Spatial Inference Using Censored Multiple Testing With FDR Control

TECHNISCHE UNIVERSITÄT DARMSTADT

2023 International Conference on Acoustics, Speech and Signal Processing

Martin Gölz, Abdelhak M. Zoubir and Visa Koivunen

SIGNAL PROCESSING GROUP

Aalto University
School of Electrical Engineering

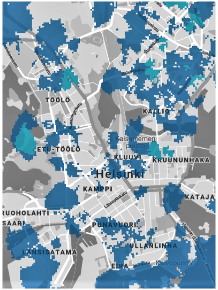
Signal Processing Group
Technische Universität Darmstadt

Dept. of Signal Processing and Acoustics
Aalto University

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Spatial signals
  - spatially smoothly varying physical phenomena
  - occur in a large number of practical applications
    - RADAR
    - wireless
    - environmental monitoring
    - smart buildings
    - cyber-physical systems (e.g. smart grids)
    - and many more



Screenshot taken from elisa.fi/kuuluvuus/

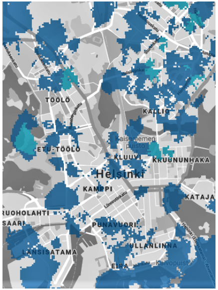
Aalto University
School of Electrical Engineering

Motivation

- Spatial signals
  - spatially smoothly varying physical phenomena
  - occur in a large number of practical applications
    - RADAR
    - wireless
    - environmental monitoring
    - smart buildings
    - cyber-physical systems (e.g. smart grids)
    - and many more

> **The spatial inference problem:** identifying *interesting* areas under strict error control

- Interesting areas: spatially smooth subregions of *anomalous*, *deviating*, *unexpected*, ..., signal behavior



Screenshot taken from elisa.fi/kuuluvuus/

Motivation

Recently proposed method

TECHNISCHE UNIVERSITÄT DARMSTADT

- Spatial signal monitored by large-scale wireless sensor network (IoT)
- Recently proposed: Multiple hypothesis testing (MHT) approach with false discovery rate (FDR) control

→ M. Gölz, et al., *Multiple Hypothesis Testing Framework for Spatial Signals*. IEEE TSIPN, 2022.

→ M. Gölz, et al. *Estimating Test Statistic Distributions for Multiple Hypothesis Testing in Sensor Networks*. CISS 2022.

→ M. Gölz, et al. *Improving Inference for Spatial Signals by Contextual False Discovery Rates*. 2022 ICASSP.

Motivation

Recently proposed method

TECHNISCHE UNIVERSITÄT DARMSTADT

- Spatial signal monitored by large-scale wireless sensor network (IoT)
- Recently proposed: Multiple hypothesis testing (MHT) approach with false discovery rate (FDR) control

  → M. Gölz, et al., *Multiple Hypothesis Testing Framework for Spatial Signals*. IEEE TSIPN, 2022.

  → M. Gölz, et al. *Estimating Test Statistic Distributions for Multiple Hypothesis Testing in Sensor Networks*. CISS 2022.

  → M. Gölz, et al. *Improving Inference for Spatial Signals by Contextual False Discovery Rates*. 2022 ICASSP.

- Finds interesting areas by the (contextual) local false discovery rate ((c)lfdr)
- Power and communication bandwidth efficient
  - works well also when quantized *p*-values with few bits are used

Aalto University School of Electrical Engineering

Motivation

Recently proposed method

- Spatial signal monitored by large-scale wireless sensor network (IoT)
- Recently proposed: Multiple hypothesis testing (MHT) approach with false discovery rate (FDR) control

In this work: Further extend of sensor lifetime & improve communication efficiency by censoring

- only sensors providing highly valuable local information transmit to the cloud or fusion center (FC)

Contribution

1. Two novel censored inference methods that considerably reduce the average number of transmissions (ANT)
   1.1 sensors make decide on transmission themselves
   1.2 FC or cloud control transmissions
2. An analytical evaluation of the impact of censoring on FDR control and detection power
3. A mathematical criterion to automatically stop transmitting when most true positives have been identified

Considerably improve communication efficiency while maintaining FDR control and detection power

Spatial Inference by Multiple Hypothesis Testing

System model

TECHNISCHE UNIVERSITÄT DARMSTADT

- Discrete spatial grid model of Q points is employed
- $H_q \in \{\mathrm{H}_0, \mathrm{H}_1\}$: the true hypothesis at grid point $q \in [Q]$

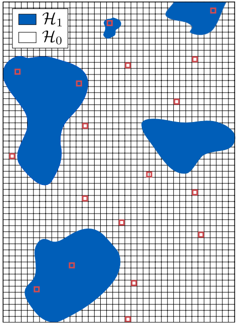
$$H_q = \mathrm{H}_0: \quad \textit{signal in nominal state at } q$$
$$H_q = \mathrm{H}_1: \quad \textit{signal in any deviating state at } q$$

- Solve the spatial inference problem by identifying

$$\mathcal{H}_0 = \{q \in [Q] : H_q = \mathrm{H}_0\}: \quad \textit{the null region}$$
$$\mathcal{H}_1 = \{q \in [Q] : H_q = \mathrm{H}_1\}: \quad \textit{the interesting region}$$

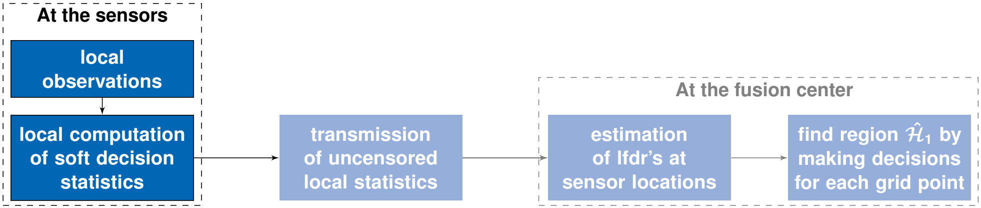
- Many hypotheses to test simultaneously $\rightarrow$ MHT
  - grid point hypotheses enable the identification of interesting areas with strict error control

  $\rightarrow$ problem is different from distributed detection



Aalto University School of Electrical Engineering

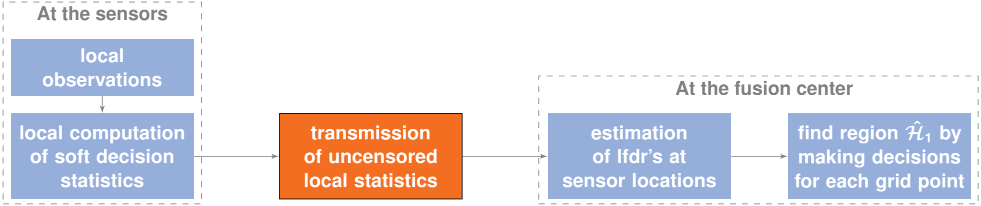
The censored MHT spatial inference framework



- Local soft decision statistic for sensor $n \in [N]$: p-value p_n
  - measurements, local signal and noise models available (or estimated) at each sensor
  - small p-values indicate weak support for H_0

Spatial Inference by Multiple Hypothesis Testing

The censored MHT spatial inference framework

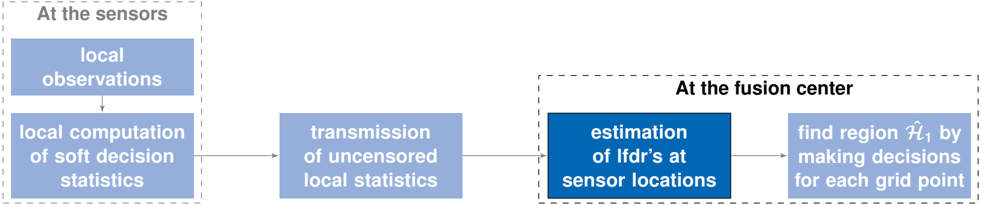


- No censoring: each node transmits p_n to the fusion center (FC)
- Idea of censoring
  - many nodes provide redundant information
  - $\rightarrow$ Let only informative nodes transmit and censor the rest
- With censoring: a fraction of all nodes transmit p_n to the FC

Spatial Inference by Multiple Hypothesis Testing

The censored MHT spatial inference framework

TECHNISCHE UNIVERSITÄT DARMSTADT



$$\mathrm{lfdr}_{n,\kappa} = \frac{\pi_0}{\pi_0 + (1-\pi_0)\cdot\left[t_n \cdot f_{P|H_1}(p_n) + (1-t_n)\cdot \frac{1-F_{P|H_1}(\kappa)}{1-\kappa}\right]}$$

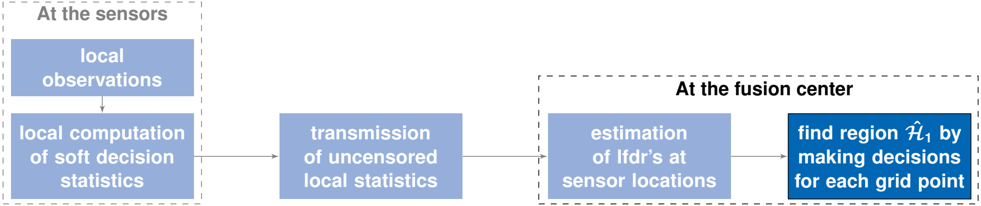
- κ : lower bound of no-send region
- π_0 : percentage of sensors where H_0 holds
- t_n, p_n: transmission indicator, p-value of sensor $n \in [N]$
- $f_{P|\mathrm{H}_1}(p), F_{P|\mathrm{H}_1}(p)$: PDF, CDF of p-values at sensors $n \in \mathcal{H}_1$

Aalto University
School of Electrical Engineering

Spatial Inference by Multiple Hypothesis Testing

The censored MHT spatial inference framework

TECHNISCHE UNIVERSITÄT DARMSTADT



- Estimate the set of sensors located in interesting regions

$$\hat{\mathcal{H}}_1^{(\mathrm{sen})} = \underset{\mathcal{H}\subseteq[N]}{\arg\max} \left\{ |\mathcal{H}| \;\middle|\; \frac{1}{|\mathcal{H}|} \cdot \sum_{n\in\mathcal{H}} \mathrm{lfdr}_{n,\kappa} \le \alpha \right\} \qquad \mathrm{FDR} = \mathrm{E}\left[\frac{\sum_n \mathbb{1}\{\hat{H}_n = \mathrm{H}_1 | H_n = \mathrm{H}_0\}}{\sum_n \mathbb{1}\{\hat{H}_n = \mathrm{H}_1\}} \right]$$

- The false discovery rate (FDR) is controlled at level α

Aalto University School of Electrical Engineering

Censoring for Multiple Hypothesis Testing

Fundamentals

TECHNISCHE UNIVERSITÄT DARMSTADT

Reduce the number of transmissions while maintaining FDR control and high detection power

- Divide the p-value domain $[0, 1]$ into a send and a no-send region
  - defined by κ, the lower bound of the no-send region $[\kappa, 1]$
  - if a sensor's p-value $p_n \geq \kappa$, it is censored (no transmission)
  - for a single sensor, κ is the probability that a p-value is transmitted under the null hypothesis
- Multiple sensors: control the total number of transmissions
  - κ can be selected such that not more than a certain number $N_T < N$ of sensors transmit
  - most informative ones should transmit $\rightarrow$ those with the smallest p_n

Aalto University School of Electrical Engineering

Censoring for Multiple Hypothesis Testing

Which nodes should transmit?

Reduce the number of transmissions while maintaining FDR control and high detection power

Ordered Transmission Censoring for MHT (OTCEN)

- Sensors transmit their *p*-value after a time-delay proportional to its informativity
  - sensors with small *p*-values transmit first
- Offers two ways to limit transmissions
  1. after a pre-defined limit N_{lim} of transmissions has been reached
  2. when allowing more transmissions is not expected to yield many more true positives

     $\rightarrow$ omitted in this video due to time constraints

Censoring for Multiple Hypothesis Testing

Ordered Transmission Censoring for MHT (OTCEN) with fixed number of transmissions

TECHNISCHE UNIVERSITÄT DARMSTADT

1. FC communicates scheduling time constant M to all nodes
2. Initialize total number of transmissions $T = 0$, stopping message $\mathrm{stop} = 0$
3. While $\mathrm{stop} = 0$, do
   - 3.1 sensor $n \in [N]$ transmits its p_n after time $M \cdot p_n$ has passed
   - 3.2 after each newly received p-value, the FC
     - 3.2.1 adds the newly arrived p-value to the existing ones s. t. $p_{(1)} \leq \ldots \leq p_{(T)}$
     - 3.2.2 sets lower bound of no-send region $\kappa = p_{(T)}$
     - 3.2.3 increments $T = T + 1$
     - 3.2.4 if $T \geq N_{\text{init}}$ estimates $\mathrm{lfdr}_{n,\kappa}\ \forall n \in [N]$
     - 3.2.5 if $T = N_{\text{lim}}$ set $\mathrm{stop} = 0$ and broadcasts stop
4. Determine the set of rejected sensors $\hat{\mathcal{H}}_1^{(\text{sen})}$ based on $\mathrm{lfdr}_{n,\kappa}\ \forall n \in [N]$

Simulation Results

Setup

TECHNISCHE UNIVERSITÄT DARMSTADT

- Identification of areas with occupied radio frequency spectrum
  - different propagation environments: line-of-sight (LOS)/non-LOS (NLOS)
  - varying shape and size of decision regions
- Observation area: 100×100 grid points
- Performance measures averaged over 200 independent repetitions

Aalto University School of Electrical Engineering

Simulation Results

Setup

TECHNISCHE UNIVERSITÄT DARMSTADT

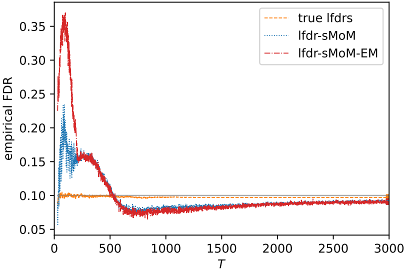
- Identification of areas with occupied radio frequency spectrum
  - different propagation environments: line-of-sight (LOS)/non-LOS (NLOS)
  - varying shape and size of decision regions
- Observation area: 100×100 grid points
- Performance measures averaged over 200 independent repetitions

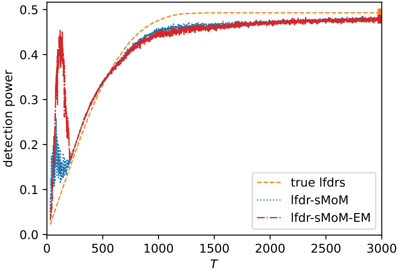
In this video (more in the paper)

- 5 sources in a LOS environment
- 3000 sensors
- nominal FDR level $\alpha = 0.1$
- evaluation of performance as function of N_T with
  - true lfdrs
  - lfdrs estimated with state-of-the art estimators `lfdr-sMoM` and `lfdr-sMoM-EM`

Simulation Results

False discovery rate (FDR) and detection power





- → Strict FDR control with newly derived true censored lfdrs
- → $T \approx 1250$: Detection power levels off
  - → $< 50\%$ of transmissions without censoring, but nearly no performance loss
- → For $T \geq 500$: performance with estimators (`lfdr-sMoM`, `lfdr-sMoM-EM`) similar to true lfdrs

Summary

Summary

- Proposed two censoring methods to minimize transmissions in MHT-based spatial inference
  - → control either by sensors or cloud/fusion center
- Analyzed the impact of censoring on FDR control and detection power
- Proposed a stopping criterion useful for ending transmissions after the bulk of true positives has been observed

TECHNISCHE UNIVERSITÄT DARMSTADT

Thank you for your attention!

Aalto University
School of Electrical Engineering