

Chapter 2

How to Utilize Sensor Network Data to Efficiently Perform Model Calibration and Spatial Field Reconstruction

Gareth W. Peters, Ido Nevat and Tomoko Matsui

Abstract This chapter provides a tutorial overview of some modern applications of the statistical modeling that can be developed based upon spatial wireless sensor network data. We then develop a range of new results relating to two important problems that arise in spatial field reconstructions from wireless sensor networks. The first new result allows one to accurately and efficiently obtain a spatial field reconstruction which is optimal in the sense that it is the Spatial Best Linear Unbiased Estimator for the field reconstruction. This estimator is obtained under three different system model configurations that represent different types of heterogeneous and homogeneous wireless sensor networks. The second novelty presented in this chapter relates to development of a framework that allows one to incorporate multiple sensed modalities from related spatial processes into the spatial field reconstruction. This is of practical significance for instance, if there are d spatial physical processes that are all being monitored by a wireless sensor network and it is believed that there is a relationship between the variability in the target spatial process to be reconstructed and the other spatial processes being monitored. In such settings it should be beneficial to incorporate these other spatial modalities into the estimation and spatial reconstruction of the target process. In this chapter we develop a spatial covariance regression framework to provide such estimation functionality. In addition, we develop a highly efficient estimation procedure for the model parameters via an Expectation Maximization algorithm. Results of the estimation and spatial field reconstructions are provided for two different real-world applications related to modeling the spatial relationships between coastal wind speeds and ocean height bathymetry measurements based on sensor network observations.

G.W. Peters (✉)

Department of Statistical Science, University College London, London, UK
e-mail: gareth.peters@ucl.ac.uk

I. Nevat

Institute for Infocomm Research, A*STAR, Singapore, Singapore
e-mail: idonevat@gmail.com

T. Matsui

The Institute of Statistical Mathematics, Tokyo, Japan
e-mail: tmatsui@ism.ac.jp

© The Author(s) 2015

G.W. Peters and T. Matsui (eds.), *Modern Methodology and Applications in Spatial-Temporal Modeling*, JSS Research Series in Statistics,
DOI 10.1007/978-4-431-55339-7_2

2.1 Introduction to Wireless Sensor Networks

Wireless sensor networks (WSN) are composed of a large numbers of low-cost, low-power, densely distributed, and possibly heterogeneous sensors. WSN increasingly attract considerable research attention due to the large number of applications, such as environmental monitoring [36], weather forecasts [14, 15, 20, 35, 36], surveillance [39], health care [22], structural safety and building monitoring [9], and home automation [3, 15]. We consider WSN which consist of a set of spatially distributed low-cost sensors that 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, etc.) and regularly communicate their observations to a Fusion Center (FC) in a wireless manner (for example, as in [4, 5, 12, 24, 38, 42]). The FC collects these observations and fuses them in order to reconstruct the signal of interest, based on which effective actions are made [3].

The majority of recent research on WSN consider problems related to addressing estimation of a single point source, such as source localization [23, 31, 32, 46, 47], or source detection (i.e., hypothesis testing) [11, 19, 26] class of problems. In [23, 31, 32, 46], location estimation algorithms of a scalar point source were developed, and in [47] the Posterior Cramér-Rao lower bound (PCRLB) for single target tracking in WSN with quantization was approximated via particle filters. In [11, 19], decision fusion algorithms for a single source detection were developed, and in [26] a vector-valued quantity of a single source was estimated in WSN with censoring and quantization.

In this chapter we explain how one can 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 nonhomogeneous spatial covariance function, depending on the statistical properties of the observed spatial field.

In general the following two fundamental problems naturally arise within this context, and they are the general focus of this chapter:

1. ***Spatial field model calibration and selection***: the task is to determine the best-fitting statistical model for the characterization of the spatial process and to perform the model parameter estimation and then model selection.
2. ***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 a variety of other out-of-sample locations.

We consider in this chapter to model the physical phenomenon being monitored by the WSN according to a Gaussian random field (GRF) with a spatial correlation structure [4, 16, 28, 42]. More generally, examples of GRFs include wireless channels [2], speech processing [33], natural phenomena (temperature, rainfall intensity, etc.) [14, 20], and recently in models developed in [27, 29, 30, 34].

The simplest form of Gaussian process model would typically assume that the spatial field observed is only corrupted by additive Gaussian noise. For example, in [16] a linear regression algorithm for GRF reconstruction in mobile wireless sensor networks was presented, but relied on the assumption of only Additive White Gaussian Noise (AWGN); in [45] an algorithm was developed to learn the parameters of nonstationary spatiotemporal GRFs again assuming AWGN; and in [21] an algorithm for choosing sensor locations in GRF assuming AWGN was developed.

In practical WSN deployments, two deviations from these simplified modeling assumptions arise and can be important in practice to consider: these include the presence of heterogeneous sensor types, i.e., sensors may have different degrees of accuracy throughout the field of spatial monitoring; and secondly quite often the sensors may employ some form of energy conservation such as quantization of analog measurements to digital for efficient and low-power wireless transmission to the FC. To further elaborate these points, one may for instance consider the scenario in which high-quality sensors may be deployed by government agencies (e.g., weather stations). These are sparsely deployed due to their high costs, limited space constraints, high power consumption, etc. Then in order to improve the coverage of the WSN, low-quality cheap sensors perhaps employing quantization may be deployed to augment the higher quality analog sensor network [36]. For instance, battery operated low-cost sensors can be deployed and use simple wireless transmission techniques for data aggregation to the FC [43]. The low-quality sensors considered in this chapter transmit 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.

Hence, the consequence of this type of practical framework is that the observations are heterogeneous and generally non-Gaussian distributed as the quantization procedure introduces a nonlinear transformation of the observations. The sensors transmit their quantized measurements to a FC over wireless channels, which introduce further distortion, due to bandwidth and power constraints. Such practical WSN were considered in [26, 31, 32, 44]. However, these works only considered the estimation of a point source and not of the entire spatial random field, the recent works of [27, 29, 30, 34] extend these frameworks to the entire field reconstruction problem, it is these frameworks that are summarized in this chapter. The intention of the chapter is to highlight and survey recent results that may be obtained for such modeling frameworks.

Notation: 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 $N(x; \mu, \sigma^2) = \phi(x; \mu, \sigma^2)$ 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 $\Phi(\lambda, \mu, \sigma^2) = \int_{-\infty}^{\lambda} \phi(x; \mu, \sigma^2) dx$. We also define $\delta(a, b, c, d) := \phi(a; c, d) - \phi(b; c, d)$ and $\Delta(a, b, c, d) := \Phi(a; c, d) - \Phi(b; c, d)$. We will utilize throughout the chapter the following notations:

- $\mathbf{x}_{\mathcal{N}}$ is the physical location (in terms of $[x, y]$ coordinates) of the N sensors deployed in the field, comprised of N_A analog and N_D digital or quantised sensors such that $N = N_A + N_D$.
- $\mathbf{Y}_{\mathcal{N}} = \{Y_1, \dots, Y_N\} \in \mathbb{R}^{N \times 1}$ 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 N_A analog sensors at the fusion center which are located at points $\mathbf{x}_i \in \mathcal{X}^A$ such that $\text{Card}(\mathcal{X}^A) = N_A$.
- $\mathbf{Y}_{\mathcal{D}} \subseteq \mathbf{Y}_{\mathcal{N}}$ is the collection of observations from all N_D lower quality quantized or digital sensors at the fusion center which are located at points $\mathbf{x}_i \in \mathcal{X}^D$ such that $\text{Card}(\mathcal{X}^D) = N_D$.
- $\mathbf{f}_{\mathcal{N}} = \{f_1, \dots, f_N\} \in \mathbb{R}^{N \times 1}$ is the realization of the random spatial field being monitored $f(\cdot)$ at the sensors located at $\mathbf{x}_{\mathcal{N}}$.
- $\mathbf{f}_{\mathcal{A}} \subseteq \mathbf{f}_{\mathcal{N}}$ is the realization of the random spatial field being monitored $f(\cdot)$ 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 the random spatial field being monitored $f(\cdot)$ at the digital sensors, located at $\mathbf{x}_{\mathcal{D}} \subseteq \mathbf{x}_{\mathcal{N}}$.
- $\mathbf{x}_{\mathcal{N} \setminus n} := [\mathbf{x}_1, \dots, \mathbf{x}_{n-1}, \mathbf{x}_{n+1}, \dots, \mathbf{x}_N]$.

Furthermore, we generically denote a location in space \mathbf{x}_* for which the lower script $*$ indicates that a sensor is not located at this point to make a measurement but for which one wishes to reconstruct the spatial process $f(\mathbf{x}_*) = f_*$.

2.2 Introduction to Spatial Gaussian Random Fields

We consider a generic system model where wireless sensors are deployed in the field. The sensors monitor a spatial physical phenomenon which is observed with measurement error, quantization error, and incomplete sampling of the spatial field. These quantized measurements are transmitted over imperfect wireless channels to the fusion center (FC) to obtain 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.

We assume that the observed phenomenon can be adequately modeled by 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 modeled as a Gaussian random field.¹ 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 2.1 (*Gaussian process* [1, 37]): 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

¹We use Gaussian Process and Gaussian random field interchangeably.

process if all its finite-dimensional distributions are Gaussian, where for any $m \in \mathbb{N}$, the random variables $(f(\mathbf{x}_1), \dots, f(\mathbf{x}_m))$ are normally distributed.

A GP in this chapter is formally defined by the following class of random functions:

$$\begin{aligned}\mathcal{F} &:= \{f(\cdot) : \mathcal{X} \mapsto \mathbb{R} \text{ s.t. } f(\cdot) \sim \mathcal{GP}(\mu(\cdot; \boldsymbol{\Theta}), \mathcal{C}(\cdot, \cdot; \boldsymbol{\Omega}))\}, \text{ with} \\ \mu(\mathbf{x}; \boldsymbol{\Theta}) &:= \mathbb{E}[f(\mathbf{x})] : \mathcal{X} \mapsto \mathbb{R}, \\ \mathcal{C}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\Omega}) &:= \mathbb{E}[(f(\mathbf{x}_i) - \mu(\mathbf{x}_i; \boldsymbol{\Theta}))(f(\mathbf{x}_j) - \mu(\mathbf{x}_j; \boldsymbol{\Theta}))] : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}\},\end{aligned}$$

where at each point the mean of the function is $\mu(\cdot; \boldsymbol{\Theta}) : \mathcal{X} \mapsto \mathbb{R}$, parameterised by $\boldsymbol{\Theta}$, and the spatial dependence between any two points is given by the covariance function (Mercer kernel) $\mathcal{C}(\cdot, \cdot; \boldsymbol{\Omega}) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, parameterised by $\boldsymbol{\Omega}$, see detailed discussion in [37].

It will be useful to make the following notational definitions:

$$\begin{aligned}k(\mathbf{x}_*, \mathbf{x}_{\mathcal{N}}) &:= \mathbb{E}[f(\mathbf{x}_*) f(\mathbf{x}_{\mathcal{N}})] \in \mathbb{R}^{1 \times N} \\ \mathcal{K}(\mathbf{x}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}) &:= \begin{bmatrix} \mathcal{C}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \mathcal{C}(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \mathcal{C}(\mathbf{x}_n, \mathbf{x}_1) & \cdots & \mathcal{C}(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \in \mathcal{S}^+(\mathbb{R}^n),\end{aligned}$$

with $\mathcal{S}^+(\mathbb{R}^n)$ is the manifold of symmetric positive definite matrices. To proceed with an understanding of this class of statistical model we need to consider the choice of the kernel functions available as these will have important implications for the ability of the GP model to capture the variability of the observed process over space.

2.2.1 Model Choices for Spatial Covariance Functions

In this section we discuss a few parametric family of kernels which characterize the covariance function in the Gaussian process. A kernel, also called a covariance function, a kernel function, or a covariance kernel, is a positive definite function of two input vectors, for instance locations in space $\mathbf{x}_i \in \mathbb{R}^2$ and $\mathbf{x}_j \in \mathbb{R}^2$. There are many possible choices of covariance function that one may consider, sometimes the choice is based upon a known physical structure for the spatial processing being monitored, and other times the choice of kernel is obtained based on a statistical model selection procedure. In this section we briefly note some common choices considered in practice and the resulting properties of their implied covariance structure.

In many settings, it may be suitable to make a simplifying assumption such as assuming a spatially isotropic covariance structure in which the spatial covariance kernel may be modeled, for instance via the popular radial basis or the squared exponential function kernel given by

$$\mathbb{C}\text{ov}(f(\mathbf{x}), f(\mathbf{x}')) = C_{\boldsymbol{\Omega}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2l^2}\right), \quad (2.1)$$

for parameter vector $\boldsymbol{\Omega} = (\sigma^2, l)$, with σ the magnitude of the covariance and l defining the characteristic length scale.

The second more flexible family of isotropic covariance function is recommended and used in a variety of application domains, see discussion in [25, 37], the Matern family of Mercer kernels which is characterized by covariance functions given by

$$\mathbb{C}\text{ov}(f(\mathbf{x}), f(\mathbf{x}')) = C_{\boldsymbol{\Omega}}(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{x}'\|_2}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} \|\mathbf{x} - \mathbf{x}'\|_2}{l} \right), \quad (2.2)$$

for $\boldsymbol{\Omega} = (\nu, l)$ with $\nu > 0, l > 0$ and the modified Bessel function given by

$$K_\nu(x) = \int_0^\infty \exp(-x \cosh t) \cosh(\nu t) dt. \quad (2.3)$$

Other possible kernel choices widely used in practice include cases in which there is a periodic structure such as characterized by the kernel,

$$\mathbb{C}\text{ov}(f(\mathbf{x}), f(\mathbf{x}')) = C_{\boldsymbol{\Omega}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{2}{l^2} \sin^2\left(\pi \frac{\mathbf{x} - \mathbf{x}'}{p}\right)\right), \quad (2.4)$$

where $\boldsymbol{\Omega} = (\sigma, l, p)$.

Though these kernels presented above are widely used, it has been argued in [37, 40] that in many problems such restrictive isotropic and smoothness assumptions may not be appropriate for modeling realistic processes, in which case one may resort to an alternative class of covariance kernel which is less restrictive in terms of their spatial symmetries. For instance, one may consider the class of quadrant symmetric kernels that make less restrictive assumptions regarding the isotropy and involves selecting a kernel choice that satisfies the ‘even’ condition for each component given by

$$\mathcal{C}(x_1, \dots, x_k, \dots, x_n) = \mathcal{C}(x_1, \dots, -x_k, \dots, x_n). \quad (2.5)$$

where, quadrant symmetry implies homogeneity in the weak sense, see discussions in [41].

Another class of kernels one may consider is given by the anisotropic family of dot product “regression” kernels in which one considers the basic regression structure $\sigma_0^2 + \mathbf{x}'\mathbf{x}$ and generalizes it with a covariance matrix and positive powers to obtain for strictly positive $\sigma > 0$ an inhomogeneous family. Typically, one considers one of three kernels in this family for the spatial covariance given by,

Linear Kernel:

$$\mathbb{C}\text{ov}(f(\mathbf{x}), f(\mathbf{x}')) = C_{\boldsymbol{\Omega}}(\mathbf{x}, \mathbf{x}') = (\sigma^2 + \mathbf{x}^T \boldsymbol{\Sigma}_1 \mathbf{x}'),$$

Quadratic Kernel:

$$\text{Cov}(f(\mathbf{x}), f(\mathbf{x}')) = \mathcal{C}_{\boldsymbol{\Omega}}(\mathbf{x}, \mathbf{x}') = (\sigma^2 + \mathbf{x}^T \boldsymbol{\Sigma}_2 \mathbf{x}')^2, \quad (2.6)$$

Cubic Kernel:

$$\text{Cov}(f(\mathbf{x}), f(\mathbf{x}')) = \mathcal{C}_{\boldsymbol{\Omega}}(\mathbf{x}, \mathbf{x}') = (\sigma^2 + \mathbf{x}^T \boldsymbol{\Sigma}_3 \mathbf{x}')^3,$$

with $\boldsymbol{\Omega} = (\sigma, \boldsymbol{\Sigma}_i, p_i)$. The linear covariance kernel can also be utilized under an alternative parameterization, it will prove to be beneficial in the context of the estimation developed in the following sections. Under this choice, one assumes that the spatial variability in the process $f(\cdot)$ can be explained by a set of covariates, denoted generically by $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{N^A}]$, where at each of a set of N^A sensors, a set of covariates $\mathbf{w}_i \in \mathbb{R}^{(q \times 1)}$ are observed and provide the following explanatory structure for the spatial processes correlations, given by

$$\text{Cov}[f(\mathbf{x}_i), f(\mathbf{x}_j) | \mathbf{w}] = \mathcal{C}_{\boldsymbol{\Omega}}(\mathbf{x}_i, \mathbf{x}_j) = (\sigma^2 + \mathbf{b}_i^T \mathbf{w} \mathbf{w}^T \mathbf{b}_j),$$

where the parameters of association of the spatial field for each local point are defined by vectors $\mathbf{b}_i \in \mathbb{R}^{(q \times 1)}$ for $i \in \{1, \dots, N^A\}$.

Having formally specified the semi-parametric class of Gaussian process models, we proceed with presenting the system model.

2.3 Wireless Sensor Network System Model

We now present the WSN system with practical quantization and imperfect wireless channels:

1. Consider a random spatial phenomenon to be monitored 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 modeled as a Gaussian Process (GP) according to

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

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

$$Z(\mathbf{x}_n) = f(\mathbf{x}_n) + V_n^s, \quad n = \{1, \dots, N\}$$

where V_n^s are i.i.d. Gaussian noise terms, i.e., $V_n^s \sim N(0, \sigma_s^2)$.

4. **Analog (high quality) sensors processing and communication model:** each of the analog high-quality N_A sensors transmits its noisy observation to the FC over

AWGN channels, as follows:

$$Y_n^A = Z(\mathbf{x}_n) + V_n^A, \quad n = \{1, \dots, N_A\},$$

where V_n^A is i.i.d. Gaussian noise $V_n^A \sim N(0, \sigma_A^2)$.

5. **Digital quantized (lower quality) sensors processing:** each of the N_D digital sensors first performs a thresholding-based decision based on its noisy observations. This step is summarized as follows for two common settings: first for the case in which an L -bit quantizer is assumed to operate at the sensor; second for the case in which a simple binary thresholding decision is performed.

Setting 1—Low-Quality Power/Bandwidth Constrained Sensors:

the quantizer explicitly maps its input $Z_n = Z(\mathbf{x}_n)$ to the output B_n through a mapping or encoder $B_n : \mathbb{R} \mapsto \{0, \dots, L-1\}$, as follows:

$$B_n = \mathcal{Q}[Z_n] := \begin{cases} 0, & \lambda_0 \leq Z_n < \lambda_1 \\ 1, & \lambda_1 \leq Z_n < \lambda_2 \\ \vdots & \vdots \\ L-2, & \lambda_{L-2} \leq Z_n < \lambda_{L-1} \\ L-1, & \lambda_{L-1} \leq Z_n < \lambda_L, \end{cases}$$

where $\lambda_0 = -\infty$ and $\lambda_L = \infty$.

Setting 2—Basic Thresholding:

the sensor simply thresholds via a binary decision rule (a special case of the $L = 1$ quantizer), with the binary decision rule given by:

$$B_n = \begin{cases} 1, & Z(\mathbf{x}_n) > \lambda \\ 0, & Z(\mathbf{x}_n) \leq \lambda. \end{cases} \quad (2.7)$$

where λ is a predefined threshold. We denote the thresholding operation by $\mathcal{Q}[\cdot]$.

6. **Digital quantized (lower quality) sensors communication model:** each of the N_D digital sensors, having first performed the quantization or thresholding-based decision on its noisy observations, then transmits the L -bit decision over imperfect wireless channels [11, 19, 26]. The decision $B_n = B(\mathbf{x}_n)$ is transmitted to the FC over imperfect binary wireless channels, as in [32]. Under this model, the statistic B_n is transmitted to the FC over imperfect wireless channels for which the conditional probability mass function (PMF) of the quantized/encoded observation from the n th sensor can be expressed, for all $m \in \{0, \dots, L-1\}$, as:

$$\mathbb{P}(Y_n = m | f(\mathbf{x}_n)) = \sum_{l=0}^{L-1} \mathbb{P}(Y_n = m | B_n = l) \mathbb{P}(B_n = l | f(\mathbf{x}_n)),$$

where $\mathbb{P}(Y_n = m | B_n = l)$ represents the channels statistics (e.g., probability of making an error).

7. **Additional modalities sensed by high-quality analog sensors:** it is assumed that for the N_A analog sensors they are capable of making observations of additional spatial covariates, related to the physical process being monitored. At the n th analog sensor location the vector of additional spatial covariates is denoted $\tilde{\mathbf{W}}_n \in \mathbb{R}^{(q \times 1)}$. The analog sensor then transmits this vector of additional covariates to the FC over AWGN channels, as follows:

$$\mathbf{W}_n = \tilde{\mathbf{W}}_n + \mathbf{V}_n^C, \quad n = \{1, \dots, N_A\},$$

where \mathbf{V}_n^C is i.i.d Gaussian noise $V_n^C \sim N(\mathbf{0}, \Sigma)$. In the remainder of this chapter we consider to stack all the N_A sensor covariates into a matrix $\mathbf{W} = [\mathbf{W}_1, \dots, \mathbf{W}_{N_A}]$ for which we denote the realization by the matrix $\mathbf{w} \in \mathbb{R}^{(q \times N_A)}$.

2.3.1 Homogeneous and Heterogeneous WSNs

Hence, having specified this system model, we now consider two classes of WSN, the first will be termed the “**homogeneous sensor networks**” in which we assume each sensor performs processing of the sensed observed spatial phenomenon via the L -bit quantization before transmission to the FC for spatial field reconstruction. We note that in the ideal case $L \rightarrow \infty$ one would obtain from such a network the optimal estimation, in the sense of information content in the reconstruction of the spatial field. The second class of WSN we consider is the “**heterogeneous sensor networks**” in which a subset of sensors have capability, wireless transmission bandwidth and battery power, to transmit unquantized observations to the FC, whilst the remainder of cheaper sensors are bandwidth constrained, battery constrained, or inaccurate enough to only transmit L -bit quantized observations to the FC. In practice these lower quality sensors typically may even be simple binary quantizations of the analog sensed signal, in such cases one has $L = 1$ binary thresholding of the observed spatial field.

It is also not unreasonable to assume that the higher quality sensors which are not battery or bandwidth constrained may have additional capabilities to also observe or sense other spatial attributes in the monitoring environment. For instance, one may be interested in monitoring wind speed as the primary target spatial process, however, these higher quality sensors may also monitor other potentially related spatial physical attributes such as barometric pressure, temperature, humidity, and bathymetry. In general these other processes being monitored will be termed alternative modalities, and these modalities can often be very informative of the spatial structure and dynamics of the target physical process that one wishes to reconstruct the spatial field for based on the sensor observations. For this reason, we demonstrate next how to incor-

porate such other sensed modality information into the covariance structure of the target spatial process as part of a specialized form of spatial covariance regression.

In the case of the heterogeneous WSN model, we assumed that one may wish to incorporate alternative sensed modalities, i.e., exogenous spatial covariates which are observed jointly at the analog (high quality) sensors. This can be achieved in two standard ways in the regression model, through the trend (mean of the GP) or through the volatility in the sensed spatial process model (covariance function of the GP). In this manuscript we focus on incorporation of the spatial covariates into explaining helping to explain the spatial variability of the target spatial process with respect to both spatial structure as well as variability in these other sensed modalities. This creates a powerful class of models that has both spatial features and explanatory power derived from incorporation of related local spatial processes that should improve the accuracy of spatial reconstructions.

In this case one may consider to develop a spatial covariance kernel comprised of the following structure for any two locations \mathbf{x}_i and \mathbf{x}_j :

$$\begin{aligned}\tilde{C}(\mathbf{x}_i, \mathbf{x}_j) &= \mathbb{E}[f(\mathbf{x}_i) f(\mathbf{x}_j)] \\ &= C_{\Omega}(\mathbf{x}_i, \mathbf{x}_j) + (\zeta_i^2 + \mathbf{b}_i^T \mathbf{w} \mathbf{w}^T \mathbf{b}_i) \mathbb{I}[\mathbf{x}_i \in \mathcal{X}^A, \mathbf{x}_j \notin \mathcal{X}^A] \\ &\quad + (\zeta_j^2 + \mathbf{b}_j^T \mathbf{w} \mathbf{w}^T \mathbf{b}_j) \mathbb{I}[\mathbf{x}_i \notin \mathcal{X}^A, \mathbf{x}_j \in \mathcal{X}^A] \\ &\quad + (\zeta_{i,j}^2 + \mathbf{b}_{ij}^T \mathbf{w} \mathbf{w}^T \mathbf{b}_{ij}) \mathbb{I}[\mathbf{x}_i \in \mathcal{X}^A, \mathbf{x}_j \in \mathcal{X}^A]\end{aligned}\tag{2.8}$$

where we denote the set of locations of the analog sensors in the WSN by the subset $\mathcal{X}^A \subseteq \mathcal{X}$. In this structure the first functional form $C_{\Omega}(\mathbf{x}_i, \mathbf{x}_j)$ represents the parameterization, via a kernel, for the spatial dependence of the target spatial field. The remaining three terms correspond to incorporated information in the spatial covariance regression structures arising from realizations of the additional modalities characterized by vector \mathbf{w} which are only available at the analog sensor locations. This is quite a generic structure since many possible choices may be made for what would go into \mathbf{w} .

The validity of construction of the spatial kernel in this manner utilizes the fact that in general the linear combination of two kernels given by

$$k_{12}(\mathbf{x}_i, \mathbf{x}_j) = c_1 k_1(\mathbf{x}_i, \mathbf{x}_j) + c_2 k_2(\mathbf{x}_i, \mathbf{x}_j)\tag{2.9}$$

is a valid Mercer kernel and will construct a covariance matrix which will be symmetric and positive definite so long as $c_1, c_2 > 0$ and kernels k_1 and k_2 are Mercer kernels.

The construction of the covariance kernel in this manner admits two different types of interpretation of the resulting spatial model. The first is based on a linear combination of two GPs, the second is based on a hybrid model which involves a linear combination of a GP and a Gaussian graphical model (GMM) of [18]. In the remainder of this chapter we adopt the first approach.

When we interpret the spatial process model as a linear combination of two Gaussian processes, then this would be like thinking that theoretically the sensed additional modalities being utilized as covariates, which can be observed over the entire spatial domain and that they have a smooth functional relationship spatially. In this case the resulting GP model would be:

$$f(\cdot) = h(\cdot) + g(\cdot) \sim \mathcal{GP}(\mu(\cdot; \boldsymbol{\theta}_h) + \mu(\cdot; \boldsymbol{\theta}_g), \mathcal{C}_h(\cdot, \cdot) + \mathcal{C}_g(\cdot, \cdot)). \quad (2.10)$$

Here, we associate $\mathcal{C}_h(\cdot, \cdot)$ to $\mathcal{C}_{\boldsymbol{\Omega}}(\cdot, \cdot)$ and we interpret the Gaussian process $h(\cdot)$ as the baseline spatial process model and we associate $\mathcal{C}_g(\cdot, \cdot)$ with the additional spatial covariate terms from the additionally sensed modalities giving the spatial covariance function for the secondary, independent spatial Gaussian process $g(\cdot)$ given by manipulating (2.8) as follows:

$$\mathcal{C}_g(\cdot, \cdot) := \tilde{\mathcal{C}}(\mathbf{x}_i, \mathbf{x}_j) - \mathcal{C}_{\boldsymbol{\Omega}}(\mathbf{x}_i, \mathbf{x}_j) \quad (2.11)$$

We note that if it is not suitable to make a smooth spatial relationship (potentially nonstationary in space) for the additional covariates variability in space, then in this case it would be more beneficial to think of the resulting model under the second interpretation of a hybrid GP and GGM model.

2.4 Model Calibration for WSN Spatial Models

For the different classes of WSN system models developed above we will require the ability to evaluate the spatial cross-correlation between observations of the target spatial process. This will be useful for both calibration purposes as well as spatial field estimation purposes.

Hence, we first consider the covariance matrix of the spatially distributed observations, given by $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}}[\mathbf{Y}_{\mathcal{N}}\mathbf{Y}_{\mathcal{N}}^T]$. The expression for the individual covariance terms in the covariance matrix need to be considered under one of three possible cases:

- Case 1 with $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^A$, i.e., both sensors are high-quality analog sensors;
- Case 2 with $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^D$, i.e., one sensor is analog and one sensor is a cheaper quantized sensor; and
- Case 3 in which $\mathbf{x}_i \in \mathcal{X}^D$ and $\mathbf{x}_j \in \mathcal{X}^D$. The resulting covariance matrix results for the (i, j) th components are specified in Theorem 2.1.

Theorem 2.1 (Covariance between Spatial Observations) *The (i, j) th term of $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}}[\mathbf{Y}_{\mathcal{N}}\mathbf{Y}_{\mathcal{N}}^T]$ is given by one of the following three cases where we define throughout the notation*

$$\begin{aligned}
c_2 &:= \frac{\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j)}{\mathcal{C}(\mathbf{x}_j, \mathbf{x}_j)}, \\
c_1 &:= \mu(\mathbf{x}_i) - c_2 \mu(\mathbf{x}_j), \\
G_1(a, b; m, s) &:= \{\Phi(a; m, s) - \Phi(b; m, s)\} \\
G_2(a, b; m, s) &:= \{\phi(a; m, s) - \phi(b; m, s)\}.
\end{aligned}$$

Case 1: $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^A$

In this case one has two high-quality analog sensors resulting in the cross-correlation given by

$$\begin{aligned}
\mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \mathbb{E}_{f_i, f_j} [\mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j]] = \mathbb{E}_{f_i, f_j} [f_i f_j] \\
&= C_g(\mathbf{x}_i, \mathbf{x}_j) + C_h(\mathbf{x}_i, \mathbf{x}_j).
\end{aligned}$$

Case 2: $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^D$

In this case one has a high-quality analog sensor and a lower quality L -bit quantized sensor resulting in the cross-correlation given by

$$\begin{aligned}
\mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \mathbb{E}_{f_i, f_j} [\mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j]] \\
&= \mathbb{E}_{f_j} \left[(c_1 + c_2 f_j) \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \Pr(Y_j = l | B_j = k) \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \right] \\
&= c_1 \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \Pr(Y_j = l | B_j = k) \mathbb{E}_{f_j} \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \\
&\quad + c_2 \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \Pr(Y_j = l | B_j = k) \mathbb{E}_{f_j} \left[f_j G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right],
\end{aligned} \tag{2.12}$$

where we obtain for the first integral the closed form expression

$$\begin{aligned}
&\mathbb{E}_{f_j} \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \\
&= \frac{1}{\sigma_{f_j}^2} \left[\Phi \left(\frac{\mu_{f_j} - \lambda_{k+1}}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) - \Phi \left(\frac{\mu_{f_j} - \lambda_k}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) \right].
\end{aligned} \tag{2.13}$$

and the second integral the closed form expression

$$\begin{aligned}
&\mathbb{E}_{f_j} \left[f_j G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \\
&= \left(\frac{\sigma_{f_j}^2 + \mu_{f_j}}{\sigma_A^2 \sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) \left[\Phi \left(\frac{\mu_{f_j} - \lambda_{k+1}}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) - \Phi \left(\frac{\mu_{f_j} - \lambda_k}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) \right].
\end{aligned} \tag{2.14}$$

Case 3: $\mathbf{x}_i \in \mathcal{X}^D$ and $\mathbf{x}_j \in \mathcal{X}^D$

In this case one has two lower quality L -bit quantized sensors resulting in the cross-correlation given by

$$\begin{aligned} \mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \mathbb{P}(Y_i = k | B_i = m) \mathbb{P}(Y_j = l | B_j = n) \\ &\times \mathbb{E}_{f_i, f_j} \left[G_1(\lambda_{m+1}, \lambda_m; f_i, \sigma_A^2) G_1(\lambda_{n+1}, \lambda_n; f_j, \sigma_A^2) \right]. \end{aligned} \quad (2.15)$$

Note: the approximation of the expectations in case 3 will be provided in detail in Sect. 2.5.1.2 where an efficient specialized form of quadrature rule will be developed based on the discrete cosine transform, known as the Clenshaw–Curtis quadrature rule.

2.4.1 Proof of Theorem 2.1

Using the law of total expectation, the (i, j) th term of $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T]$ is expressed as $\mathbb{E}_{Y_i, Y_j} [Y_i Y_j] = \mathbb{E}_{f_i, f_j} [\mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j]]$. Deriving this quantity for Case 1 is trivial, so we focus on Case 2 and Case 3 below.

Case 2: $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^D$

In this case one has a high-quality analog sensor and a lower quality L -bit quantized sensor resulting in the cross-correlation given by

$$\begin{aligned} \mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \mathbb{E}_{f_i, f_j} [\mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j]] \\ &= \mathbb{E}_{f_i, f_j} \left[\int \sum_{l=0}^{L-1} y_j l \mathbb{P}(Y_j = l | f_j) f_{Y_i}(y_i | f_i) dy_i \right] \\ &= \mathbb{E}_{f_i, f_j} \left[\int y_i \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \mathbb{P}(Y_j = l | B_j = k) \mathbb{P}(B_j = k | f_j) f_{Y_i}(y_i | f_i) dy_i \right] \\ &= \mathbb{E}_{f_i, f_j} \left[f_i \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \mathbb{P}(Y_j = l | B_j = k) \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \right] \\ &= \mathbb{E}_{f_j} \left[\int f_i f_{f_i | f_j}(f_i) df_i \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \mathbb{P}(Y_j = l | B_j = k) \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \right] \\ &= \mathbb{E}_{f_j} \left[(c_1 + c_2 f_j) \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \mathbb{P}(Y_j = l | B_j = k) \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \right] \end{aligned}$$

We may now work out these integrals for case 2 given generically by

$$\begin{aligned}
\mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \mathbb{E}_{f_i, f_j} [\mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j]] \\
&= \mathbb{E}_{f_j} \left[(c_1 + c_2 f_j) \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \Pr(Y_j = l | B_j = k) \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \right] \\
&= c_1 \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \Pr(Y_j = l | B_j = k) \mathbb{E}_{f_j} \left[G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right] \\
&\quad + c_2 \sum_{l=0}^{L-1} \sum_{k=0}^{L-1} l \Pr(Y_j = l | B_j = k) \mathbb{E}_{f_j} \left[f_j G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2) \right],
\end{aligned} \tag{2.16}$$

and we need to evaluate the two expectations given by $\mathbb{E}_{f_j} [G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2)]$ and $\mathbb{E}_{f_j} [f_j G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2)]$. We start by considering the first integral

$$\begin{aligned}
&\mathbb{E}_{f_j} [G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2)] \\
&= \mathbb{E}_{f_j} \left[\Phi(\lambda_{k+1}; f_j, \sigma_A^2) - \Phi(\lambda_k; f_j, \sigma_A^2) \right] \\
&= \mathbb{E}_{f_j} \left[\Phi(f_j; \lambda_{k+1}, \sigma_A^2) - \Phi(f_j; \lambda_k, \sigma_A^2) \right] \\
&= \mathbb{E}_{f_j} \left[\Phi\left(\frac{f_j - \lambda_{k+1}}{\sigma_A^2}\right) - \Phi\left(\frac{f_j - \lambda_k}{\sigma_A^2}\right) \right] \\
&= \int_{-\infty}^{\infty} \left\{ \phi\left(\frac{f_j - \mu_{f_j}}{\sigma_{f_j}^2}\right) \Phi\left(\frac{f_j - \lambda_{k+1}}{\sigma_A^2}\right) - \phi\left(\frac{f_j - \mu_{f_j}}{\sigma_{f_j}^2}\right) \Phi\left(\frac{f_j - \lambda_k}{\sigma_A^2}\right) \right\} df_j.
\end{aligned} \tag{2.17}$$

Now denote $x = \frac{f_j - \mu_{f_j}}{\sigma_{f_j}}$ with $dx = \frac{1}{\sigma_{f_j}} df_j$ and

$$\begin{aligned}
&\mathbb{E}_{f_j} [G_1(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2)] \\
&= \frac{1}{\sigma_{f_j}^2} \int_{-\infty}^{\infty} \left\{ \phi(x) \Phi\left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_{k+1}}{\sigma_A^2}\right) - \phi(x) \Phi\left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_k}{\sigma_A^2}\right) \right\} dx.
\end{aligned} \tag{2.18}$$

Now we can use the identity given by

$$\int_{-\infty}^{\infty} \phi(x) \Phi(a + bx) dx = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right), \tag{2.19}$$

to obtain for the first expectation

$$\begin{aligned}
& \mathbb{E}_{f_j} \left[G_1 \left(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2 \right) \right] \\
&= \frac{1}{\sigma_{f_j}^2} \int_{-\infty}^{\infty} \left\{ \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_{k+1}}{\sigma_A^2} \right) - \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_k}{\sigma_A^2} \right) \right\} dx \\
&= \frac{1}{\sigma_{f_j}^2} \left[\Phi \left(\frac{\mu_{f_j} - \lambda_{k+1}}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) - \Phi \left(\frac{\mu_{f_j} - \lambda_k}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) \right].
\end{aligned} \tag{2.20}$$

Now we consider the second integral $\mathbb{E}_{f_j} [f_j G_1 (\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2)]$ which can be rewritten as

$$\begin{aligned}
& \mathbb{E}_{f_j} \left[f_j G_1 \left(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2 \right) \right] \\
&= \int_{-\infty}^{\infty} f_j \left\{ \phi \left(\frac{f_j - \mu_{f_j}}{\sigma_{f_j}^2} \right) \Phi \left(\frac{f_j - \lambda_{k+1}}{\sigma_A^2} \right) - \phi \left(\frac{f_j - \mu_{f_j}}{\sigma_{f_j}^2} \right) \Phi \left(\frac{f_j - \lambda_k}{\sigma_A^2} \right) \right\} df_j.
\end{aligned} \tag{2.21}$$

Now denote $x = \frac{f_j - \mu_{f_j}}{\sigma_{f_j}^2}$ with $dx = \frac{1}{\sigma_{f_j}^2} df_j$ and

$$\begin{aligned}
& \mathbb{E}_{f_j} \left[f_j G_1 \left(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2 \right) \right] \\
&= \frac{1}{\sigma_{f_j}^2} \int_{-\infty}^{\infty} (\sigma_{f_j}^2 x + \mu_{f_j}) \left\{ \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_{k+1}}{\sigma_A^2} \right) - \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_k}{\sigma_A^2} \right) \right\} dx \\
&= \int_{-\infty}^{\infty} x \left\{ \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_{k+1}}{\sigma_A^2} \right) - \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_k}{\sigma_A^2} \right) \right\} dx \\
&\quad + \frac{\mu_{f_j}}{\sigma_{f_j}^2} \int_{-\infty}^{\infty} \left\{ \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_{k+1}}{\sigma_A^2} \right) - \phi(x) \Phi \left(\frac{\sigma_{f_j}^2 x + \mu_{f_j} - \lambda_k}{\sigma_A^2} \right) \right\} dx.
\end{aligned} \tag{2.22}$$

Now we utilize the identity in Eq. 2.19 and the following additional identity

$$\int_{-\infty}^{\infty} x \phi(x) \Phi(a + bx) dx = \frac{b}{\sqrt{1 + b^2}} \phi \left(\frac{a}{\sqrt{1 + b^2}} \right), \tag{2.23}$$

to obtain the result

$$\begin{aligned}
& \mathbb{E}_{f_j} \left[f_j G_1 \left(\lambda_{k+1}, \lambda_k; f_j, \sigma_A^2 \right) \right] \\
&= \left(\frac{\sigma_{f_j}^2 + \mu_{f_j}}{\sigma_A^2 \sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) \left[\Phi \left(\frac{\mu_{f_j} - \lambda_{k+1}}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) - \Phi \left(\frac{\mu_{f_j} - \lambda_k}{\sqrt{1 + \frac{\sigma_{f_j}^2}{\sigma_A^2}}} \right) \right].
\end{aligned} \tag{2.24}$$

Case 3: $\mathbf{x}_i \in \mathcal{X}^D$ and $\mathbf{x}_j \in \mathcal{X}^D$

The conditional expectation, $\mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j]$, can be expressed as:

$$\begin{aligned} \mathbb{E}_{Y_i, Y_j} [Y_i Y_j | f_i, f_j] &= \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \mathbb{P}(Y_i = k, Y_j = l | f_i, f_j) \\ &= \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \sum_{m=0}^{L-1} \mathbb{P}(Y_i = k | B_i = m) G_1(\lambda_{m+1}, \lambda_m; f_i, \sigma_A^2) \\ &\quad \times \sum_{n=0}^{L-1} \mathbb{P}(Y_j = l | B_j = n) G_1(\lambda_{n+1}, \lambda_n; f_j, \sigma_A^2). \end{aligned}$$

Next we derive the unconditional expectation of $\mathbb{E}_{f_i, f_j} [Y_i Y_j | f_i, f_j]$:

$$\begin{aligned} \mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \mathbb{E}_{f_i, f_j} \left[\sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \sum_{m=0}^{L-1} \left(\mathbb{P}(Y_i = k | B_i = m) G_1(\lambda_{m+1}, \lambda_m; f_i, \sigma_A^2) \right) \right. \\ &\quad \left. \times \sum_{n=0}^{L-1} \left(\mathbb{P}(Y_j = l | B_j = n) G_1(\lambda_{n+1}, \lambda_n; f_j, \sigma_A^2) \right) \right] \\ &= \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \mathbb{P}(Y_i = k | B_i = m) \mathbb{P}(Y_j = l | B_j = n) \\ &\quad \times \mathbb{E}_{f_i, f_j} \left[G_1(\lambda_{m+1}, \lambda_m; f_i, \sigma_A^2) G_1(\lambda_{n+1}, \lambda_n; f_j, \sigma_A^2) \right]. \end{aligned} \quad \blacksquare$$

Having derived the spatial cross-correlation between the observations, our next goal is to consider model calibrations. In the context of the spatial WSN models developed this will correspond to addressing the issue of parameter estimation, in particular hyperparameter estimation of the parameters in the covariance kernel functions given the observed data. To achieve this we will consider calibration based on the high-quality sensor information, given in Case 1.

To achieve the model calibration for the kernel parameters in an efficient manner we will develop a special representation of the problem in the form of a regression model through the introduction of an additional auxiliary variable for each observation, i.e., per sensor location. In doing this it will allow us to avoid directly trying to perform maximum likelihood estimation in the models, which can be very difficult, especially when it comes to the matrices of parameters given by each \mathbf{b}_i for each analog sensor location. Instead, through the use of auxiliary variables we may write a random effects regression model, which preserves the conditional covariance structure developed above, whilst admitting an efficient estimation procedure comprised of simple expectation and maximization stages of the EM algorithm. In the models considered we will see that the expectation stage is closed form and analytic and the maximization stage is simply a least squares problem after a change of parameterization. Making estimation both guaranteed to converge to a maxima and highly computationally efficient.

2.4.2 Random Effects WSN Spatial Model Reinterpretation

We begin with the scenario in which the majority of the sensor are analog, i.e., the spatial distribution of such high-quality sensors is distributed in some manner over the entire field of interest when performing spatial field reconstruction. These high-quality sensors can be sparse and will still be supplemented by cheaper sensors as discussed above, however, in this stage we will concentrate on the calibration of the model based solely on the high-quality analog sensors.

The advantage of this approach is that we will be able to utilize an interesting result for the estimation of the model parameters which is based on results known for covariance regressions, see [17]. Consider the observation covariance at the analog sensors given in this case by

$$\begin{aligned}
 \text{Cov}[\mathbf{Y}_{\mathcal{N}}|\mathbf{w}] &= \mathbb{E}_{\mathbf{Y}_{\mathcal{N}}}[\mathbf{Y}_{\mathcal{N}}\mathbf{Y}_{\mathcal{N}}^T|\mathbf{w}] \\
 &= \mathbb{E}_{\mathbf{h}}[\mathbf{h}\mathbf{h}^T] + \mathbb{E}_{\mathbf{g}}[\mathbf{g}\mathbf{g}^T|\mathbf{w}] + \text{diag}(\sigma_A^2, \sigma_A^2, \dots, \sigma_A^2) \\
 &= \begin{bmatrix} C_h(\mathbf{x}_1, \mathbf{x}_1) & \cdots & C_h(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ C_h(\mathbf{x}_n, \mathbf{x}_1) & \cdots & C_h(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} + \begin{bmatrix} C_g(\mathbf{x}_1, \mathbf{x}_1) & \cdots & C_g(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ C_g(\mathbf{x}_n, \mathbf{x}_1) & \cdots & C_g(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} + \begin{bmatrix} \sigma_A^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_A^2 \end{bmatrix} \\
 &= K_h + B\mathbf{w}\mathbf{w}^T B^T + \text{diag}(\sigma_A^2, \sigma_A^2, \dots, \sigma_A^2).
 \end{aligned} \tag{2.25}$$

We note that under our model formulations, typically we would select ζ_i^2 , ζ_j^2 and $\zeta_{i,j}^2$ all to zero, since we already have a baseline covariance function given by the independent spatial GP $h(\cdot)$.

We may now reinterpret the model covariance as a form of covariance regression which admits a representation as a random effects model, making it an extension of the framework proposed in [17]. The random effects representation is given for m realizations of the spatial process, i.e., $\mathbf{y}_1, \dots, \mathbf{y}_m$ with $\mathbf{y}_k = \mathbf{y}_{1:N^A, k} = [y_k(\mathbf{x}_1), \dots, y_k(\mathbf{x}_{N^A})]$ and $\boldsymbol{\mu}_g = \boldsymbol{\mu}_{g, 1:N^A} = [\mu_g(\mathbf{x}_1), \dots, \mu_g(\mathbf{x}_{N^A})]$ is the spatial mean function of the first baseline spatial GP $g(\cdot)$, for each of the analog sensor locations $\mathbf{x}_i \in \mathcal{X}^A$. This then gives the random effect model given by

$$\mathbf{Y}_k = \boldsymbol{\mu}_g + B\mathbf{w}_k\Gamma_k + \mathbf{U}_k, \tag{2.26}$$

where one defines

$$\begin{aligned}
 \mathbb{E}[\mathbf{U}_i] &= \mathbf{0}, \quad \mathbb{E}[\Gamma_i \mathbf{U}_i] = \mathbf{0}, \quad \mathbb{E}[\Gamma_i] = 0, \quad \mathbb{V}\text{ar}[\Gamma_i] = 1, \\
 \text{Cov}[\mathbf{U}_i] &= K_h + \text{diag}(\sigma_A^2, \sigma_A^2, \dots, \sigma_A^2).
 \end{aligned} \tag{2.27}$$

To see that this random effects formulation of the spatial model indeed produces the correct spatial covariance structure we consider the following:

$$\begin{aligned}
& \mathbb{E} \left[(Y_k - \mu_g) (Y_k - \mu_g)^T \right] \\
&= \mathbb{E} \left[\gamma_k^2 B \mathbf{w} \mathbf{w}^T B^T + \gamma_k (B \mathbf{w} \mathbf{u}_k^T + \mathbf{u}_k \mathbf{w}^T B^T) + \mathbf{u}_k \mathbf{u}_k^T \right] \\
&= B \mathbf{w} \mathbf{w}^T B^T + K_h + \text{diag} (\sigma_A^2, \sigma_A^2, \dots, \sigma_A^2)
\end{aligned} \tag{2.28}$$

2.4.3 Random Effects WSN Spatial Model Estimation via EM Algorithm

We can now perform the estimation of the spatial field using the information from the high-quality analog sensors to make estimation via an EM algorithm using the reinterpreted random effects model form from Sect. 2.4.2. To achieve this, we make the following additional statistical assumptions regarding the random effects reinterpretation, in particular we assume that the regression errors are Gaussian random vectors, independent of the Gaussian random variables for the random effect:

$$\begin{aligned}
\mathbf{u}_k &\stackrel{iid}{\sim} N(\mathbf{0}, A), \forall k \in \{1, \dots, m\} \\
\Gamma_k &\stackrel{iid}{\sim} N(0, 1),
\end{aligned} \tag{2.29}$$

with $A := K_h + \text{diag} (\sigma_A^2, \sigma_A^2, \dots, \sigma_A^2)$.

The resulting log-likelihood of the random effects model can be rewritten by subtracting the mean from the observations to obtain the matrix of mean adjusted residuals, given by $E = (\mathbf{e}_1^T, \dots, \mathbf{e}_m^T)^T$, with residual vectors for the k th spatial map observation given by $\mathbf{e}_k = [Y_k - \hat{\mu}_g]$. This results in the following log-likelihood for the model parameter matrices A and B , given the observation matrix of residuals E and covariate matrix W from the other sensed modalities, producing for a constant c the log-likelihood:

$$\begin{aligned}
l(A, B; E, W) &= c - \frac{1}{2} \sum_{k=1}^m \log |A + B \mathbf{w}_k \mathbf{w}_k^T B^T| \\
&\quad - \frac{1}{2} \sum_{k=1}^m \text{tr} \left[(A + B \mathbf{w}_k \mathbf{w}_k^T B^T)^{-1} \mathbf{e}_k \mathbf{e}_k^T \right]
\end{aligned} \tag{2.30}$$

It is clear that direct maximization of this log-likelihood with respect to the matrices A and B will be a very challenging non-convex optimization problem. This arises since the matrix A must be optimized with respect to constraints that ensure that it remains symmetric and positive definite in order for it to be a well-defined covariance matrix.

Therefore, instead of attempting this difficult direct likelihood-based inference, we will adopt an alternative two-stage expectation maximization (EM) algorithm-based approach. The EM algorithm developed will be even more efficient and numerically robust, since both the expectation and maximization stages will be obtainable in closed form. In addition, we can be sure that such a procedure will find an optimum.

The ability to obtain a closed form expression for the expectation stage of the EM algorithm arises from the structure of the random effects model specified and the distributional assumptions made. One can show the following result given in Lemma 2.1 for the conditional distribution of the auxiliary random effects variable, conditional on the observations and covariates (other sensed modalities at each analog sensor location). Deriving this conditional distribution is important for the expectation step of the EM algorithm.

Lemma 2.1 (Conditional Distribution of the Random Effects) *The conditional distribution of the random effects given the data and covariates according to*

$$[\Gamma_k | \mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{w}_1, \dots, \mathbf{w}_m, A, B] \sim N(m_i, v_i) \quad (2.31)$$

with $\mathbf{w}_i \in \mathbb{R}^{(q \times 1)}$ and

$$v_i = (1 + \mathbf{w}_i^T B^T A^{-1} B \mathbf{w}_i)^{-1} \in \mathbb{R}^+,$$

$$m_i = v_i (\mathbf{y}_i - \mathbf{m} \mathbf{u})^T A^{-1} B \mathbf{w}_i \in \mathbb{R}.$$

Proof The derivation of this conditional distribution for the random effect follows trivially from the standard multivariate Gaussian properties since the joint distribution for the N^A auxiliary variables and observations is multivariate Gaussian:

$$p(\gamma_1, \dots, \gamma_m, \mathbf{y}_1, \dots, \mathbf{y}_m | \mathbf{w}_1, \dots, \mathbf{w}_m, A, B) = N(\mathbf{m}, C) \quad (2.32)$$

with $\mathbf{m} = [\mathbf{m}_1, \mathbf{m}_2]$ where \mathbf{m}_1 is a vector of m zeros and \mathbf{m}_2 a $1 \times mN^A$ vector given by $\mathbf{m}_2 = [\mu_g, \dots, \mu_g]$; and $C = \oplus_{i=1}^2 C_i$ where C_1 is a $m \times m$ matrix $C_1 = \text{diag}(1, \dots, 1)$ and $C_2 = \oplus_{j=1}^m A$. Then one can use the following properties of a multivariate normal to obtain the conditional distribution, where if μ and Σ are the mean and covariance of a Gaussian random vector, which is partitioned as follows:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad (2.33)$$

with sizes $q \times 1$ and $(N - q) \times 1$ and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad (2.34)$$

with sizes $q \times q$, $q \times (N - q)$, $(N - q) \times q$ and $(N - q) \times (N - q)$ then, the distribution of x_1 conditional on $x_2 = a$ is multivariate normal $(x_1 | x_2 = a) \sim N(\mu, \Sigma)$

where

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{a} - \boldsymbol{\mu}_2) \quad (2.35)$$

and covariance matrix

$$\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}. \quad (2.36)$$

This decomposition completes the required proof. \blacksquare

It will be assumed for now that the mean process is already estimated and given by $\hat{\boldsymbol{\mu}}_g$. In this section we discuss the more challenging aspect of estimation of the hyperparameters that make up the specifications of the covariance functions for process $g(\cdot)$ and $h(\cdot)$. This is achieved by the EM algorithm as follows:

We first write the complete data log-likelihood $\ln p(E|A, B, W, \gamma_{1:m})$ with respect to the matrix of $p \times m$ residuals E and random effects $\gamma_1, \dots, \gamma_m$ as follows:

$$l(A, B; E, W, \gamma_{1:m}) = -\frac{1}{2} \left(mp \ln(2\pi) + m \ln |A| + \sum_{k=1}^m (e_i - \gamma_i B \mathbf{w}_i)^T A^{-1} (e_i - \gamma_i B \mathbf{w}_i) \right) \quad (2.37)$$

Then from the complete data likelihood we consider the expectation step with respect to the random effect (nuisance parameters) as obtained in Lemma 2.2.

Lemma 2.2 (Integrated Complete Data Likelihood) *The following conditional expectation of the complete data likelihood with respect to the conditional distribution of the random effects nuisance parameters is obtained:*

$$\begin{aligned} & -2\mathbb{E}_{\gamma_{1:m}} [l(A, B; E, W, \gamma_{1:m}) | \hat{A}, \hat{B}] \\ &= mp \ln(2\pi) + m \ln |A| + \sum_{k=1}^m (e_i - \hat{m}_i B \mathbf{w}_i)^T A^{-1} (e_i - \hat{m}_i B \mathbf{w}_i) \\ &+ \sum_{k=1}^m \hat{s}_i \mathbf{w}_i^T B^T A^{-1} B \mathbf{w}_i \hat{s}_i, \end{aligned} \quad (2.38)$$

with $s_i = \sqrt{\hat{v}_i}$ and

$$\begin{aligned} \hat{v}_i &= (1 + \mathbf{w}_i^T \hat{B}^T \hat{A}^{-1} \hat{B} \mathbf{w}_i)^{-1}, \\ \hat{m}_i &= \hat{v}_i (\mathbf{y}_i - \mathbf{m} \mathbf{u})^T \hat{A}^{-1} \hat{B} \mathbf{w}_i. \end{aligned}$$

Proof Here the previous estimates for the target model parameters, denoted \hat{A}, \hat{B} are conditioned upon in the expectation in the sense that they are used to calculate the sufficient statistics for the distribution of the random effects $\gamma_{1:m}$ given by

$$\begin{aligned} \hat{v}_i &= (1 + \mathbf{w}_i^T \hat{B}^T \hat{A}^{-1} \hat{B} \mathbf{w}_i)^{-1} \\ \hat{m}_i &= \hat{v}_i (\mathbf{y}_i - \mathbf{m} \mathbf{u})^T \hat{A}^{-1} \hat{B} \mathbf{w}_i. \end{aligned}$$

One takes the conditional expectations of the complete data likelihood as follows:

$$\begin{aligned}
 & -2\mathbb{E}_{\gamma_{1:m}} [l(A, B; E, W, \gamma_{1:m}) | \widehat{A}, \widehat{B}] \\
 & = mp \ln(2\pi) + m \ln |A| + \sum_{k=1}^m \mathbb{E}_{\gamma_{1:m}} [(e_i - \gamma_i B \mathbf{w}_i)^T A^{-1} (e_i - \gamma_i B \mathbf{w}_i) | \widehat{A}, \widehat{B}]
 \end{aligned} \tag{2.39}$$

Next observe that γ_i 's are i.i.d. hence we can consider the individual expectations

$$\begin{aligned}
 & \mathbb{E}_{\gamma_i} [(e_i - \gamma_i B \mathbf{w}_i)^T A^{-1} (e_i - \gamma_i B \mathbf{w}_i) | \widehat{A}, \widehat{B}] \\
 & = \mathbb{E}_{\gamma_i} [e_i^T A^{-1} e_i - \gamma_i^2 \mathbf{w}_i^T B^T A^{-1} B \mathbf{w}_i | \widehat{A}, \widehat{B}] \\
 & = \mathbb{E}_{\gamma_i} [e_i^T A^{-1} e_i | \widehat{A}, \widehat{B}] - \mathbb{E}_{\gamma_i} [\gamma_i^2 \mathbf{w}_i^T B^T A^{-1} B \mathbf{w}_i | \widehat{A}, \widehat{B}] \\
 & = (e_i - \widehat{m}_i B \mathbf{w}_i)^T A^{-1} (e_i - \widehat{m}_i B \mathbf{w}_i) + \widehat{s}_i \mathbf{w}_i^T B^T A^{-1} B \mathbf{w}_i \widehat{s}_i.
 \end{aligned} \tag{2.40}$$

Then one simply rewrites the expression using this mean and variance expressions to complete the proof. \blacksquare

Having obtained a closed form expression for the expectation step, we next need to obtain the maximization step of the EM algorithm which involves the maximization of

$$\begin{aligned}
 & \arg \min_{A, B} -2\mathbb{E}_{\gamma_{1:m}} [l(A, B; E, W, \gamma_{1:m}) | \widehat{A}, \widehat{B}] \\
 & = \arg \min_{A, B} \left\{ mp \ln(2\pi) + m \ln |A| + \sum_{k=1}^m (e_i - \widehat{m}_i B \mathbf{w}_i)^T A^{-1} (e_i - \widehat{m}_i B \mathbf{w}_i) \right. \\
 & \quad \left. + \sum_{k=1}^m \widehat{s}_i \mathbf{w}_i^T B^T A^{-1} B \mathbf{w}_i \widehat{s}_i \right\}.
 \end{aligned} \tag{2.41}$$

Finally, one observes that this maximization can be easily implemented through a least squares solution by rewriting the argument in the form of a single quadratic with a change of representation given by constructing:

- \widetilde{W} as a $2m \times q$ matrix with i th row given by $m_i \mathbf{w}_i$ and whose $(n+i)$ th row is given by $s_i \mathbf{w}_i$;
- \widetilde{E} as a $2m \times p$ matrix of residuals given by $[E^T, \mathbf{0}]$ with the matrix of $\mathbf{0}$ the same dimension as matrix E , i.e., $m \times p$.

This produces the new argument for the optimization as follows:

$$\begin{aligned}
 & \arg \min_{A, B} -2\mathbb{E}_{\gamma_{1:m}} [l(A, B; E, W, \gamma_{1:m}) | \widehat{A}, \widehat{B}] \\
 & = \arg \min_{A, B} \{ mp \ln(2\pi) + m \ln |A| + \text{tr} [(\widetilde{E} - B \widetilde{W})(\widetilde{E} - B \widetilde{W})^T A^{-1}] \}.
 \end{aligned} \tag{2.42}$$

Rewriting the problem in this manner makes it appear directly as a least squares optimization problem which admits a solution given by:

$$\begin{aligned}\widehat{B} &= \widetilde{E}^T \widetilde{W} (\widetilde{W}^T \widetilde{W})^{-1}, \\ \widehat{A} &= \frac{1}{n} (\widetilde{E} - \widetilde{W} \widehat{B})^T (\widetilde{E} - \widetilde{W} \widehat{B})\end{aligned}\quad (2.43)$$

The EM algorithm proceeds as follows:

- Initialize the parameters making matrices \widehat{A} and \widehat{B} , where A is comprised of kernel hyperparameters and noise variance terms.
- Calculate the conditional estimators:

$$\begin{aligned}m_i &= \mathbb{E} [I_i | \widehat{A}, \widehat{B}, e_i] \\ v_i &= \mathbb{V}\text{ar} [I_i | \widehat{A}, \widehat{B}, e_i]\end{aligned}\quad (2.44)$$

- Construct new matrices \widetilde{W} and \widetilde{E} based on the data $\mathbf{y}_{1:m}$ and covariates $\mathbf{w}_{1:m}$.
- Evaluate the updated model parameters via the following least squares solutions for updated \widehat{A} and \widehat{B} according to

$$\begin{aligned}\widehat{B} &= \widetilde{E}^T \widetilde{W} (\widetilde{W}^T \widetilde{W})^{-1} \\ \widehat{A} &= \frac{1}{n} (\widetilde{E} - \widetilde{W} \widehat{B})^T (\widetilde{E} - \widetilde{W} \widehat{B})\end{aligned}\quad (2.45)$$

where matrix \widetilde{E} is the $2m \times 1$ matrix given by $(E^T, 0 \times E^T)^T$ and \widetilde{W} is a $2m \times d$ matrix with i th row given by $m_i \mathbf{w}_i$ and whose $(m+i)$ th is $\sqrt{v_i} \mathbf{w}_i$.

- Having solved for the matrix \widehat{A} , one then solves the system of equations given by

$$\begin{bmatrix} \widehat{A}_{11} & \cdots & \widehat{A}_{1N^A} \\ \vdots & \ddots & \vdots \\ \widehat{A}_{N^A 1} & \cdots & \widehat{A}_{N^A N^A} \end{bmatrix} = \begin{bmatrix} C_h(\mathbf{x}_1, \mathbf{x}_1) & \cdots & C_h(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ C_h(\mathbf{x}_n, \mathbf{x}_1) & \cdots & C_h(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} + \begin{bmatrix} \sigma_A^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_A^2 \end{bmatrix} \quad (2.46)$$

for the variance and hyperparameter terms in the kernels.

- repeat the above procedure until convergence

Having outlined an estimation procedure, the remainder of the chapter focuses on what can be done for spatial field reconstruction given an estimated model.

2.5 Spatial Field Reconstruction: Analytic Solutions

In this section we address the estimation problem known as *spatial field reconstruction* in the case of either a homogeneous WSN sensor model with all sensors

performing an L -bit quantization, which was recently studied in [29, 30], and then in the second case based on a heterogeneous WSN framework with a mixture of analog and binary sensors. In general the target criterion in developing the spatial estimator of the field reconstruction is achieve the minimum mean squared error (MMSE) in the estimation. This involved the following distortion metric:

$$D(\hat{f}_*, f_*) := \mathbb{E} \left[(f_* - \hat{f}_*)^2 \right]. \quad (2.47)$$

Under this framework, one may develop two closed form approximations for the estimators of the spatial field, in [27, 30] approximate series expansions based on saddle point and Laplace types were developed for nonlinear estimators which are optimal in the sense of minimizing the distortion metric in (2.47) [6].

In this chapter we wish to emphasize a computationally very efficient alternative class of estimators that we denote the spatial best linear unbiased estimators (S-BLUE) linear Bayes estimators. Such estimators are characterized by the following formal estimation objective (Objective 1):

Objective 1: spatial field reconstruction via best linear unbiased (S-BLUE) estimate, given by the solution to the following problem:

$$\hat{f}_* := \hat{a} + \hat{\mathbf{B}}\mathbf{Y}_{\mathcal{N}} = \arg \min_{a, \mathbf{B}} \mathbb{E} \left[(f_* - (a + \mathbf{B}\mathbf{Y}_{\mathcal{N}}))^2 \right], \quad (2.48)$$

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

The S-BLUE estimators are optimal in the sense that it achieves minimum variance among all linear estimators and have the desirable properties of being unbiased and efficient.

Theorem 2.2 (Spatial Best Linear Unbiased Estimator (S-BLUE)) *The optimal linear estimator of the spatial field \hat{f}_* at location \mathbf{x}_* in the class of all linear estimators taking the form $\hat{f}_* = a + \mathbf{B}\mathbf{Y}_{\mathcal{N}}$ for some scalar $a \in \mathbb{R}$, vector $\mathbf{B} \in \mathbb{R}^{1 \times N}$ at a location \mathbf{x}_* that solves (2.48) is given by*

$$\hat{f}_* = \mu(\mathbf{x}_*) + \mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T] \mathbb{E}_{\mathbf{Y}_{\mathcal{N}}}^{-1} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T] (\mathbf{Y}_{\mathcal{N}} - \mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}}]). \quad (2.49)$$

In addition, one may derive the estimation accuracy of the S-BLUE in closed form according to the result in Corollary [27, 30].

Corollary 2.1 *The associated MSE of the S-BLUE is given by*

$$\sigma_*^2 = \mathcal{C}(\mathbf{x}_*, \mathbf{x}_*) - \mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T] \mathbb{E}_{\mathbf{Y}_{\mathcal{N}}}^{-1} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T] \mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} f_*]. \quad (2.50)$$

The following sequence of algorithmic steps are then required to perform estimation of the S-BLUE, see Algorithm 1.

Algorithm 1 S-BLUE Field reconstruction

Input: $\mathbf{Y}_{\mathcal{N}}, \mathbf{x}_{\mathcal{N}}, \mathbf{x}_*, \sigma_A^2, \sigma^2, \mu(\cdot)$

Output: \hat{f}_*

- 1: Calculate the cross-correlation vector, $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$, where its i th element is implemented according to Lemma 2.3.
- 2: Calculate the covariance matrix, $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T]$, where its (i, j) th element is implemented according to Proposition 2.1 and the Clenshaw–Curtis coefficients in (2.57).
- 3: Calculate the S-BLUE of the intensity of the spatial field at a location \mathbf{x}_* as follows:

$$\hat{f}_* = \mu(\mathbf{x}_*) + \mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T] \mathbb{E}_{\mathbf{Y}_{\mathcal{N}}}^{-1} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T] (\mathbf{Y}_{\mathcal{N}} - \mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}}]).$$

The key components of the S-BLUE estimator that must be obtained for any form of WSN design involve the following components:

- the cross-correlation $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$; and
- the covariance $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T]$, (detailed in Lemma 2.1).

In the following set of results we will derive these quantities and subsequently the S-BLUE estimators for two classes of WSN: the L -bit homogeneous quantized/digitized WSN; and the heterogeneous L -bit digital/quantized and analog WSN. We begin with a detailed account of the result for the heterogeneous case.

Note: the covariance matrix for case 1 is derived in Lemma 2.1, however, for case 2 and case 3 we provide in Sect. 2.5.1.2 an accurate and efficient approximation for the expectations.

2.5.1 S-BLUE Spatial Field Estimator for Heterogeneous L -bit and Analog WSNs

In this section we consider the development of the S-BLUE class of spatial field reconstruction estimator to the Heterogeneous WSN setting in which we incorporate also additional sensed modalities, included as regressors into the spatial covariance structure through the kernel developed in (2.8) for the analog sensors.

2.5.1.1 Deriving the Cross-Correlation $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$

We now derive the cross-correlation between the spatial phenomenon predictive response, f_* at \mathbf{x}_* , and the observation vector $\mathbf{Y}_{\mathcal{N}}$. We prove that this quantity can be obtained exactly in closed form in the following Lemma 2.3.

Lemma 2.3 (Cross-Correlation between Spatial Process and Observations) *The i th element of $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$ is given by one of two cases.*

Case 1 - $\mathbf{x}_i \in \mathcal{X}^D$: where one has Cross-Correlation terms given by

$$\begin{aligned} & \mathbb{E}_{f_*, Y_i} [f_* Y_i] \\ &= c_1 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \mathbb{P}(Y_i = l | B_i = j) G_1 \left(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) \right) \\ &+ c_2 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \mathbb{P}(Y_i = l | B_i = j) \left\{ \mu(\mathbf{x}_i) G_1 \left(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) \right) \right. \\ &\quad \left. - \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) G_2 \left(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) \right) \right\}. \end{aligned}$$

where $c_2 := \frac{\mathcal{C}(\mathbf{x}_*, \mathbf{x}_i)}{\mathcal{C}(\mathbf{x}_i, \mathbf{x}_i)}$, $c_1 := \mu(\mathbf{x}_*) - c_2 \mu(\mathbf{x}_i)$ and

$$\begin{aligned} G_1(a, b; m, s) &= \{\Phi(a; m, s) - \Phi(b; m, s)\} \\ G_2(a, b; m, s) &= \{\phi(a; m, s) - \phi(b; m, s)\}. \end{aligned}$$

Case 2 - $\mathbf{x}_i \in \mathcal{X}^A$: where one has Cross-Correlation terms given by

$$\mathbb{E}_{f_*, Y_i} [f_* Y_i] = c_1 \mu(\mathbf{x}_i) + c_2 [\mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) - \mu(\mathbf{x}_i)^2],$$

Proof To make the proof we consider the i th term of $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$ which has its expectation decomposed via the tower property as follows:

$$\mathbb{E}_{f_*, Y_i} [f_* Y_i] = \mathbb{E}_{f_i} [\mathbb{E}_{f_*, Y_i} [f_* Y_i | f_i]]. \quad (2.51)$$

We then consider for each of the possible cases, i.e., Case 1 $\mathbf{x}_i \in \mathcal{X}^D$ and Case 2 $\mathbf{x}_i \in \mathcal{X}^A$, the analytic calculation of this cross-correlation. It will be useful to first make the following definitions used throughout the proof:

$$\begin{aligned} c_2 &:= \frac{\mathcal{C}(\mathbf{x}_*, \mathbf{x}_i)}{\mathcal{C}(\mathbf{x}_i, \mathbf{x}_i)}, \\ c_1 &:= \mu(\mathbf{x}_*) - c_2 \mu(\mathbf{x}_i), \\ G_1(a, b; m, s) &:= \{\Phi(a; m, s) - \Phi(b; m, s)\} \\ G_2(a, b; m, s) &:= \{\phi(a; m, s) - \phi(b; m, s)\}. \end{aligned}$$

Case 1:

The conditional expectation, $\mathbb{E}_{f_*, Y_i} [f_* Y_i | f_i]$, can be expressed as:

$$\begin{aligned}
 \mathbb{E}_{f_*, Y_i} [f_* Y_i | f_i] &= \int_{-\infty}^{\infty} \sum_{l=0}^{L-1} f_* l p(f_*, Y_i = l | f_i) df_* \\
 &= \int_{-\infty}^{\infty} f_* \phi(f_*; c_1 + c_2 f_i, \sigma_{f_{\mathcal{N}} | \mathbf{Y}_{\mathcal{N}}}^2) \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} (\mathbb{P}(Y_i = l | B_i = j) \mathbb{P}(B_i = j | f_i)) df_* \\
 &= (c_1 + c_2 f_i) \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \mathbb{P}(Y_i = l | B_i = j) G_1(\lambda_{j+1}, \lambda_j; f_i, \sigma_A^2),
 \end{aligned}$$

The expectation with respect to f_i of the first term is given by

$$\begin{aligned}
 &\mathbb{E}_{f_i} \left[c_1 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \left(\mathbb{P}(Y_i = l | B_i = j) G_1(\lambda_{j+1}, \lambda_j; f_i, \sigma_A^2) \right) \right] \\
 &= c_1 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \left(\mathbb{P}(Y_i = l | B_i = j) \mathbb{E}_{f_i} \left[G_1(\lambda_{j+1}, \lambda_j; f_i, \sigma_A^2) \right] \right) \quad (2.52) \\
 &= c_1 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \left(\mathbb{P}(Y_i = l | B_i = j) G_1(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + C(\mathbf{x}_i, \mathbf{x}_i)) \right).
 \end{aligned}$$

The expectation of the second term is given by:

$$\begin{aligned}
 &\mathbb{E}_{f_i} \left[c_2 f_i \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \left(\mathbb{P}(Y_i = l | B_i = j) G_1(\lambda_{j+1}, \lambda_j; f_i, \sigma_A^2) \right) \right] \\
 &= c_2 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \left(\mathbb{P}(Y_i = l | B_i = j) \mathbb{E}_{f_i} \left[\left(f_i \Phi(\lambda_{j+1}, f_i, \sigma_A^2) - f_i \Phi(\lambda_j, f_i, \sigma_A^2) \right) \right] \right) \\
 &= c_2 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \left(\mathbb{P}(Y_i = l | B_i = j) \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\lambda_j} f_i \phi(a; f_i, \sigma_A^2) \phi(f_i; \mu(\mathbf{x}_i), C(\mathbf{x}_i, \mathbf{x}_i)) da df_i \right. \right. \\
 &\quad \left. \left. - \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda_j} f_i \phi(a; f_i, \sigma_A^2) \phi(f_i; \mu(\mathbf{x}_i), C(\mathbf{x}_i, \mathbf{x}_i)) da df_i \right) \right) \quad (2.53) \\
 &= c_2 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \mathbb{P}(Y_i = l | B_i = j) \\
 &\quad \times \left\{ \mu(\mathbf{x}_i) \Phi(\lambda_{j+1}, \mu(\mathbf{x}_i), \sigma_A^2 + C(\mathbf{x}_i, \mathbf{x}_i)) - C(\mathbf{x}_i, \mathbf{x}_i) \phi(\lambda_{j+1}, \mu(\mathbf{x}_i), \sigma_A^2 + C(\mathbf{x}_i, \mathbf{x}_i)) \right. \\
 &\quad \left. - \left(\mu(\mathbf{x}_i) \Phi(\lambda_j, \mu(\mathbf{x}_i), \sigma_A^2 + C(\mathbf{x}_i, \mathbf{x}_i)) - C(\mathbf{x}_i, \mathbf{x}_i) \phi(\lambda_j, \mu(\mathbf{x}_i), \sigma_A^2 + C(\mathbf{x}_i, \mathbf{x}_i)) \right) \right\}.
 \end{aligned}$$

Combining (2.52) and (2.53), we obtain that the i th term of $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$ is expressed as:

$$\begin{aligned} \mathbb{E}_{f_*, Y_i} [f_* Y_i] &= c_1 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \mathbb{P}(Y_i = l | B_i = j) G_1 \left(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) \right) \\ &+ c_2 \sum_{l=0}^{L-1} l \sum_{j=0}^{L-1} \mathbb{P}(Y_i = l | B_i = j) \left\{ \mu(\mathbf{x}_i) G_1 \left(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) \right) \right. \\ &\quad \left. - \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) G_2 \left(\lambda_{j+1}, \lambda_j; \mu(\mathbf{x}_i), \sigma_A^2 + \mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) \right) \right\}. \end{aligned}$$

Case 2:

The conditional expectation, $\mathbb{E}_{f_*, Y_i} [f_* Y_i | f_i]$, can be expressed as:

$$\begin{aligned} \mathbb{E}_{f_*, Y_i} [f_* Y_i | f_i] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_* y_i p_{Y_i | f_i}(y_i | f_i) p_{f_* | f_i}(f_* | f_i) dy_i df_* \\ &= f_i (c_1 + c_2 f_i), \end{aligned}$$

The expectation with respect to f_i is then given by

$$\mathbb{E}_{f_i} [f_i (c_1 + c_2 f_i)] = c_1 \mu(\mathbf{x}_i) + c_2 [\mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) - \mu(\mathbf{x}_i)^2] \quad (2.54)$$

Hence in Case 2 one obtains that the i th term of $\mathbb{E}_{f_*, \mathbf{Y}_{\mathcal{N}}} [f_* \mathbf{Y}_{\mathcal{N}}^T]$ is expressed as:

$$\mathbb{E}_{f_*, Y_i} [f_* Y_i] = c_1 \mu(\mathbf{x}_i) + c_2 [\mathcal{C}(\mathbf{x}_i, \mathbf{x}_i) - \mu(\mathbf{x}_i)^2] \quad \blacksquare$$

2.5.1.2 Deriving the Covariance Matrix $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T]$ Estimators

We have already derived the covariance matrix, $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T]$ completely in case one and case two, what remains is the expectations in case three. Recall, Case 1 involved $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^A$, i.e., both sensors are high-quality analog sensors; Case 2 with $\mathbf{x}_i \in \mathcal{X}^A$ and $\mathbf{x}_j \in \mathcal{X}^D$, i.e., one sensor is analog and one sensor is a cheaper quantized sensor; and Case 3 in which $\mathbf{x}_i \in \mathcal{X}^D$ and $\mathbf{x}_j \in \mathcal{X}^D$. These were given in Case 3 up to an expectation which would need to be approximated. We briefly explain in this section an efficient manner to perform such approximation using a form of quadrature.

Case 3: $\mathbf{x}_i \in \mathcal{X}^D$ and $\mathbf{x}_j \in \mathcal{X}^D$

$$\begin{aligned} \mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &= \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \mathbb{P}(Y_i = k | B_i = m) \mathbb{P}(Y_j = l | B_j = n) \\ &\quad \times \mathbb{E}_{f_i, f_j} \left[\left(\Phi(\lambda_{m+1}, f_i, \sigma_A^2) - \Phi(\lambda_m, f_i, \sigma_A^2) \right) \left(\Phi(\lambda_{n+1}, f_j, \sigma_A^2) - \Phi(\lambda_n, f_j, \sigma_A^2) \right) \right]. \end{aligned}$$

This involves an intractable integral which we solve via an efficient numerical procedure, based on the Clenshaw–Curtis quadrature rule [10]. We begin by solving the first integral with respect to f_i , thus reducing the dimension of the problem:

$$\begin{aligned} &\mathbb{E}_{f_i, f_j} \left[\left(\Phi(\lambda_{m+1}, f_i, \sigma_A^2) - \Phi(\lambda_m, f_i, \sigma_A^2) \right) \left(\Phi(\lambda_{n+1}, f_j, \sigma_A^2) - \Phi(\lambda_n, f_j, \sigma_A^2) \right) \right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(f_i, f_j) \left(\Phi(\lambda_{m+1}, f_i, \sigma_A^2) - \Phi(\lambda_m, f_i, \sigma_A^2) \right) \left(\Phi(\lambda_{n+1}, f_j, \sigma_A^2) - \Phi(\lambda_n, f_j, \sigma_A^2) \right) \\ &\quad - \Phi(\lambda_n, f_j, \sigma_A^2) \, df_i df_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(f_i | f_j) p(f_j) \left(\Phi(\lambda_{m+1}, f_i, \sigma_A^2) - \Phi(\lambda_m, f_i, \sigma_A^2) \right) \left(\Phi(\lambda_{n+1}, f_j, \sigma_A^2) - \Phi(\lambda_n, f_j, \sigma_A^2) \right) \\ &\quad - \Phi(\lambda_n, f_j, \sigma_A^2) \, df_i df_j \\ &= \mathbb{E}_{f_j} \left[\Delta(\lambda_m, \lambda_{m+1}, c_1 + c_2 f_j, \sigma_A^2 + \sigma^2) \Delta(\lambda_n, \lambda_{n+1}, f_j, \sigma_A^2) \right], \end{aligned} \tag{2.55}$$

where we define $c_2 := \frac{\mathcal{C}(\mathbf{x}_*, \mathbf{x}_i)}{\mathcal{C}(\mathbf{x}_i, \mathbf{x}_i)}$, and $c_1 := \mu(\mathbf{x}_*) - c_2 \mu(\mathbf{x}_i)$.

This integral with respect to f_j does not admit a closed form representation, and we utilize a numerical procedure to solve it. We now develop an efficient numerical solution via the Clenshaw–Curtis quadrature [10].

The Clenshaw–Curtis quadrature only works on finite integral domains, while (2.55) has infinite support. We shall first use a generic coordinate transformation which will transform the integral in (2.55) from an infinite interval into a finite one, presented in Lemma 2.4 and then utilize the Clenshaw–Curtis quadrature in Lemma 2.6 and finally calculate the covariance matrix in Proposition 2.1.

Lemma 2.4 (Generic Coordinate Transformation for Integration on Infinite Intervals) *Consider the generic coordinate transformation for the integrand and terminals via the mapping $x = \frac{t}{1-t^2}$ giving the mapped definite integral*

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-1}^{+1} f\left(\frac{t}{1-t^2}\right) \frac{1+t^2}{(1-t^2)^2} dt.$$

When Lemma 2.4 is applied to (2.55), one obtains

$$\begin{aligned} &\mathbb{E}_{f_j} \left[\Delta(\lambda_m, \lambda_{m+1}, c_1 + c_2 f_j, \sigma_A^2 + \sigma^2) \Delta(\lambda_n, \lambda_{n+1}, f_j, \sigma_A^2) \right] \\ &= \int_{-\infty}^{\infty} \Delta(\lambda_m, \lambda_{m+1}, c_1 + c_2 f_j, \sigma_A^2 + \sigma^2) \Delta(\lambda_n, \lambda_{n+1}, f_j, \sigma_A^2) p(f_j) df_j \end{aligned}$$

$$\begin{aligned}
& \left(f_j := \frac{t}{1-t^2} \right) \int_{-1}^1 \Delta \left(\lambda_m, \lambda_{m+1}, c_1 + c_2 \frac{t}{1-t^2}, \sigma_A^2 + \sigma^2 \right) \Delta \left(\lambda_n, \lambda_{n+1}, \frac{t}{1-t^2}, \sigma_A^2 \right) \\
& \quad \times p \left(\frac{t}{1-t^2} \right) \frac{1+t^2}{(1-t^2)^2} dt.
\end{aligned} \tag{2.56}$$

Next, we solve this integral via the Clenshaw–Curtis Quadrature rule.

Lemma 2.5 (Clenshaw–Curtis Quadrature Rule [10]) *Consider the closed form approximation of the integral*

$$\int_0^\pi g(\cos \theta) \sin(\theta) d\theta \simeq a_0 + \sum_{k=1}^{M/2-1} \frac{2a_{2k}}{1-(2k)^2} + \frac{a_M}{1-M^2}.$$

which involves finding a subset of the coefficients $\{a_k\}_{k \geq 0}$ given by a_{2k} , due to aliasing arguments in [8]. These coefficients are solution to integrals involving periodic functions $f(\cos \theta)$, then the Fourier series can be computed efficiently and accurately up to Nyquist frequency $k = M$, through a $(M+1)$ equally spaced and equally weighted points $\theta_m = m\pi/M$ for $m = 0, \dots, M$. At the endpoints of the domain the weights are given by $1/2$ to ensure double-counting is avoided. This is equivalent to a discrete cosine transform (DCT) approximation given by

$$a_k = \frac{2}{M} \left[\frac{g(1)}{2} + \frac{g(-1)}{2} (-1)^k + \sum_{m=1}^{M-1} g(\cos[n\pi/M]) \cos(mk\pi/M) \right], \quad \forall k \in \{0, \dots, M\}. \tag{2.57}$$

We now apply the Clenshaw–Curtis quadrature rule to our integral in (2.55).

Lemma 2.6 *The expectation in (2.55) can be evaluated by applying the Clenshaw–Curtis quadrature to the transformed integral in (2.56), as follows:*

$$\begin{aligned}
& \mathbb{E}_{f_j} \left[\Delta \left(\lambda_m, \lambda_{m+1}, c_1 + c_2 f_j, \sigma_A^2 + \sigma^2 \right) \Delta \left(\lambda_n, \lambda_{n+1}, f_j, \sigma_A^2 \right) \right] \\
& = \underbrace{\int_{-1}^1 \Delta \left(\lambda_m, \lambda_{m+1}, c_1 + c_2 \frac{t}{1-t^2}, \sigma_A^2 + \sigma^2 \right) \Delta \left(\lambda_n, \lambda_{n+1}, \frac{t}{1-t^2}, \sigma_A^2 \right) p \left(\frac{t}{1-t^2} \right) \frac{1+t^2}{(1-t^2)^2} dt}_{:=g(t)} \\
& \simeq a_0 + \sum_{k=1}^{M/2-1} \frac{2a_{2k}}{1-(2k)^2} + \frac{a_M}{1-M^2},
\end{aligned}$$

with a_k defined in (2.57).

Now that we have evaluated the expectation term, we derive the (i, j) th term of the covariance matrix.

Proposition 2.1 *The (i, j) th term of $\mathbb{E}_{\mathbf{Y}_{\mathcal{N}}} [\mathbf{Y}_{\mathcal{N}} \mathbf{Y}_{\mathcal{N}}^T]$ can be approximated as:*

$$\begin{aligned} \mathbb{E}_{Y_i, Y_j} [Y_i Y_j] &\simeq \sum_{k=0}^{L-1} \sum_{l=0}^{L-1} kl \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} \mathbb{P}(Y_i = k | B_i = m) \mathbb{P}(Y_j = l | B_j = n) \\ &\times \left(a_0 + \sum_{k=1}^{M/2-1} \frac{2a_{2k}}{1 - (2k)^2} + \frac{a_M}{1 - M^2} \right). \end{aligned}$$

2.6 Simulations

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 modeling in Europe, under the class of weather events known in insurance modeling 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 [7, 13].

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

$$\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\Omega}) := \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{l}\right). \quad (2.58)$$

A realization from the GP is shown in Fig. 2.1. 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 configurations, 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 $\sigma_v^2 = 0$ and define $\text{SNR} = 10 \log \sigma_w^2$. The prediction mean squared errors (PMSE) are presented in the right side of Fig. 2.1. 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.

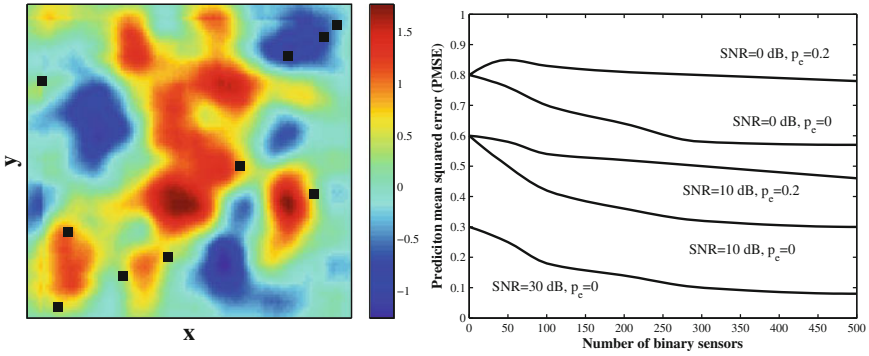


Fig. 2.1 Realization from a 2-d Gaussian process. The *black* markers denote the locations of the 10 high-quality sensors

2.6.2 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. This database is comprised of 23 storms which caused high insurance losses known as ‘insurance storms’ and 27 storms which were selected because they are the top ‘noninsurance’ storms as ranked by the storm severity index, see details on the web site.

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° , latitude = 37.5° . As discussed in the data description provided with the dataset, 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° . The data for each of the storms provide a list of grid number and maximum 3-s 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 Dagmar (Patrick or Tapani) which took place on 26/12/2011 and affected are Finland and Norway; and the second was the storm known as Ulli taking place on 03/01/2012 which affected the UK.

To understand the significance of this analysis, we note that the Dagmar-Patrick storm is reported to have caused damage worth 40 Million USD and reached a maximum wind speed of 70 mph over land. The storm Ulli is reported to have

²<http://www.met.reading.ac.uk/~extws/database/dataDesc>.

caused even more damage of 200 Million USD and reached a wind speed of 87 mph. In our study, we reconstruct the spatial map of these wind speeds for a given instant of time.

2.6.2.1 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:

$$\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j; \boldsymbol{\Omega}) := \sigma_x^2 \exp\left(-\frac{\|x_i - x_j\|}{l_x}\right) \times \sigma_y^2 \exp\left(-\frac{\|y_i - y_j\|}{l_y}\right), \quad (2.59)$$

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-Patrick 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 hyperparameters can be found in [Chap. 5] [37]. The covariance function estimates are presented in Fig. 2.4 for the Dagmar-Patrick (left panel) and Ulli (right panel) wind storms. These plots show the spatial dependence over UK and Europe between wind speeds during the peak of the storm fronts as they transited across different regions of the English channel. It is clear that the correlation of the Dagmar-Patrick storm is much stronger than of the Ulli storm in both axes. This should have an impact on the quality of the field reconstruction estimation that we will demonstrate next.

2.6.2.2 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. 2.2 and 2.3, for the Dagmar-Patrick and Ulli storms, respectively. We set the region of interest (ROI) as shown in the upper left of Figs. 2.2 and 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

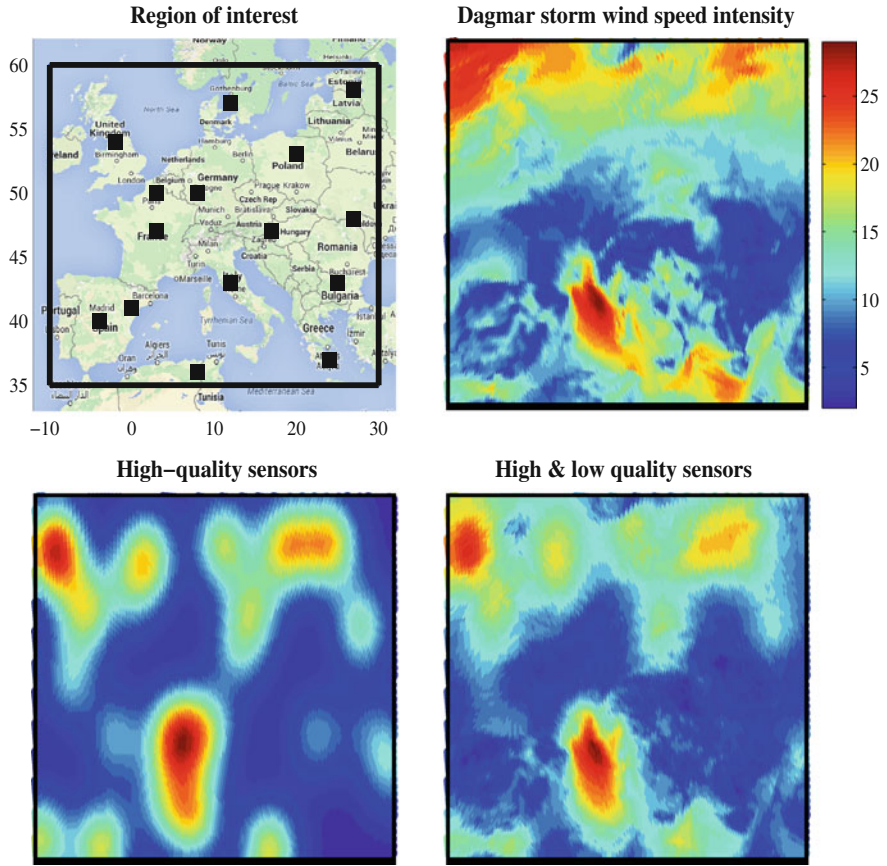


Fig. 2.2 Wind speed prediction of the Dagmar-Patrick storm. The *rectangular* in the *upper left* figure represents the region of interest which contains 15 high-quality sensors. The *upper right* figure represents the “true” data wind speed intensities (m/s). The *lower left* figure shows the field 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

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 Table 2.1. As expected the prediction performance for the Dagmar-Patrick storm is better than for the Ulli storm. This can be explained by the higher spatial correlation exhibited by the Dagmar-Patrick storm as shown in Fig. 2.4.

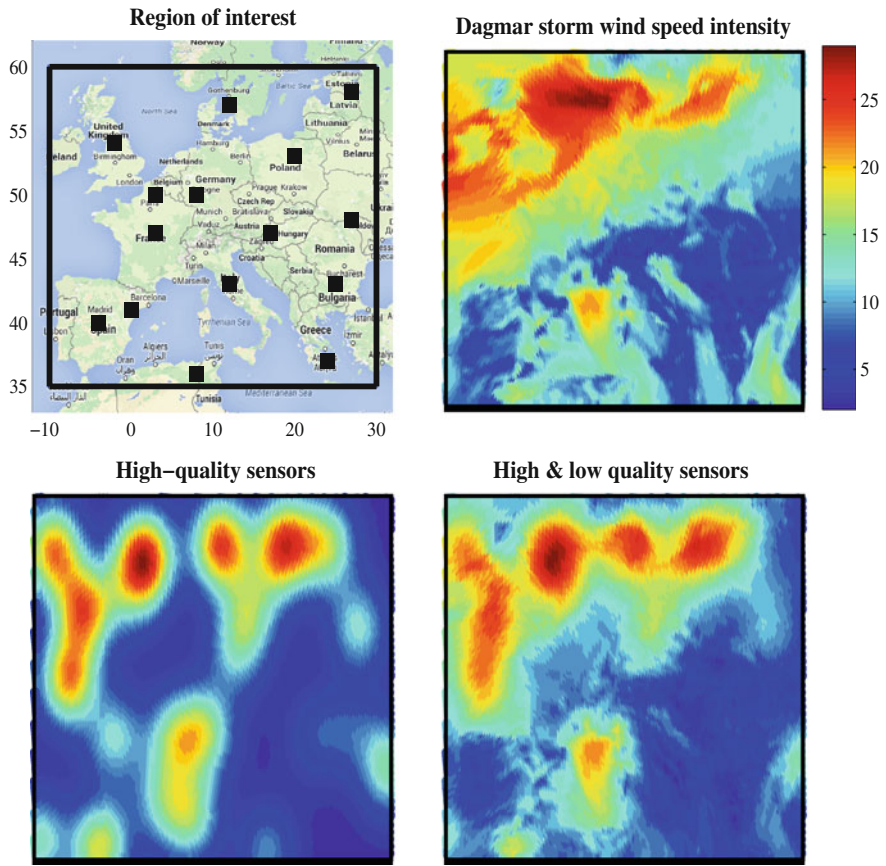


Fig. 2.3 Wind speed prediction of the Ulli storm. The *rectangular* in the *upper left* figure represents the region of interest which contains 15 high-quality sensors. The *upper right* figure represents the “true” data wind speed intensities (m/s). The *lower left* figure shows the field reconstruction 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 0.85 and based on both high- and low-quality sensors is 0.37

Table 2.1 Field reconstruction performance for the two storms

Normalized prediction mean squared error		
Reconstruction method	Dagmar-Patrick storm	Ulli storm
15 high-quality sensors	0.67	0.85
15 high-quality and 100 low-quality sensors	0.25	0.37

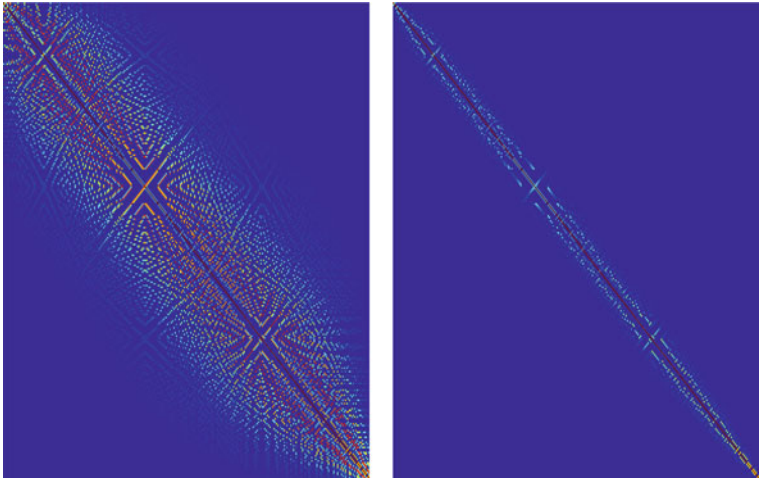


Fig. 2.4 The covariance function estimation of the Dagmar-Patrick (*left panel*) and Ulli (*right panel*) storms. These results show that the spatial correlation of the Dagmar-Patrick storm is larger than the Ulli storm

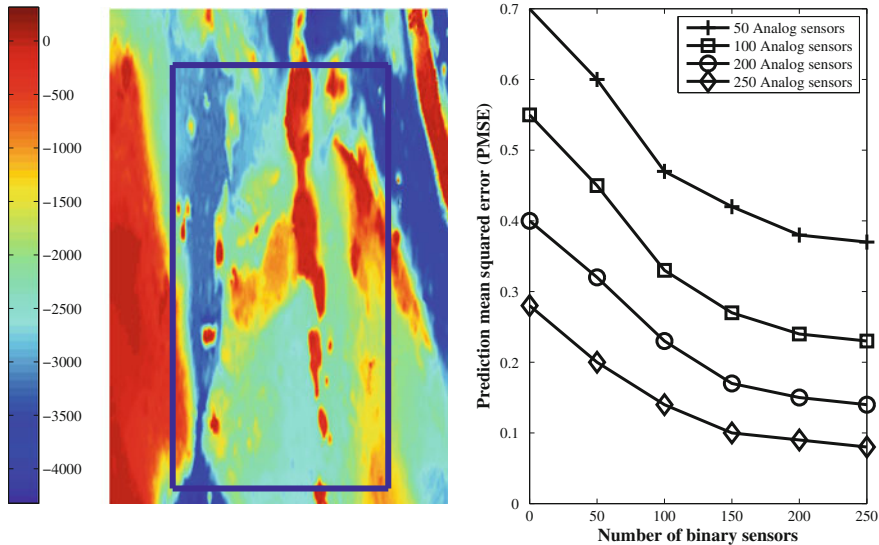


Fig. 2.5 Ocean depth estimation results. The *rectangular* in the *left figure* represents the region of interest and the intensity represents the true ocean’s depth. The *right figure* presents the prediction MSE of our algorithm based on the heterogeneous network with 50, 100, 200, 250 high-quality sensors and varying number of low-quality sensors

2.6.3 Bathymetry Example

In this example we use the heterogeneous sensor network to estimate the spatial field for the depth of the ocean floor based on measurements known as Bathymetry. This type of analysis is also directly relevant to the wind speed and storm surge modeling done in the first example, as bathymetric measurements are known to vary significantly during storm surge events and cyclones. This makes the spatial field reconstruction of such a feature directly relevant to modeling practically important spatial features. The ocean depth can help to provide an indication of the likely event of a flooding event from a storm front.

We use a publicly available database known as the eSurge.³ We selected to analyze a square region in the north-east corner of Australia at the South-Pacific ocean presented in the left panel of Fig. 2.5. This region is known to be frequently hit by cyclones which cause a change in the topography of the ocean floor.

We performed the depth estimation using our algorithm and compared it to the case where only high-quality sensors are utilized. In each simulation the sensors were deployed on a regular grid and we changed the number of high-quality and low-quality sensors deployed. We then calculated the prediction mean squared error (PMSE) which is presented in the right panel of Fig. 2.5. Similarly to the synthetic example in Sect. 2.6.1, there was a diminishing improvement when the number of low-quality sensors was above 200.

References

1. Adler, R., Taylor, J.: Random Fields and Geometry, vol. 115. Springer, New York (2007)
2. Agrawal, P., Patwari, N.: Correlated link shadow fading in multi-hop wireless networks. *IEEE Trans. Wirel. Commun.* **8**(8), 4024–4036 (2009)
3. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. *Comput. Netw.* **38**(4), 393–422 (2002)
4. Akyildiz, I., Vuran, M., Akan, O.: On exploiting spatial and temporal correlation in wireless sensor networks. In: *Proceedings of WiOpt'04: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks* pp. 71–80 (2004)
5. Anastasi, G., Conti, M., Di Francesco, M., Passarella, A.: Energy conservation in wireless sensor networks: a survey. *Ad Hoc Netw.* **7**(3), 537–568 (2009)
6. Berger, J.: Statistical Decision Theory and Bayesian Analysis. Springer, New York (1985)
7. Berz, G.: Windstorm and storm surges in Europe: loss trends and possible counter-actions from the viewpoint of an international reinsurer. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **363**(1831), 1431–1440 (2005)
8. Boyd, J.: Chebyshev and Fourier Spectral Methods. Dover Publications, New York (2001)
9. Chintalapudi, K., Fu, T., Paek, J., Kothari, N., Rangwala, S., Caffrey, J., Govindan, R., Johnson, E., Masri, S.: Monitoring civil structures with a wireless sensor network. *IEEE Internet Comput.* **10**(2), 26–34 (2006)
10. Clenshaw, C., Curtis, A.: A method for numerical integration on an automatic computer. *Numer. Math.* **2**(1), 197–205 (1960)

³<http://www.storm-surge.info/project>.

11. Cohen, K., Leshem, A.: Energy-efficient detection in wireless sensor networks using likelihood ratio and channel state information. *IEEE J. Sel. Areas Commun.* **29**(8), 1671–1683 (2011)
12. Fazel, F., Fazel, M., Stojanovic, M.: Random access sensor networks: field reconstruction from incomplete data. In: *IEEE Information Theory and Applications Workshop (ITA)*, pp. 300–305 (2012)
13. Flather, R., Smith, J., Richards, J., Bell, C., Blackman, D.: Direct estimates of extreme storm surge elevations from a 40-year numerical model simulation and from observations. *Global Atmos. Ocean Syst.* **6**(2), 165–176 (1998)
14. Fonseca, C., Ferreira, H.: Stability and contagion measures for spatial extreme value analyses. [arXiv:1206.1228](https://arxiv.org/abs/1206.1228) (2012)
15. French, J.P., Sain, S.R.: Spatio-Temporal Exceedance Locations and Confidence Regions. *Annals of Applied Statistics*. Prepress (2013)
16. Gu, D., Hu, H.: Spatial Gaussian process regression with mobile sensor networks. *IEEE Trans. Neural Netw. Learn. Syst.* **23**(8), 1279–1290 (2012)
17. Hoff, P.D., Niu, X.: A Covariance Regression Model. [arXiv:1102.5721](https://arxiv.org/abs/1102.5721) (2011)
18. Højsgaard, S., Edwards, D., Lauritzen, S.: Gaussian graphical models. In: *Graphical Models with R*, pp. 77–116. Springer, New York (2012)
19. Katenka, N., Levina, E., Michailidis, G.: Local vote decision fusion for target detection in wireless sensor networks. *IEEE Trans. Signal Process.* **56**(1), 329–338 (2008)
20. Kottas, A., Wang, Z., Rodriguez, A.: Spatial modeling for risk assessment of extreme values from environmental time series: a Bayesian nonparametric approach. *Environmetrics* **23**(8), 649–662 (2012). doi:[10.1002/env.2177](https://doi.org/10.1002/env.2177)
21. Krause, A., Singh, A., Guestrin, C.: Near-optimal sensor placements in gaussian processes: theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.* **9**, 235–284 (2008)
22. Lorincz, K., Malan, D.J., Fulford-Jones, T.R., Nawoj, A., Clavel, A., Shnayder, V., Mainland, G., Welsh, M., Moulton, S.: Sensor networks for emergency response: challenges and opportunities. *IEEE Pervasive Comput.* **3**(4), 16–23 (2004)
23. Masazade, E., Niu, R., Varshney, P., Keskinos, M.: Energy aware iterative source localization for wireless sensor networks. *IEEE Trans. Signal Process.* **58**(9), 4824–4835 (2010)
24. Matamoros, J., Fabbri, F., Antón-Haro, C., Dardari, D.: On the estimation of randomly sampled 2D spatial fields under bandwidth constraints. *IEEE Trans. Wirel. Commun.* **10**(12), 4184–4192 (2011)
25. Matern, B.: Spatial variation. meddelanden fraan statens skogsforskningsinstitut, **49**(5), 1–144. Also appeared as *Lecture Notes in Statistics*, vol. 36 (1986)
26. Msechu, E., Giannakis, G.: Sensor-centric data reduction for estimation with WSNs via censoring and quantization. *IEEE Trans. Signal Process.* **60**(1), 400–414 (2012)
27. Nevat, I., Peters, G., Collings, I.: Location-aware cooperative spectrum sensing via Gaussian processes. In: *IEEE Australian Communications Theory Workshop (AusCTW)*, pp. 19–24 (2012)
28. Nevat, I., Peters, G.W., Collings, I.B.: Location-aware cooperative spectrum sensing via gaussian processes. In: *Communications Theory Workshop (AusCTW)*, 2012 Australian, pp. 19–24. IEEE (2012)
29. Nevat, I., Peters, G.W., Collings, I.B.: Estimation of correlated and quantized spatial random fields in wireless sensor networks. In: *2013 IEEE International Conference on Communications (ICC)*, pp. 1931–1935. IEEE (2013)
30. Nevat, I., Peters, G.W., Collings, I.B.: Random field reconstruction with quantization in wireless sensor networks. *IEEE Trans. Signal Process.* **61**, 6020–6033 (2013)
31. Niu, R., Varshney, P.K.: Target location estimation in sensor networks with quantized data. *IEEE Trans. Signal Process.* **54**(12), 4519–4528 (2006)
32. Ozdemir, O., Niu, R., Varshney, P.K.: Channel aware target localization with quantized data in wireless sensor networks. *IEEE Trans. Signal Process.* **57**(3), 1190–1202 (2009)
33. Park, S., Choi, S.: Gaussian processes for source separation. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1909–1912 (2008)

34. Peters, G., Nevat, I., Lin, S., Matsui, T.: Modelling threshold exceedence levels for spatial stochastic processes observed by sensor networks. In: 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1–7. IEEE (2014)
35. Rajasegarar, S., Havens, T.C., Karunasekera, S., Leckie, C., Bezdek, J.C., Jamriska, M., Gunatilaka, A., Skvortsov, A., Palaniswami, M.: High-resolution monitoring of atmospheric pollutants using a system of low-cost sensors. *IEEE Trans. Geosci. Remote Sens.* **52**, 3823–3832 (2014)
36. Rajasegarar, S., Zhang, P., Zhou, Y., Karunasekera, S., Leckie, C., Palaniswami, M.: High resolution spatio-temporal monitoring of air pollutants using wireless sensor networks. In: 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1–6. IEEE (2014)
37. Rasmussen, C., Williams, C.: *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press (2005)
38. Schabenberger, O., Pierce, F.J.: *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton (2002)
39. Sohraby, K., Minoli, D., Znati, T.: *Wireless Sensor Networks: Technology, Protocols, and Applications*. Wiley, Hoboken (2007)
40. Stein, M.L.: *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York (1999)
41. Vanmarcke, E.: *Random Fields: Analysis and Synthesis*. World Scientific Publishing Company Inc., Singapore (2010)
42. Vuran, M.C., Akan, O.B., Akyildiz, I.F.: Spatio-temporal correlation: theory and applications for wireless sensor networks. *Comput. Netw. J.*, Elsevier **45**, 245–259 (2004)
43. Werner-Allen, G., Lorincz, K., Ruiz, M., Marcillo, O., Johnson, J., Lees, J., Welsh, M.: Deploying a wireless sensor network on an active volcano. *IEEE Internet Comput.* **10**(2), 18–25 (2006)
44. Wu, T., Cheng, Q.: Distributed estimation over fading channels using one-bit quantization. *IEEE Trans. Wirel. Commun.* **8**(12), 5779–5784 (2009)
45. Xu, Y., Choi, J.: Adaptive sampling for learning Gaussian processes using mobile sensor networks. *Int. J. Sens.* **11**(3), 3051–3066 (2011)
46. Zheng, Y., Niu, R., Varshney, P.: Closed-form performance for location estimation based on quantized data in sensor networks. In: 13th Conference on Information Fusion (FUSION), pp. 1–7. IEEE (2010)
47. Zhou, Y., Li, J., Wang, D.: Posterior cramer-rao lower bounds for target tracking in sensor networks with quantized range-only measurements. *IEEE Signal Process. Lett.* **17**(2), 157–160 (2010)

Modern Methodology and Applications in
Spatial-Temporal Modeling

Peters, G.W.; Matsui, T. (Eds.)

2015, XV, 111 p. 17 illus., 4 illus. in color., Softcover

ISBN: 978-4-431-55338-0