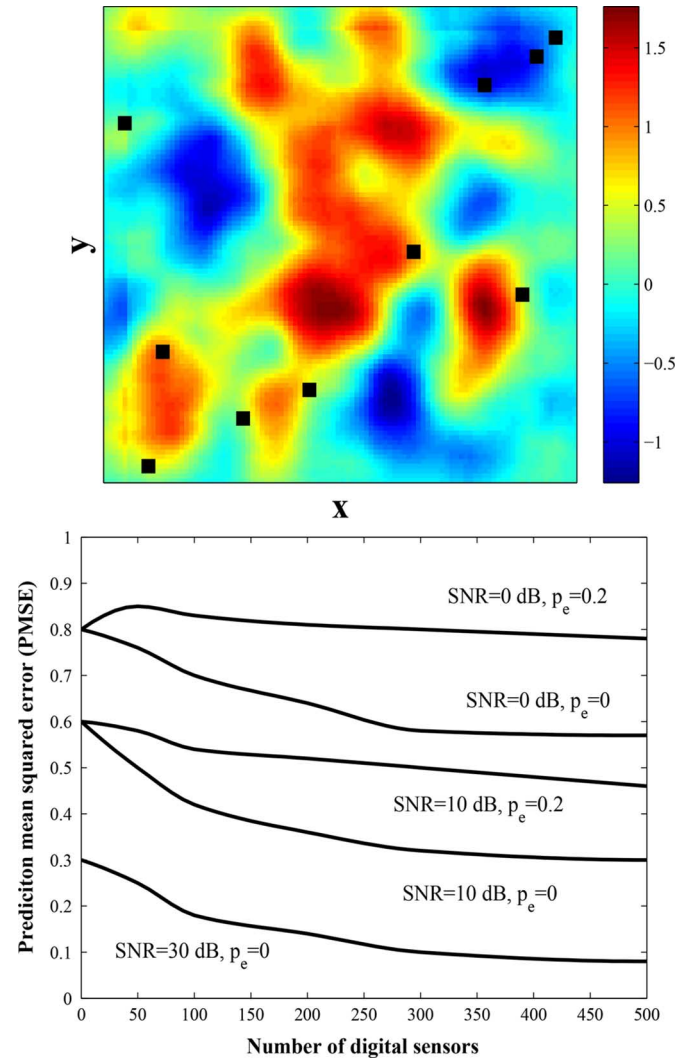

Fig. 1. Upper panel: realization from a 2-d Gaussian Process. The black markers denote the locations of the 10 high-quality sensors. Lower panel: PMSE results vor various configurations.

is $\mu(\mathbf{x}) = 0$ and the kernel is a radial basis function with length scale, $l = 2$.

$$\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\Psi}) := \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{l}\right).$$

A realization from the GP is shown in Fig. 1. In this example we placed 10 high-quality sensors which are marked by the black markers. We then tested the field reconstruction algorithm for various system configuration, changing the number of low-quality sensors, the SNR and the probabilities of incorrect wireless channels transmission, denoted p_e. To obtain the same measure of SNR for both types of sensors, we set $\sigma^2_v = 0$ and define $\mathrm{SNR} = 10 \log \sigma^2_w$. The Prediction Mean Squared Errors (PMSE) are presented in the right side of Fig. 1. The results show that substantial improvements can be obtained by adding low-quality sensors. This is especially true in the cases of high SNR and perfect wireless channels communications, where the PMSE of the heterogeneous network is roughly 1/3 of the PMSE based only on high-quality sensors.
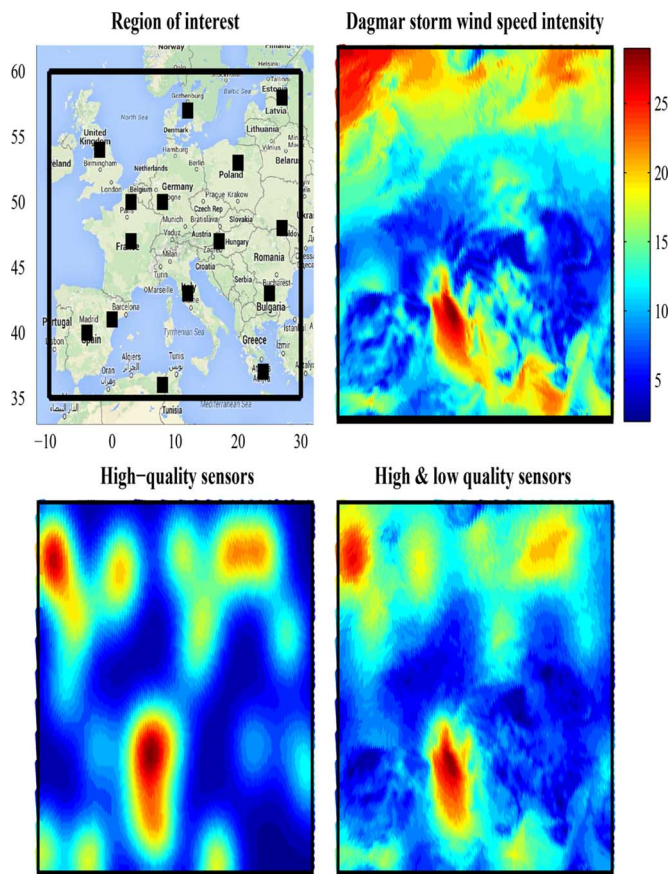
Fig. 2. Wind speed prediction of the Dagmar storm. The rectangular in the upper left figure represents the region of interest which contains 15 high-quality sensors. The upper right figure represents the "true" data wind speed intensities [m/s]. The lower left figure shows the field reconstruction based solely on the 15 high-quality sensors via Gaussian Process regression. The lower right figure shows the field reconstruction of our algorithm based on the heterogeneous network with 15 high-quality sensors and 100 low-quality sensors. The Normalized Mean Squared Error based on the high-quality sensors is 0.67 and based on both high and low quality sensors is 0.25.

B. Sensor Networks for Insurance: Wind Speed and Insurance Storms

In this study we use a publicly available insurance storm surge database known as the Extreme Wind Storms Catalogue[2]. The data is available for research as the XWS Datasets: (c) Copyright Met Office, University of Reading and University of Exeter. Licensed under Creative Commons CC BY 4.0 International License. The data provided is comprehensive and provides features such as the footprint of the observations on a location grid with a rotated pole at longitude = 177.5 degrees, latitude = 37.5 degrees. As discussed in the data description provided with the data-set, this is a standard technique used to ensure that the spacing in km between grid points remains relatively consistent. The footprints are on a regular grid in the rotated coordinate system, with horizontal grid spacing 0.22 degrees. The data for each of the storms provides a list of grid number and maximum 3-second gust speed in meters per second. The true locations (longitude and latitude) of the grid points are given in grid locations file. We selected two storms to analyze, the first is known as
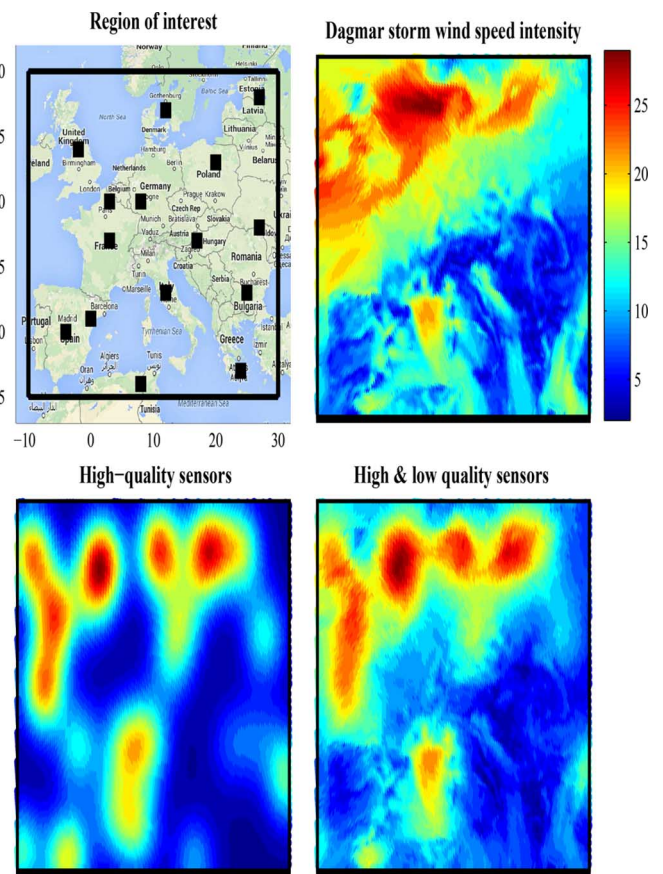
Fig. 3. Wind speed prediction of the Ulli storm. The Normalized Mean Squared Error based on the high-quality sensors is 0.85 and based on both high and low quality sensors is 0.37.

Dagmar which took place on 26/12/2011 and affected Finland and Norway; and the second was the storm known as Ulli taking place on 3/1/2012 which affected the UK.

1) Model Calibration Wind Speed Data: To calibrate the model we first fit the hyperparameters of the model via Maximum Likelihood Estimation (MLE) procedure. We used a 2-D radial basis function, of the following form

$$\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\Psi}) := \sigma_x^2 \exp\left(-\frac{\|x_i - x_j\|}{l_x}\right)\sigma_y^2$$
$$\times \exp\left(-\frac{\|y_i - y_j\|}{l_y}\right),$$

thus decomposing the kernel into orthogonal coordinates which we found provided a much more accurate fit. The reason for this is it allows for inhomogeneity through differences in spatial dependence in vertical and horizontal directions, which is highly likely to occur in the types of wind speed data studied. The MLE of the length and scale parameters obtained are given by:

- Dagmar storm: $\sigma_x^2 = 0.1, l_x = 0.5$ and $\sigma_y^2 = 10, l_y = 0.1$.
- Ulli storm: $\sigma_x^2 = 0.5, l_x = 0.1$ and $\sigma_y^2 = 1, l_y = 0.1$.

We note that details on how to estimate the GP hyperparamarets can be found in [Chapter 5] [48].

2) Wind Field Intensity Estimation for Insurance Wind Storms: We performed wind field intensity estimation using our algorithm and compared it to the case where only high-quality sensors are utilized. The results are presented in Figs. 2–3, for

TABLE I
FIELD RECONSTRUCTION PERFORMANCE FOR THE TWO STORMS

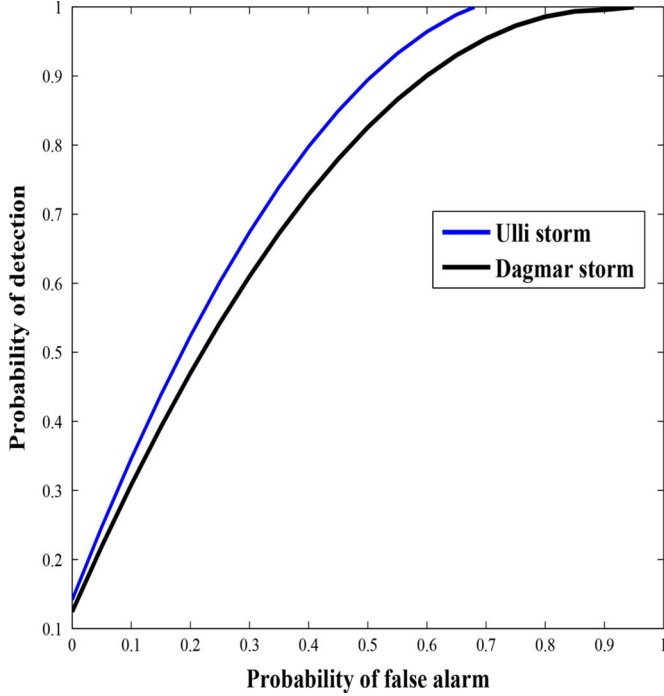| Normalised Prediction Mean Squared Error		
Reconstruction method	Dagmar storm	Ulli storm
15 Analog sensors	0.67	0.85
15 Analog and 100 Digital sensors	0.25	0.37

Fig. 4. Wind speed detection vs. false alarm probabilities of Dagmar and Ulli storms.

the Dagmar and Ulli storms, respectively. We set the region of interest (ROI) as shown in the upper left of Figs. 2–3. We then chose 15 locations to place high-quality sensors. These locations are depicted with black square markers. The actual wind speed field intensity is shown in the upper right figures. The lower left figures show the estimated field based only on the 15 high-quality sensors. The lower right figures show the estimated field based on the 15 high-quality and 100 low-quality sensors. To illustrate the impact of adding low-quality sensors make, we set the error probability of the wireless channels to zero. The figures show that a significant improvement can be obtained by augmenting the high-quality sensor network with many cheap low-quality sensors. The field reconstruction (the ROI contains 14006 spatial points) prediction mean squared error for the two storms is given in the Table I. As expected the prediction performance for the Dagmar storm is better than for the Ulli storm. This can be explained by the higher spatial correlation exhibited by the Dagmar storm.

3) Wind Field Intensity Classification: We now turn our attention to the classification problem presented in Objective III. We used the same set up as the previous example and performed spatial classification where we set the threshold λ to be the mean value of each of the storms plus 5. We plot the Receiver Operating Characteristic (ROC) curve for both storms in Fig. 4.

The results show that good classification performance can be obtained for both storms. It is also clear that the detection performance of the Dagmar storm is better than the Ulli storm. This is because the spatial correlation of the Dagmar storm is much stronger than of the Ulli storm in both axes, resulting in better detection performance.

VII. CONCLUSION

We developed a new framework for *spatial field reconstruction, exceedance level estimation* and *classification* in heterogeneous (mixed analog & digital sensors) wireless sensor networks. We developed novel algorithms that are based on a multivariate series expansion approach resulting in a Saddle-point type approximation. We then studied the performance gain that can be obtained by augmenting the high-quality sensors with low-quality sensors using real data of insurance storm surge database. The results show a significant improvement for both field reconstruction and spatial classification.

APPENDIX A
PROOF OF LEMMA 2

$$
\begin{aligned}
p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) &= \frac{p(\mathbf{f}_\mathcal{A}, \mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})}{p(\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N})} \\
&= \frac{p(\mathbf{Y}_\mathcal{N}|\mathbf{f}_\mathcal{A}, \mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})p(\mathbf{f}_\mathcal{A}, \mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N})}{p(\mathbf{Y}_\mathcal{N}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})p(\mathbf{f}_\mathcal{D}|\mathbf{x}_\mathcal{N})} \\
&= \frac{p(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{A}, \mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})}{p(\mathbf{Y}_\mathcal{A}|\mathbf{Y}_\mathcal{D}, \mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})} \\
&= \frac{p(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{A}, \mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})}{\int p(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{A}, \mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N})d\mathbf{f}_\mathcal{A}}.
\end{aligned}
$$

Using the fact that the spatial phenomenon is modelled a-priori as a GP, and utilizing the properties of conditional Normal distribution, the second term in the nominator is given by:

$$
\begin{aligned}
p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}) &= N\left(\mathbf{f}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}, \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}\right), \\
\boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}} &= \mu(\mathbf{x}_\mathcal{A}) + K(\mathbf{x}_\mathcal{A}, \mathbf{x}_\mathcal{D})K^{-1} \\
&\quad \times (\mathbf{x}_\mathcal{D}, \mathbf{x}_\mathcal{D})(\mathbf{f}_\mathcal{D} - \mu(\mathbf{x}_\mathcal{D})), \\
\boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}} &= K(\mathbf{x}_\mathcal{A}, \mathbf{x}_\mathcal{A}) - K(\mathbf{x}_\mathcal{A}, \mathbf{x}_\mathcal{D})K^{-1} \\
&\quad \times (\mathbf{x}_\mathcal{D}, \mathbf{x}_\mathcal{D})K(\mathbf{x}_\mathcal{D}, \mathbf{x}_\mathcal{A}). \quad (8)
\end{aligned}
$$

Next, using (8) and invoking Normal-Normal conjugacy properties, we obtain

$$
\begin{aligned}
p(\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}) &= N\left(\mathbf{f}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}}, \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}}\right), \\
\boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}} &:= \left(\boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}^{-1} + \sigma_W^{-2}\mathbf{I}\right)^{-1} \\
&\quad \times \left(\boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}^{-1}\boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}} + \sigma_W^{-2}\mathbf{Y}_\mathcal{A}\right), \\
\boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{Y}_\mathcal{N}} &:= \left(\boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}^{-1} + \sigma_W^{-2}\mathbf{I}\right)^{-1}.
\end{aligned}
$$

APPENDIX B
PROOF OF LEMMA 4

The MAP estimate is given by (9) at the bottom of the page.

APPENDIX C
PROOF OF LEMMA 5

To prove that the objective function in Lemma 4 is quasi-convex we first note that the second and third terms are log quasi-convex since these are multivariate Normal distributions. Therefore, we only need to show that $\sum_{l=0}^{1} \mathbb{P}(Y_n^D|B_n = l)\mathbb{P}(B_n = l|f_n)$ is log quasi-convex in $f(\mathbf{x}_n), \forall n \in \{1, \ldots, N_D\}$.

$$
\begin{aligned}
g(f_n) &:= \sum_{l=0}^{1} \mathbb{P}\left(Y_n^D|B_n = l\right) \mathbb{P}\left(B_n = l|f_n\right) \\
&= \mathbb{P}\left(Y_n^D|B_n = 0\right) \mathbb{P}\left(B_n = 0|f_n\right) \\
&\quad + \mathbb{P}\left(Y_n^D|B_n = 1\right) \mathbb{P}\left(B_n = 1|f_n\right) \\
&= \left(\mathbb{P}\left(Y_n^D|B_n=0\right) - \mathbb{P}\left(Y_n^D|B_n=1\right)\right) \Phi\left(\lambda, f(\mathbf{x}_n), \sigma_W^2\right) \\
&\quad + \mathbb{P}\left(Y_n^D|B_n = 1\right).
\end{aligned}
\tag{10}
$$

To show that (10) is log quasi-convex we use the following property: if $h'(x) = 0$, and $h''(x) \geq 0$, then $h(x)$ is quasi-convex in x [56]. The first derivative of $\log g(f_n)$ is given by:

$$
\begin{aligned}
\frac{d \log g(f_n)}{df_n} &= \frac{1}{g(f_n)}\phi\left(\lambda, f_n, \sigma_W^2\right) \\
&\quad \times (\mathbb{P}(Y_n|B_n = 0) - \mathbb{P}(Y_n|B_n = 1)).
\end{aligned}
$$

Next, we evaluate the second derivative and set the first derivative to 0 [see the equation at the bottom of the page]:

APPENDIX D
PROOF OF LEMMA 6

The MAP estimate for the n-th component of $\mathbf{f}_D$, $n \in \{1, \ldots, N_D\}$, conditional on all other components $\widehat{\mathbf{f}}_{1:N_D \backslash n} := [\widehat{f}_1, \ldots, \widehat{f}_{n-1}, \widehat{f}_{n+1}, \ldots, \widehat{f}_D]$ and the observations $\mathbf{Y}_\mathcal{N}$, is obtained by expressing the full conditional posterior for the n-th

$$
\begin{aligned}
\widehat{\mathbf{f}}_D^{\text{MAP}} &= \arg\max_{\mathbf{f}_D} p\left(\mathbf{f}_D|\mathbf{Y}_\mathcal{N}, \mathbf{x}_\mathcal{N}\right) \\
&= \arg\max_{\mathbf{f}_D} \mathbb{P}(\mathbf{Y}_\mathcal{N}|\mathbf{f}_D, \mathbf{x}_\mathcal{N})p(\mathbf{f}_D|\mathbf{x}_\mathcal{N}) \\
&= \arg\max_{\mathbf{f}_D} \mathbb{P}(\mathbf{Y}_D|\mathbf{f}_D, \mathbf{x}_\mathcal{N})p(\mathbf{Y}_\mathcal{A}|\mathbf{f}_D, \mathbf{x}_\mathcal{N})p(\mathbf{f}_D|\mathbf{x}_\mathcal{N}) \\
&= \arg\max_{\mathbf{f}_D} \mathbb{P}(\mathbf{Y}_D|\mathbf{f}_D, \mathbf{x}_\mathcal{N}) \left(\int p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{A}, \mathbf{x}_\mathcal{N}\right) p(\mathbf{f}_\mathcal{A}|\mathbf{f}_D, \mathbf{x}_\mathcal{N})d\mathbf{f}_\mathcal{A}\right) p(\mathbf{f}_D|\mathbf{x}_\mathcal{N}) \\
&= \arg\max_{\mathbf{f}_D} \left(\log\left(\prod_{n=1}^{N_D}\sum_{l=0}^{1} \mathbb{P}\left(Y_n^D|B_n = l\right) \mathbb{P}(B_n = l|f_n)\right)\right. \\
&\quad \left. + \log N\left(\mathbf{Y}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_D}, \left(\sigma_V^2 + \sigma_W^2\right) \mathbf{I}_{N_A} + \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_D}\right) + \log N\left(\mathbf{f}_D; \boldsymbol{\mu}(\mathbf{x}_D), \mathbf{K}(\mathbf{x}_D, \mathbf{x}_D)\right)\right) \\
&= \arg\max_{\mathbf{f}_D} \left(\sum_{n=1}^{N_D}\log\left(\sum_{l=0}^{1} \mathbb{P}\left(Y_n^D|B_n = l\right) \mathbb{P}(B_n = l|f_n)\right)\right. \\
&\quad \left. + \log N\left(\mathbf{Y}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_D}, \left(\sigma_V^2 + \sigma_W^2\right) \mathbf{I}_{N_A} + \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_D}\right) + \log N\left(\mathbf{f}_D; \boldsymbol{\mu}(\mathbf{x}_D), \mathbf{K}(\mathbf{x}_D, \mathbf{x}_D)\right)\right).
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\frac{d^2 \log g(f(\mathbf{x}_n))}{df^2(\mathbf{x}_n)}&\bigg|_{\frac{dg(f(\mathbf{x}_n))}{df(\mathbf{x}_n)}=0} \\
&= \frac{(\mathbb{P}(Y_n|B_n = 0) - \mathbb{P}(Y_n|B_n = 1))(\lambda - f(\mathbf{x}_n))\phi\left(\lambda, f(\mathbf{x}_n), \sigma_W^2\right)}{g^2(f(\mathbf{x}_n))\sigma_W^2} \\
&\quad - \frac{\phi\left(\lambda, f(\mathbf{x}_n), \sigma_W^2\right)(\mathbb{P}(Y_n|B_n = 0) - \mathbb{P}(Y_n|B_n = 1))}{g^2(f(\mathbf{x}_n))}\underbrace{\frac{dg(f(\mathbf{x}_n))}{df(\mathbf{x}_n)}}_{0} \\
&= \frac{(\lambda - f(\mathbf{x}_n))}{g(f(\mathbf{x}_n))\sigma_W^2}\underbrace{\frac{dg(f(\mathbf{x}_n))}{df(\mathbf{x}_n)}}_{0} = 0.
\end{aligned}
$$

component, $p(f(\mathbf{x}_n)|\mathbf{x}_\mathcal{N}, \widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{Y}_\mathcal{N})$, setting its derivative to 0 and solving for f_n. The MAP estimate is given by:

$$\widehat{\mathbf{f}}_\mathrm{n}^{\mathrm{MAP}} = \arg\max_{f_n} p\left(f\left(\mathbf{x}_n\right)|\mathbf{x}_\mathcal{N}, \widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{Y}_\mathcal{N}\right).$$

The log-posterior is given by

$$\log p\left(f\left(\mathbf{x}_n\right)|\mathbf{Y}_\mathcal{N}, \widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)$$
$$\propto \log \mathbb{P}\left(\mathbf{Y}_\mathcal{D}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right) + \log p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right)$$
$$+ \log p\left(f\left(\mathbf{x}_n\right)|\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)$$
$$= \sum_{j=1}^{N_\mathrm{D}} \log \left(\sum_{l=0}^{1} \mathbb{P}\left(Y_j^D|B_j = l\right) \mathbb{P}\left(B_j = l|f_j\right)\right)$$
$$+ \log p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right) + \log p\left(f\left(\mathbf{x}_n\right)|\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)$$
$$= \sum_{j=1, j\neq n}^{N_\mathrm{D}} \log \left(\sum_{l=0}^{1} \mathbb{P}\left(Y_j^D|B_j = l\right) \mathbb{P}\left(B_j = l|f_j\right)\right)$$
$$+ \log \left(\sum_{l=0}^{1} \mathbb{P}\left(Y_n^D|B_n = l\right) \mathbb{P}\left(B_n = l|f_n\right)\right)$$
$$+ \log p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right) + \log p\left(f\left(\mathbf{x}_n\right)|\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)$$
$$\propto \log \left(\sum_{l=0}^{1} \mathbb{P}\left(Y_n^D|B_n = l\right) \mathbb{P}\left(B_n = l|f_n\right)\right)$$
$$+ \log p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right) + \log p\left(f\left(\mathbf{x}_n\right)|\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)$$
$$= \log \left(\underbrace{\mathbb{P}\left(Y_n|B_n = 1\right) + \Phi\left(\lambda, f\left(\mathbf{x}_n\right), \sigma_\mathrm{W}^2\right)}_{\mathrm{T1}a}\right.$$
$$\left.\times \underbrace{\left(\mathbb{P}\left(Y_n|B_n = 0\right) - \mathbb{P}\left(Y_n|B_n = 1\right)\right)}_{\mathrm{T1}b}\right)$$
$$+ \underbrace{\log p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right)}_{\mathrm{T2}} + \underbrace{\log p\left(f\left(\mathbf{x}_n\right)|\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)}_{\mathrm{T3}}.$$

Taking the derivative of $\log(\mathrm{T1}a \times \mathrm{T1}b)$, we obtain:

$$\frac{\mathrm{d}\log\left(\mathrm{T1}a \times \mathrm{T1}b\right)}{\mathrm{d}f\left(\mathbf{x}_n\right)} = \frac{\phi\left(\lambda, f\left(\mathbf{x}_n\right), \sigma_\mathrm{W}^2\right)\mathrm{T1}b}{\mathrm{T1}a \times \mathrm{T1}b}.$$

To calculate the derivative of T2, we express the second term (see (9) and (8)):

$$p(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}) = N\left(\mathbf{Y}_\mathcal{A}; \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}, \left(\sigma_\mathrm{V}^2 + \sigma_\mathrm{W}^2\right)\mathbf{I}_{N_\mathrm{A}} + \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}\right)$$

The derivative is given by:

$$\frac{\partial \log p\left(\mathbf{Y}_\mathcal{A}|\mathbf{f}_\mathcal{D}, \mathbf{x}_\mathcal{N}\right)}{\partial f\left(\mathbf{x}_n\right)} = \left(\mathbf{Y}_\mathcal{A} - \boldsymbol{\mu}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}\right)^T$$
$$\times \left(\left(\sigma_\mathrm{V}^2 + \sigma_\mathrm{W}^2\right)\mathbf{I}_{N_\mathrm{A}} + \boldsymbol{\Sigma}_{\mathbf{f}_\mathcal{A}|\mathbf{f}_\mathcal{D}}\right)^{-1}$$
$$\times K(\mathbf{x}_\mathcal{A}, \mathbf{x}_\mathcal{D}) K^{-1}(\mathbf{x}_\mathcal{D}, \mathbf{x}_\mathcal{D})\mathbb{1}_n.$$

where $\mathbb{1}_n$ is the indicator function for the n-th element. The derivative of T3, which is conditionally Normal, is given by:

$$\frac{\partial \log p\left(f\left(\mathbf{x}_n\right)|\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n}, \mathbf{x}_\mathcal{N}\right)}{\partial f\left(\mathbf{x}_n\right)} = \frac{-\left(f(\mathbf{x}_n) - \mu_{\mathbf{x}_{\mathrm{n}|\mathbf{f}_\mathcal{D}\backslash n}}\right)}{\sigma_{\mathbf{x}_{\mathrm{n}|\mathbf{f}_\mathcal{D}\backslash n}}^2},$$

where

$$\mu_{\mathbf{x}_{\mathrm{n}|\mathbf{f}_\mathcal{D}\backslash n}} := \mu(\mathbf{x}_n) + k\left(\mathbf{x}_n, \mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}\right) K^{-1}\left(\mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}, \mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}\right)$$
$$\times \left(\widehat{\mathbf{f}}_{1:\mathrm{N_D}\backslash n} - \mu\left(\mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}\right)\right),$$
$$\sigma_{\mathbf{x}_{\mathrm{n}|\mathbf{f}_\mathcal{D}\backslash n}}^2 := k\left(\mathbf{x}_n, \mathbf{x}_n\right) - k\left(\mathbf{x}_n, \mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}\right) K^{-1}\left(\mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}, \mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}\right)$$
$$\times k\left(\mathbf{x}_{\mathcal{N}\backslash \mathrm{n}}, \mathbf{x}_n\right).$$

Finally, we set the derivative to 0 to obtain the one-dimensional equation which can be easily solved via a one-dimensional line search.

APPENDIX E
PROOF OF THEOREM 2

$$\mathbb{P}\left(B_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right)$$
$$= \int \mathbb{P}\left(B_*|f_*, \mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) p\left(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \mathrm{d}f_*$$
$$\underset{(6)}{\simeq} \int \mathbb{P}\left(B_*|f_*, \mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \widehat{p}\left(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \mathrm{d}f_*$$
$$= \mathbb{I}_{[B_*=0]} \int \mathbb{P}\left(B_* = 0|f_*, \mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right)$$
$$\times \widehat{p}\left(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \mathrm{d}f_*$$
$$+ \mathbb{I}_{[B_*=1]} \int \mathbb{P}\left(B_* = 1|f_*, \mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right)$$
$$\times \widehat{p}\left(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \mathrm{d}f_*$$
$$= \mathbb{I}_{[B_*=0]} \int \Phi\left(\lambda, f_*, \sigma_\mathrm{W}^2\right) \widehat{p}\left(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \mathrm{d}f_*$$
$$+ \mathbb{I}_{[B_*=1]} \int \left(1 - \Phi\left(\lambda, f_*, \sigma_\mathrm{W}^2\right)\right)$$
$$\times \widehat{p}\left(f_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right) \mathrm{d}f_*$$
$$= \mathbb{I}_{[B_*=0]} \int_{-\infty}^{\lambda} \int N\left(t; f_*, \sigma_\mathrm{W}^2\right) N\left(f_*; \mu_{f_*|\mathbf{Y}_\mathcal{N}}, \sigma_{f_*|\mathbf{Y}_\mathcal{N}}^2\right) \mathrm{d}f_* \mathrm{d}t$$
$$+ \mathbb{I}_{[B_*=1]} \left(1 - \int_{-\infty}^{\lambda} \int N\left(t; f_*, \sigma_\mathrm{W}^2\right)\right.$$
$$\left.\times N\left(f_*; \mu_{f_*|\mathbf{Y}_\mathcal{N}}, \sigma_{f_*|\mathbf{Y}_\mathcal{N}}^2\right) \mathrm{d}f_* \mathrm{d}t\right).$$

Using convolution properties, it is easy to show that

$$\int N\left(t; f_*, \sigma_\mathrm{W}^2\right) N\left(f_*; \mu_{f_*|\mathbf{Y}_\mathcal{N}}, \sigma_{f_*|\mathbf{Y}_\mathcal{N}}^2\right) \mathrm{d}f_*$$
$$= N\left(t; \mu_{f_*|\mathbf{Y}_\mathcal{N}}, \sigma_\mathrm{W}^2 + \sigma_{f_*|\mathbf{Y}_\mathcal{N}}^2\right).$$

Finally, we obtain

$$\mathbb{P}\left(B_*|\mathbf{x}_*, \mathbf{x}_\mathcal{N}, \mathbf{Y}_\mathcal{N}\right)$$
$$= \begin{cases} \Phi\left(\lambda, \mu_{f_*|\mathbf{Y}_\mathcal{N}}, \sigma_\mathrm{W}^2 + \sigma_{f_*|\mathbf{Y}_\mathcal{N}}^2\right), & B_* = 0, \\ 1 - \Phi\left(\lambda, \mu_{f_*|\mathbf{Y}_\mathcal{N}}, \sigma_\mathrm{W}^2 + \sigma_{f_*|\mathbf{Y}_\mathcal{N}}^2\right), & B_* = 1. \end{cases}$$

REFERENCES

[1] J. K. Hart and K. Martinez, "Environmental sensor networks: A revolution in the earth system science?," *Earth-Sci. Rev.*, vol. 78, no. 3, pp. 177–191, 2006.

[2] S. Rajasegarar, T. C. Havens, S. Karunasekera, C. Leckie, J. C. Bezdek, M. Jamriska, A. Gunatilaka, A. Skvortsov, and M. Palaniswami, "High-resolution monitoring of atmospheric pollutants using a system of low-cost sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 3823–3832, 2014.

[3] S. Rajasegarar, P. Zhang, Y. Zhou, S. Karunasekera, C. Leckie, and M. Palaniswami, "High resolution spatio-temporal monitoring of air pollutants using wireless sensor networks," in *Proc. IEEE 9th Int. Conf. Intell. Sensors, Sensor Netw., Inf. Process. (ISSNIP)*, 2014, pp. 1–6.

[4] A. Kottas, Z. Wang, and A. Rodrguez, "Spatial modeling for risk assessment of extreme values from environmental time series: A Bayesian nonparametric approach," *Environmetrics*, vol. 23, no. 8, pp. 649–662, 2012.

[5] J. P. French and S. R. Sain *et al.*, "Spatio-temporal exceedance locations and confidence regions," *Ann. Appl. Statist.*, vol. 7, no. 3, pp. 1421–1449, 2013.

[6] K. Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks: Technology, Protocols, and Applications*. New York, NY, USA: Wiley, 2007.

[7] K. Lorincz, D. J. Malan, T. R. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton, "Sensor networks for emergency response: Challenges and opportunities," *IEEE Pervasive Comput.*, vol. 3, no. 4, pp. 16–23, 2004.

[8] K. Chintalapudi, T. Fu, J. Paek, N. Kothari, S. Rangwala, J. Caffrey, R. Govindan, E. Johnson, and S. Masri, "Monitoring civil structures with a wireless sensor network," *IEEE Internet Comput.*, vol. 10, no. 2, pp. 26–34, 2006.

[9] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[10] F. Fazel, M. Fazel, and M. Stojanovic, "Random access sensor networks: Field reconstruction from incomplete data," in *Proc. IEEE Inf. Theory Appl. Workshop (ITA)*, 2012, pp. 300–305.

[11] J. Matamoros, F. Fabbri, C. Antón-Haro, and D. Dardari, "On the estimation of randomly sampled 2d spatial fields under bandwidth constraints," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4184–4192, 2011.

[12] O. Schabenberger and F. J. Pierce, *Contemporary Statistical Models for the Plant and Soil Sciences*. Boca Raton, FL, USA: CRC Press, 2002.

[13] I. Akyildiz, M. Vuran, and O. Akan, "On exploiting spatial and temporal correlation in wireless sensor networks," *Proc. WiOpt04: Model. Optim. Mobile, Ad Hoc, Wireless Netw.*, pp. 71–80, 2004.

[14] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: Theory and applications for wireless sensor networks," *Comput. Netw. J.*, vol. 45, pp. 245–259, 2004.

[15] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, 2005.

[16] Y. Xia and M. S. Kamel, "Cooperative learning algorithms for data fusion using novel estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1083–1095, 2008.

[17] W.-Y. Chiu and B.-S. Chen, "Multisource prediction under nonlinear dynamics in wsns using a robust fuzzy approach," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 1, pp. 137–149, 2011.

[18] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4519–4528, 2006.

[19] O. Ozdemir, R. Niu, and P. K. Varshney, "Channel aware target localization with quantized data in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1190–1202, 2009.

[20] E. Masazade, R. Niu, P. Varshney, and M. Keskinoz, "Energy aware iterative source localization for wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4824–4835, 2010.

[21] Y. Zheng, R. Niu, and P. Varshney, "Closed-form performance for location estimation based on quantized data in sensor networks," in *Proc. 13th IEEE Conf. Inf. Fusion (FUSION)*, 2010, pp. 1–7.

[22] Y. Zhou, J. Li, and D. Wang, "Posterior Cramér-Rao lower bounds for target tracking in sensor networks with quantized range-only measurements," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 157–160, 2010.

[23] N. Katenka, E. Levina, and G. Michailidis, "Local vote decision fusion for target detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 329–338, 2008.

[24] K. Cohen and A. Leshem, "Energy-efficient detection in wireless sensor networks using likelihood ratio and channel state information," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1671–1683, 2011.

[25] E. Msechu and G. Giannakis, "Sensor-centric data reduction for estimation with WSNS via censoring and quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 400–414, 2012.

[26] P. Zhang, G. Peters, I. Nevat, G. Xiao, and H.-P. Tan, "Distributed event detection in sensor networks under random spatial deployment," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, 2014, pp. 623–629.

[27] I. Nevat, G. W. Peters, and I. B. Collings, "Distributed detection in sensor networks over fading channels with multiple antennas at the fusion centre," *IEEE Trans. Signal Process.*, vol. 62, no. 1–4, pp. 671–683, 2014.

[28] P. Zhang, J. Y. Koh, S. Lin, and I. Nevat, "Distributed event detection under byzantine attack in wireless sensor networks," in *Proc. IEEE 9th Int. Conf. Intell. Sensors, Sensor Netw., Inf. Process. (ISSNIP)*, 2014, pp. 1–6.

[29] Environment Protection Authority Victoria Sensor Locations, 2012 [Online]. Available: http://www.epa.vic.gov.au/air/airmap

[30] T. Watkins, Draft Roadmap for Next Generation Air Monitoring, 2013 [Online]. Available: http://www.epa.gov/airscience/docs/next-generation-air-monitoring-region4.pdf

[31] G. Werner-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, and M. Welsh, "Deploying a wireless sensor network on an active volcano," *IEEE Internet Comput.*, vol. 10, no. 2, pp. 18–25, 2006.

[32] M. Wang, K. D. Knobelspiesse, and C. R. McClain, "Study of the sea-viewing wide field-of-view sensor (SeaWiFS) aerosol optical property data over ocean in combination with the ocean color products," *J. Geophys. Res.: Atmospheres (1984–2012)*, vol. 110, no. D10, 2005.

[33] S. B. Hooker, "An overview of SeaWiFS and ocean color," Nat. Aeronaut. Space Admin., Goddard Space Flight Center, 1992.

[34] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for seawifs," *J. Geophys. Res.: Oceans*, vol. 103, pp. 24 937–24 953, 1998.

[35] K. Baith, R. Lindsay, G. Fu, and C. R. McClain, "Data analysis system developed for ocean color satellite sensors," *Eos, Trans. Amer. Geophys. Union*, vol. 82, no. 18, pp. 202–202, 2001.

[36] P. D. Sampson, A. A. Szpiro, L. Sheppard, J. Lindström, and J. D. Kaufman, "Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data," *Atmos. Environ.*, vol. 45, no. 36, pp. 6593–6606, 2011.

[37] C. Fonseca and H. Ferreira, "Stability and contagion measures for spatial extreme value analyses," 2012, arXiv preprint arXiv:1206.1228 [Online]. Available: http://arxiv.org/abs/1206.1228

[38] I. Nevat, G. W. Peters, and I. B. Collings, "Location-aware cooperative spectrum sensing via Gaussian processes," in *Proc. IEEE Austral. Commun. Theory Workshop (AusCTW)*, 2012, pp. 19–24.

[39] R. P. Haining, R. Kerry, and M. A. Oliver, "Geography, spatial data analysis, and geostatistics: An overview," *Geographic. Anal.*, vol. 42, no. 1, pp. 7–31, 2010.

[40] G. Jun and J. Ghosh, "Spatially adaptive classification of land cover with remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2662–2673, 2011.

[41] H. Daniels, "Saddlepoint approximations in statistics," *Ann. Math. Statist.*, vol. 25, no. 4, pp. 631–650, 1954.

[42] J. Besag, "Statistical analysis of dirty pictures," *J. Appl. Statist.*, vol. 20, no. 5, pp. 63–87, 1993.

[43] G. W. Peters, I. Nevat, Y. Jinhong, and I. B. Collings, "System identification in wireless relay networks via a gaussian process," *IEEE Trans. Veh. Technol.*, vol. 61, no. 9, pp. 3969–3983, 2012.

[44] I. Nevat, G. W. Peters, and I. B. Collings, "Random field reconstruction with quantization in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, pp. 6020–6033, 2013.

[45] D. Gu and H. Hu, "Spatial Gaussian process regression with mobile sensor networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1279–1290, 2012.

[46] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4024–4036, 2009.

[47] S. Park and S. Choi, "Gaussian processes for source separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 1909–1912.

[48] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA, USA: MIT Press, 2005.

[49] R. Adler and J. Taylor, *Random Fields and Geometry*. New York, NY, USA: Springer-Verlag, 2007, vol. 115.

[50] J. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer, 1985.

[51] S. W. Raudenbush, M.-L. Yang, and M. Yosef, "Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation," *J. Comput. Graphic. Statist.*, vol. 9, no. 1, pp. 141–157, 2000.

[52] J. Stewart, *Multivariable Calculus*. Pacific Grove, CA, USA: Brooks/Cole, 2011.

[53] L. Tierney and J. B. Kadane, "Accurate approximations for posterior moments and marginal densities," *J. Amer. Statist. Assoc.*, vol. 81, no. 393, pp. 82–86, 1986.

[54] G. Berz, "Windstorm and storm surges in Europe: Loss trends and possible counter-actions from the viewpoint of an international reinsurer," *Philos. Trans. Roy. Soc. London A, Math., Phys., Eng. Sci.*, vol. 363, no. 1831, pp. 1431–1440, 2005.

[55] R. Flather, J. Smith, J. Richards, C. Bell, and D. Blackman, "Direct estimates of extreme storm surge elevations from a 40-year numerical model simulation and from observations," *Global Atmosph. Ocean Syst.*, vol. 6, no. 2, pp. 165–176, 1998.

[56] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
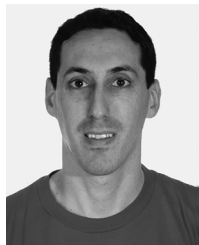
Ido Nevat received the B.Sc. degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel in 1998 and the Ph.D. degree in electrical engieering from the University of NSW, Sydney, Australia in 2010. From 2010–2013 he was a research fellow at the Wireless and Networking Technologies Laboratory at CSIRO, Australia. Currently he is a scientist at the Institute for Infocomm Research (I2R), Singapore. His main areas of interests include statistical signal processing and Bayesian statistical modeling.

Gareth W. Peters received the B.Sc. and B.Eng. (first Hons.) in mathematics and electrical engineering from The University of Melbourne, Victoria, Australia, and the M.Sc. (by research) from the University of Cambridge, Cambridge, U.K., where he worked in the Statistical Signal Processing Laboratory. He is currently working as a lecturer in the Department of Statistical Science, University College London and also holds a position as a lecturer in the Department of Mathematics and Statistics at the University of NSW, Sydney, Australia, where he received his Ph.D. degree in statistics. In addition he holds a visiting Scientist position at the Commonwealth Scientific and Industrial Research Organisation (CSIRO-Australia). His main areas of interest include Bayesian statistical modelling and sampling theory (Markov chain Monte Carlo and sequential Monte Carlo).

François Septier received the Engineer Degree in electrical engineering and signal processing in 2004 from Télécom Lille France, in 2008. From March 2008 to August 2009, he was a Research Associate in the Signal Processing and Communications Laboratory, Cambridge University, Engineering Department, UK. From August 2009, he is an Associate Professor with the Institut Mines-Télécom Lille/CRIStAL UMR CNRS 9189, France. His research focuses on Bayesian computational methodology with a particular emphasis on the development of Monte Carlo based approaches for complex and high-dimensional problems.

Tomoko Matsui (M'91) received the Ph.D. degree from the Computer Science Department, Tokyo Institute of Technology, Tokyo, Japan, in 1997. From 1988 to 2002, she was with NTT, where she worked on speaker and speech recognition. From 1998 to 2002, she was with the Spoken Language Translation Research Laboratory, ATR, Kyoto, Japan, as a Senior Researcher and worked on speech recognition. From January to June 2001, she was an Invited Researcher in the Acoustic and Speech Research Department, Bell Laboratories, Murray Hill, NJ, working on finding effective confidence measures for verifying speech recognition results. She is currently a Professor in the Institute of Statistical Mathematics, Tokyo, working on statistical modeling for speech and speaker recognition applications. Prof. Matsui received the paper award of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICE) in 1993.