PREDICTIVE STATE VECTOR ENCODING FOR DECENTRALIZED FIELD ESTIMATION IN SENSOR NETWORKS

Florian Xaver[⋆] *Gerald Matz*[⋆] *Peter Gerstoft*[†] *Christoph Mecklenbräuker*[⋆]

[⋆] Institute of Telecommunications, Vienna University of Technology, Vienna, Austria

[†] Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, California

ABSTRACT

Decentralized physics-based field estimation in clustered sensor networks requires the exchange of state vectors between neighboring clusters. We reduce the communication overhead between clusters by using a differential encoding of state vectors that exploits the spatio-temporal field dependencies. This encoding involves a Kalman prediction step that builds on the state-space equations governing the field's spatio-temporal evolution. The Kalman step keeps the computational complexity low. Simulation results for an acoustic field demonstrate the approach.

Index Terms— distributed parameter estimation, Kalman filter, differential encoding, linear prediction, acoustic field

1. INTRODUCTION

1.1. Background

In this paper we address the communication overhead in sensor networks (SN) which sequentially infer states of a physical field in a decentralized fashion, i.e., without fusion center. This type of estimation problems has been studied in [1–7]. From these, [6,7] use numerical methods to approximate the underlying partial differential equation (PDE) and reformulate the resulting difference equation as a state space model. For decentralized field estimation, the state space model is partitioned into subsystems that correspond to sensor clusters. In the field estimation algorithm, the clusters exchange elements of the state vector that correspond to the cluster boundaries. The communication overhead (analyzed in [7]) resulting from this state vector exchange is addressed by our paper. As an illustrative example we use a 2D acoustic field in a hallway along with an estimator for the position of an acoustic source.

Specifically, we propose to use differential encoding based on a Kalman predictor that exploits the spatio-temporal field correlation via the underlying state space model. For details on linear prediction of discrete-time vector processes see e.g. [8,9]. With differential encoding, only the difference between the predicted signal and the measurement is transmitted and the receiver reconstructs the original signal using the prediction error. Ideally, the prediction error is a white innovation signal and hence has a flat power spectrum.

This work was funded by Grant ICT08-44 of the *Viennese Science and Technology Fund (WWTF)*.

1.2. Contributions and Outline

Our paper proposes a differential encoding scheme that merges the field estimation algorithm with the state prediction and exploits the strong spatial and temporal correlations to whiten/decorrelate the signal to be communicated between clusters.Due to the discretized PDE, the global field is modeled by an autoregressive process of order one and thus is predictable. However, using the state space model to this end is not trivial in the context of decentralized estimation. In this case, the hyperbolic structure of the PDE becomes relevant and the order increases with the iteration of the sequential estimator. Our main contributions is a modified Kalman filter to whiten the state vectors based on the state space model. The white innovations/prediction error signal has a smaller dynamic range and hence can be sent using less transmit power. This is particularly desirable in battery-operated wireless sensor networks. However, the actual source encoding (i.e., quantization and bit allocation) of the innovations signal is beyond the scope of this paper.

2. PROBLEM DEFINITION

2.1. Centralized system

In the sequel, we focus on a 2D acoustic field in a hallway described by the scalar wave equation [10]

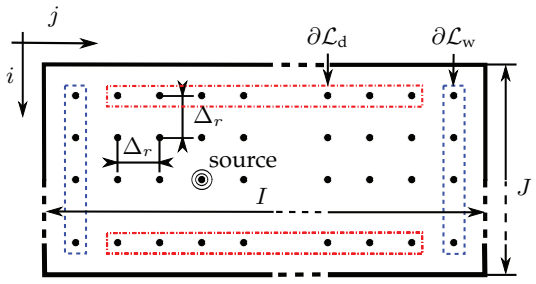
$$\frac{1}{c^2}\partial_t^2 p(\boldsymbol{r},t) - \nabla^2 p(\boldsymbol{r},t) = s(\boldsymbol{r},t), \quad \boldsymbol{r}\in\Omega. \tag{1}$$

This is a linear hyperbolic second-order partial differential equation (PDE), where $p(\boldsymbol{r},t)$ denotes pressure dependent on location $\boldsymbol{r}$ and time t, ∂_t is the partial derivative with respect to time, ∇^2 is the Laplace operator, c is the sound speed, $s(\boldsymbol{r},t)$ is a (random) source, and $\Omega\subset\mathbb{R}^2$ is the 2-dimensional region of interest. For the boundary and initial conditions we refer to Section 4.

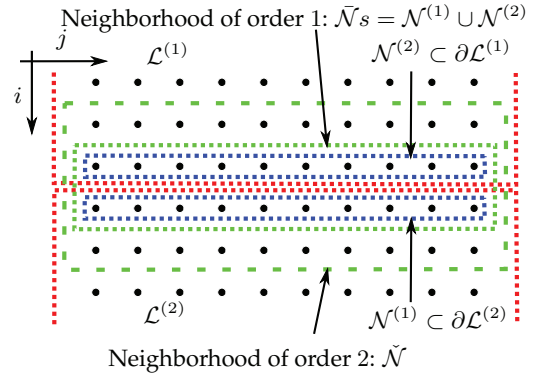
We define $q(\boldsymbol{r},t)=\partial_t p(\boldsymbol{r},t)$ and approximate the wave equation via a finite difference method (FDM) [10]. This results in the state transition model

$$\begin{bmatrix}\boldsymbol{q}_{k+1}\\ \boldsymbol{p}_{k+1}\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12}\\ \Delta_t\boldsymbol{I} & \boldsymbol{I}\end{bmatrix}}_{\boldsymbol{\Phi}_{\text{FDM}}}\begin{bmatrix}\boldsymbol{q}_k\\ \boldsymbol{p}_k\end{bmatrix} + \Delta_t\, c^2\begin{bmatrix}\boldsymbol{s}_k\\ \boldsymbol{0}\end{bmatrix}, \tag{2}$$

where the pressure vector is defined as $\boldsymbol{p}_k=\text{vec}\{\boldsymbol{P}_k\}$ with $[\boldsymbol{P}_k]_{ij}=p(i\Delta_r,j\Delta_r,k\Delta_t), i,j,k\in\mathbb{N}$ and similar for $\boldsymbol{q}_k$ and source $\boldsymbol{s}_k$. The set of tuples (i,j), termed nodes, is denoted by $\mathcal{L}=\{(i,j)\in\mathbb{N}^2 : (i\Delta_r, j\Delta_r)\in\Omega\}$ (cf. Figure 1a).



(a)



(b)

Fig. 1. The discretized area $\mathcal{L}$ with disjoint boundary $\partial\mathcal{L} = \bigcup_{\ell=1}^{4} \partial\mathcal{L}_\ell$ is shown in (a). The nodes correspond to sample points of the field. In (b) the area is decomposed into two clusters corresponding to $\mathcal{L}^{(1)}$ and $\mathcal{L}^{(2)}$. Various neighbor sets are defined in this sketch.

The model (2) consists of two parts:

- The matrix $\boldsymbol{\Phi}_\mathrm{FDM}$ maps the state $[\boldsymbol{q}_k^\mathrm{T}\ \boldsymbol{p}_k^\mathrm{T}]^\mathrm{T}$ deterministically to its *prediction* $[\boldsymbol{q}_{k+1|k}^\mathrm{T}\ \boldsymbol{p}_{k+1|k}^\mathrm{T}]^\mathrm{T}$ in the following time step. (Note that an optimal predictor additionally considers the redundancy of the source.)
- The second term on the right-hand side involving $\boldsymbol{s}_k$ perturbs the state vectors and represents the *innovations* process.

2.2. Decentralized System

We now partition the nodes $\mathcal{L}$ into two disjoint clusters $\mathcal{L}^{(1)}, \mathcal{L}^{(2)} \subset \mathcal{L}$. The symbol $\bar{}$ will be used to denote the elements of a vector that correspond to nodes along the boundary between both the clusters and $\check{}$ signifies that second-order boundary nodes, i.e., neighbors of boundary nodes, are also included in the respective subvector. Figure 1b specifies the various neighborhood sets used in what follows.

Assuming the source to be far away from the boundary, we neglect the source term in the decentralized setup. Due to the sparse structure of the global matrix $\boldsymbol{\Phi}_{12}$ (nodes depend only on their neighbors), only the elements of $\bar{\boldsymbol{p}}_k$ need to be exchanged between the clusters for the decentralized estimation of the field [7]. For this state subvector, (2) implies

$$\bar{\boldsymbol{p}}_{k+1} = \bar{\boldsymbol{p}}_k + \Delta_t \bar{\boldsymbol{\Phi}}_{11} \bar{\boldsymbol{q}}_{k-1} + \Delta_t \check{\boldsymbol{\Phi}}_{12} \check{\boldsymbol{p}}_{k-1} \,. \tag{3}$$

Here $\boldsymbol{s}_k$ does not denote the source of innovation as in the central case, but rather the unknown pressure in the neighborhood $\check{\mathcal{N}} \setminus \bar{\mathcal{N}}$. Note that no approximation is performed. The main idea of this paper is to signal only the innovations vector

$$\bar{\boldsymbol{\varepsilon}}_{k+1} = \bar{\boldsymbol{p}}_{k+1} - \bar{\boldsymbol{p}}_{k+1|k} \tag{4}$$

rather than the actual state vector $\bar{\boldsymbol{p}}_{k+1}$. This is advantageous since $\bar{\boldsymbol{\varepsilon}}_{k+1}$ can be better compressed because it is white and has smaller power. The receiving cluster can revert the differential encoding by adding the received innovation vector to the local prediction, i.e. $\bar{\boldsymbol{p}}_{k+1} = \bar{\boldsymbol{\varepsilon}}_{k+1} + \bar{\boldsymbol{p}}_{k+1|k}$.

3. PREDICTION BY THE KALMAN FILTER

Linear prediction theory [9] exploits the statistical structure of the source which in our context corresponds to the stochastic transition model (2). But due to (3) it does not seem feasible without high prediction order and with the unknown second order neighbor states. Instead we use the computationally efficient Kalman filtering (KF).

The KF [9,11–14] falls within the scope of *Bayesian estimators* and is a combination of a sequential *linear minimum mean square error* estimator combined with a state space model, ie. a state transition model of random states as in (2) with additive noise and a observation model, respectively. It estimates the states through observations $\boldsymbol{y}_k$ and is optimum if and only if the priors and noise are Gaussian.

3.1. Model of the decentralized system

We now propose the use of the KF to predict the states $\bar{\boldsymbol{p}}_{k+1}$. For this, let the states of the KF be $[\bar{\boldsymbol{q}}_k^\mathrm{T}\ \check{\boldsymbol{p}}_k^\mathrm{T}]^\mathrm{T}$ with the associated transition model

$$\begin{bmatrix} \bar{\boldsymbol{q}}_{k+1} \\ \check{\boldsymbol{p}}_{k+1} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\Phi}}_{11} & \check{\boldsymbol{\Phi}}_{12} \\ \Delta_t \bar{\boldsymbol{I}} & \check{\boldsymbol{I}} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{q}}_k \\ \check{\boldsymbol{p}}_k \end{bmatrix} + \boldsymbol{G}\boldsymbol{w}_k \,, \tag{5a}$$

where $\boldsymbol{w}_k$ is the driving noise. The matrix $\boldsymbol{G} : \mathbb{R}^{|\check{\mathcal{N}}\setminus\bar{\mathcal{N}}|} \to \mathbb{R}^{|\check{\mathcal{N}}|}$ ensures that the driving noise is only added to pressure states in the second-order cluster boundary $\check{\mathcal{N}} \setminus \bar{\mathcal{N}}$. These states are modeled as unknowns and are estimated by the KF in both clusters in the same way. We note that the linear approach underlying the KF is optimal only if $\boldsymbol{w}_k$ is Gaussian.

The aim of matrix $\boldsymbol{G}$ combined with an observation model is to maintain correct state subvectors $\bar{\boldsymbol{p}}_k$. This in turn improves the estimates of their neighbors $\check{\mathcal{N}} \setminus \bar{\mathcal{N}}$. With this in mind, let the observation model of cluster m be

$$\boldsymbol{y}_{k+1}^{(m)} = \bar{\boldsymbol{p}}_{k+1}\big(\bar{\boldsymbol{p}}_{k+1}^{(m)}, \boldsymbol{\varepsilon}_{k+1}^{(\overline{m})}, \bar{\boldsymbol{p}}_{k+1|k}^{(\overline{m})}\big). \tag{5b}$$

where cluster $\overline{m}$ is the neighbor of cluster m. Here $\bar{\boldsymbol{p}}_{k+1}(\cdot)$ is viewed as a vector-valued linear function depending on the own pressure states $\bar{\boldsymbol{p}}_{k+1}^{(m)}$ at the boundary and the sum $\boldsymbol{\varepsilon}_{k+1}^{(\overline{m})} + \bar{\boldsymbol{p}}_{k+1|k}^{(\overline{m})}$ (received signal plus prediction). Its sole purpose is to shift the elements of the vectors appropriately. As a direct

Our discussion extends straightforwardly to the case of more than two clusters.

consequence of the measurement model, the correction step of the KF adjusts $\bar{\boldsymbol{p}}_{k+1|k}$ to the exact values.

3.2. Decentralized predictive encoding algorithm

In the following, the subscript $k|k$ emphasizes that estimates at time k are based on observations up to time k. The complete state vector $[\boldsymbol{q}_1^{\mathrm{T}}\ \boldsymbol{p}_1^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{p}_0$ with $\boldsymbol{q}_1 = \boldsymbol{q}_1(\boldsymbol{p}_0)$ are assumed to be known as prior for the KF and the covariance matrices are estimated via the *empirical covariance function*. The latter allows the computation of the Kalman gain matrix $\boldsymbol{K}_k$ for every iteration $k > 1$.

With our method, cluster m performs the following steps starting at time $k = 1$ (decompositions and compositions of vectors are not stated explicitly):

1. Compute the prediction
$$\begin{bmatrix}\bar{\boldsymbol{q}}_{k+1|k}\\ \check{\boldsymbol{p}}_{k+1|k}\end{bmatrix} = \begin{bmatrix}\bar{\boldsymbol{\Phi}}_{11} & \check{\boldsymbol{\Phi}}_{12}\\ \Delta_t\bar{\boldsymbol{I}} & \check{\boldsymbol{I}}\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{q}}_{k|k}\\ \check{\boldsymbol{p}}_{k|k}\end{bmatrix}. \tag{6a}$$
2. Determine the innovation vector
$$\boldsymbol{\varepsilon}_{k+1}^{(m)} = \bar{\boldsymbol{p}}_{k+1}^{(m)} - \bar{\boldsymbol{p}}_{k+1|k}^{(m)}. \tag{6b}$$
3. Send $\boldsymbol{\varepsilon}_{k+1}^{(m)}$ to cluster $\overline{m}$ and receive $\boldsymbol{\varepsilon}_{k+1}^{(\overline{m})}$ from cluster $\overline{m}$.
4. Correct the prediction via
$$\begin{bmatrix}\bar{\boldsymbol{q}}_{k+1|k+1}\\ \check{\boldsymbol{p}}_{k+1|k+1}\end{bmatrix} = \begin{bmatrix}\bar{\boldsymbol{q}}_{k+1|k}\\ \check{\boldsymbol{p}}_{k+1|k}\end{bmatrix} + \boldsymbol{K}_k\boldsymbol{\varepsilon}_{k+1}^{(\overline{m})}. \tag{6c}$$
5. Use $\bar{\boldsymbol{p}}_{k+1} = \bar{\boldsymbol{p}}_{k+1|k+1}$ for the decentralized estimation of the field $[\boldsymbol{p}_{k+2}^{(m)\mathrm{T}}\ \boldsymbol{p}_{k+2}^{(m)\mathrm{T}}]^{\mathrm{T}}$.
6. Increase k by one and go to step 1.

The predictor (6) scales straightforwardly with the number $\check{\mathcal{N}}$ of the neighboring clusters. A larger number of neighborhoods entails a larger number of unknown pressure states which have to be estimated in eeach time step. This increases the dimension of the state vector $[\bar{\boldsymbol{q}}_k^{\mathrm{T}}\ \check{\boldsymbol{p}}_k^{\mathrm{T}}]^{\mathrm{T}}$ in (5a) but still exploits the structure of the global $\boldsymbol{\Phi}_{12}$.

4. NUMERICAL RESULTS

The level of redundancy in a random signal is specified by the autocorrelation and, equivalently, the power spectral density (PSD). For a white signal, the former equals the delta function while the latter is a constant. Beyond that, several methods are used to show the flatness of a PSD and to define a distance between two of them. We recall briefly those definitions.

Let $S(\omega)$ denote the PSD of a discrete time process defined on $[-\pi, pi)$. In the sequel we use the notion of the distance between two PSDs S_1 and S_2 from [15]. There, Georgiou's distance $d(S_1, S_2)$ is defined by
$$\ln\left(\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{S_1(\omega)}{S_2(\omega)}\mathrm{d}\omega\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{S_2(\omega)}{S_1(\omega)}\mathrm{d}\omega\right)\right). \tag{7a}$$
It induces a metric tensor in a manifold $\mathfrak{P}$ of PSDs (up to scaling factors). Integration of the metric tensor over a geodesic line between two PSDs gives the path length between them in the manifold,
$$\ell(S_1, S_2) = \operatorname*{var}_{\omega}\{\ln S_1(\omega) - \ln S_2(\omega)\}. \tag{7b}$$
To measure the whiteness of a signal we choose one of the PSDs to be constant. In addition, we consider a traditional metric used to measure the flatness of a PSD [8],
$$\mathrm{fl}(S_1) = \frac{e^{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln S_1(\omega)\mathrm{d}\omega}}{\frac{1}{2\pi}\int_{-\pi}^{\pi}S_1(\omega)\mathrm{d}\omega}. \tag{8}$$

As illustrative example, a 2-D rectangular hallway from [7] is simulated using the FDM from above. This hallway and thus the nodes are portioned into two clusters in common with Figure 1. The source $\boldsymbol{s}_k$ is modelled by a Ricker wavelet with additive white Gaussian noise. All parameter values are summarized in Table 1.

In this example, 10^4 time steps were simulated with the goal of comparing the statistics of the original signal $\bar{\boldsymbol{p}}_k^{(1)}$ and the whitened signal $\boldsymbol{\varepsilon}_k^{(1)}$, $k = 1, \ldots, 10^4$. Figure 2 presents the temporal and spatial empirical *autocorrelation functions* (ACF) averaged over space and time, respectively for the actual state vector and the innovations vector (in the temporal case, the source ACF is also shown). The (small) residual noise in Figure 2a corresponds to estimation errors of the KF and is neglected in the following. Clearly, our KF-based predictor succeeds in decorrelating the state vector both temporally and spatially. At time lag $k = \pm 1$ there are two minima that stem from the structure of the state transition model (2).

The distance and flatness metrics introduced above are applied to the original state vector and to the innovations vector, averaged over all boundary nodes in cluster 1. The results are summarized in Figure 3. In particular, S_{i} is the discrete-time Fourier transform of a modified innovation ACF where the amplitudes between ± 0.1 are truncated. Hence the high-frequency component in Figure 2a are removed. S_0 = constant denotes the reference PSD of a perfect white signal while S_{p} is the PSD of the actual state vector.

Furthermore, the improvement achieved by our approach strongly impacts the error resulting from the subsequent quantization. Consider an 8-bit quantizer whose dynamic range is matched to minimum and maximum of the corresponding signals. In the example considered, the quantization mean square error (MSE) incurred with the innovations approach is -100 dB. In comparison, the MSE resulting from quantization of the original state equals -13.5 dB.

5. CONCLUSIONS

Exploiting the spatio-temporal field dependencies of the field reduces the communication overhead in a clustered sensor network. The differential field state vector encoding builds on a Kalman prediction step governing the field's spatio-temporal evolution. Only the innovations/prediction error vector needs to be exchanged among clusters. The resulting decentralized sequential algorithm uses the empirical covariances and Kalman gain matrix $\boldsymbol{K}_k$ which are pre-computed and stored in a look-up table. As verified for the example of an acoustic field, our method succeeds in decorrelating the relevant state subvector. The approach reduces transmit power and quantization errors

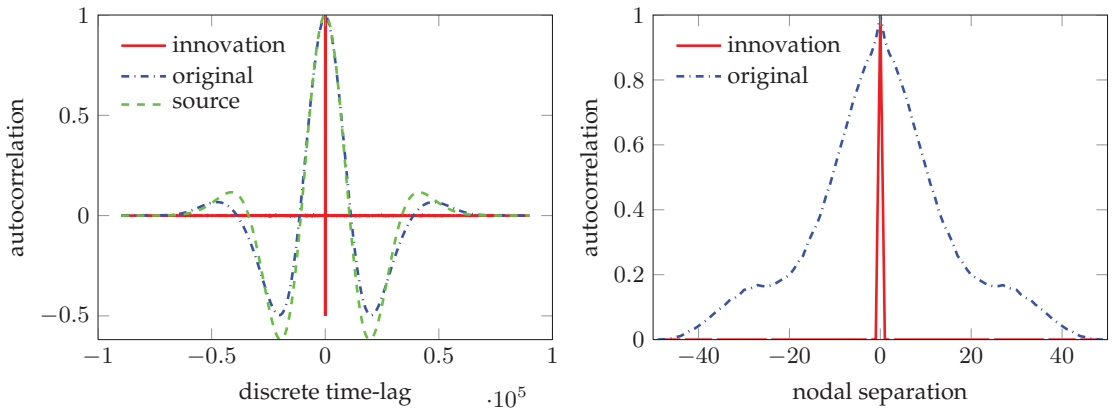


Fig. 2. Empirical autocorrelation functions (ACF) normalized to unit power. The temporal ACF (a) is averaged over nodes $\partial\mathcal{L}^{(1)}$. The spatial ACF (b) is a function of nodes $\partial\mathcal{L}^{(1)}$ and is averaged over time.

Table 1. **Settings for the simulated hallway**

quantity	*notation*	*value*
rectangular area	$I \times J$	50×50
	Δ_t	371 ns
	Δ_r	12.24 cm
rectangular area	$I \times J$	50×50
acoustic speed	c	340 m/s
source shape	$s_0(t)$	$\mathrm{ricker}(t - 16.7\,\mathrm{ms})$
source location	(i_0, j_0)	$(25, 25)$
source noise	$\sigma_{\boldsymbol{s}}$	$0.001/(\Delta_t c^2)$
8 sensors		$\{(i,j) : i = 1, \cdots, 4; j = 1, 4\}$
cluster boundary		betw. $i = 10$ and 11
rigid walls $\boldsymbol{r} \in \partial\mathcal{L}_\mathrm{w}$		$\partial_t p(\boldsymbol{r}, t) = 0$
open doors $\boldsymbol{r} \in \partial\mathcal{L}_\mathrm{d}$		$\frac{1}{c}\partial_t p(\boldsymbol{r}, t) - \boldsymbol{\nabla} p(\boldsymbol{r}, t) \cdot \boldsymbol{n} = 0$
initial conditions		$\partial_t p(\boldsymbol{r}, 0) = 0, p(\boldsymbol{r}, 0) = 0$

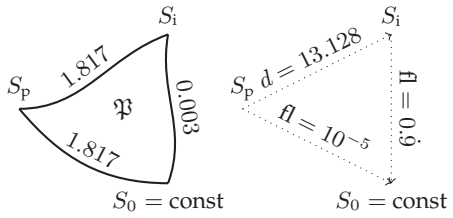


Fig. 3. Length ℓ of geodesics in a manifold $\mathfrak{P}$ of PSDs (left) vs. flatness (right) related to the empirical ACF over time (cf. with Figure 2a). S_0 = constant is the PSD of a perfect white signal, S_p is the PSD of the pressure states and S_i is the modified PSD of the innovation.

in clustered sensor networks for physics-based field estimation.

6. REFERENCES

[1] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, Boston, 2004.

[2] L.A. Rossi, B. Krishnamachari, and C.C.J. Kuo, "Distributed parameter estimation for monitoring diffusion phenomena using physical models," in *IEEE SECON*, 2004.

[3] T. Zhao and A. Nehorai, "Distributed sequential Bayesian estimation of a diffusive source in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1511, 2007.

[4] O. Hlinka, O. Slučiak, F. Hlawatsch, P. M. Djurić, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," Available online: arXiv:1108.6214v2 [stat.AP], Dec. 2012.

[5] M. Coates, "Distributed particle filters for sensor networks," in *Proceedings of the 3rd international symposium on Information Processing in Sensor Networks*. ACM, 2004, pp. 99–107.

[6] F. Sawo, *Nonlinear State and Parameter Estimation of Spatially Distributed Systems*, Ph.D. thesis, Universität Karlsruhe, 2009.

[7] F. Xaver, G. Matz, P. Gerstoft, and C. Mecklenbräuker, "Localization of acoustic sources using a decentralized particle filter," *EURASIP JWCN*, 2011.

[8] P. P. Vaidyanathan, *The Theory of Linear Prediction*, Synthesis Lectures on Signal Processing. Morgan & Claypool Publishers, 2008.

[9] Brian D. O. Anderson and John B. Moore, *Optimal Filtering*, Rentice-hall, Inc., Englewood Cliffs, New Jersey 07632, information and system science series edition, 1979.

[10] F. Jensen, W.A. Kuperman, M.B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, American Institute of Physics Press, New York, 1 edition, 1994.

[11] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[12] D. Simon, *Optimal state estimation: Kalman, H [infinity] and nonlinear approaches*, Wiley-Interscience, 2006.

[13] S.M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*, vol. 1, Pearson Education, NJ, 1993.

[14] C Yardim, Z-H Michalopoulou, and P Gerstoft, "An overview of sequential bayesian filtering in ocean acoustics," *IEEE Oceanic Eng*, vol. 36(1), pp. 73–91, 2011.

[15] T.T. Georgiou, "Distances between power spectral densities," *IEEE Transactions on Signal Processing*, vol. 55(8), pp. 3995–4003, August 2007.