DISTRIBUTED ROBUST BAYESIAN CLUSTER ENUMERATION CRITERION FOR UNSUPERVISED LEARNING

Yani Zhang, Freweyni K. Teklehaymanot, Michael Muma, Abdelhak M. Zoubir

Signal Processing Group
Technische Universität Darmstadt
Merckstraße 25, 64283 Darmstadt, Germany
{muma, zoubir}@spg.tu-darmstadt.de

ABSTRACT

Wireless sensor networks have been widely deployed for industrial and consumer applications. The amount of data in such applications is large, and as a results a result, the automatic discovery of the underlying structure in the data (cluster analysis) becomes of prominent interest. A challenging task in cluster analysis is the estimation of the number of clusters. To this end, we propose a robust decentralized diffusion-based cluster enumeration method that enables distributed sensor nodes to estimate the number of clusters in their respective data sets through cooperation with their immediate neighbors. The proposed method is robust to the presence of heavy-tailed noise and outliers, which is useful for sensor networks as outliers can occur due to measurement errors or sensor failure. Through experiments, we show that the proposed method is promising, and achieves the performance of a centralized network using a fusion center.

Index Terms— Distributed Robust Cluster Enumeration, Diffusion, Clustering, Outlier, Sensor Networks

1. INTRODUCTION

The ubiquitous use of electronic devices with sensing and communication capabilities has created a Sensornet-of-Things (SoT) that extends physical functions through a digital space created by networked devices. Rapid advancements in cellular network technology further increase local massive machine type communication. Distributed signal processing and statistical learning for ad hoc sensor networks without a fusion center have become a topic of large interest [1, 2, 3, 4, 5, 6]. Unsupervised cooperative exploratory data mining, such as distributed cluster analysis [7, 8, 6, 9, 10, 11, 12, 13] finds the underlying groupings (or clusters) in a set of unlabeled data that is dispersed over a sensor network. Existing distributed algorithms, such as those which utilize K-means or Expectation Maximization (EM) require estimation of the number of clusters that best describe the underlying structure of the data. This task is referred to as cluster enumeration. Recently, Bayesian cluster enumeration approaches based on maximizing the posterior probability of a mixture model that best represents the data, given a set of candidate models have been proposed by Teklehaymanot *et al.* [14]. Assuming Gaussian Mixture Models (GMMs), distributed and adaptive Bayesian cluster enumeration algorithms have been proposed [15]. In real-world applications, the observed data is often subject to heavy-tailed noise and outliers [16] which obscure the underlying data structure and lead to drastic performance loss of GMM or Euclidian distance based approaches. Therefore, robust Bayesian cluster enumeration criteria that maximize the posterior probability of a heavy-tailed mixture model, such as the t_ν mixture with a small degree-of-freedom parameter ν have been proposed [17]. This paper describes a robust distributed adaptive Bayesian cluster enumeration method as an extension of the criteria proposed in [17] to ad hoc sensor networks based on the diffusion adaptation strategy [18]. Through simulations and a real remote sensing example, we demonstrate the superiority of robust criteria compared to the existing GMM based competitors. We also show the gain that can be obtained through cooperation compared to non-cooperative networks. In our experiments, the proposed decentralized diffusion-based cluster enumeration method achieves the performance of a centralized network using a fusion center, in which data from all nodes is available for cluster enumeration. Finally, we show, that calculating the exact penalty term based on maximizing the posterior probability leads to an improvement compared to using asymptotic approximations, especially when the available number of observations is small.

The paper is organized as follows. Sec. 2 formulates the problem, while Sec. 3 describes the proposed method. Simulated and real-data experiments are reported in Sec. 4. Finally, conclusions are drawn in Sec. 5.

2. PROBLEM FORMULATION

Consider a wireless sensor network with J nodes whose topology is described by a graph. The neighborhood of node $j \in \mathcal{J} \triangleq \{1, \ldots, J\}$, denoted as $\mathcal{B}_j$, is the set of

The work of M. Muma is supported by the 'Athene Young Investigator Programme' of Technische Universität Darmstadt.

all nodes, including j, that node j can exchange information with. At time instant t, where $t = 1, 2, \ldots$, each node $j \in \mathcal{J}$ collects r-dimensional data vectors and stores them in $\mathcal{X}_{jt} \triangleq \{\boldsymbol{x}_{j1}, \ldots, \boldsymbol{x}_{jN_t}\} \in \mathbb{R}^{r \times N_t}$, where N_t is the number of data vectors observed by node $j \in \mathcal{J}$ at time instant t. As time progresses, each node $j \in \mathcal{J}$ stores its data in $\mathcal{S}_{jt} \triangleq \{\mathcal{X}_{j1}, \ldots, \mathcal{X}_{jt}\} \in \mathbb{R}^{r \times N_{jt}}$, where $N_{jt} = \sum_{i=1}^{t} N_i$. $\mathcal{S}_{jt}$ contains K_t independent, mutually exclusive, and non-empty clusters. Assume that a set of candidate models $\mathcal{M}_j \triangleq \{M_{jL_{\min}}, \ldots, M_{jL_{\max}}\}$, is given, where $L_{\min}$ and $L_{\max}$ are the specified minimum and maximum number of clusters, respectively. Each candidate model $M_{jl} \in \mathcal{M}_j$ represents a partitioning of $\mathcal{S}_{jt}$ into $l \in \{L_{\min}, \ldots, L_{\max}\}$ clusters, where $l \in \mathbb{Z}^+$. Each data vector $\boldsymbol{x}_{jn} \in \mathcal{S}_{jt}, n = 1, \ldots, N_{jt}$, has an associated class label $k \in \mathcal{K} \triangleq \{1, \ldots, K_t\}$. Our goal is to enable each node $j \in \mathcal{J}$ to adaptively and robustly estimate the number of clusters in the data set $\mathcal{S}_{jt}$ by cooperating with its neighbors in $\mathcal{B}_j$.

Algorithm 1 Distributed and adaptive robust Bayesian cluster enumeration algorithm

Inputs: $L_{\min}$ and $L_{\max}$
for $t = 1, 2, \ldots$ **do**
  for $j = 1, \ldots, J$ **do**
    Collect N_t data vectors
    Store the data vectors in $\mathcal{X}_{jt}$
    Update $\mathcal{S}_{jt}$
  end for
  for $l = L_{\min}, \ldots, L_{\max}$ **do**
    for $j = 1, \ldots, J$ **do**
      for $m = 1, \ldots, l$ **do**
        Estimate N_{jml}, $\boldsymbol{\mu}^0_{jml}$, and $\boldsymbol{\Psi}^0_{jml}$ via EM
      end for
    end for
    for $j = 1, \ldots, J$ **do**
      Exchange $\hat{\boldsymbol{\mu}}^0_{jl}$ and $\hat{\boldsymbol{\Psi}}^0_{jl}$ within $\mathcal{B}_j$
      Synchronize $\hat{\boldsymbol{\mu}}^0_{jl}$ and $\hat{\boldsymbol{\Psi}}^0_{jl}$ within $\mathcal{B}_j$
      Adapt scatter matrix estimates using (1)
      Calculate BIC via (2)
    end for
  end for
  for $j = 1, \ldots, J$ **do**
    Estimate K^0_{jt} using (3)
  end for
  for $j = 1, \ldots, J$ **do**
    Exchange $\hat{K}^0_{jt}$ within $\mathcal{B}_j$
    Combine $\hat{K}^0_{jt}$ and $\hat{K}^0_{bt}, b \in \mathcal{B}_j/\{j\}$, using (4)
  end for

3. PROPOSED ROBUST DIFFUSION-BASED CLUSTER ENUMERATION METHOD

The working principle of the proposed method is summarized in Algorithm 1 and is detailed as follows:

1. *Data Collection:* each node $j \in \mathcal{J}$ collects N_t data vectors at time instant t and stores them in $\mathcal{X}_{jt}$. The accumulated data vectors at node j at time instant t are stored in $\mathcal{S}_{jt} \triangleq \{\mathcal{X}_{j1}, \ldots, \mathcal{X}_{jt}\}$.

2. *Parameter Estimation:* each node $j \in \mathcal{J}$ estimates cluster parameters for each candidate model $M_{jl} \in \mathcal{M}_j$ using the t_ν EM algorithm detailed in [17]. The estimated parameters are the cluster centroids $\hat{\boldsymbol{\mu}}^0_{jml} \in \mathbb{R}^{r \times 1}$, scatter matrices $\hat{\boldsymbol{\Psi}}^0_{jml} \in \mathbb{R}^{r \times r}$, and the number of data vectors per cluster $N_{jml} \in \mathbb{Z}^+$ for $m = 1, \ldots, l$ and $l = L_{\min}, \ldots, L_{\max}$, where $N_{jt} = \sum_{m=1}^{l} N_{jml}$.

3. *Exchange of Parameter Estimates:* each node $j \in \mathcal{J}$ exchanges $\hat{\boldsymbol{\mu}}^0_{jml}$ and $\hat{\boldsymbol{\Psi}}^0_{jml}$ for $m = 1, \ldots, l$ and $l = L_{\min}, \ldots, L_{\max}$ within its neighborhood $\mathcal{B}_j$.

4. *Synchronize parameter estimates:* nodes assign labels to each cluster in an arbitrary manner. Hence, each node $j \in \mathcal{J}$ synchronizes the labels based on the Euclidean norm between its own cluster centroid estimates $\hat{\boldsymbol{\mu}}^0_{jml}$ and the estimates of its neighbors $\hat{\boldsymbol{\mu}}^0_{bml}$, where $b \in \mathcal{B}_j/\{j\}$, as detailed in [15].

5. *Adaptation of Parameter Estimates:* the own and received cluster scatter matrix estimates are adapted via

$$\hat{\boldsymbol{\Psi}}_{jml} = \alpha \hat{\boldsymbol{\Psi}}^0_{jml} + (1 - \alpha) \sum_{b \in \mathcal{B}_j/\{j\}} a_{bml} \hat{\boldsymbol{\Psi}}^0_{bml} \quad (1)$$

at each node $j \in \mathcal{J}$ for each candidate model $M_{jl} \in \mathcal{M}_j$. Herein, α denotes the tradeoff between the weight given to the own and neighbors' estimates. For a_{bml}, we use uniform combination weights [18].

6. *Model Order Selection:* using the adapted parameter estimates, each node $j \in \mathcal{J}$ selects the model $M_{j\hat{K}^0_{jt}} \in \mathcal{M}_j$, with $\hat{K}^0_{jt} \in \{L_{\min}, \ldots, L_{\max}\}$, that maximizes the posterior probability given $\mathcal{S}_{jt}$. For this purpose, each node calculates

$$\text{D-BIC}_{Ft_\nu}(M_{jl}) \approx \log \mathcal{L}(\hat{\boldsymbol{\Theta}}_{jl}|\mathcal{X}) - \frac{1}{2} \sum_{m=1}^{l} \log |\hat{\boldsymbol{J}}_{jm}|, \quad (2)$$

where the likelihood function $\mathcal{L}(\hat{\boldsymbol{\Theta}}_{jl}|\mathcal{X})$ and the determinant of the Fisher information matrix $|\hat{\boldsymbol{J}}_{jm}|$ are derived in [17], and $\boldsymbol{\Theta}_{jl} = [\boldsymbol{\theta}_{j1}, \ldots, \boldsymbol{\theta}_{jl}]$ with $\boldsymbol{\theta}_{jm} = [\boldsymbol{\mu}_{jm}, \boldsymbol{\Psi}_{jm}]^\top$. The second term in Eq. (2) is called the penalty term, as it penalizes model complexity. Once

each node $j \in \mathcal{J}$ computed D-BIC$_{Ft_\nu}(M_{jl})$ for each candidate model $M_{jl} \in \mathcal{M}_j$, the number of clusters in $\mathcal{S}_{jt}$ is determined by

$$\hat{K}_{jt}^0 = \underset{l=L_{\min},\ldots,L_{\max}}{\arg\max} \text{D-BIC}_{Ft_\nu}(M_{jl}) \quad (3)$$

7. *Exchange of Cluster Number Estimates:* each node $j \in \mathcal{J}$ exchanges its preliminary estimate of the number of clusters, $\hat{K}_{jt}^0$, in $\mathcal{S}_{jt}$ at time instant t within its neighborhood $\mathcal{B}_j$.

8. *Adaptation of Cluster Number Estimates:* finally, each node $j \in \mathcal{J}$ adapts its cluster number estimate using

$$\hat{K}_{jt} = \text{median}\left(\hat{K}_{jt}^0, \hat{K}_{bt}^0\right), \quad (4)$$

where $\hat{K}_{bt}^0, b \in \mathcal{B}_j/\{j\}$, denotes the cluster number estimates that node j received from its neighbors.

4. RESULTS

4.1. Simulation Setup

We consider a wireless sensor network with $J = 10$ nodes and $\#\mathcal{B}_j = 5$. All simulation results are an average of 150 Monte Carlo (MC) experiments and the minimum and maximum number of clusters in the candidate models is set to $L_{\min} = 1$ and $L_{\max} = 2K$, respectively, where K is the true number of clusters in the data set $\mathcal{S}_{jt}$ at the final time instant. Equal weight is given to the own and neighborhood-based estimates by setting $\alpha = 0.5$. Results are reported for different cooperation modes, i.e., cooperative (coop), non-cooperative (non-coop), and centralized. In a centralized network, the fusion center solves the cluster enumeration task after receiving data vectors from all nodes in the network. Each cluster k contains $N_k = 100$ data vectors for Simulation 1 and $N_k = 40$ data vectors for Simulation 2. For all simulations, each node $j \in \mathcal{J}$ observes $N_t = 20$ data vectors, which are randomly drawn from the data set, at time instant t.
The proposed D-BIC$_{Ft_\nu}$, is compared to a robust version of Schwarz's information criterion [19] (D-BIC$_{Ot_\nu}$) that uses the same t_ν EM algorithm to estimate the parameters of the likelihood function, but adds a penalty equal to $(ql)/2\log(N_j)$, instead of the one given in Eq. (2). Here, q is the dimension of $\boldsymbol{\theta}_{jm}$ in Step 6 of the algorithm. The GMM-based Bayesian Information Criteria D-BIC$_{NF}$ and D-BIC$_{N}$ presented in [15], and the Schwarz' information criterion D-BIC$_{ON}$, adding a penalty of $(ql)/2\log(N_j)$ to the Gaussian likelihood function, use the same Gaussian EM algorithm.

4.2. Simulation 1: ε-Contaminated Mixture With Outlier-Probability Varying Across the Network

For Data-1, at each node, the $(1-\varepsilon_j)$ fraction of data points contains realizations of $\boldsymbol{x}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, for $k = 1,\ldots,4$ and $j \in \mathcal{J}$, with cluster centroids $\boldsymbol{\mu}_1 = [-2,0]^\top$, $\boldsymbol{\mu}_2 = [5,0]^\top$, $\boldsymbol{\mu}_3 = [0,7]^\top$, $\boldsymbol{\mu}_4 = [8,4]^\top$, and covariance matrices $\boldsymbol{\Sigma}_1 = \text{diag}(0.2,0.2)$, $\boldsymbol{\Sigma}_2 = \text{diag}(0.6,0.6)$, $\boldsymbol{\Sigma}_3 = \text{diag}(0.4,0.4)$, $\boldsymbol{\Sigma}_4 = \text{diag}(0.2,0.2)$. The outliers probability varies randomly over the nodes according to $\varepsilon_j = \mathcal{U}(0,3)$. The contaminating distribution generates replacement outliers at random positions within the feature space in a given interval, in our case, $\mathcal{U}(-20,20)$.
Data-2, contains $K = 5$ clusters. The outlier probability is increased to $\varepsilon_j = \mathcal{U}(0,5)$. The cluster centroids are defined by $\boldsymbol{\mu}_1 = [-5,-5]^\top$, $\boldsymbol{\mu}_2 = [5,-5]^\top$, $\boldsymbol{\mu}_3 = [0,12]^\top$, $\boldsymbol{\mu}_4 = [15,4]^\top$, $\boldsymbol{\mu}_5 = [-10,5]^\top$. The covariance matrices $\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_4$ are the same as in Data-1, and $\boldsymbol{\Sigma}_5 = \text{diag}(0.3,0.3)$.

4.3. Simulation 2: Heavy-tailed Clusters With Degree-of-Freedom Parameter Varying Across the Network

Data-3, contains realizations of random variables $\boldsymbol{x}_k \sim t_{\nu_k}(\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k)$, where $\boldsymbol{\mu}_k \in \mathbb{R}^{r\times 1}$, $\boldsymbol{\Psi}_k \in \mathbb{R}^{r\times r}$, and $\nu_k \in \mathbb{R}^+$ represent the centroid, the scatter matrix, and the degree of freedom of the kth cluster, respectively, for $k = 1,\ldots,6$, with cluster centroids $\boldsymbol{\mu}_1 = [-8,0]^\top$, $\boldsymbol{\mu}_2 = [3,0]^\top$, $\boldsymbol{\mu}_3 = [0,5]^\top$, $\boldsymbol{\mu}_4 = [9,4]^\top$, $\boldsymbol{\mu}_5 = [-5,-9]^\top$, $\boldsymbol{\mu}_6 = [7,-7]^\top$, and scatter matrices $\boldsymbol{\Psi}_1 = \text{diag}(0.6,1.2)$, $\boldsymbol{\Psi}_2 = \text{diag}(1.8,0.9)$, $\boldsymbol{\Psi}_3 = \text{diag}(1.2,0.6)$,$\boldsymbol{\Psi}_4 = \text{diag}(0.9,0.9)$, $\boldsymbol{\Psi}_5 = \text{diag}(0.9,1.5)$, $\boldsymbol{\Psi}_6 = \text{diag}(1.2,1.2)$. The cluster impulsiveness differs among the nodes, as the degree-of-freedom parameter for each node is randomly selected according to $\nu_j \in \{2,3,...,10\}$.

4.4. Real Data Experiment: Multi-temporal Remote Sensing Data of a Forested Area in Japan

The remote sensing data is a subset of the data studied by [20] that mapped different forest types based on their spectral characteristics at visible-to-near infrared wavelengths, using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) satellite imagery. We combine training and testing data to create a single (unlabled) data set ($N = 507$). A subset of $K = 2$ classes are chosen, which are, the 'Hinoki' (Japanese Cypress) planted forest, and the 'other land use' (e.g. agricultural fields, roads, bare soil) classes. The 15 m spatial resolution multispectral ASTER image bands containing spectral information in the green (0.52-0.60 μm) and red bands (0.63-0.69 μm) are used as features. To create a cooperative sensor network setting, the features, which are measured at multiple time instances to provide adequate spectral for coniferous and deciduous tree species, are randomly distributed for each MC experiment among a network consisting of $J = 7$ nodes with $\#\mathcal{B}_j = 4$ neighbors. The presented results are averages over 1000 MC experiments.

4.5. Simulation and Real-Data Results

Tables 1–4 report the cluster enumeration performance based on the network-wide empirical probability of detection and

mean absolute error, which are defined as

$$p_{\text{det}}^{\text{net}} = \frac{1}{JIT}\sum_{j=1}^{J}\sum_{i=1}^{I}\sum_{t=1}^{T}\mathbb{1}_{\{\hat{K}_{jt}^{(i)}=K_t\}} \tag{5}$$

$$\text{MAE}^{\text{net}} = \frac{1}{JIT}\sum_{j=1}^{J}\sum_{i=1}^{I}\sum_{t=1}^{T}\left|K_t - \hat{K}_{jt}^{(i)}\right|, \tag{6}$$

where I is the total number of MC experiments, T is the total number of time instances, $\hat{K}_{jt}^{(i)}$ is the estimated number of clusters by the jth node at time instant t and the ith MC experiment, and $\mathbb{1}_{\{\hat{K}_{jt}^{(i)}=K_t\}}$ is the indicator function.

A considerable gain is obtained compared to nonrobust methods, and the decentralized method achieved the performance of its centralized counterpart. For the remote sensing data, Fig. 1 shows the insensitivity of the results to the degree of freedom parameter ν, and Tab. 5 displays the labeling performance in terms of the confusion matrix.

Table 1. Network-wide average results for Data-1.

		non-coop	coop	centralized
$p_{\text{det}}^{\text{net}}(\%)$	D-BIC$_{\text{Ft}_3}$	96.57	**98.26**	97.28
	D-BIC$_{\text{Ot}_3}$	52.38	93.96	93.43
	D-BIC$_{\text{N}}$	24.23	0.64	0.72
	D-BIC$_{\text{NF}}$	26.85	0.74	0.82
	D-BIC$_{\text{ON}}$	26.03	0.67	0.78
MAE$^{\text{net}}$	D-BIC$_{\text{Ft}_3}$	0.036	**0.017**	0.027
	D-BIC$_{\text{Ot}_3}$	0.768	0.064	0.075
	D-BIC$_{\text{N}}$	2.435	2.906	2.799
	D-BIC$_{\text{NF}}$	1.413	2.912	2.862
	D-BIC$_{\text{ON}}$	2.003	1.811	1.405

Table 2. Network-wide average results for Data-2.

		non-coop	coop	centralized
$p_{\text{det}}^{\text{net}}(\%)$	D-BIC$_{\text{Ft}_3}$	93.61	**97.14**	85.84
	D-BIC$_{\text{Ot}_3}$	48.29	80.55	76.16
	D-BIC$_{\text{N}}$	15.39	0.11	0.07
	D-BIC$_{\text{NF}}$	18.37	0.12	0.13
	D-BIC$_{\text{ON}}$	17.56	0.12	0.11
MAE$^{\text{net}}$	D-BIC$_{\text{Ft}_3}$	0.070	**0.029**	0.146
	D-BIC$_{\text{Ot}_3}$	1.037	0.197	0.277
	D-BIC$_{\text{N}}$	3.912	3.285	3.243
	D-BIC$_{\text{NF}}$	1.854	3.425	3.736
	D-BIC$_{\text{ON}}$	2.762	1.537	1.355

5. CONCLUSION

A robust distributed and adaptive cluster enumeration method was proposed that showed promising performance for simulated and real-data experiments. Due to the diffusion adaptation strategy the method is scalable to large networks, and has many potential applications for cluster analysis in the big data setting.

Table 3. Network-wide average results for Data-3.

		non-coop	coop	centralized
$p_{\text{det}}^{\text{net}}(\%)$	D-BIC$_{\text{Ft}_3}$	98.03	99.93	**100**
	D-BIC$_{\text{Ot}_3}$	97.80	99.60	99.00
	D-BIC$_{\text{N}}$	40.96	35.37	9.58
	D-BIC$_{\text{NF}}$	54.10	42.75	10.83
	D-BIC$_{\text{ON}}$	69.43	42.75	13.25
MAE$^{\text{net}}$	D-BIC$_{\text{Ft}_3}$	0.020	0.001	**0**
	D-BIC$_{\text{Ot}_3}$	0.247	0.004	0.014
	D-BIC$_{\text{N}}$	3.141	1.068	1.813
	D-BIC$_{\text{NF}}$	0.937	1.051	1.293
	D-BIC$_{\text{ON}}$	0.605	1.767	1.224

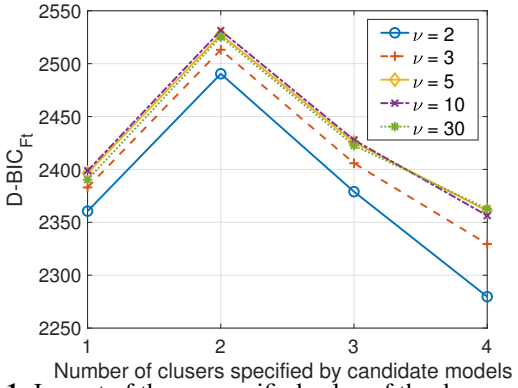

Fig. 1. Impact of the prespecified value of the degree of freedom parameter ν on the performance of D-BIC$_{Ft_\nu}$.

Table 4. Network-wide average results: remote sensing data.

		non-coop	coop	centralized
$p_{\text{det}}^{\text{net}}(\%)$	D-BIC$_{\text{Ft}_3}$	0	**100**	**100**
	D-BIC$_{\text{Ot}_3}$	75.07	43.63	**100**
	D-BIC$_{\text{N}}$	17.97	51.99	47.80
	D-BIC$_{\text{NF}}$	82.21	51.03	46.30
	D-BIC$_{\text{ON}}$	57.91	44.20	94.4
MAE$^{\text{net}}$	D-BIC$_{\text{Ft}_3}$	1.000	**0**	**0**
	D-BIC$_{\text{Ot}_3}$	0.346	0.629	**0**
	D-BIC$_{\text{N}}$	1.372	0.530	0.979
	D-BIC$_{\text{NF}}$	0.193	0.536	0.994
	D-BIC$_{\text{ON}}$	0.586	0.646	0.056

Table 5. Network-wide confusion matrix (%) for the remote sensing data set when $M_{\hat{K}} = M_K$. 'Est' stands for estimated.

		Est: Forrest	Est: Non-forrest
centralized	Forrest	99.22	0.78
	Non-forrest	7.24	92.76
coop	Forrest	99.12	0.88
	Non-forrest	7.20	92.80

6. REFERENCES

[1] J. Plata-Chaves, A. Bertrand, M. Moonen, S. Theodoridis, and A. M. Zoubir, "Heterogeneous and multi-task wireless sensor networks - algorithms, applications and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 450–465, Apr. 2017.

[2] N. Bogdanovic, J. Plata-Chaves, and K. Berberidis, "Distributed incremental-based LMS for node-specific adaptive parameter estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5382–5397, Oct 2014.

[3] A. Bertrand and M. Moonen, "Distributed signal estimation in sensor networks where nodes have different interests," *Signal Process.*, vol. 92, pp. 1679–1690, July 2012.

[4] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, June 2015.

[5] J. Plata-Chaves, A. Bertrand, and M. Moonen, "Distributed signal estimation in a wireless sensor network with partially-overlapping node-specific interests or source observability," in *Proc. 40th IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Brisbane, Australia, 2015, pp. 5808–5812.

[6] S. Chouvardas, M. Muma, L. K. Hamaidi, S. Theodoridis, and A. M. Zoubir, "Distributed robust labeling of audio sources in heterogeneous wireless sensor networks," in *Proc. 40th IEEE Int. Conf. Acoust., Speech and Signal Process.(ICASSP)*, Brisbane, Australia, 2015, pp. 5783–5787.

[7] F. K. Teklehaymanot, M. Muma, B. Béjar-Haro, P. Binder, A. M. Zoubir, and M. Vetterli, "Robust diffusion-based unsupervised object labelling in distributed camera networks," in *Proc. 12th IEEE AFRICON*, Addis Ababa, Ethiopia, 2015.

[8] F. Teklehaymanot, M. Muma, and A. M. Zoubir, "Adaptive diffusion-based track assisted multi-object labeling in distributed camera networks," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Kos, Greece, 2017, pp. 2363–2367.

[9] A. Bertrand M. H. Bahari, J. Plata-Chaves and M. Moonen, "Distributed labelling of audio sources in wireless acoustic sensor networks using consensus and matching," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Budapest, Hungary, 2016, pp. 2345–2349.

[10] P. Binder, M. Muma, and A. M. Zoubir, "Robust and computationally efficient diffusion-based classification in distributed networks," in *Proc. 40th IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Brisbane, Australia, 2015, pp. 3432–3436.

[11] P. Binder, M. Muma, and A. M. Zoubir, "Robust and adaptive diffusion-based classification in distributed networks," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 34, pp. 1–13, 2016.

[12] R. Altilio, P. Di Lorenzo, M. Panella, "Distributed data clustering over networks," *Pattern Recognition*, vol. 93, pp. 603–620, 2019.

[13] M. Hassani, "Overview of Efficient Clustering Methods for High-Dimensional Big Data Streams," Clustering Methods for Big Data Analytics, pp. 25-42, Springer, 2019.

[14] Teklehaymanot, F. K. and Muma, M. and Zoubir, A. M., "Bayesian Cluster Enumeration Criterion for Unsupervised Learning," *IEEE Trans. Signal Process.,* vol. 66, no. 20, pp. 5392–5406, 2018.

[15] Teklehaymanot, F. K. and Muma, M. and Zoubir, A. M., "Diffusion-Based Bayesian Cluster Enumeration in Distributed Sensor Networks," *IEEE Statist. Signal Process. Workshop (SSP)*, pp. 1–5, 2018.

[16] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing.* Cambridge University Press, Cambridge, UK, 2018.

[17] Teklehaymanot, F. K. and Muma, M. and Zoubir, A. M., "Robust Bayesian Cluster Enumeration," *arXiv:1811.12337*, 2019.

[18] A. H. Sayed, *Diffusion Adaptation over Networks, Elsevier Academic Press Library in Signal Processing: Array and Statistical Signal Processing*, vol. 3, ch. 9, 2014.

[19] G. Schwarz, "Estimating the Dimension of a Model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[20] Johnson, B., Tateishi, R., Xie, Z., 2012. "Using geographically-weighted variables for image classification". *Remote Sensing Lett.*, vol. 3, no. 6, pp. 491–499.