Golz, Martin; Zoubir, Abdelhak M.; Koivunen, Visa

Estimating Test Statistic Distributions for Multiple Hypothesis Testing in Sensor Networks

Published in:
2022 56th Annual Conference on Information Sciences and Systems, CISS 2022

DOI:
10.1109/CISS53076.2022.9751186

Published: 01/01/2022

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Golz, M., Zoubir, A. M., & Koivunen, V. (2022). Estimating Test Statistic Distributions for Multiple Hypothesis Testing in Sensor Networks. In *2022 56th Annual Conference on Information Sciences and Systems, CISS 2022* (pp. 90-95). IEEE. https://doi.org/10.1109/CISS53076.2022.9751186

Estimating Test Statistic Distributions for Multiple Hypothesis Testing in Sensor Networks

Martin Gölz†, Abdelhak M. Zoubir† and Visa Koivunen*
TU Darmstadt†, Germany and Aalto University*, Finland
Email: {goelz, zoubir}@spg.tu-darmstadt.de, visa.koivunen@aalto.fi

***Abstract*—We recently proposed a novel approach to perform spatial inference using large-scale sensor networks and multiple hypothesis testing [6]. It identifies the regions in which a spatial phenomenon of interest exhibits different behavior from its nominal statistical model. To reduce the intra-sensor-network communication overhead, the raw data is pre-processed at the sensors locally and a summary statistic is send to the cloud or fusion center where the actual spatial inference using multiple hypothesis testing and false discovery control takes place. Local false discovery rates (lfdrs) are estimated to express local believes in the state of the spatial signal. In this work, we extend our approach by proposing two novel lfdr estimators stemming from the Expectation-Maximization method. The estimation bias is considered to explain the differences in performance among the compared lfdr estimators.**

***Index Terms*—Large-scale inference, local false discovery rate, density estimation, sensor networks, information fusion**

I. Introduction

Large-scale wireless sensor networks are a central technology in the Internet of Things (IoT). Formed by interconnecting a large number of heterogeneous devices with different sensing capabilities, large-scale sensor networks can be used to monitor various physical phenomena across a spatial domain of interest [1]. By a *spatial signal* or *spatial phenomenon*, we mean any physical process that varies smoothly as a function of location. Spatial signals occur in many areas, such as monitoring the environment, weather or radio spectrum. A common task in many practical applications is the identification of the *alternate region* $\mathcal{H}_1$, in which the spatial phenomenon exhibits anomalous, interesting or simply different behavior from its nominal one [2]. $\mathcal{H}_1$ may be composed of multiple locally continuous sub-regions, as the sketch in Fig. 1 illustrates.

Dedicated inference algorithms are needed to extract the relevant information from the large amount of data acquired by distributed sensors. Due to the large number of sensors and the congested radio spectrum, the communication between the sensors and the cloud or fusion center (FC) must be constrained. Moreover, the sensors are typically battery-powered and reduced communication will expand the lifespan of each sensor and the entire network [3], [4].

 The work of M. Gölz is supported by the German Research Foundation (DFG) under grant ZO 215/17-2. Author for correspondence: M. Gölz.

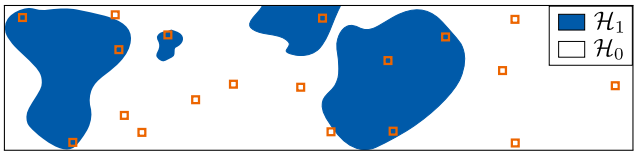


Figure 1: A spatial phenomenon whose alternate region $\mathcal{H}_1$ is composed of four distinct, locally continuous sub-regions. Orange squares indicate sensor locations.

In this work, we propose a novel method to learn the behavior of a spatial phenomenon of interest using a large-scale sensor network and the expectation-maximization (EM) maximum likelihood estimator (MLE) [5]. This method expands our previously proposed multiple hypothesis testing (MHT) approach for spatial signals [6], which can be summarized as follows. First, the observation area is represented by a discrete grid. Each grid point is assigned a binary local state. If the spatial signal behaves as expected at a certain grid point, we say that it is locally in its nominal state, i.e., the null hypothesis H_0 holds at this grid point. If this is not the case, i.e., if the spatial field exhibits different, anomalous or interesting behavior, the *alternative* H_1 holds at this grid point. The alternate region $\mathcal{H}_1$ is thus the subset of all grid points at which H_1 holds. While the statistical properties of the spatial signal can be very different at different locations under H_1, the common property is that the behavior of the spatial signal is statistically different from the nominal local probability model. To identify the region $\mathcal{H}_1$, a binary decision between H_0 and H_1 has to be made for each point of the grid individually. This is a MHT problem, since a large number of simultaneous binary decisions are made and local statistical models may differ over space. The sensors are assumed to be located at a (potentially sparse) subset of distinct grid points. No assumptions on the geometry or configuration of the sensor locations are made. Each sensor condenses its the local observations into a single scalar local summary statistic, which is then communicated to the FC. Based on these local statistics, the FC or cloud estimates the lfdr [7]–[10] at each sensor location. By appropriate interpolation using the lfdrs at the sensor locations, inference can be made also in areas between the sensor locations. Finally, the estimated alternate region $\hat{\mathcal{H}}_1$ is found by grouping the grid points depending on whether a decision in favor or H_0 or H_1 was made. This approach provides Type I error control by guaranteeing that the false discovery rate (FDR), i.e., the expected proportion of grid points within $\hat{\mathcal{H}}_1$ where actually the null hypothesis H_0

is in place, is below a nominal level α.

By pre-processing the sensor-level data locally, but performing inference in the cloud or at the FC, this framework reduces the communication overhead while gaining from fusing the information from all sensors. In comparison to existing lfdr-based methods, the lfdr-based inference method proposed in [6] was shown to be computationally more efficient. Finally, the framework enables the incorporation of spatially varying empirical Bayes priors in different locations [11]. This allows for taking advantage of spatial smoothness assumption without the need to deploy a specific mathematical signal model that would limit the range of practical applications.

In this work, we introduce a novel EM-based lfdr estimator. In addition, we provide an extensive comparison of the proposed method with a variety of existing lfdr estimators. Moreover, we extend the lfdr estimator from [6] with the EM algorithm for mixture model estimation. It yields a very stable lfdr estimator for all sizes of sensor networks that scales well to networks with a large number of sensors.

The paper is structured as follows. In the next section, we introduce the spatial inference framework formally and discuss the impact of lfdr estimation bias on the inference performance. In Sec. III, we present our proposed method. We conclude with a performance comparison of different lfdr estimators on radio signals in Sec. IV.

II. Problem Formulation

In this work, we focus on the estimation of lfdrs and decision making at the sensor locations, since the interpolation of the estimated lfdrs and hypothesis testing at regions in between sensors can be done by a subsequent step, similar to the approach in [6]. Thus, the goal in this work is to estimate the mutually exclusive sets $\mathcal{H}_1^N$ and $\mathcal{H}_0^N$, which respectively contain all sensor that are located in the regions $\mathcal{H}_1$ and $\mathcal{H}_0$.

A. System model

The observation area is discretized by a fine grid of Q grid points. The sensors are located at a subset of $N \leq Q$ distinct grid points. Per sensor, the true local hypothesis $H_n = \{\mathrm{H}_0, \mathrm{H}_1\}$ encodes the local state of the spatial phenomenon. If the null hypothesis holds, $H_n = \mathrm{H}_0$, the spatial phenomenon is in its nominal state at the location of $n \in [N]$. If the spatial phenomenon deviates from its nominal state at n, the alternative holds $H_q = \mathrm{H}_1$. Grouping the sensors according to their true local hypothesis yields two mutually exclusive sets $\mathcal{H}_0^N = \{n \in [N] : H_n = \mathrm{H}_0\}$ and $\mathcal{H}_1^N = \{n \in [N] : H_n = \mathrm{H}_1\}$. Applying the estimators

$$\hat{\mathcal{H}}_1^N = \underset{\mathcal{H}\subseteq[N]}{\operatorname{argmax}} \left\{ |\mathcal{H}| : \frac{1}{|\mathcal{H}|} \cdot \sum_{n\in\mathcal{H}} \Pr(H_n = \mathrm{H}_0) \leq \alpha \right\}, \quad (1)$$

$$\hat{\mathcal{H}}_0^N = [N] \setminus \hat{\mathcal{H}}_1^N,$$

controls the Type I errors by FDR control such that

$$\mathrm{FDR} = \mathrm{E}\left[\frac{\sum_{n\in\mathcal{H}_0^N} \mathbb{1}\{n \in \hat{\mathcal{H}}_1^N \,|\, n \in \mathcal{H}_0^N\}}{\max\{|\hat{\mathcal{H}}_1^N|, 1\}} \right], \quad (2)$$

the fraction of false discoveries among $\hat{\mathcal{H}}_1^N$, does not surpass α [10]. A false discovery occurs, if the alternative H_1 is accepted for a sensor n with $H_n = \mathrm{H}_0$. $\mathbb{1}\{\cdot\}$ is the indicator function.

In [6], the lfdrs serve as the local null probabilities $\Pr(H_n = \mathrm{H}_0)$. To this end, each sensor $n \in [N]$ records observation data locally and condenses its local believe in $H_n = \mathrm{H}_0$ into p_n, the local p-value. The statistical properties of the local p-values depend on the local state of the field. Thus, the probability density function (PDF) differ under H_0 and H_1,

$$f_{P_n}(p) = \begin{cases} f_{P|\mathrm{H}_0}(p) = \mathcal{U}[0,1], & H_n = \mathrm{H}_0, \\ f_{P_n|\mathrm{H}_1}(p) & H_n = \mathrm{H}_1, \end{cases} \quad (3)$$

$\forall\, n \in [N]$. The PDF for p-values from all sensors $n \in \mathcal{H}_0^N$ follow the uniform distribution $\mathcal{U}[0,1]$. The p-values from sensors in $\mathcal{H}_1^N$, however, can differ across sensors.

Each sensor forwards its p-value to the FC or cloud. The marginal p-value PDF of the elements of $\mathcal{P}^N = \{p_n\}_{n\in[N]}$ is

$$f_P(p) = \pi_0 + (1-\pi_0) f_{P|\mathrm{H}_1}(p), \quad (4)$$

where $f_{P|\mathrm{H}_1}(p) = |\mathcal{H}_1^N|^{-1} \sum_{n\in\mathcal{H}_1^N} f_{P_n|\mathrm{H}_1}(p)$ and $\pi_0 = |\mathcal{H}_0^N|/N$. Hence, the lfdr is

$$\mathrm{lfdr}_n = \frac{\pi_0}{f_P(p_n)} = \frac{\pi_0}{\pi_0 + (1-\pi_0)\cdot f_{P|\mathrm{H}_1}(p_n)}. \quad (5)$$

π_0 can be interpreted as the prior probability of the null hypothesis, while lfdr_n is the posterior null probability [10].

B. The impact of lfdr estimation bias on the FDR

Estimating the alternate region with Eq. (1) controls the FDR at nominal level α when using the true lfdrs. The lfdrs lfdr_n have to be estimated $\forall n \in [N]$, since π_0 and $f_P(p)$ are unknown. Their estimates $\hat{\mathrm{lfdr}}_n$ are found by replacing $f_P(p)$, π_0 in Eq. (5) with estimates $\hat{f}_P(p)$, $\hat{\pi}_0$ [7]–[10], [12], [13].

To analyze the impact of bias on lfdr-based decision making, we distinguish between upward and downward bias at sensors in $\mathcal{H}_1$ and $\mathcal{H}_0$. Upward lfdr bias at sensors in $\mathcal{H}_1$ means that the lfdr is overestimated at sensors located in the alternate region, $b_{\mathcal{H}_1} = \mathrm{E}_{n\in\mathcal{H}_1}[\hat{\mathrm{lfdr}}_n - \mathrm{lfdr}_n] > 0$. Downward bias means that the lfdr is underestimated, $b_{\mathcal{H}_1} < 0$. Upwards (downwards) bias for lfdrs at sensors where H_0 holds is defined equivalently, $b_{\mathcal{H}_0} = \mathrm{E}_{n\in\mathcal{H}_0}[\hat{\mathrm{lfdr}}_n - \mathrm{lfdr}_n] > 0$ ($b_{\mathcal{H}_0} < 0$). Distinguishing up- and downwards bias across $\mathcal{H}_1$ and $\mathcal{H}_0$ is important, since positive or negative $b_{\mathcal{H}_1}$, $b_{\mathcal{H}_0}$ have a different impact on FDR control and detection power.

Assume that $b_{\mathcal{H}_0} > 0$, i.e., the lfdr is biased upwards for sensors in $\mathcal{H}_0$. In Eq. (1), summing the estimated lfdrs over sensors located in $\mathcal{H}_0$ yields a larger value than summing over the true lfdrs. Hence, the FDR control limit α may be reached for a smaller subset of sensors, i.e., the number of false discoveries in $\hat{\mathcal{H}}_1^N$ may be reduced. Thus, $b_{\mathcal{H}_0} > 0$ can lead to a smaller empirical FDR than the nominal α. In contrast, if $b_{\mathcal{H}_0} < 0$, $\hat{\mathcal{H}}_1^N$ potentially contains more false discoveries than if the true lfdrs were used. Thus, a downward lfdr bias at sensors located in $\mathcal{H}_0$ deteriorates FDR control. In contrast, $b_{\mathcal{H}_1}$ affects the detection power. Smaller lfdr values at sensors in $\mathcal{H}_1$ lead to a smaller value of the sum in Eq. (1), which allows a larger subset of sensors into $\hat{\mathcal{H}}_1^N$. Hence, downward lfdr bias at sensors in $\mathcal{H}_1$ can increase the detection power.

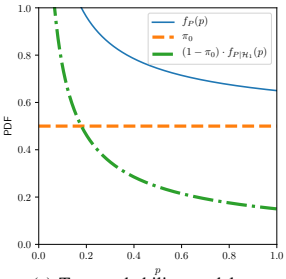
The ideal lfdr estimator would hence exhibit a different bias for sensors in different regions, $b_{\mathcal{H}_0} > 0, b_{\mathcal{H}_1} < 0$. Since the true local hypotheses are unknown, the design of such an lfdr estimator becomes very challenging. By analyzing the

(a) FDR

$\mu_{\mathcal{H}_1} \backslash \mu_{\mathcal{H}_0}$	0	.075	−.075
0	.11	.06	.16
.075	.11	.06	.17
.0325	.11	.06	.16
−.075	.11	.07	.16

(b) Detection power

$\mu_{\mathcal{H}_1} \backslash \mu_{\mathcal{H}_0}$	0	.075	−.075
0	.38	.38	.37
.075	.25	.25	.25
.0325	.31	.32	.31
−.075	.48	.49	.47



(c) True probability model

Figure 2: The impact of lfdr bias on $\hat{\mathcal{H}}_1^N$. A non-positive bias for sensors in the null region $\mathcal{H}_0$ can lead to violation of the nominal FDR level $\alpha = .1$.

differences in $b_{\mathcal{H}_0}, b_{\mathcal{H}_1}$ a better understanding of the disparity in FDR and detection power for different methods is obtained.

To illustrate the effects of lfdr bias on FDR and detection power, we consider a toy example. We generate $N = 1\,000$, p-values such that the prior null probability $\pi_0 = .5$ and $f_{P_n|\mathrm{H}_1}(p_n) = f_{P|\mathrm{H}_1}(p) = a \cdot p^a$, i.e., a beta distribution with $a = .3$. Then, $f_P(p)$ obeys the popular beta-and-uniform mixture (BUM) [12] depicted in Fig. 2c. We simulate differently biased lfdr estimators by adding a random distortion δ to the true lfdrs. We use $\hat{\mathrm{lfdr}}_n = \min\{\max\{\mathrm{lfdr}_n + \delta, 0\}, 1\}$,

$$\delta \sim \begin{cases} \mathcal{N}(\mu_{\mathcal{H}_0}, \sigma^2), & n \in \mathcal{H}_0 \\ \mathcal{N}(\mu_{\mathcal{H}_1}, \sigma^2), & n \in \mathcal{H}_1. \end{cases} \tag{6}$$

While max and min guarantee that $\hat{\mathrm{lfdr}}_n$ is in $[0, 1]$, the model in Eq. (6) allows to compare biases at sensors located in the null and alternative regions. In Fig. 2a and 2b, we show the resulting empirical FDR and detection power for different cases and $\sigma = .1$. The results in Fig. 2 are averaged over 100 realizations. Fig. 2 confirms our observations. A downward lfdr bias for sensors where H_0 is in place leads to a violation of the desired FDR control level, as the the last column shows. In addition, a downward lfdr bias for sensors where H_1 is in place increases detection power, as the last row shows. A comparison of the second and the third row justify the observation that already a small reduction in $\mu_{\mathcal{H}_1}$ can lead to a significant difference in detection power. However, the results reveal another interesting aspect. In particular, the top left entry of the table shows that even a non-biased lfdr estimator can lead to a violation of the desired FDR level, due to the variance of the estimation error. To avoid FDR violation, traditional FDR control methods often aim to slightly overestimate π_0 [10], which induces a small non-zero upward lfdr bias, see Eq. (5). However, keeping this bias small, in particular for $\hat{\mathrm{lfdr}}_n, n \in [N]$ is essential to avoid a loss in detection power.

III. EM-based lfdr Estimation

In this section, we introduce our proposed EM-based lfdr estimator. This method estimates the lfdrs by estimating the components $\hat{\pi}_0$ and $f_P(p)$, which are then inserted into Eq. (5).

A. Probability density model

EM requires a parametric model for $f_P(p)$. The true $f_{P|\mathrm{H}_1}(p)$ is a mixture of $|\mathcal{H}_1^N|$ local PDFs that can be of different shape and parametrization. Since only a single p-value per sensor is available at the FC, we propose to use a multi single-parameter beta distribution mixture model (MBM),

$$f_P^{\mathrm{MBM}}(p) = \sum_{k=1}^{K} w^{(k)} \cdot a^{(k)} \cdot p^{a^{(k)}-1}, \tag{7}$$

where the beta distribution parameters and mixture component weights are $a^{(k)} \geq 0$ and $w^{(k)} \in [0, 1]$, $k \in [K]$. The beta distribution is a very flexible parametric PDF model on $[0, 1]$. Its single parameter version is frequently used in p-value PDF modeling, for example in the BUM [12]. While the BUM model is sometimes [6] too rigid to model $f_P(p)$ with high accuracy, the MBM model uses K mixture components to allow more accurate modeling.

If $f_P^{\mathrm{MBM}}(p)$ is employed in lfdr estimation, $2K - 1$ model parameters and the model order K have to be found. A method proposed in [6] subdivides the total set of p-values $\mathcal{P}^N$ into subsets and fits a multivariate p-value distribution to vectors of the subsets. This multivariate PDF is then averaged over its elements to yield a one-dimensional PDF similar to the model in Eq. (7). In this work, we apply EM [5] to directly estimate $a^{(k)} \, \forall \, k \in [K]$.

B. Proposed algorithm

EM iterates an expectation (E) and a maximization (M) step [5], until a (local) maximum of the log-likelihood function (LLF)

$$\ell^{\mathcal{P}^N} = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} w^{(k)} \cdot a^{(k)} \cdot p_n^{a^{(k)}-1} \right) \tag{8}$$

is reached. The model order $K \in \mathcal{K}$ can be determined by information theoretic criteria such as the Bayesian information criterion (BIC) [14]. $\mathcal{K}$ denotes the set of candidate model orders.

Once the PDF model parameters $a^{(k)}$ and $w^{(k)}$ have been estimated, we extract the estimate $\hat{\pi}_0$ for the relative fraction of sensors where H_0 is in place and then estimate the part of p-value density associated with the alternative, $\hat{f}_{P|\mathrm{H}_1}(p)$, from $f_P^{\mathrm{MBM}}(p)$ using Pound's estimator [12]. Then, $\hat{\pi}_0$ and $\hat{f}_{P|\mathrm{H}_1}(p)$ are inserted into Eq. (5) to obtain the lfdr estimates $\hat{\mathrm{lfdr}}_n \, \forall \, n \in [N]$. Finally, the estimated lfdrs are used in Eq. (1) to determine $\hat{\mathcal{H}}_1^N$ and $\hat{\mathcal{H}}_1^N$. We refer to this method as lfdr-MBM-EM. Details are provided in Alg. 1.

C. EM-Initialization with lfdr-sMoM

EM aims at finding the MLE by maximizing the LLF. However, the problem may be highly non-convex with many local maxima, in particular, if the number of model parameters is large. EM only guarantees convergence to a local optimum. Thus, selecting poor starting points might result in bad parameter estimates or slow convergence due to EM getting stuck in a local maximum. To overcome this issue, lfdr-MBM-EM runs the EM M times with different random starting points. However, this comes at an increase in computational cost. Moreover, the necessity to optimize the LLF for a series of candidate model orders $\mathcal{K}$ adds to the method's overall computational costliness. A stopping condition for the iteration is defined such that no reduction of the BIC within two iterations terminates the loop.

Algorithm 1 lfdr-MBM-EM

Input: $\mathcal{P}^N = \{p_1, \dots, p_N\}$, $\mathcal{K}$, M, $a_{\max}^{(k)}$, ϵ, α
Output: $\hat{\mathcal{H}}_1^N$, $\hat{\mathcal{H}}_0^N$

PDF model fitting

for $K \in \mathcal{K}$ **do**
  for $m \in [M]$ **do**
    $\ell_{-1}^{\mathcal{P}^N} = \infty$, $v_n^{(k)} = 0 \, \forall \, n \in [N], k \in [K]$
    Sample $\mathcal{U}[0, a_{\max}^{(k)}]$, $\mathcal{U}[0,1]$ to initialize $\hat{a}^{(k)}$, $\hat{w}^{(k)}$, $\forall \, k \in [K]$
    Compute $\ell^{\mathcal{P}^N}$ by inserting $\hat{a}^{(k)}, \hat{w}^{(k)}$ into Eq. (8)
    while $\left(\ell^{\mathcal{P}^N} - \ell_{-1}^{\mathcal{P}^N}\right) / \ell_{-1}^{\mathcal{P}^N} < \epsilon$ **do**
      E: $v_n^{(k)} = \frac{\hat{w}^{(k)} \cdot \hat{a}^{(k)} \cdot p_n^{\hat{a}^{(k)}-1}}{\sum_{l=1}^{K} \hat{w}^{(l)} \cdot \hat{a}^{(l)} \cdot p_n^{\hat{a}^{(l)}-1}}$
      M: $\hat{a}^{(k)} = -\frac{\sum_{n=1}^{N} v_n^{(k)}}{\sum_{n=1}^{N} v_n^{(k)} \cdot \log(p_n)}$, $\hat{w}^{(k)} = \frac{\sum_{n=1}^{N} v_{n,i+1}^{(k)}}{N}$
      $\ell_{-1}^{\mathcal{P}^N} = \ell^{\mathcal{P}^N}$, update $\ell^{\mathcal{P}^N}$ with $\hat{a}^{(k)}, \hat{w}^{(k)}$ and Eq. (8)
    Store the estimates $\hat{a}_m^{(k)} = \hat{a}^{(k)}$, $\hat{w}_m^{(k)} = \hat{w}^{(k)}$ and $\ell_m^{\mathcal{P}^N} = \ell^{\mathcal{P}^N}$
  $\hat{a}_K^{(k)} = \hat{a}_{m^*}^{(k)}$, $\hat{w}_K^{(k)} = \hat{w}_{m^*}^{(k)}$, $\ell_K^{\mathcal{P}^N} = \ell_{m^*}^{\mathcal{P}^N}$ with $m^* = \underset{m \in [M]}{\operatorname{argmax}} \, \ell_m^{\mathcal{P}^N}$
  Compute $\mathrm{BIC}(K) = (2K-1) \cdot \log(N) - 2 \cdot \ell_K^{\mathcal{P}^N}$
  if $\min \mathrm{BIC}(\cdot) \notin \{\mathrm{BIC}(K), \mathrm{BIC}(K-1)\}$ **then** break
Model order selection $K^* = \underset{K \in \mathcal{K}}{\operatorname{argmin}} \, \mathrm{BIC}(K)$
Obtain $\hat{f}_P(p)$ by inserting $\hat{a}_{K^*}^{(k)}, \hat{w}_{K^*}^{(k)} \, \forall \, k \in [K^*]$ into Eq. (7)

lfdr estimation

$\hat{\pi}_0 = \min(\hat{f}_P(p))$, $\hat{f}_{P|\mathrm{H}_1}(p) = (1 - \hat{\pi}_0)^{-1}(\hat{f}_P(p) - \hat{\pi}_0)$
Insert $\hat{\pi}_0$, $\hat{f}_{P|\mathrm{H}_1}(p)$ into Eq. (5) to obtain $\widehat{\mathrm{lfdr}}_q \, \forall \, n \in [N]$

Estimation of $\mathcal{H}_1^N$, $\mathcal{H}_0^N$

Use $\Pr(H_n = \mathrm{H}_0) = \widehat{\mathrm{lfdr}}_n \, \forall n \in [N]$ in Eq. 1

On the other hand, EM's computation time can be reduced significantly by initializing the parameters with good starting values. In [6], the spectral method of moments (sMoM)-based PDF estimator lfdr-sMoM was proposed. This method divides $\mathcal{P}^N$ into equally sized subsets and fits a multivariate mixture PDF model to the resulting p-value vectors using sMoM. Finally, lfdr-sMoM averages over the elements of the multivariate mixture to obtain a p-value mixture PDF estimate that follows a similar structure as Eq. (7). It has been shown [6] to yield well-fitting probability models at lower computational cost than comparable methods from the literature. Also, lfdr-sMoM automatically selects a model order K_{sMoM}. We extend lfdr-sMoM with the EM agorithm resulting to lfdr-sMoM-EM in Alg. 1. Hence, as a fast alternative to lfdr-MBM-EM, we propose lfdr-sMoM-EM in Alg. 2. Using the mixture model parameter estimates from [6, Alg. 1, Step 1] as starting values, it consists of a single execution of EM for a fix model order. The additional inputs U, G and $\mathrm{d}(\cdot,\cdot)$ needed for lfdr-sMoM are chosen according to the recommendations in [6].

IV. Simulation Results

In this section, we compare the different lfdr estimators from the literature to our proposed methods. We first describe simulation scenarios and briefly introduce the competitors. The employed quantitative performance criteria are the FDR, the detection power and the lfdr bias for a sensor network of a size $N = 3\,000$. Finally, we compare the behavior of the methods with varying N.

Algorithm 2 lfdr-sMoM-EM

Input: $\mathcal{P}^N = \{p_1, \dots, p_N\}$, $\mathcal{K}$, U, G, $\mathrm{d}(\cdot,\cdot)$, ϵ, α
Output: $\hat{\mathcal{H}}_1^N$, $\hat{\mathcal{H}}_0^N$

PDF model fitting

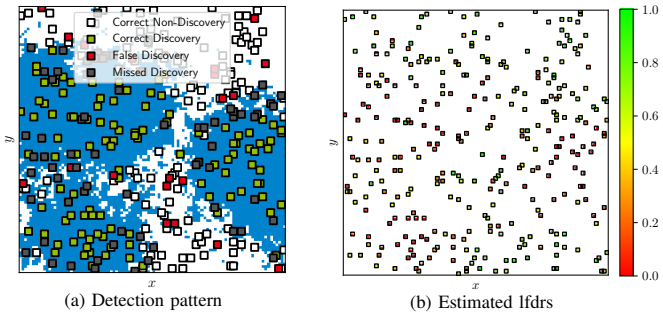
Find $\hat{f}_P^{\mathrm{init}}(p)$ from [6, Alg. 1, Step 1], inputs $\mathcal{P}^N, \mathcal{K}, G, \mathrm{d}(\cdot,\cdot), U$
Initialize $\hat{a}^{(k)}$, $\hat{w}^{(k)}$, $\forall \, k \in [K]$ with the parameters of $\hat{f}_P^{\mathrm{init}}(p)$
$\ell_{-1}^{\mathcal{P}^N} = \infty$, $v_n^{(k)} = 0 \, \forall \, n \in [N], k \in [K]$
Compute $\ell^{\mathcal{P}^N}$ by inserting $\hat{a}^{(k)}, \hat{w}^{(k)}$ into Eq. (8)
while $\left(\ell^{\mathcal{P}^N} - \ell_{-1}^{\mathcal{P}^N}\right) / \ell_{-1}^{\mathcal{P}^N} < \epsilon$ **do**
  E: $v_n^{(k)} = \frac{\hat{w}^{(k)} \cdot \hat{a}^{(k)} \cdot p_n^{\hat{a}^{(k)}-1}}{\sum_{l=1}^{K} \hat{w}^{(l)} \cdot \hat{a}^{(l)} \cdot p_n^{\hat{a}^{(l)}-1}}$
  M: $\hat{a}^{(k)} = -\frac{\sum_{n=1}^{N} v_n^{(k)}}{\sum_{n=1}^{N} v_n^{(k)} \cdot \log(p_n)}$, $\hat{w}^{(k)} = \frac{\sum_{n=1}^{N} v_{n,i+1}^{(k)}}{N}$
  $\ell_{-1}^{\mathcal{P}^N} = \ell^{\mathcal{P}^N}$, update $\ell^{\mathcal{P}^N}$ with $\hat{a}^{(k)}, \hat{w}^{(k)}$ and Eq. (8)
Obtain $\hat{f}_P(p)$ by inserting $\hat{a}^{(k)}, \hat{w}^{(k)} \, \forall \, k \in [K]$ into Eq. (7)

lfdr estimation

$\hat{\pi}_0 = \min(\hat{f}_P(p))$, $\hat{f}_{P|\mathrm{H}_1}(p) = (1 - \hat{\pi}_0)^{-1}(\hat{f}_P(p) - \hat{\pi}_0)$
Insert $\hat{\pi}_0$, $\hat{f}_{P|\mathrm{H}_1}(p)$ into Eq. (5) to obtain $\widehat{\mathrm{lfdr}}_q \, \forall \, n \in [N]$

Estimation of $\mathcal{H}_1^N$, $\mathcal{H}_0^N$

Use $\Pr(H_n = \mathrm{H}_0) = \widehat{\mathrm{lfdr}}_n \, \forall n \in [N]$ in Eq. 1



(a) Detection pattern (b) Estimated lfdrs

Figure 4: Exemplary run for ScA, $N = 310$, lfdr-sMoM-EM is applied. For a), the nominal FDR is $\alpha_{\mathrm{FDR}} = 10\%$. Green, red and gray indicate true, false and missed discoveries, blue the entire alternate region $\mathcal{H}_1$. Boxes mark sensor locations. b) shows the estimated sensor-level lfdrs.

A. Setup

The spatial phenomenon we monitor is simulated radio frequency electromagnetic field. In this context, the alternate region $\mathcal{H}_1$ is composed of all grid points where a radio frequency signal is present, i.e., where the spectrum is occupied. The signals are generated using the nonuniform sampling method from [15], which simulates the path-loss and shadow fading. The shape and size of the null and alternate regions are influenced by the number of sources and the propagation conditions. In a suburban environment, line-of-sight (LOS) propagation dominates, which results in rather smooth region boundaries. Non-LOS (NLOS) propagation caused by the large number of scatterers in an urban environment lead to rougher boundaries.

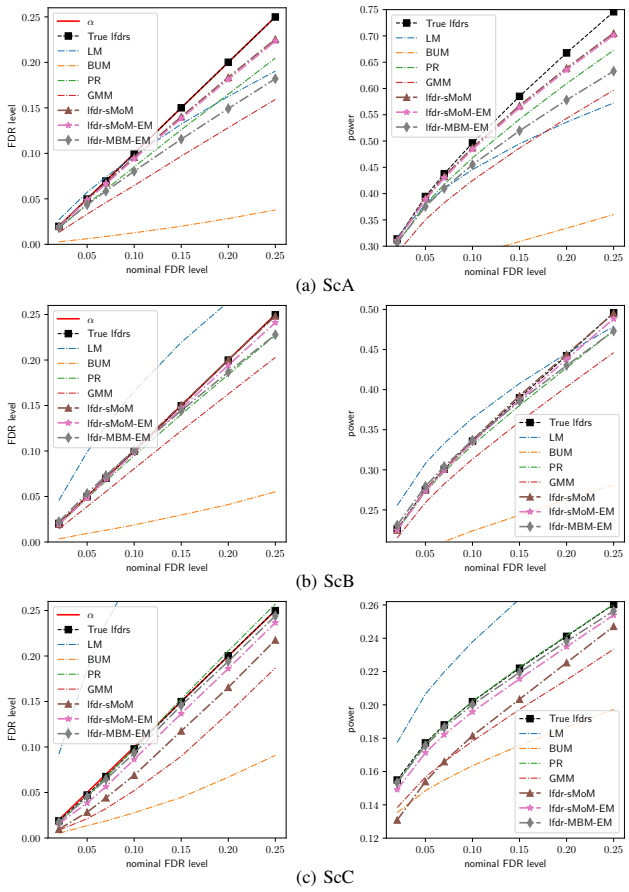
The observation area is represented by a 100×100 grid. We provide results for three scenarios.

ScA: 5 sources, suburban environment, high signal strength.

ScB: 8 sources, suburban environment, low signal strength.

ScC: 2 sources, urban environment, low signal strength.

In ScA, $\pi_0 \approx 55\%$, in ScB $\pi_0 \approx 34\%$ and in ScC $\pi_0 \approx 10\%$.



(a) ScA

(b) ScB

(c) ScC

Figure 5: The FDR and detection power for different scenarios with $N = 3\,000$ with all methods. Both proposed methods outperform the competitors in ScC. In ScA, the proposed lfdr-sMoM-EM performs as good as lfdr-sMoM. In ScB, the proposed methods and the best competitors perform similarly.

The smaller the null proportion π_0, the more challenging. Each sensor computes its local p-value based on the received signal energy, which is observed for $T = 256$ samples in all scenario. The size of the sensor network N is varied. We consider smaller to medium sized large-scale sensor networks, with $N \in [110, 10\,000]$. The results are averaged over 200 independent realizations in each experiment. The code to reproduce the result is available on github.

B. Compared methods and their parameters

For lfdr-MBM-EM, we set $\mathcal{K} = [20]$ to allow for a very flexible MBM PDF. The algorithm most often selects a model order $K \ll 20$. The number of starting points in EM is $M = 5$. We found that choosing $M > 5$ does not improve the results much but adds to the computational complexity. The maximum initial value of $\hat{a}^{(k)}$ is $a_{\max}^{(k)} = 2$ in all simulation examples. The relative convergence threshold for EM is $\epsilon = 1e{-}5$, which yields a good trade-off between run time and performance. For lfdr-sMoM-EM, we additionally use the previously [6] well-working $U = G = 10$ along the Wasserstein (WS) distance as the difference function $\mathrm{d}(\cdot, \cdot)$.

Competitors: We use sMoM-lfdr [6] with the same parameters as lfdr-sMoM-EM. In addition, we apply the traditional BUM model-based MLE approach from [12]. We also apply a set of z-score-based lfdr estimators. There exists one-to-one mapping between z-scores and p-values, hence conversion can be made both ways. We show results obtained with Lindsey's method (LM) [10]. LM fits an exponential family model to the z-scores. Also, we fit a Gaussian mixture model (GMM) to the z-scores. Finally, predictive recursion (PR) recursion [16], [17] is applied. PR is frequently used, for example in [18], [19], to estimate lfdrs based on z-scores, since it leads highly accurate results at comparably low computational cost.

C. Numerical results

The accuracy of estimates $\hat{\mathcal{H}}_1^N$ and $\hat{\mathcal{H}}_0^N$ can be evaluated by assessing the FDR, i.e., the fraction of sensors assigned to $\mathcal{H}_1^N$ that are actually located in $\mathcal{H}_0$ and the detection power. The latter is percentage of sensors located in $\mathcal{H}_1$ that is assigned to $\hat{\mathcal{H}}_1^N$ by the applied procedure. Fig. 5 displays the FDR (in the left column) and detection powers (right column) obtained with the different procedures for different nominal FDR levels α with a network of moderate size $N = 3\,000$.

The competitors from the literature perform quite differently. BUM provides very low detection power in all scenarios. LM performs acceptable for small α in ScA. but violates the nominal level in ScB, ScC. GMM controls the FDR, but provides the second lowest detection power. The best performing competitor is PR. Our proposed EM with random initialization, lfdr-MBM-EM, lacks detection power in comparison to PR, lfdr-sMoM and lfdr-sMoM-EM for ScA. For ScC, on the other hand, lfdr-MBM-sMoM performs similarly to PR, but better than lfdr-sMoM and lfdr-sMoM-EM. Our second proposed method, lfdr-sMoM-EM, performs in general similar to lfdr-sMoM. However, running the additional EM does provide a performance gain for ScC.

In Sec. II-B, we discussed the impact of lfdr estimation bias on detection results. The lfdr bias achieved with the best performing methods is shown in Tab. I. As the results for ScA show, the methods that yield a larger detection power also exhibit reduced lfdr bias. In particular, we notice the difference in bias at sensors in the null regions $b_{\mathcal{H}_0}$ and the bias at sensors in the alternate region $b_{\mathcal{H}_1}$. While $b_{\mathcal{H}_0}$ is very similar for the different methods, significant differences in $b_{\mathcal{H}_1}$ are observed. lfdr-sMoM and lfdr-sMoM-EM with smallest $b_{\mathcal{H}_1}$ also provide the largest detection power. In ScB, the present bias difference translates into a detection power difference at larger nominal FDR α. The bias differences in ScC are marginal.

We now compare the FDR, detection power and run time as a function of network size. In ScA, ScB, lfdr-sMoM and the proposed lfdr-sMoM-EM provide the largest detection power. The proposed lfdr-sMoM-EM has strictly controls the FDR at the nominal level, while lfdr-sMoM slightly exceeds α for small network sizes. Thus, the proposed method has the advantage of yielding very stable inference results with the FDR below nominal level even for small network sizes. PR is in general more conservative, i.e., yields lower power and lower FDR. In ScC, PR is the only method to violate the nominal FDR level while the second proposed method, lfdr-MBM-EM, provides slighly larger detection power than lfdr-sMoM-

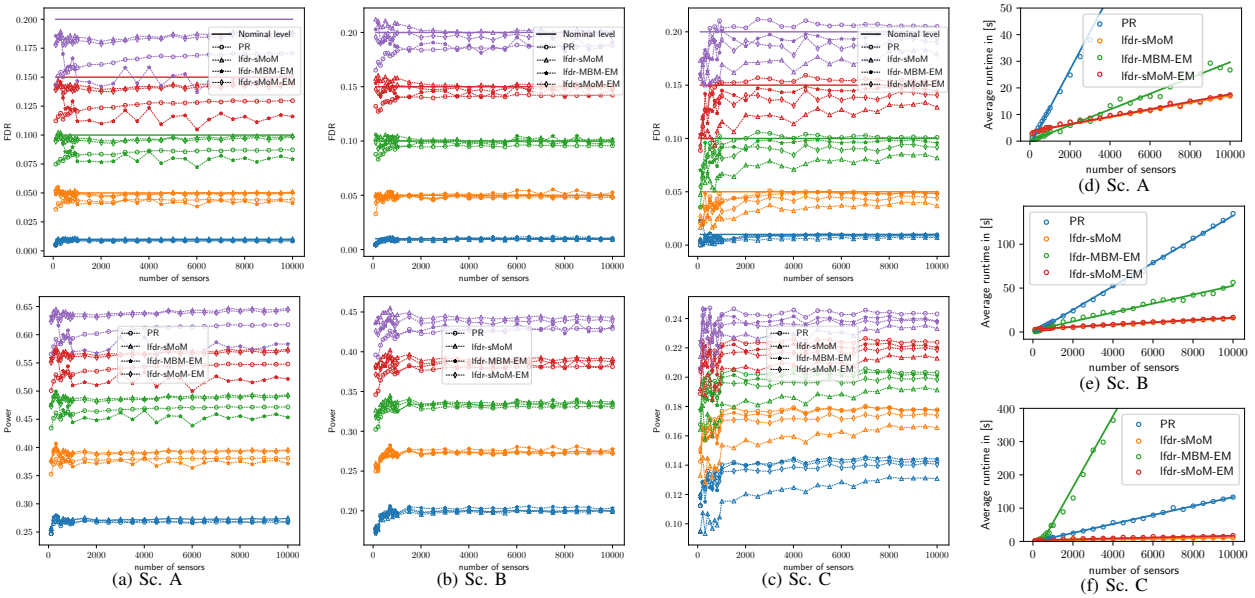


Figure 6: The first three columns display the FDR (top row) and detection power (bottom row) over different network sizes N and nominal FDR levels α. The fourth column compares the average run times per method.

Table I: The bias for the estimated lfdrs with different methods, $N = 3000$. A smaller $b_{\mathcal{H}_1}$, leads to a larger detection power in Fig. 5

		b	$b_{\mathcal{H}_0}$	$b_{\mathcal{H}_1}$
Sc. A	PR	.088	.111	.068
	lfdr-sMoM	.079	.110	.053
	lfdr-sMoM-EM	.081	.112	.055
	lfdr-MBM-EM	.157	.126	.194
Sc. B	PR	.06	.066	.05
	lfdr-sMoM	.055	.066	.033
	lfdr-sMoM-EM	.059	.07	.039
	lfdr-MBM-EM	.101	.112	.08
Sc. C	PR	$-.004$	$-.002$	.003
	lfdr-sMoM	.011	.012	.009
	lfdr-sMoM-EM	.016	.016	.012
	lfdr-MBM-EM	.016	.016	.014

EM and lfdr-sMoM. lfdr-sMoM-EM is only marginally slower than lfdr-sMoM, which indicates, that lfdr-sMoM provides good initial parameter values for EM. PR scales considerably worse with N. lfdr-MBM-EM exhibits very different scaling properties for different scenarios.

V. Conclusion

Methods for detecting regions where a spatial signal behaves differently from what is expected were proposed. lfdr estimators were developed and their statistical performance such as the impact of lfdr bias were studied. In addition, two novel Expectation-Maximization lfdr estimators were proposed. Inference with the proposed Expectation-Maximization method and sophistically determined initial parameter values maintains FDR control and yields high detection power at moderate computational cost for a wide range of sensor network sizes.

VI. References

[1] E. Arias-de-Reyna, P. Closas, D. Dardari, and P. M. Djuric, "Crowd-based learning of spatial fields for the internet of things: From harvesting of data to inference," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 130–139, Sep. 2018.

[2] R. Nowak and U. Mitra, "Boundary estimation in sensor networks: Theory and methods," in *Information Processing in Sensor Networks*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 80–95.

[3] X. Wang, G. Li, and P. K. Varshney, "Detection of sparse signals in sensor networks via locally most powerful tests," *IEEE Signal Process. Lett.*, vol. 25, no. 9, pp. 1418–1422, Sep. 2018.

[4] E. Nitzan, T. Halme, and V. Koivunen, "Bayesian methods for multiple change-point detection with reduced communication," *IEEE Trans. Signal Process.*, vol. 68, pp. 4871–4886, 2020.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.

[6] M. Gölz, A. M. Zoubir, and V. Koivunen, "Multiple hypothesis testing framework for spatial signals," Aug. 27, 2021. arXiv: 2108.12314.

[7] B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher, "Empirical Bayes analysis of a microarray experiment," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1151–1160, 2001.

[8] B. Efron, "Local false discovery rates," Tech. Rep., 2005.

[9] ——, "Microarrays, empirical Bayes and the two-groups model," *Statist. Sci.*, vol. 23, no. 1, pp. 1–22, Feb. 2008.

[10] ——, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, ser. Institute of Mathematical Statistics Monographs. Cambridge, UK: Cambridge University Press, 2010.

[11] M. Gölz, A. Zoubir, and V. Koivunen, "Improving inference for spatial signals by contextual false discovery rates," in *Proc. 2022 IEEE Int. Conf. Acoust. Speech Signal Process.*, IEEE, May 2022.

[12] S. Pounds and S. Morris, "Estimating the occurence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values," *Bioinformatics*, vol. 19, pp. 1236–1242, Jan. 2003.

[13] X. Chen, D. G. Robinson, and J. D. Storey, "The functional false discovery rate with applications to genomics," *Biostatistics*, vol. 22, no. 1, pp. 68–81, May 2019.

[14] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, Mar. 1978.

[15] X. Cai and G. Giannakis, "A two-dimensional channel simulation model for shadowing processes," *IEEE Trans. Veh. Technol.*, vol. 52, no. 6, pp. 1558–1567, Nov. 2003.

[16] M. A. Newton, "On a nonparametric recursive estimator of the mixing distribution," *Sankhyā Ser. A*, vol. 64, no. 2, pp. 306–322, 2002.

[17] R. Martin and S. T. Tokdar, "A nonparametric empirical Bayes framework for large-scale multiple testing," *Biostatistics*, vol. 13, no. 3, pp. 427–439, Nov. 2011.

[18] J. G. Scott, R. C. Kelly, M. A. Smith, P. Zhou, and R. E. Kass, "False discovery rate regression: An application to neural synchrony detection

in primary visual cortex," *J. Amer. Statist. Assoc.*, vol. 110, no. 510, pp. 459–471, Apr. 2015.

[19] R. Martin, "On nonparametric estimation of a mixing density via the predictive recursion algorithm," 2018. arXiv: 1812 . 02149 [stat.ME].