Mathematical Statistics
Stockholm University

Bayesian Distributed Detection by Sensor Networks with Missing Data

Ola Hössjer
Tommy Nyberg
Stefan Petrovic
Frank Sjöberg
Tommy Öberg

Research Report 2012:6

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2012:6**,
http://www.math.su.se/matstat

Bayesian Distributed Detection by Sensor Networks with Missing Data

Ola Hössjer*
Tommy Nyberg†
Stefan Petrovic, Tommy Öberg‡
Frank Sjöberg

April 28, 2012

Abstract

We consider detection and identification of a moving target, using a network of sensors. The target emits a signal; a stationary stochastic process corrupted by additive noise, independently across sensors. Before inter-sensor communication, all sensors reduce external data as energy over disjoint frequency bands and time blocks. One sensor, the internal fusion center (IFC), gathers feature vectors from the other sensors, possibly after message passing. Using Bayesian decision theory, it decides for presence or absence of the target and computes a maximum posterior estimate of target (trajectory and spectral) parameters. The main novelties of the paper are: 1) To apply statistical theory of missing data to an inter-sensor communication protocol which censors weak signals before transmission and an imperfect channel in which some transmitted signals are lost. A Naive Bayes approximate detector is defined, which requires recursive computation of reception probabilities. 2) To derive asymptotic approximations of the distribution of the spectral feature vectors and a Laplace type approximation of the detector. The performance of the proposed Bayesian detector is shown in a simulation study to be only slightly inferior to that of an ideal Bayesian detector (with no missing data), as well as superior to a naive detector.

Key words: Bayesian detection, data reduction, distributed detection, internal fusion center, missing data, routing, periodogram, sensor network.

*Dept. of Mathematics, Stockholm University
†Clinical Cancer Epidemiology, Dept. of Oncology and Pathology, Karolinska Institutet
‡Swedish Defence Research Agency

1 INTRODUCTION

Over the last decades, decentralized signal processing with fusing of information has gained considerable attention, as reviewed in [2], [30], [32] and [33]. The original applications were in radar ([28]), driven by the need to decrease communication cost. More recent applications involve wireless sensor networks ([5]). A distributed network consists of a number of sensors which quantize/compress received data locally before transmitting it to a common internal fusion center (IFC), which in this paper is one of the sensors. The IFC gathers information in order to detect a signal from a target, so called **distributed detection** ([29],[30]) and possibly also estimates attributes of the target, so called **distributed estimation** ([7]). This may include position and velocity (target tracking) and emitted frequencies (target identification). An alternative strategy, when no IFC exists, is to exchange messages between sensors in order to reach consensus about a common decision ([27]).

Our motivating application is underwater surveillance, although the methodology applies more widely. In more detail, the objective is short-term time-critical underwater exploration and protection of an area surrounding e.g. a harbor or other installation, using a network of wireless acoustic sensors. The target represents a potential threat, such as a submarine or diver. Other potential applications are detection and localization of chemical spills or detection of changes in wildlife populations.

To this end, we propose a model for combined distributed detection and estimation of a moving target by means of a network of synchronous sensors that individually lack array-processing capacity. The sensors receive attenuated time delayed versions of a possible stationary signal emitted from a target, corrupted by additive noise. The IFC operates within a time window, during which the target is assumed to move linearly, and employs Bayesian decision theory, incorporating the possibility to discriminate between several possible types of targets. The framework is quite general, since it involves both data compression of raw sensor measurements, data communication, detection and estimation. Data compression is achieved by dividing the time window into short time blocks, and sensors convert raw data into feature vectors during each time block. These are functions of the periodogram and contain estimated signal energies over disjoint frequency bands in the spectral domain. For the underwater application, raw signals are typically sampled at rate 10^3 Hz; time blocks are of the order 1 second and the time window of the order 10 seconds.

After feature vector extraction, distributed detection incorporates two types of missing data:

a) In order to increase life-time of power supply, feature vectors are censored and transmitted only when they signify presence of a target.

b) The communication channel is unreliable due to additive noise, multiplicative fading or lost messages.

The censoring problem a) has been studied, for instance, in [1] and [26]. For distributed detection, each sensor should optimally censor messages based on the likelihood ratio when the pure noise and target signals are independent across sensors, with a known distribution. If the target signal distribution involves unknown parameters, a robust minimax or a generalized likelihood ratio approach can be employed, provided data is independent across sensors. We apply a somewhat simpler and intuitively appealing censoring rule: Signals with large enough energy over the studied spectral domain are transmitted, which under certain assumptions is asymptotically optimal. In [25], locally optimal and feedback censoring rules are derived for hierarchical sensor networks.

For distributed estimation, [10] and references therein consider optimal linear data compression of raw data, using various types of regression models. Some rather general results on optimal data reduction are obtained in terms of minimizing sum of variances of estimated parameters. Optimal censoring rules are derived in [21] for linear models, within a maximum likelihood and Bayesian maximum aposteriori framework, using optimization with the amount of transmitted messages as side constraint. These results are not directly transferable to our setting though, since our spectral feature vectors are nonlinear functions of raw data and we consider distributed detection and estimation jointly.

As reviewed in [6], there are various ways of handling b), depending on whether sensors have complete, partial or no knowledge of the channel. In our case we assume that the additive noise distribution is estimated from training data, with some lost messages. We therefore employ message passing within the network, not only for transmitting messages to the IFC, but also for for reducing transmission losses. To this end, we use a simple routing protocol with four rules by which no sensor forwards the same feature vector twice, too old messages are not forwarded and a binary routing matrix prescribes which pairs of sensors that forward each others feature vectors.

The first novelty of our approach is to apply the statistical theory of missing data, as outlined in [18] and [11], to distributed detection. Since our routing rules are independent of transmitted feature vectors, data with no censoring is Missing Completely at Random (MCAR). Then the missingness mechanism can be ignored and it suffices to know the distribution of feature vectors. Alternatively, if some messages are censored, data is Not Missing at Random (NMAR). Then reception probabilities of feature vectors have to be included in the likelihood. To this end, we devise a recursive algorithm for computing these probabilities in Appendix E. Dependence across sensors makes the Bayesian detector much more complicated, requiring summation over all possible sent but lost messages. For this reason we make a simplifying so called Naive Bayes independence assumption. Even though this involves some loss of information, it is known that NB classifiers often perform surprisingly well ([8],[13]).

As a second novelty we provide new results on the asymptotic multivariate normal distribution of periodogram based feature vectors, as detailed in Appendix C. These are used to define an approximation of the optimal Bayes detector and maximum a posteriori (MAP) estimator. We then introduce a second layer of approximation: Within our joint detection and estimation framework, the null hypothesis (no target) is simple, whereas each of the the alternative hypotheses (presence of a target of one of several possible types) is composite. The Bayes detector therefore requires numerical evaluation of a high-dimensional integral, which is facilitated using a Laplace type approximation (Appendix D).

In a simulation study, we compare detection and estimation performance of the Bayes detector with 5% sent feature vectors, the ideal detector with a full data set at the IFC and a naive detector, also with 5% sensor activity, but sensors only communicate individual decisions to the IFC, not feature vectors.

In Section 2 we introduce the statistical model for emitted and received signals. Data compression, message passing and data transmission is presented in Section 3, whereas distributed detection and estimation is treated in Section 4. The simulation study is presented in Section 5, possible extensions are discussed in Section 6 and the technical results are gathered in the appendices.

2 STOCHASTIC MODEL FOR OBSERVATIONS

2.1 Emitted and Received Signals

Let X_t be the signal emitted from the target at **time points** $t \in \mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ sampled at $10^3 - 10^4$ Hz. We assume that X_t is a zero mean stationary Gaussian process with covariance function $C_X(n) = \mathrm{Cov}(X_t, X_{t+n})$

and symmetric spectral density

$$S_X(\omega) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\omega} C_X(n), \quad -\pi \le \omega \le \pi. \tag{1}$$

Stationarity tacitly requires that important characteristics of the target, such as speed, are more or less constant during the time window along which the detector operates. The target passes a network with N sensors during some time interval including time point τ and

$$Z_{jt} = A_j X_{t-D_j} + \varepsilon_{jt} \tag{2}$$

is the raw data signal received by Sensor $j \in \{1, \ldots, N\}$ at time point t. The first term in (2) is the target signal, attenuated by a factor A_j and delayed by an amount D_j, and the second term ε_j is noise, representing the received signal from other sources than the target. The latter is a stationary Gaussian process with covariance function C_{ε_j} and spectral density S_{ε_j}. Moreover, all processes $X, \varepsilon_1, \ldots, \varepsilon_N$ are independent, requiring that the sensors are far apart compared to the communicative range of the other noise sources.

2.2 Target Trajectory Parameters

Let $\boldsymbol{x}_t = (x_{t1}, x_{t2}, x_{t3})$ denote the target position at time point t, containing three space coordinates. For t close to τ, we will approximate the target trajectory by a straight line

$$\boldsymbol{x}_t = (\psi_1, \psi_2, \psi_3) + (t - \tau)(\psi_4, \psi_5, \psi_6), \tag{3}$$

corresponding to a position (ψ_1, ψ_2, ψ_3) at time point τ, and constant velocity vector $\boldsymbol{v} = (\psi_4, \psi_5, \psi_6)$. The vector $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_6)$ thus contains the six **target trajectory parameters** from this linear representation.

We assume that the network has gone through an initial phase of cooperative localization ([34]), so that $\boldsymbol{z}_j = (z_{j1}, z_{j2}, z_{j3})$, the position of Sensor j, is known by all sensors. The Euclidean distance between j and the target at time t is

$$\delta_j = \delta_{j,t-\tau}(\boldsymbol{\psi}) = \|\boldsymbol{x}_t - \boldsymbol{z}_j\| = \sqrt{\sum_{a=1}^{3} (\psi_a + (t-\tau)\psi_{a+3} - z_{ja})^2}.$$

Since the emitted signal is assumed to be isotropic, the attenuation term only depends on δ_j, adjusted for delay D_j. In the simulations we focus on

$$A_j = A_{j,t-D_j-\tau}(\boldsymbol{\psi}) = \frac{A}{(\delta_{j,t-D_j-\tau}(\boldsymbol{\psi})/\delta)^2 + 1}, \tag{4}$$

where δ and A as fixed constants. Here δ represents the size of the target, whereas A depends on the size and effectiveness of the sensors' microphones. The attenuation model (4) is very much related to the one used in [23], although we add A as an extra parameter. The exponent 2 of the denominator is typically used in shallow water. The theoretical results only require that A_j is a known function of δ_j though. In Appendix A, we derive the Doppler effect delay D_j.

2.3 Spectral Parameters

Introduce M disjoint subintervals $\Omega_1, \ldots, \Omega_M$ of $[0, \pi]$ and define corresponding **target spectral parameters** $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)$ and noise spectral parameters $\boldsymbol{\eta}_j = (\eta_{j1}, \ldots, \eta_{jM})$ as

$$\xi_m = 2 \int_{\Omega_m} S_X(\omega) d\omega \tag{5}$$

and

$$\eta_{jm} = 2 \int_{\Omega_m} S_{\varepsilon_j}(\omega) d\omega, \tag{6}$$

for $m = 1, \ldots, M$. These parameters define how the expected power of the target signal and noise is decomposed into various frequency components. In practice, M and $\{\Omega_j\}_{j=1}^{M}$ should be chosen based on some apriori knowledge of the target spectral density S_X. This makes it possible adjust the detector settings for targets known to have high energy in certain frequency bands.

3 DATA COMPRESSION AND COMMUNICATION

3.1 Feature Vector

Let $\mathrm{TB}_q = \{t;\, (q-1)r+1 \le t \le qr\}$, for $q \in \mathbb{Z}$ denote non-overlapping **time blocks** consisting of r time points. The length of a time block is typically of the order one second, so that r is of the order $10^3 - 10^4$. This is small enough so that target movement within a time block may be neglected. The data $\boldsymbol{Z}_{jq} = \{Z_{jt}; t \in \mathrm{TB}_q\}$ collected by Sensor j during the q^{th} time block is to be compressed to a feature vector Y_{jq} and sent away if Y_{jq} signifies presence of a target. In this paper we use

$$Y_{jq} = (Y_{jq1}, \ldots, Y_{jqM}) \tag{7}$$

where Y_{jqm} is the energy of $\boldsymbol{Z}_{jq}$ within frequency band Ω_m, see Appendix B for details.

The censoring rule for j is to send Y_{jq} if

$$Y_{jq\cdot} = \sum_{m=1}^{M} Y_{jqm} \geq \lambda_j, \tag{8}$$

where λ_j is a threshold and the dotted index indicates summation over m. It is defined in Section 4.2 (cf. (27)) based on the requirement

$$P(Y_{jq} \text{ sent}|\text{no target present}) = p, \tag{9}$$

so that a proportion $0 < p \leq 1$ of feature vectors are sent in absence of a target. A transmission rule (8) is natural since $Y_{jq\cdot}$ represents the average power of the part of $\boldsymbol{Z}_{jq}$ which belongs to the frequency band of interest; $\cup_{m=1}^{M}\Omega_m$.

3.2 Transmission and Routing

When a message with feature vector Y_{jq} has been sent by Sensor j, it will eventually reach the IFC with some probability, possibly after routing by other sensors. In order to model the missingness mechanism, we introduce P_{ji}, the probability that a message sent by Sensor j is successfully transferred to Sensor i without any forwarding by other sensors. We regard P_{ji} as known and estimated from training data. When communication and data sampling is conducted by means of acoustic microphones and speakers, see [12], [19] and [24], it is reasonable to choose an isotropic angular distribution of sent messages,

$$P_{ji} = \begin{cases} 1, & j = i, \\ P1_{\{\|\boldsymbol{z}_j - \boldsymbol{z}_i\| \leq \delta_{\max}\}}, & j \neq i, \end{cases} \tag{10}$$

where 1_A is the indicator function for the event A, which equals one if A is true and zero if it is not. It follows from (10) that all messages within a sphere of radius $\delta_{\max}$ are received with probability P and those outside the sphere are never received. Here $\delta_{\max}$ depends on the transmission power of the sensors as well as the amount of absorption and scattering ([19]).

Messages are forwarded or routed using a binary $N \times N$ **routing matrix** $\boldsymbol{F} = (F_{ji})$ and a **routing protocol** consisting of the following four rules:

F1 A message originally from j is forwarded by i only if $F_{ji} = 1$.

F2 $F_{jj} = 0$, i.e. j does not forward any of its own messages.

F3 No feature vectors are forwarded twice.

F4 No feature vectors older than $\bar{w} - 2$ time blocks are forwarded, where $\bar{w} = w_1 + w_3$ is related to the length of a time window and defined in Section IV.

Based on (10) and F1-F4, we devise in Appendix E an algorithm for computing reception probabilities at the IFC.

4 BAYESIAN MODEL AND DETECTOR

In this section, we consider detection of a target as well as parameter estimation during a fixed **time window** $\mathrm{TW} = \{q; -w_1 + 1 \leq q \leq w_2\}$ of $w = w_1 + w_2$ time blocks TB_q, without loss of generality centered around a reference time block TB_0. We choose w small enough so that the target trajectory within TW can be well approximated by a linear function (3). Typically w is of the order 10.

4.1 Hypotheses and Priors

Our model is Bayesian (cf. [31]), with prior distribution on the various hypotheses (target types) as well as the parameters. We formulate the hypotheses to choose between as

$$H_u : \text{Target of type } u \text{ at time point } \tau = [(1-r)/2] \tag{11}$$

for $u = 0, \ldots, \bar{u}$, with $[(1-r)/2]$ the mid point of TB_0. Here $u = 0$ means absence of a target and $1 \leq u \leq \bar{u}$ presence of a target of type u. We let

$$\pi_u = P(H_u \text{ is true}) \tag{12}$$

denote the prior probabilities of the $\bar{u} + 1$ hypotheses.

Let

$$\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\xi}) \in \Theta = \mathbb{R}^6 \times [0, \infty)^M. \tag{13}$$

denote the random vector of **target parameters** that include information about trajectory as well as spectral density, cf. Sections 2.2-2.3. In absence of a target, zero power is emitted within the frequency range $\cup_m \Omega_m$, leading to $\boldsymbol{\xi} = 0$. The target parameters $\boldsymbol{\psi}$ are then unidentifiable, although it will be convenient (and no loss of generality) to assume $\boldsymbol{\psi} = 0$. As a consequence,

$$P(\boldsymbol{\theta} = 0 | H_0) = 1. \tag{14}$$

For $u = 1, \ldots, \bar{u}$, we assume a priori independence between target and spectral parameters under H_u, giving a prior density

$$f_u(\boldsymbol{\theta}) = f_u(\boldsymbol{\psi}) f_u(\boldsymbol{\xi}). \tag{15}$$

For the spectral parameter vector we choose independent and gamma distributed components $\xi_m \sim \Gamma(c_{um}^2/\sigma_{um}^2, \sigma_{um}^2/c_{um})$ under H_u, where $\{c_{um}\}$ and $\{\sigma_{um}\}$ are fixed hyperparameters. This gives a prior density

$$f_u(\boldsymbol{\xi}) = \prod_{m=1}^{M} \frac{(c_{um}/\sigma_{um}^2)^{c_{um}^2/\sigma_{um}^2} (\xi_m)^{c_{um}^2/\sigma_{um}^2 - 1}}{\Gamma(c_{um}^2/\sigma_{um}^2)} \exp(-c_{um}\xi_m/\sigma_{um}^2), \tag{16}$$

with $\Gamma(\cdot)$ the Gamma function, $c_{um} = E(\xi_m)$ the average and $\sigma_{um}^2 = \mathrm{Var}(\xi_m)$ the standard deviation of the power of the signal emitted from a target of type u within the m^{th} frequency band. Alternatively, we could have used a lognormal or inverse gamma prior or some other non-negative distribution. However, the exact form of the prior is less important, given that the mean and variance have been specified. If no apriori knowledge about $\boldsymbol{\xi}$ is available, a flat improper prior (constant f_u) may be used.

For the trajectory parameters we use a multivariate normal distribution prior $N(\boldsymbol{m}_u, \boldsymbol{\Sigma}_u)$, with density

$$f_u(\boldsymbol{\psi}) = (2\pi)^{-3} |\boldsymbol{\Sigma}_u|^{-1/2} \exp(-0.5(\boldsymbol{\psi} - \boldsymbol{m}_u)\boldsymbol{\Sigma}_u^{-1}(\boldsymbol{\psi} - \boldsymbol{m}_u)'), \tag{17}$$

under H_u, where $|\boldsymbol{\Sigma}_u|$ is the determinant of $\boldsymbol{\Sigma}_u$.

We assume $\boldsymbol{m}_u = (0, 0, 0, m_{u4}, m_{u5}, m_{u6})$ and a Block Diagonal covariance matrix $\boldsymbol{\Sigma}_u = \mathrm{BD}(\boldsymbol{\Sigma}_{1:3}, \boldsymbol{\Sigma}_{u,4:6})$. In this case $(0, 0, 0)$ is a reference position of the network, where the target is most likely positioned at time point τ, with prior covariance matrix $\boldsymbol{\Sigma}_{1:3}$. The a priori most likely velocity vector (m_{u4}, m_{u5}, m_{u6}) is allowed to depend on u, since different targets may travel at different speed and the geometry may favour certain directions. We put $(m_{u4}, m_{u5}, m_{u6}) = (0, 0, 0)$ when there is no apriori knowledge of target direction. When specifying the covariance matrices $\boldsymbol{\Sigma}_{1:3}$ and $\boldsymbol{\Sigma}_{u,4:6}$ (of which the latter is allowed to depend on u), the geometry of the network and perhaps also the surrounding area should be taken into account, as well as apriori information about target speed. In conclusion, one should make the entries of $\boldsymbol{\Sigma}$ so large that f_u is fairly flat over all possible trajectories of the target, making the choice of reference position and apriori most likely velocity vector less important. Notice that the speed $\sqrt{\psi_4^2 + \psi_5^2 + \psi_6^2}$ is guaranteed to be positive, although some components of the velocity vector may be negative.

4.2 Full Data Set

Ideally, the IFC would perform detection and estimation based on all feature vectors

$$\boldsymbol{Y} = \{Y_{jq};\ 1 \le j \le N, q \in \mathrm{TW}\} = \{Y_{jqm};\ 1 \le j \le N, q \in \mathrm{TW}, 1 \le m \le M\}. \tag{18}$$

In this subsection, we will find useful approximations of the distribution of $\boldsymbol{Y}$. We assume that a time block corresponds to a duration short enough so that movement of the target within the time block can be neglected. This motivates the assumptions that A_i and D_i are constant for $t \in \mathrm{TB}_q$, The constant values are set to

$$d_{iq} = d_{iq}(\boldsymbol{\psi}) = D_{i,rq}(\boldsymbol{\psi})/r \tag{19}$$

and

$$a_{iq} = a_{iq}(\boldsymbol{\psi}) = A_{i,r(q-d_{iq})}(\boldsymbol{\psi}). \tag{20}$$

The division by r in (19) assures that d_{iq} is given in block time rather than time point units. Given these two approximations, it is shown in Appendix C that asymptotically for large r

$$\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{V}) \text{ given } \tau = [(1-r)/2] \text{ and } \boldsymbol{\theta}. \tag{21}$$

Thus we assume that the target and noise spectra consist of frequencies large enough to warrant that r, the sampling frequency per time block, is large enough for a Central Limit Theorem result (21) to hold (typically, r is of the order $10^3 - 10^4$). The mean vector $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_{jqm})$ has components

$$\mu_{jqm} = a_{jq}^2 \xi_m + \eta_{jm}, \tag{22}$$

the covariance matrix $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\theta}) = (V_{jqm,j'q'm'})$ has entries

$$V_{jqm,jq'm'} = 1_{\{m=m'\}} 1_{\{q=q'\}} \frac{8\pi}{r} \int_{\Omega_m} \left(a_{jq}^2 S_X(\omega) + S_{\varepsilon_i}(\omega)\right)^2 d\omega \tag{23}$$

and

$$V_{jqm,j'q'm'} = 1_{\{m=m'\}} \rho\left(q' - d_{j'q'} - (q - d_{jq})\right) \frac{8\pi}{r} a_{jq}^2 a_{j'q'}^2 \int_{\Omega_m} S_X^2(\omega) d\omega \tag{24}$$

when $j \ne j'$, where

$$\rho(d) = (1 - |d|)_+ \tag{25}$$

can be interpreted as a correlation coefficient, with $x_+ = \max(0, x)$. It is the fraction of overlap of two time blocks of length r at distance rd.

Without further assumptions, we need additional parameters to characterize the integrals in (23) and (24). However, we will make the simplifying assumption that S_X as well as S_{ε_i} can be regarded as constant within each Ω_m. Then

$$\begin{aligned} V_{jqm,jq'm'} &= 1_{\{m=m'\}}1_{\{q=q'\}}2\pi\left(a_{jq}^2\xi_m+\eta_{jm}\right)^2/(|\Omega_m|r), \\ V_{jqm,j'q'm'} &= 1_{\{m=m'\}}2\pi\rho\left(q-d_{jq}-(q'-d_{j'q'})\right)a_{jq}^2a_{j'q'}^2\xi_m^2/(|\Omega_m|r), \end{aligned} \tag{26}$$

making use of the definitions (5)-(6) of ξ_m and η_{jm}.

We can now use (21) and (23) to determine the threshold λ_j in (9). Indeed, since

$$Y_{jq\cdot} \sim N\left(\sum_{m=1}^{M}\mu_{jqm}, \sum_{m=1}^{M}V_{jqm,jqm}\right)$$

and $a_{jq}=0$ in (26) under the null hypothesis, we obtain

$$\lambda_j = \sum_{m=1}^{M}\eta_{jm} + \Phi^{-1}(1-p)\sqrt{2\pi\sum_{m=1}^{M}\eta_{jm}^2/(|\Omega_m|r)}, \tag{27}$$

from (9), (21) and (26), where Φ is the cumulative distribution function of a standard normal random variable.

Under certain simplifying conditions; $M=1$, asymptotic normality (21) of feature vectors and the Naive Bayes independence assumption of Section 4.6.2, a censoring rule (9) with threshold (27) is derived from the likelihood ratio, which, according to [1], is optimal. More generally, the optimal censoring rule will depend on the unknown model parameter $\boldsymbol{\theta}$. Then (27) is a robust choice that doesn't require any sensor to know (an estimate of) $\boldsymbol{\theta}$ or data from other sensors. In addition, such a censoring rule makes the Naive Bayes approximation more accurate.

4.3 Received Data

Let i denote the IFC for some fixed $1 \le i \le N$. The data received by i is incomplete because of unsent, lost or not yet arrived feature vectors at the time of detection. We assume that detection starts just after completion of time block w_3, with $w_3 \ge w_2$, allowing for the possibility of an additional $w_3 - w_2$ time blocks to elapse outside TW before classification starts. The choice of w_3 is a compromise between detector speed and detector performance. We describe this mathematically by introducing the reception indicator variables

$$R_{jq} = 1_{\{Y_{jq}\text{ sent and received by } i \text{ at time block } w_3\}}. \tag{28}$$

Then the data available to i from TW is $(\boldsymbol{R}, \boldsymbol{Y}^{\text{rec}})$, where $\boldsymbol{R} = \{R_{jq};\ 1 \le j \le N, q \in \text{TW}\}$ and $\boldsymbol{Y}^{\text{rec}} = \{Y_{jq};\ 1 \le j \le N, q \in \text{TW}, R_{jq}=1\}$.

4.4 The Detector

Let U denote the unknown target type during TW. According to (12) we have $P(U = u) = \pi_u$. The detector operating on TW is a function $\hat{U} = \hat{U}(\boldsymbol{R}, \boldsymbol{Y}^{\text{rec}})$ of received data at the IFC. In order to analyze $\hat{U}$, we assume that C_u is the cost of not choosing H_u when this is the true hypothesis. We wish to keep the average cost or

$$\begin{aligned}\text{Risk} &= E(C_U 1_{\{\hat{U} \neq U\}}) \\ &= \textstyle\sum_{u=0}^{\bar{u}} C_u \pi_u P(\hat{U} \neq u | H_u) \\ &= C_0 \pi_0 P(\hat{U} \neq 0 | \boldsymbol{\theta} = \boldsymbol{0}) + \textstyle\sum_{u=1}^{\bar{u}} C_u \pi_u \int_{\Theta} P(\hat{U} \neq u | \boldsymbol{\theta}) f_u(\boldsymbol{\theta}) d\boldsymbol{\theta},\end{aligned} \tag{29}$$

as small as possible, where Θ is the parameter space defined in (13) and in the last step we used (14) and (15). The optimal Bayes detector (see e.g. [14], Ch. 4.2) minimizing Risk is

$$\hat{U} = \arg\max\{C_0 \pi_0, C_u \pi_u \int_{\Theta} \Lambda_u(\boldsymbol{\theta}) d\boldsymbol{\theta};\ u = 1, \ldots, \bar{u}\}, \tag{30}$$

with

$$\Lambda_u(\boldsymbol{\theta}) = f_u(\boldsymbol{\theta}) L(\boldsymbol{\theta}) / L(\boldsymbol{0}) \tag{31}$$

and

$$L(\boldsymbol{\theta}) = P(\boldsymbol{R}, \boldsymbol{Y}^{\text{rec}} | \tau = [(1-r)/2], \boldsymbol{\theta}) \tag{32}$$

the apriori weighted likelihood ratio and likelihood function of data received by Sensor i. The Bayesian detector (30) takes both cost, prior information and observed data into account, and an algorithm for computing the integral in (30) based on a Laplace type approximation, is outlined in Appendix D. Whether Sensor i detects a target or not during TW, a Maximum a posteriori estimate

$$\hat{\boldsymbol{\theta}}_u = (\hat{\boldsymbol{\psi}}_u, \hat{\boldsymbol{\xi}}_u) = \arg\max_{\boldsymbol{\theta} \in \Theta} \Lambda_u(\boldsymbol{\theta}) \tag{33}$$

of $\boldsymbol{\theta}$ is computed for all types of targets $u = 1, \ldots, \bar{u}$.

Asymptotically, as the number of time points in TW grows, the Bayes detector (30) is consistent ($P(\hat{U} = U) \to 1$). Therefore, asymptotically, the estimator (33) behaves as if U was known, and then it is asymptotically optimal both in a Bayesian and frequentistic (see for instance Section 6.8 of [17]) setting.

4.5 Likelihood Function

In this subsection, we give an explicit expression of the likelihood function (32), which is quite complicated because of the missing data problem. We start by describing the model for missing data. To this end, define send indicator variables

$$S_{jq} = 1_{\{Y_{jq} \text{ sent}\}} = 1_{\{Y_{jq\cdot} > \lambda_j\}}$$

and variables indicating successful transfer,

$$I_{qnji} = \begin{cases} 1, & i = j, \\ 1_{\{Y_{jq} \text{ received by } i \text{ after time block } n | Y_{jq} \text{ sent}\}}, & i \neq j. \end{cases}$$

We assume that all I_{qnji} are independent for fixed (q, n) but varying (j, i), with

$$P(I_{qnji} = 1) = P_{n-q,ji}, \tag{34}$$

i.e. the probability of successful transfer only depends on the pair of sensors (j, i) and the time delay $n - q$. In Appendix E we show how to compute $P_{n-q,ji}$. The independence assumption on $\{I_{qnji}\}$ can of course be questioned. It corresponds to a situation when the channel state between sensors varies slowly with time.

The reception indicator variables (28) are now defined as

$$R_{jq} = S_{jq} I_{q,w_3,ji}, \tag{35}$$

where w_3 is the time block after which detection starts (cf. Section 4.3). In vector form, we rewrite (35) as $\boldsymbol{R} = \boldsymbol{S} \cdot \boldsymbol{I}$, where $\boldsymbol{S} = \{S_{jq};\, 1 \leq j \leq N, q \in \text{TW}\}$, $\boldsymbol{I} = \{I_{q,w_3,ji};\, 1 \leq j \leq N, q \in \text{TW}\}$ and $\cdot$ denotes elementwise multiplication. Whereas $\boldsymbol{R}$ is known, both $\boldsymbol{S}$ and $\boldsymbol{I}$ are unknown. Indeed, a non-received feature vector is either not sent or sent and then lost during transfer. We therefore compute the likelihood by summing over all $\boldsymbol{S} \geq \boldsymbol{R}$, where $\boldsymbol{S} \geq \boldsymbol{R}$ is interpreted componentwise, i.e. $S_{jq} \geq R_{jq}$ for $1 \leq j \leq N$ and $q \in \text{TW}$. This yields

$$\begin{aligned} L(\boldsymbol{\theta}) &= P(\boldsymbol{R}, \boldsymbol{Y}^{\text{rec}}) \\ &= \textstyle\sum_{\boldsymbol{S};\boldsymbol{S} \geq \boldsymbol{R}} P(\boldsymbol{S}) P(\boldsymbol{R}|\boldsymbol{S}) P(\boldsymbol{Y}^{\text{rec}}|\boldsymbol{S}, \boldsymbol{R}) \\ &= \textstyle\sum_{\boldsymbol{S};\boldsymbol{S} \geq \boldsymbol{R}} P(\boldsymbol{Y} \in A_{\boldsymbol{S}}) P(\boldsymbol{R}|\boldsymbol{S}) P(\boldsymbol{Y}^{\text{rec}}|\boldsymbol{Y} \in A_{\boldsymbol{S}}, \boldsymbol{R}) \end{aligned} \tag{36}$$

where

$$P(\boldsymbol{R}|\boldsymbol{S}) = \prod_{j,q} P_{w_3-q,ji}^{R_{jq}} (1 - P_{w_3-q,ji})^{S_{jq}(1-R_{jq})}, \tag{37}$$

$w_3 - q$ is the time delay between sending and detection and

$$A_{\boldsymbol{S}} = \{\boldsymbol{y} = (y_{jqm});\, \sum_{m=1}^{M} y_{jqm} \begin{array}{c} \geq \\ < \end{array} \lambda_j \text{ if } S_{jq} = \begin{array}{c} 1 \\ 0 \end{array} \}$$

is a subset of $\mathbb{R}^{NwM}$, corresponding to all full data sets (18) consistent with $\boldsymbol{S}$. Given that a feature vector is sent, we assume in (37) quite naturally that transmission losses are independent of the feature vectors sent.

The statistical model, with all parameters, hidden variables and observed data, is summarized in a Directed Acyclic Graph in Fig. 1.

4.6 Simplifications of the Likelihood

In most applications the likelihood (36) is intractable, due to the summation over $\boldsymbol{S}$. This has to do with the way data is missing. The statistical theory of missing data, as outlined in the book [18], can be applied to describe this. On one hand, transmission losses $\boldsymbol{I}$ result in data Missing Completely at Random (MCAR). This is known not to complicate statistical analysis. On the other hand, losses due to censoring depend on the unsent feature vectors, cf. (8). This is referred to as data Not Missing at Random (NMAR), which is known to complicate statistical analysis. In order to obtain a manageable detector, we will consider two different simplifications of the likelihood.

4.6.1 All Feature Vectors Sent

By putting $p = 1$, we get $P(\boldsymbol{S} = \boldsymbol{1}) = 1$, where $\boldsymbol{1} = (1, \ldots, 1)$. Then data is only MCAR due to transmission losses, meaning that the missing data can be ignored in the analysis. This is reflected in the likelihood (36), which simplifies to

$$\begin{aligned} L(\boldsymbol{\theta}) &\overset{p=1}{=} P(\boldsymbol{R}|\boldsymbol{S} = \boldsymbol{1})P(\boldsymbol{Y}^{\mathrm{rec}}|\boldsymbol{R}) \\ &\propto P(\boldsymbol{Y}^{\mathrm{rec}}|\boldsymbol{R}) \\ &\propto |\boldsymbol{V}^{\mathrm{rec}}|^{-1/2} \exp\left(-0.5(\boldsymbol{Y}^{\mathrm{rec}} - \boldsymbol{\mu}^{\mathrm{rec}})(\boldsymbol{V}^{\mathrm{rec}})^{-1}(\boldsymbol{Y}^{\mathrm{rec}} - \boldsymbol{\mu}^{\mathrm{rec}})'\right), \end{aligned} \tag{38}$$

where the proportionality constants are independent of $\boldsymbol{\theta}$ and cancel out in the likelihood ratio (31). In the last equation, $\boldsymbol{\mu}^{\mathrm{rec}}$ is the subvector of $\boldsymbol{\mu}$ and $\boldsymbol{V}^{\mathrm{rec}}$ the submatrix of $\boldsymbol{V}$ corresponding to received feature vectors $\boldsymbol{R}$. As a result, the transmission probabilities of Appendix E are not needed (the first proportionality constant in (38)) and this considerably increases the robustness of the detector/estimator. The disadvantage of $p = 1$ is the increased communication and energy resources.

4.6.2 Ignoring Dependence Between Feature Vectors

Assuming a diagonal $\boldsymbol{V}$, we ignore dependence between feature vectors, a so called Naive Bayes approach. This implies that all Y_{jq} (and hence also all

S_{jq}) are regarded as independent, which involves some loss of information. However, much of the information on coexistence of target signals at different sensors is still retained in the mean vector components (22), since they involve the attenuation terms a_{jq}. Let

$$Q_{jq} = Q_{jq}(\boldsymbol{\theta}) = P(Y_{jq\cdot} \geq \lambda_j | \tau = [(1-r)/2], \boldsymbol{\theta}) = 1-\Phi\left((\lambda_j - \sum_{m=1}^{M} \mu_{jqm})/\sqrt{\sum_{m=1}^{M} V_{jqm}}\right)$$

denote the probability of sending Y_{jq}. The independence assumption implies that $P(\boldsymbol{Y} \in A_{\boldsymbol{S}})P(\boldsymbol{Y}^{\text{rec}}|\boldsymbol{Y} \in A_{\boldsymbol{S}}, \boldsymbol{R})$ in (36) can be written as a product

$$\prod_{jq} (1 - Q_{jq})^{(1-S_{jq})} \left(Q_{jq}(1 - P_{w_3-q,ji})\right)^{S_{jq}(1-R_{jq})} \left(P_{w_3-q,ji}P(Y_{jq})\right)^{S_{jq}R_{jq}}.$$

By inserting this expression into (36) and interchanging the order of summation over $\boldsymbol{S}$ and product over $1 \leq j \leq N$ and $q \in$ TW, we arrive at the approximation

$$\begin{array}{lcl} L(\boldsymbol{\theta}) & \overset{V \text{ diagonal}}{=} & \prod_{jq}(1 - Q_{jq}P_{w_3-q,ji})^{(1-R_{jq})} \left(P_{w_3-q,ji}P(Y_{jq})\right)^{R_{jq}} \\ & \propto & \prod_{jq}(1 - Q_{jq}P_{w_3-q,ji})^{(1-R_{jq})} P(Y_{jq})^{R_{jq}} \end{array} \tag{39}$$

of the likelihood, where

$$P(Y_{jq}) = (2\pi)^{-M/2}|V_{jq}|^{-1/2} \exp\left(-0.5(Y_{jq} - \mu_{jq})V_{jq}^{-1}(Y_{jq} - \mu_{jq})'\right), \tag{40}$$

and $\mu_{jq} = \mu_{jq}(\boldsymbol{\theta})$ and $V_{jq} = V_{jq}(\boldsymbol{\theta})$ are the relevant $1 \times M$ subvector and $M \times M$ submatrix of $\boldsymbol{\mu}$ and $\boldsymbol{V}$ respectively. Notice that data is NMAR in (39), implying that the transmission probabilities of Appendix E have to be taken into account.

5 SIMULATIONS

In this section we present simulation results showing detection performance and estimation accuracy for a network during one time window TW. Some general system parameters included in all simulations are listed in Table 1. We have used a network with 5 sensor nodes and maximal amount of routing ($F_{ij} = 1$ when $i \neq j$), cf. Fig. 2. Although the network is small we believe the results are quite relevant for a larger network since a weak target rarely will be observed by more than 5 sensors at a time and stronger targets are easy to detect anyhow. Fig. 2 also shows the tracks of three different targets (numbered 1,2, and 3) during one time window. The average speed

Table 1: General systems parameters used in all the simulations.

Symbol	Value	Description
r	10 000	Sampling rate per time block
w	10	Number of time blocks in a time window ($w_1 = 10, w_2 = w_3 = 0$)
M	1	Number of frequency bands
N	5	Number of sensors
δ	1 m	Target size
p	0.05	Communication activity
$\delta_{\max}$	50 m	Maximum internode distance for direct comm.
P	0.95	Probability of correct internode communication.

is relatively low, in the range 0.1-0.14 m/s, and therefore we neglected the Doppler effect. The number of the target indicates how many nearest sensors that are located at approximately the same distance from the target, and the other sensors are located further away. The target is restricted to move in two dimensions, corresponding to shallow water. Otherwise the prior of the trajectory is fairly flat (see Table 2) with a standard deviation of 40 m and 2 m/s of the position and velocity along each of the first two coordinates.

We assume that all noise processes have the same distribution, $\boldsymbol{\eta}_1 = \ldots = \boldsymbol{\eta}_N =: \boldsymbol{\eta} = (\eta_1, \ldots, \eta_M)$, kept fixed in all simulations. The spectral parameters have a rather flat prior, with a coefficient of variation $\sigma/c = 0.82$ (see Table 2). Define signal strength by means of the **Target Signal-to-noise Ratio**

$$\mathrm{TSNR}(dB)(\boldsymbol{\xi}) = 10 \log_{10} \left(\frac{\sum_{m=1}^{M} \xi_m}{\sum_{m=1}^{M} \eta_m} \right)$$

and the **Sensor Received Signal-to-noise Ratio**

$$\mathrm{SNR}_{jq}(dB)(\boldsymbol{\theta}) = 20 \log_{10}(a_{jq}(\boldsymbol{\psi})) + \mathrm{TSNR}(dB)(\boldsymbol{\xi})$$

of Sensor j during time block q. Since SNR_{jq} includes both target signal strength and closeness, it varies with time and between sensors. Since the targets in all three scenarios do not move so much during one time window, the SNR is almost constant. The SNR-values in the results below refer to the largest value over all sensors and measurements periods in each simulation.

5.1 Detection Performance

In order to evaluate the performance of the proposed detection method, Monte-Carlo simulations have been performed on the three scenarios in Fig. 2. Data from 10 consecutive time blocks are collected and routed to the upper middle node, the IFC, with a packet loss probability of 5% within a sphere of

Table 2: Target a priori distribution parameters.

Symbol	Value	Description
$\bar{u}$	1	Number of possible target types
$\boldsymbol{m}$	$\mathbf{0}$	Position and velocity mean.
$\boldsymbol{\Sigma}$	$\mathrm{diag}(40^2, 40^2, 0, 2^2, 2^2, 0)$	Position and velocity covariance matrix.
c	1.8	Mean signal power.
σ^2	2.2	Variance of signal power.

radius 50 m, i.e. $P = 0.95$ and $\delta_{\max} = 50$. The best SNR (over all sensors and time blocks) varies between -25 dB and -10 dB, and sensors located further away will of course experience lower SNR-values. Detection performance is presented using receiver operating characteristic (ROC) curves, where detection probability $P_D = P(\hat{U} = 1|H_1)$ is plotted versus probability of false alarm $P_{\mathrm{FA}} = P(\hat{U} = 1|H_0)$. Arbitrary P_{FA} can be achieved by adjusting $C_1\pi_1/(C_0\pi_0)$, i.e. by varying the cost of false alarm and/or missed target

For hypothesis H_0 50000 simulations have been performed, which gives a satisfying accuracy when estimating the false alarm rate P_{FA}. For H_1, *i.e.*, when a target is present, 1000 simulations have been made. The results are presented as ROC-curves in Fig. 3-4 for different SNR-values and target scenarios. Two curves are plotted for each scenario. One curve shows the performance of an *ideal* detector without data censoring where the sensors always transmit data from all time blocks with no delay or packet loss. The energy efficient detector has $p = 0.05$, which means that it only transmits 5% of the data packets under H_0. For this detector we use the independence approximation of Section 4.6.2, with reception probabilities calculated as in Appendix E.

Another way of presenting the results is to plot P_D as a function of SNR for a fixed P_{FA}. This is shown in Fig. 5-7 for $P_{\mathrm{FA}} = 0.001$. These graphs also include a third curve, which represents a trivial detector that decides H_1 if the total signal energy exceeds a certain threshold more than once during a time window (counted over all sensors and time blocks). The trivial detector also has 5% transmission activity under H_0. In these figures we can see that the price for saving 95% of the communication energy consumption is a 1 dB loss in detection performance. The trivial detector is 2 dB worse than our (approximation of the) optimal detector.

Fig. 7 shows the detection performance for all three target scenarios in the same graph. This graph also shows the performance for a case where the channel deviates from the $1/\delta_j^2$ model in (4) by a log-normal variation with 2 dB standard deviation. This deviation can represent shadow fading, nonisotropic sources etc. A very interesting result is that the detector, designed for a $1/\delta_j^2$ channel, performs just as well or even better when the channel deviates from the model with a moderate log-normal fading variation. This can be interpreted as robustness of the detector against model misspecifications, and is particularly apparent for weak target signals, when occasional increase of received SNRs due to fading is particularly beneficial. We can also see that a target that is close to only one sensor requires 2 dB higher receiver SNR than a target located close to three sensors in order to achieve the same detection probability.

5.2 Localization Performance

The detection method indirectly yields a MAP-estimate (33) of the target position (and power spectral parameters). Fig. 8-13 show how the accuracy of the position estimate is affected by: SNR, the number of closely located sensors, log-normal channel variations, and the data censoring (energy saving). Only estimates that correspond to a detected target have been used in the simulations. Observe that the SNR-scale ranges from -20 dB to +5 dB in these simulations, and that the scale in the graphs that show mean square error (MSE) is in decibel. Hence, in absence of bias, an MSE of -20, 0 and 20 dB correspond to a standard deviation of 0.1, 1 and 10 meters respectively.

In Fig. 8 and 9 we can see that the energy saving does not cause any significant degradation to the position estimate. In general, accuracy improves with increasing SNR and number of nearest sensors. The slight non-monotonicity of some curves for small SNRs is due to the conditioning on target detection. However, if the channel has a log-normal deviation from the $1/\delta_j^2$-model, the position estimate does not improve with increasing SNR, which can be seen in Fig. 10-11. This is because log-normal channel variation is a multiplicative perturbation, so a stronger signal yields a larger disturbance caused by the channel mismatch. However, more observations of the target, either over time or by more sensors, will give better estimates even if the channel is not an ideal $1/\delta_j^2$-channel.
Fig. 12 shows the bias of the position estimate, and Fig. 13 shows the ratio of estimated and true signal power. For low SNR-values the method tends to over estimate the signal power, and this is strongly correlated with the bias of the position estimate. When the method believes that the target is farther away than in reality (positive bias) the estimated signal power must also be higher than in reality if the measured signal powers shall match the modeled channel attenuation.

6 CONCLUSIONS

We have proposed a framework for Bayesian distributed detection and estimation of a linearly moving target. In particular, our method handles missing data that is either not sent (to save energy) or sent and lost (due to imperfect communication between sensors). In a simulation study, we studied the impact of missing data compared to an ideal (approximately) optimal detector with no missing data and a naive detector with much higher data reduction. The proposed detector performed well, not very much worse than the ideal one. This indicates that the simplifying Naive Bayes assumption of Section 4.6.2 did not severely reduce performance. In addition, our detector was robust towards multipath fading.

Within the proposed model, several extensions are possible: It is of interest to conduct more extensive simulations based on multiple targets ($\bar{u} > 1$) and/or multiple frequency windows ($M > 1$). Moreover, the performance of detector under various real data scenarios (such as for instance described in [9]) should be investigated.

A number of model extensions are also possible. The assumed independence of $\varepsilon_1, \ldots, \varepsilon_N$ could be relaxed, implying that the covariances (24) include noise terms as well. We conjecture that this would not degrade the perfomance of the Naive Bayes detector a lot, since it assumes independence across sensors anyway. For target identification, we have concentrated on feature vectors (B.2) based on the periodogram, which is simple and well-established in signal processing. However, other feature vectors can easily be incorporated into the detector (to handle e.g. cyclostationarity), once the (asymptotic) distribution (21) is worked out. Simultaneous detection and parameter estimation of several targets could be achieved by enlarging the dimensionality of the parameter vector $\boldsymbol{\theta}$. In order to make such an approach computationally feasible one could for instance drop the velocity parameters, assuming each target to be fixed during a time window. We have ignored sending, reception and computational times in comparison to transmission times. If these assumptions fail, the transmission probabilities in Appendix E are affected, because of increased delays. We have also assumed that feature vectors are observed with perfect accuracy, ignoring the impact of quantization ([4]). Optimal fusion rules are derived in [15] and [23] for binary quantized data when complete or partial channel information is available to the sensors. Another extension is explicit modeling of lognormal channel variation due to multipath propagation ([16]). The simulations suggest that this could improve parameter estimation, whereas detection performance is marginally affected by channel variation. Finally, sequential detection is possible, using decisions and parameter estimates from several possibly overlapping time windows $\{\mathrm{TW}_k\}$, where parameter estimates from TW_{k-1} are used as input for the IFC at TW_k. This would allow for nonlinear target trajectories, using different linear approximations within different time windows. See [20] and [21] for Kalman filtering approaches when some information is lost e.g. due to inter-sensor communication.

APPENDICES

A Signal Delay

We will verify that

$$D_j = D_{j,t-\tau}(\boldsymbol{\psi}) = \frac{\boldsymbol{v}\cdot(\boldsymbol{z}_j - \boldsymbol{x}_t)}{v_{\mathrm{water}}^2 - \|\boldsymbol{v}\|^2} + \sqrt{\left(\frac{\boldsymbol{v}\cdot(\boldsymbol{z}_j - \boldsymbol{x}_t)}{v_{\mathrm{water}}^2 - \|\boldsymbol{v}\|^2}\right)^2 + \frac{\delta_{j,t-\tau}^2}{v_{\mathrm{water}}^2 - \|\boldsymbol{v}\|^2}}, \tag{A.1}$$

where $\cdot$ is the scalar product in $\mathbb{R}^3$, $\boldsymbol{v}$ the velocity vector of the target and v_{water} the sound speed in water. To this end, we solve $Dv_{\mathrm{water}} = \|\boldsymbol{x}_{t-D} - \boldsymbol{z}_j\|$ with respect to $D > 0$. Squaring both sides of this identity and rewriting,

$$\begin{array}{rcl} D^2 v_{\mathrm{water}}^2 & = & \|\boldsymbol{x}_{t-D} - \boldsymbol{z}_j\|^2 \\ & \Updownarrow & \\ D^2 v_{\mathrm{water}}^2 & = & \|(\boldsymbol{x}_t - \boldsymbol{z}_j) - D\boldsymbol{v}\|^2 \\ & \Updownarrow & \\ D^2 v_{\mathrm{water}}^2 & = & D^2\|\boldsymbol{v}\|^2 + 2D(\boldsymbol{z}_j - \boldsymbol{x}_t)\cdot\boldsymbol{v} + \delta_{j,t-\tau}^2, \end{array}$$

we get a quadratic equation in D whose only positive solution is (A.1).

B Definition of Feature Vectors

The feature vector (7) of the received signal at Sensor j and time block q is based on the energy spectrum of the received signal after FFT, although other choices are possible. In more detail, we start estimating $S_{Z_j}(\omega) = S_{A_jX}(\omega) + S_{\varepsilon_j}(\omega)$ from time block q by $(2\pi)^{-1}$ times the periodogram ([3]) at the Fourier frequency $\omega_u = 2\pi u/r$ closest to ω, where $u \in \{-[(r-1)/2], \ldots, [r/2]\}$ is an integer. That is,

$$\hat{S}_{Z_j,q}(\omega) = \frac{1}{2\pi r}\left|\sum_{t\in\mathrm{TB}_q} Z_{jt}e^{-it\omega_u}\right|^2, \tag{B.1}$$

for all $\omega \in (\omega_u - \pi/r, \omega_u + \pi/r]$. Then define

$$Y_{jqm} = 2\int_{\Omega_m} \hat{S}_{Z_j,q}(\omega)d\omega = \frac{4\pi}{r} \sum_{u;\omega_u \in \Omega_m} \hat{S}_{Z_j,q}(\omega_u). \tag{B.2}$$

In the last equality we assumed for simplicity that the boundary points of each Ω_m are located halfway between two neighboring Fourier frequencies.

C Distribution of Feature Vectors

We will concentrate on (23)-(24), since (22) is much easier to establish and asymptotic normality (21) can be deduced from the Central Limit Theorem as $r \to \infty$. The main technical result is:

Proposition 1 *Assume that $\{W_t; t \in \mathbb{Z}\}$ is a stationary Gaussian process with $E(W_t) = 0$ and covariance function C_w satisfying*

$$\sum_{n \in \mathbb{Z}} |n||C_W(n)| = C_1 < \infty. \tag{C.1}$$

Let Ω^1 and Ω^2 be two interval subsets of $[0, \pi]$ and put $Y_j = 2\int_{\Omega^j} \hat{S}_{W_j}(\omega)d\omega$, $j = 1, 2$, where $\hat{S}_j(\omega) = \left|\sum_{t=1}^r W_{t+rd_j} e^{-it\omega_u}\right|^2 / (2\pi r)$ is an estimated spectral frequency at ω_u, the Fourier frequency closest to ω and d_1, d_2 are fixed numbers. Then

$$Cov(Y_1, Y_2) = \frac{8\pi}{r}\rho(d) \int_{\Omega^1 \cap \Omega^2} S_W^2(\omega)d\omega + o(r^{-1}), \tag{C.2}$$

as $r \to \infty$, where $d = d_2 - d_1$ and $\rho(\cdot)$ is defined in (25).

In the proof of Proposition 1 we will use the following result:

Lemma 1 *Assume that $\boldsymbol{W} = (W_1, W_2, W_3, W_4) \sim N(\boldsymbol{0}, \boldsymbol{C})$, where $\boldsymbol{C} = (C_{ij})_{i,j=1}^4$. Then*

$$Cov(W_1W_2, W_3W_4) = C_{13}C_{24} + C_{14}C_{23}. \tag{C.3}$$

Proof. Let $\boldsymbol{A} = (A_{ij})$ any square root of $\boldsymbol{C}$, i.e. $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T$. Write $\boldsymbol{W} = \boldsymbol{A}\boldsymbol{X}^T$, where $\boldsymbol{X} = (X_1, X_2, X_3, X_4) \sim N(\boldsymbol{0}, \boldsymbol{I}_4)$ has a four-dimensional standard normal distribution. Then

$$\begin{aligned} E(W_1W_2W_3W_4) &= E\left(\prod_{i=1}^4 \sum_{j=1}^4 A_{ij}X_j\right) \\ &= \sum_{j_1,\ldots,j_4=1}^4 A_{1j_1}A_{2j_2}A_{3j_3}A_{4j_4}E(X_{j_1}X_{j_2}X_{j_3}X_{j_4}) \\ &= C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23}, \end{aligned}$$

where the last step follows after some computations using that $E(X_{j_1}X_{j_2}X_{j_3}X_{j_4})$ equals 3 if $j_1 = j_2 = j_3 = j_4$, 1 if $j_1 = j_2 \neq j_3 = j_4$, $j_1 = j_3 \neq j_2 = j_4$ or $j_1 = j_4 \neq j_2 = j_3$ and 0 otherwise. Finally, the lemma follows from

$$\mathrm{Cov}(W_1W_2, W_3W_3) = E(W_1W_2W_3W_4) - C_{12}C_{34}.$$

□

Proof of Proposition 1. Without loss of generality we put $d_1 = 0$ and $d_2 = d$. Notice that

$$\begin{aligned}
&\mathrm{Cov}(\hat{S}_1(\omega_u), \hat{S}_2(\omega_v)) = 1/(4\pi r^2) \textstyle\sum_{t_1,\dots,t_4=1} \mathrm{Cov}\left(W_{t_1}W_{t_2}, W_{t_3}W_{t_4}\right) e^{i(t_1-t_2)\omega_u + i(t_3-t_4)\omega_v} \\
&\quad = 1/(4\pi r^2) \textstyle\sum_{t_1,\dots,t_4=1} C_W(t_4 - t_1)e^{i(t_1\omega_u - t_4\omega_v)} C_W(t_3 - t_2)e^{i(t_3\omega_v - t_2\omega_u)} \\
&\quad +1/(4\pi r^2) \textstyle\sum_{t_1,\dots,t_4=1} C_W(t_3 - t_1)e^{i(t_1\omega_u + t_3\omega_v)} C_W(t_4 - t_2)e^{-i(t_2\omega_u + t_4\omega_v)} \\
&\quad = |F_{uv}|^2 + |G_{uv}|^2,
\end{aligned} \tag{C.4}$$

where $1 \leq t_1, t_2 \leq r$, $rd + 1 \leq t_3, t_4 \leq rd + r$ and in the second equality we used Lemma 1. In the last equality

$$F_{uv} = \frac{1}{2\pi r} \sum_{t_1,t_4} C_W(t_4 - t_1)e^{i(t_1\omega_u - t_4\omega_v)}$$

and

$$G_{uv} = \frac{1}{2\pi r} \sum_{t_1,t_3} C_W(t_3 - t_1)e^{i(t_1\omega_u + t_3\omega_v)},$$

where $1 \leq t_1 \leq r$, $rd+1 \leq t_3, t_4 \leq rd+r$ and $|z| = \sqrt{x^2 + y^2}$ is the modulus of the complex number $z = x+iy$. Since S_W is piecewise constant on intervals $(\omega_u - \pi/r, \omega_u + \pi/r)$, we find that

$$\begin{aligned}
\mathrm{Cov}(Y_1, Y_2) &= (4\pi/r)^2 \textstyle\sum_{u,v} \mathrm{Cov}(\hat{S}_1(\omega_u), \hat{S}_2(\omega_v)) \\
&=: \textstyle\sum_a I(a) + II,
\end{aligned} \tag{C.5}$$

where the double sum is over all (u, v) such that $\omega_u \in \Omega^1$, $\omega_v \in \Omega^2$ and

$$\begin{aligned}
I(a) &= (16\pi^2/r^2) \textstyle\sum_{u,v;v-u=a} |F_{uv}|^2, \\
II &= (16\pi^2/r^2) \textstyle\sum_{u,v} |G_{uv}|^2.
\end{aligned}$$

In order to analyze $I(0)$, we notice that

$$\begin{aligned}
F_{uu} &= (1/(2\pi r)) \textstyle\sum_{t_1,t_4} C_W(t_4 - t_1)e^{-i\omega_u(t_4-t_1)} \\
&= (1/(2\pi)) \textstyle\sum_{n=-r}^{r} (1 - |n|/r) C_W(rd + n)e^{-i(rd+n)\omega_u} \\
&= (1 - |d|)_+ S_W(\omega_u) + O(r^{-1}),
\end{aligned}$$

where $1 \leq t_1 \leq r$, $1+rd \leq t_4 \leq r+rd$, $t_4 - t_1 = rd+n$ and in the last step we used the definition (2) of spectral density. Because of (C.1), the remainder term is $O(r^{-1})$ uniformly in u. Summing over all u such that $\omega_u \in \Omega^1 \cap \Omega^2$ we get

$$\begin{aligned} I(0) &= \tfrac{16\pi^2(1-|d|)_+^2}{r^2}\left(\sum_u |S_W(\omega_u)|^2 + O(1)\right) \\ &= \tfrac{8\pi}{r}\cdot(1-|d|)_+^2 \cdot \int_{\Omega^1\cap\Omega^2} S_W(\omega)^2 d\omega + o(r^{-1}). \end{aligned} \tag{C.6}$$

In order to analyze $I(a)$ for $a \neq 0$, we notice that

$$F_{uv} = \frac{1}{2\pi r}\sum_{n=-r+1}^{r-1} f_{uv}(n) C_W(rd+n), \tag{C.7}$$

when $u \neq v$, where

$$\begin{aligned} &f_{uv}(n) = \sum_{t_1,t_4;t_4-t_1=rd+n} e^{i(t_1\omega_u - t_4\omega_v)} \\ &\quad = e^{i(\omega_u-\omega_v)} e^{-i\omega_v rd}(e^{i(\omega_u-\omega_v)}-1)^{-1}(-1)^{\{n<0\}}(e^{-in\omega_u} - e^{-in\omega_v}). \end{aligned} \tag{C.8}$$

Inserting (C.8) into (C.7), we get

$$\begin{aligned} &|F_{uv}|^2 = (2\pi r|e^{i(\omega_u-\omega_v)}-1|)^{-2}\left|\sum_{n=-r+1}^{r-1}(-1)^{\{n<0\}}(e^{-ni\omega_u} - e^{-ni\omega_v})C_W(rd+n)\right|^2 \\ &\quad = 1_{\{|d|\le 1\}}(r|e^{i(\omega_u-\omega_v)}-1|)^{-2}\left(|e^{rdi\omega_v}S_W(\omega_v) - e^{rdi\omega_u}S_W(\omega_u)| + O(r^{-1})\right)^2 \\ &\quad = 1_{\{|d|\le 1\}}(r|e^{i(\omega_u-\omega_v)}-1|)^{-2}|e^{rdi\omega_v} - e^{rdi\omega_u}|^2 S_W^2(\omega_u)(1+O(|u-v|/r)), \end{aligned} \tag{C.9}$$

where in the last step we used $|S(\omega_u)-S(\omega_v)| \le C_1|\omega_u-\omega_v|$, with C_1 the same constant as in (C.1). When summing over all u, v such that $v-u=a$, it turns out that the remainder term on the right-hand side of (C.9) is asymptotically negligible as $r \to \infty$. Since $\omega_v - \omega_u = \omega_a$ and $|e^{-i\omega_a}-1|| \sim 2\pi a/r$ as $r\to\infty$, we thus find that

$$\begin{aligned} I(a) &\approx 1_{\{|d|\le 1\}} 16\pi^2 |e^{rdi\omega_a}-1|^2/(r^4|e^{-i\omega_a}-1||^2)\cdot \sum_u S_W^2(\omega_u) \\ &= 1_{\{|d|\le 1\}} 16\sin^2(da\pi)/(ar)^2 \cdot \left(\sum_u S_W^2(\omega_u) + o(r)\right) \\ &= (8\pi/r)\cdot 1_{\{|d|\le 1\}}(\sin(da\pi)/(\pi a))^2 \cdot \int_{\Omega^1\cap\Omega^2} S_W^2(\omega)d\omega + o(r^{-1}) \end{aligned} \tag{C.10}$$

when $a \neq 0$, summing over all u such that $\omega_u \in \Omega^1$ and $\omega_{u+a} \in \Omega^2$. In order to show that II is asymptotically negligible as $r \to \infty$, we notice that

$$G_{uv} = \frac{1}{2\pi r}\sum_{n=-r+1}^{r-1} g_{uv}(n) C_W(rd+n), \tag{C.11}$$

where

$$\begin{aligned} g_{uv}(n) &= \sum_{t_1,t_3;t_3-t_1=rd+n} e^{i(t_1\omega_u + t_3\omega_v)} \\ &= e^{i(\omega_u+\omega_v)} e^{i\omega_v rd}(e^{i(\omega_u+\omega_v)}-1)^{-1}(-1)^{\{n<0\}}(e^{-in\omega_u} - e^{-in\omega_v}), \end{aligned} \tag{C.12}$$

and the second equality is valid when $0 < u+v < r/2$. Inserting (C.12) into (C.11) we get

$$\begin{array}{rcl} |G_{uv}|^2 & = & (2\pi r)^{-2}|e^{i(\omega_u+\omega_v)}-1|^{-2}\cdot\left|\sum_{n=-r+1}^{r}(-1)^n(e^{-in\omega_u}-e^{-in\omega_v})C_W(rd+n)\right|^2 \\ & \leq & (2\pi r)^{-2}\cdot(r/(4a))^2\cdot(2C_0)^2 \\ & = & C_0^2/(16\pi^2 a^2). \end{array} \tag{C.13}$$

for $0 < u+v < r/2$, with $a = \min(u+v, r-(u+v))$ and $C_0 = \sum_n |C_W(n)| < \infty$ according to (C.1). In the inequality of (C.13) we utilized $|e^{ix}-1| \geq 2|x|/\pi$ for $0 \leq x \leq \pi$. Let $II(a) = (16\pi^2/r^2)\sum_{u,v}|G_{uv}|^2$, the sum taken over all u, v such that $\min(u+v, r-(u+v)) = a$. It is easy to see that $|G_{00}|^2 = |G_{r/2,r/2}|^2 = C_0^2/(4\pi^2)$ and hence $II(0) = 8C_0^2/r^2$. Combining this with (C.13) and the definition of II, we get

$$\begin{array}{rcl} II & = & \sum_{a=0}^{r/2} II(a) \\ & \leq & 8C_0^2/r^2 + (C_0^2/r^2)\sum_{a=1}^{r/2} n(a)/a^2 \\ & \leq & 8C_0^2/r^2 + (2C_0^2/r^2)\log(r/2+1) = o(r^{-1}). \end{array} \tag{C.14}$$

where $n(a) = |\{(u,v);\, \omega_u \in \Omega^1, \omega_v \in \Omega^2, \min(u+v, r-(u+v)) = a\}| \leq 2a$. Combining (C.5), (C.6), (C.10) and (C.14), we find that (C.2) holds, with

$$\rho(d) = (1-|d|)_+^2 + 1_{\{|d|\leq 1\}}\sum_{a\neq 0}\frac{\sin^2(d\pi a)}{\pi^2 a^2} = (1-|d|)_+,$$

where the last equality follows from $\sum_{a>0}\sin^2(\pi da)/(\pi a)^2 = d(1-d)/2$ for $0 \leq d \leq 1$. This is a consequence of Parseval's identity, applied to the Fourier expansion of $f(x) = 1_{\{|x|\leq d\}}$, $-1 \leq x \leq 1$. This completes the proof. □

Corollary 1 *Assume that the covariance functions of the target signal and noise processes satisfy*

$$\begin{array}{rcl} \sum_n |n||C_X(n)| & < & \infty, \\ \sum_n |n||C_{\varepsilon_i}(n)| & < & \infty. \end{array} \tag{C.15}$$

Then, up to remainder terms of order $o(r^{-1})$ as $r \to \infty$, $V_{jqm,j'q'm'} = Cov(Y_{jqm}, Y_{j'q'm'})$ is given by (26) when $j \neq j'$ and by (25) when $j = j'$ and $q = q'$. If also

$$d_{jq} - d_{j,q+1} = o(1) \text{ as } r \to \infty, \tag{C.16}$$

then (25) holds for $q \neq q'$ as well.

Proof. Recall the definition of Y_{jqm} in (B.1)-(B.2). We apply Proposition 1 in order to find $\mathrm{Cov}(Y_{jqm}, Y_{j'q'm'})$, with $\Omega^1 = \Omega_m$ and $\Omega^2 = \Omega_{m'}$. An important observation, used throughout the proof, is that the process observed by Sensor j within TB_q is

$$Z_{j,t+r(q-1)} = a_{jq} X_{t+r(q-1)-rd_{jq}} + \varepsilon_{j,t+r(q-1)}$$

for $t = 1, \ldots, r$. This is a consequence of the simplifying assumptions (19) and (20).

We treat three cases separately. First, when $(j, q) = (j', q')$, we let $Y_1 = Y_{jqm}$, $Y_2 = Y_{j'q'm'}$, $W_t = Z_{j,t+r(q-1)}$ and $d_1 = d_2 = 0$. Independence of X and ε_j implies $S_W = a_{jq} S_X + S_{\varepsilon_i}$ and $\rho(d_2 - d_1) = \rho(0) = 1$. Hence (25) follows from (C.2), noticing that $\Omega^1 \cap \Omega^2$ equals Ω_m if $m = m'$ and $\emptyset$ if $m \neq m'$.

Secondly, when $j \neq j'$, we put $W_t = X_t$, $d_1 = q-1-d_{jq}$ and $d_2 = q'-1-d_{j'q'}$. Then $Z_{j,r(q-1)+t} = a_{jq} W_{t+rd_1} + \varepsilon_{j,t+r(q-1)}$ and $Z_{j',r(q'-1)+t} = a_{j'q'} W_{t+rd_2} + \varepsilon_{j',t+r(q'-1)}$. Independence of the ε_j and $\varepsilon_{j'}$-processes implies that

$$\mathrm{Cov}(Y_{jqm}, Y_{j'q'm'}) = a_{jq}^2 a_{j'q'}^2 \mathrm{Cov}(Y_1, Y_2).$$

Hence (26) follows from (C.2), since $S_W = S_X$.

Finally, when $j = j'$ and $q \neq q'$, $Z_{j,r(q-1)+t}$ and $Z_{j,r(q'-1)+t}$ for $t = 1, \ldots, r$ involve time indexes of the X- and ε_j-processes with difference of magnitude at least $|q' - q|r \geq r$ between Sensors j and j'. For the X-process, this is a consequence of (C.16). Similarly as in the proof of Proposition 1, one then proves that $\mathrm{Cov}(Y_{jqm}, Y_{j'q'm'}) = o(r^{-1})$. □

D Laplace Type Approximation of Integrand in (30)

Our goal is to find an approximation of integral $\int_\Theta \Lambda_u(\boldsymbol{\theta}) d\boldsymbol{\theta}$ of dimension $M' = 6 + M$ appearing in the Bayesian detector (30). We use a quadratic Taylor expansion of $\log(\Lambda_u)$ and get

$$\Lambda_u(\boldsymbol{\theta}) \approx \Lambda_u(\tilde{\boldsymbol{\theta}}_u) \exp\left(-0.5(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u) \boldsymbol{J}_u (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u)'\right),$$

with $\tilde{\boldsymbol{\theta}}_u = \arg\max_{\boldsymbol{\theta} \in \mathbb{R}^p} \Lambda_u(\boldsymbol{\theta})$ and the $M' \times M'$ matrix $\boldsymbol{J}_u = -(\log \Lambda_u)''(\tilde{\boldsymbol{\theta}}_u)$ gives the curvature of $-\log \Lambda_u$ at $\tilde{\boldsymbol{\theta}}_u$. Introducing the multivariate normal random variable $\boldsymbol{\vartheta} = (\vartheta_a)_{a=1}^{M'} \sim N(\tilde{\boldsymbol{\theta}}_u, \boldsymbol{J}_u^{-1})$, we get the Laplace type approximation

$$\begin{aligned}
\int_\Theta \Lambda_u(\boldsymbol{\theta}) d\boldsymbol{\theta} &\approx \Lambda_u(\tilde{\boldsymbol{\theta}}_u) \int_\Theta \exp\left(-0.5(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u) \boldsymbol{J}_u (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u)'\right) d\boldsymbol{\theta} \\
&= (2\pi)^{M'/2} |\boldsymbol{J}_u|^{-1/2} \Lambda_u(\tilde{\boldsymbol{\theta}}_u) \int_\Theta f_{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= (2\pi)^{M'/2} |\boldsymbol{J}_u|^{-1/2} \Lambda_u(\tilde{\boldsymbol{\theta}}_u) P(\boldsymbol{\vartheta} \in \Theta),
\end{aligned}$$

where $f_{\boldsymbol{\vartheta}}$ is the density of $\boldsymbol{\vartheta}$ and

$$P(\boldsymbol{\vartheta} \in \Theta) = P(\vartheta_a > 0,\, a = 7, \ldots, M'). \tag{D.1}$$

An analytical expression

$$P(\boldsymbol{\vartheta} \in \Theta) = \Phi\left(\frac{\tilde{\boldsymbol{\theta}}_u \cdot \boldsymbol{u}}{|\boldsymbol{u}\boldsymbol{J}_u^{-1/2}|}\right)$$

for (D.1) is possible when $M = 1$, with $\boldsymbol{u} = (0, \ldots, 0, 1)$ the unit vector in $\mathbb{R}^7$ with last entry one. When $M > 1$ we use a Monte Carlo approximation

$$\widehat{P(\boldsymbol{\vartheta} \in \Theta)} = \frac{1}{B}\sum_{b=1}^{B} 1_{\{\boldsymbol{\vartheta}^{(b)} \in \Theta\}}$$

of (D.1), with $\{\boldsymbol{\vartheta}^{(b)}\}_{b=1}^{B}$ an i.i.d. sample from $N(\tilde{\boldsymbol{\theta}}_u, \boldsymbol{J}_u^{-1})$ of size B.

E Reception Probabilities

It will be convenient to introduce $\boldsymbol{I}_{qn} = (I_{qnji})_{i,j=1}^{N}$ and $\boldsymbol{P}_{n-q} = (P_{n-q,ji})_{j,i=1}^{N}$. Then (34) may be written in matrix form as

$$E(\boldsymbol{I}_{qn}) = \boldsymbol{P}_{n-q}. \tag{E.1}$$

Below, we describe an algorithm of simulating $\{\boldsymbol{I}_{qn}\}_{n=q}^{n+\bar{w}-1}$ recursively with respect to n, where $\bar{w} = w_1 + w_3$, as in routing rule F4. Repeating this B times we get sequences $\{\boldsymbol{I}_{qn}^{(b)}\}_{n=q}^{n+\bar{w}-1}$ for $b = 1, \ldots, B$ and Monte Carlo estimates

$$\hat{\boldsymbol{P}}_{n-q} = \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{I}_{qn}^{(b)}.$$

Introduce the reception indicator matrices $\boldsymbol{I}_q = (I_{q,ji})_{j,i=1}^{N}$, where

$$I_{q,ji} = \begin{cases} 0, & j = i, \\ 1_{\{\text{SM}_{jq} \text{ received (without routing) by } i \text{ at some time} > qr | \text{SM}_{jq} \text{ sent}\}}, & j \neq i. \end{cases}$$

Notice that all $I_{q,ji}$ are independent, with

$$P(I_{q,ji} = 1) = \begin{cases} 0, & j = i, \\ P_{ji}, & j \neq i. \end{cases} \tag{E.2}$$

We also need the $N \times N$ 'delay' matrices $\boldsymbol{D}_v = (D_{v,ji})$, for $v = 0, 1, 2, \ldots$, where $D_{v,ji}$ is one if it takes between between $v - 1$ and v time blocks for a message sent from j to arrive at i without routing, and zero otherwise. We recall from the introduction that only the finite sound propagation in water is assumed do cause delays. Hence, the time delay for pair ji is $t_{ji} = \delta_{ji}/v_{\text{water}}$ time point units, and consequently

$$D_{v,ji} = 1_{\{(v-1)<t_{ji}\leq v\}}.$$

Notice in particular that $\boldsymbol{D}_0$ is the identity matrix of order N.

The algorithm for generating $\boldsymbol{I}_{qn}$ starts by first generating all matrices $\boldsymbol{I}_q, \dots, \boldsymbol{I}_{q+w-1}$ from (E.2). Then put $\boldsymbol{I}_{qq} = \boldsymbol{D}_0$ and recursively compute

$$\boldsymbol{I}_{qn} = \min\left(\mathbf{1}, \boldsymbol{I}_{q,n-1} + \boldsymbol{I}_q \cdot \boldsymbol{D}_{n-q} + \sum_{\nu=q+1}^{n-1} \left((\boldsymbol{I}_{q\nu} - \boldsymbol{I}_{q,\nu-1}) \cdot \boldsymbol{F}\right)(\boldsymbol{I}_\nu \cdot \boldsymbol{D}_{n-\nu})\right) \tag{E.3}$$

for $n = q+1, \dots, q+\bar{w}-1$, where $\mathbf{1}$ is an $N\times N$-matrix with 1 entries and $\boldsymbol{A}\cdot\boldsymbol{B}$ and $\min(\boldsymbol{A}, \boldsymbol{B})$ refer to elementwise multiplication and minimum of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The matrix $\boldsymbol{I}_q \cdot \boldsymbol{D}_{n-q}$ corresponds to messages transmitted at time qr that arrive without forwarding between time $(n-1)r$ and nr. The term $\left((\boldsymbol{I}_{q\nu} - \boldsymbol{I}_{q,\nu-1}) \cdot \boldsymbol{F}\right)(\boldsymbol{I}_\nu \cdot \boldsymbol{D}_{n-\nu})$ corresponds to messages sent at time qr that arrive for the first time between $(\nu-1)r$ and νr, are forwarded, and the forwarded messages arrive between time $(n-1)r$ and nr. When no routing occurs ($\boldsymbol{F} = 0$), (E.3) simplifies to the explicit expression

$$\boldsymbol{I}_{qn} = \boldsymbol{D}_0 + \boldsymbol{I}_q \cdot \sum_{\nu=1}^{n-q} \boldsymbol{D}_\nu,$$

so that

$$P_{n-q,ji} = P_{ji} 1_{\{t_{ji}\leq(n-q)r\}}.$$

ACKNOWLEDGEMENTS

Ola Hössjer's research was financially supported by the Swedish Defence Research Agency, the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine. The authors thank Ron Lennartsson, David Lindgren for valuable comments on the model.

References

[1] Appadwedula, S., Veeravalli, V.V. and Jones, D.L. (2008). Decentralized detection with censoring sensors. *IEEE Trans. Signal Processing*, **56**(4), 1362-1373.

[2] Blum, R.S., Kassam, S.A. and Poor, H.V. (1997). Distributed detection with multiple sensors: Part II Advanced topics. *Proc. IEEE* **85**(1), 64-79.

[3] Brockwell, P.J. and Davis, R.A. (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York.

[4] Chamberland, J-F. and Veeravalli, V.V. (2003). Decentralized detection in sensor networks. *IEEE Trans. Signal Processing*, **51**(2), 407-416.

[5] Chamberland, J-F. and Veeravalli, V.V. (2007). Wireless sensors in distributed detection applications. *IEEE Signal Proc. Mag.*, May, 16-25.

[6] Chen, B., Tong, L. and Varshney, P.K. (2006). Channel-aware distributed detection in wireless sensor networks. *IEEE Sign. Proc. Mag.*, 16-26, July 2006.

[7] Choi, S., Berger, R., Zhou, S. and Willett, P. (2008). Estimation of target trajectories based on distributed channel energy measurements, *11th International Conference on Information Fusion*, 1273-1278.

[8] Domingos, P. and Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* **29**, 103Ű137.

[9] Duarte, M.F. and Hu, Y.H. (2003). Vehicle classification in distributed sensor networks. Univ. of Wisconsin, Dept. of Electrical Engineering. `http://www.ece.wisc.edu/~hu/bio/pub/marco04a.pdf`.

[10] Fan, J. and Li, H. (2010). Optimal/Near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios. *IEEE Trans. Signal Processing*, **58**(8), 1362-1373.

[11] Grünewald, M. and Hössjer, O. (2012). A general statistical framework for multistage designs. To appear in *Scandinavian Journal of Statistics*.

[12] Grönkvist, J., Nilsson, J. and Sangfeldt, E. (2007). Wireless underwater sensor networks. Rep. FOI-R-2408-SE, FOI, Stockholm, 2007. `http://www.foi.se/FOI/templates/PublicationPage____171.aspx`.

[13] Hand, D.J. and Yu, K. (2001). Idiot's Bayes - not so stupid after all? *International Statistical Review*, **69**(3), 385-399.

[14] Hippenstiel, R.D. (2002). *Detection theory: Applications and digital signal processing.* CRC Press, Boca Raton, FL, USA.

[15] Jiang, R. and Chen, B. (2005). Fusion of censored decisions in wireless sensor networks. *IEEE Trans. Wireless. Comm.* **4**(6), 2668-2673.

[16] Kanchumarthy, V.R., Viswanathan, R. and Madishetty, M. (2008). Impact on channel errors on decentralized detection performance of wireless sensor networks.: A study of binary modulations, Rayleigh fading and nonfading channels, and fusion-combiners. *IEEE Trans. Signal Processing*, **56**(5), 1761-1769

[17] Lehmann, E.T. and Casella, G. (2001). *Theory of point estimation*, 2nd edition, corrected third printing, Springer, New York.

[18] Little, R.J.A. and Rubin, D.B. (2002). *Statistical analysis with missing data*, Wiley, Hoboken, New Jersey.

[19] Liu, L., Zhou, S. and Cui, J.H. (2008). Prospects and problems of wireless communication for underwater sensor networks. *Wireless Communications and Mobile Computing* **8**(8), 977-994.

[20] Liu, X. and Goldsmith, A. (2004). Kalman filtering with partial information losses. 43rd IEEE conference on decision and control, CDC, Volume **4**, 4180-4186.

[21] Msechu, E.J. (2011). Estimation with wireless sensor networks: Censoring and quantization perspectives. PhD Thesis, Dept. of Electrical and Computer Engineering, Univ. of Minnesota, Minneapolis, USA.

[22] Niu, R., Chen, B. and Varshney, P.K. (2006). Fusions of decisions transmitted over Rayleigh fading channels in wireless sensor networks. *IEEE Trans. Sign. Proc.* **54**(3), 1018-1027.

[23] Niu, R., Varshney, P.K. and Cheng, Q. (2006). Distributed detection in a large wireless sensor network. *Information Fusion* **7**, 380-394.

[24] Nyberg, T. (2008). Fully decentralized detection in an underwater setting. Stockholm Univ., Div. of Math. Statist., Master Thesis 2008:20.

[25] Patwari, N. and Hero, A.O. (2003). Hierarchical censoring for distributed detection in wireless sensor networks, in *Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03)*, **4**, 848-851. April 2003.

[26] Rago, C., Willett, P. and Bar-Shalom, Y. (1996). Censoring sensors: A low-communication-rate scheme for distributed detection. *IEEE Trans. Areosp. Electron. Syst.* **32**(2), 554-568.

[27] Saligrama, V., Alanyali, M. and Savas, O. (2006). Distributed detection in sensor networks with packet losses and finite capacity links. *IEEE Trans. Sign. Proc* **54**(11), 4118-4132.

[28] Tenney, R.R. and Sandell, N.R. (1981). Detection with distributed sensors, *IEEE Trans. Aerosp. Electron. Syst.* **17**(4), 501-510.

[29] Tsitsiklis, J.N. (1988). Decentralized detection with a large number of detectors. *Math. Contr., Signals Syst.* **1**, 167-182.

[30] Tsitsiklis, J.N. (1993). Decentralized detection, in *Advances in Statistical Signal Processing, Signal Detection*, vol. 2, eds. Poor, H.V. and Thomas, J.B. Greenwich, CT: JAI.

[31] Van Trees, H.L. (1968). *Detection, estimation and modulation theory. Part I.* Wiley. Republished in 2001.

[32] Varshney, P.K. (1997). *Distributed detection with data fusion*, Springer.

[33] Viswanathan, R. and Varshney, P.K. (1997). Distributed detection with multiple sensors: Part I Fundamentals. *Proc. IEEE* **85**(1), 54-63

[34] Wymeersch, H., Lien, J. and Win, M.Z. (2009). Cooperative localization in wireless networks. *Proc. IEEE* **97**(2), 427-450.

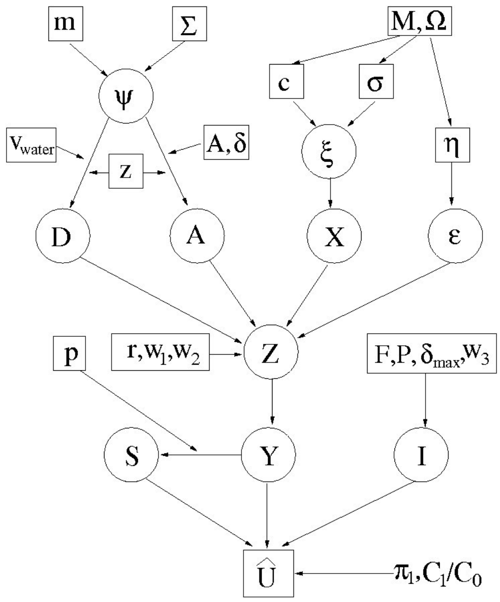

Figure 1: A Directed Acyclic Graph of the distributed detection problem during a time window TW. Random quantities (random parameters or hidden variables) are depicted as circles, and fixed quantities (hyperparameters or data known to the IFC) as squares. Arrows indicate probabilistic or deterministic relationships, $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ denote target trajectory and spectral parameters, $\boldsymbol{\eta}$ noise spectral parameters, $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ trajectory hyperparameters, $\boldsymbol{c} = \{c_m\}$, $\boldsymbol{\sigma} = \{\sigma_m\}$, M and $\Omega = \{\Omega_m\}$ spectral hyperparameters, $\boldsymbol{z} = \{z_j\}$ sensor locations, A, δ signal attenuation parameters, $D = \{D_{j,t-\tau}\}$, $A = \{A_{j,t-\tau}\}$ delays and attenuation of received signals, r time block length, w_1, w_2 time window length parameters, $F, P, \delta_{\max}$ data transmission parameters, w_3 detector delay, $X = \{X_t\}$ emitted target signal, $\boldsymbol{\varepsilon} = \{\varepsilon_{jt}\}$ received noise at sensors, $\boldsymbol{Z} = \{\boldsymbol{Z}_{jq}; 1 \leq j \leq N, q \in \mathrm{TW}\}$ received raw signals at sensors, $\boldsymbol{Y}$ compressed feature vector data for ideal detector, p proportion of sent feature vectors, $\boldsymbol{S}$ and $\boldsymbol{I}$ sent and successfully transferred feature vectors, $\hat{U}$ final decision on presence/type of target and $\{\pi_u C_u/(C_0\pi_0)\}_{u=1}^{\bar{u}}$ the false alarm thresholds (shown for $\bar{u} = 1$).

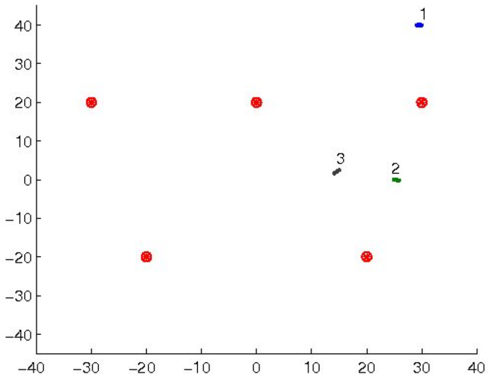

Figure 2: Simulation scenario: The red markers indicates the position of the sensor nodes (numbered from upper left to lower right) and the tracks of 3 different targets during one measurement window are indicated with the number 1-3. The scale is in meters.

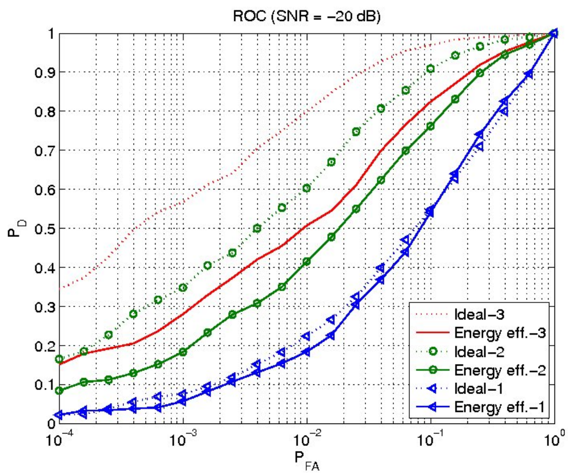

Figure 3: ROC-curves for all three scenarios, with -20 dB receiver SNR.

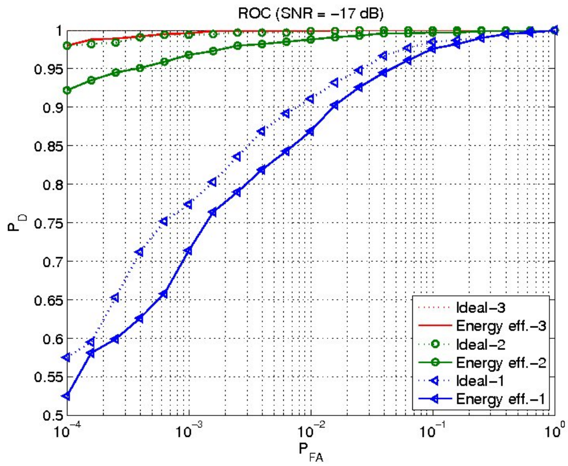

Figure 4: ROC-curves for all three scenarios, with -17 dB receiver SNR.

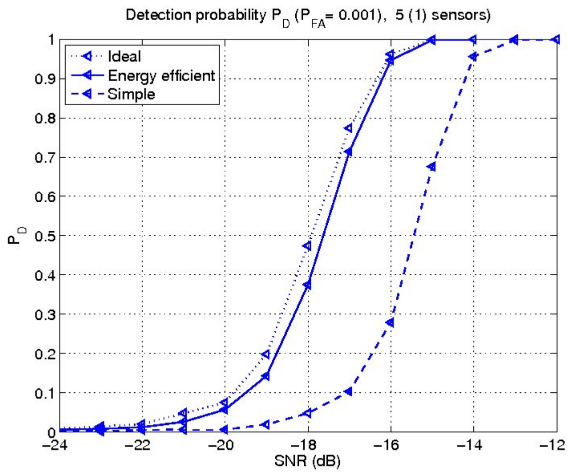

Figure 5: Detection performance as function of receiver SNR with $P_{\mathrm{FA}} = 0.001$ for Scenario 1.

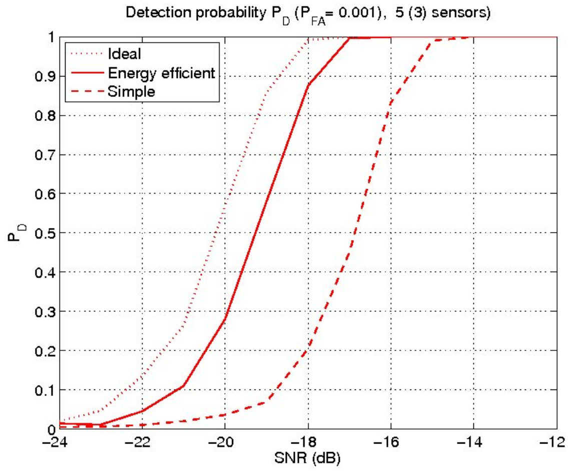

Figure 6: Detection performance as function of receiver SNR with $P_{\mathrm{FA}} = 0.001$ for Scenario 3.

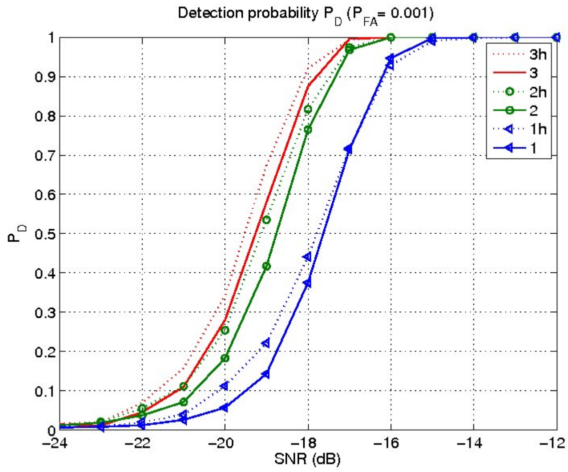

Figure 7: Detection performance as function of receiver SNR with $P_{\mathrm{FA}} = 0.001$, for all three different target scenarios. The curves labeled 1h, 2h, and 3h show the performance for a scenario where the channel exhibits a log-normal variation with 2 dB standard deviation, and the number indicates which scenario it is.

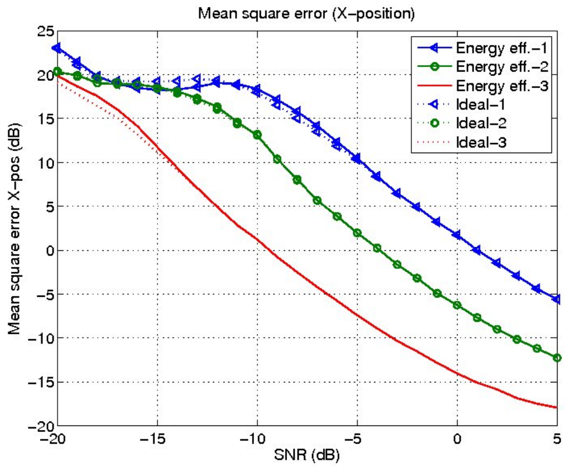

Figure 8: Mean square error of estimated position in the x-direction (unit 1 m^2) as function of receiver SNR for all three target scenarios. Here the ideal detector is compared with the energy efficient (that uses data censoring).

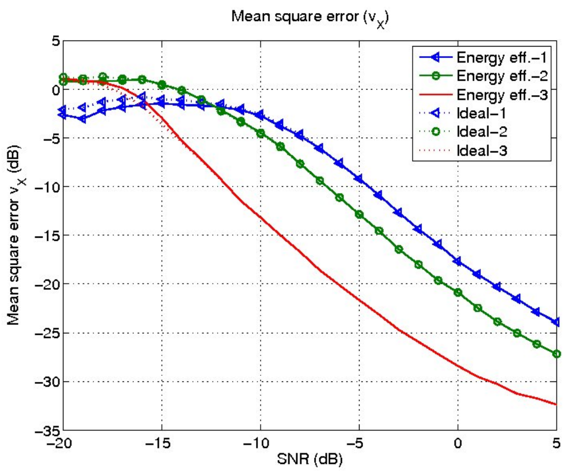

Figure 9: Mean square error of estimated velocity in the x-direction (unit 1 $(m/s)^2$) as function of receiver SNR for all three target scenarios. Here the ideal detector is compared with the energy efficient (that uses data censoring).

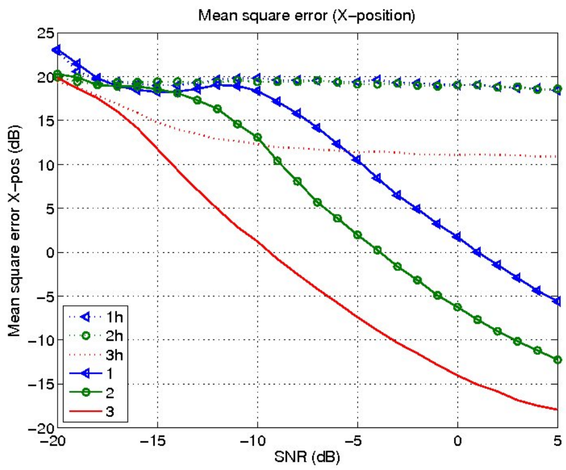

Figure 10: Mean square error of estimated position in the x-direction (unit 1 m^2) as function of receiver SNR. The curves labeled 1h, 2h, and 3h show the performance for a scenario where the channel exhibits a log-normal variation with 2 dB standard deviation, and the number indicates which scenario it represents.

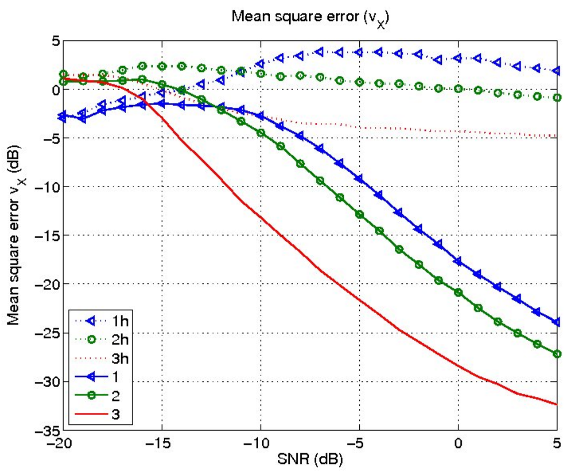

Figure 11: Mean square error of estimated velocity in the x-direction (unit 1 $(m/s)^2$) as function of receiver SNR. The curves labeled 1h, 2h, and 3h exhibit channel fading (see Figure 10).

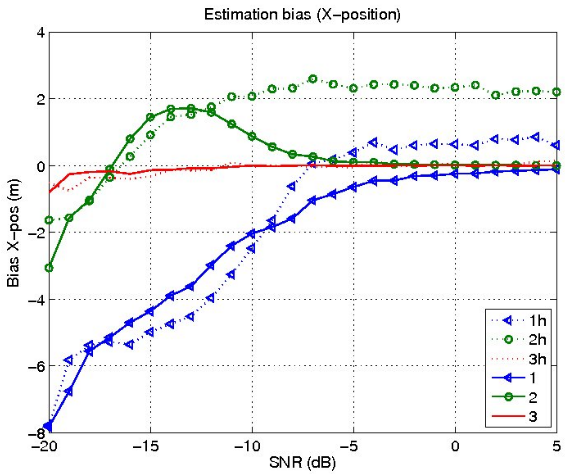

Figure 12: Bias of estimated position in the x-direction as function of receiver SNR. The curves labeled 1h, 2h, and 3h exhibit channel fading (see Figure 10).

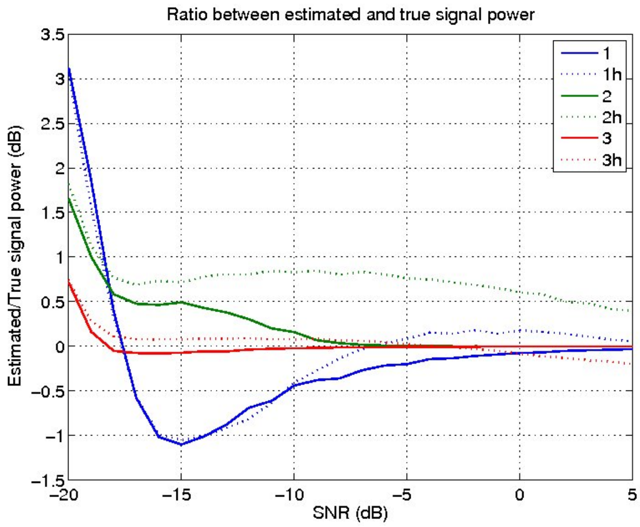

Figure 13: Ratio between estimated and true signal power as function of receiver SNR. The curves labeled 1h, 2h, and 3h exhibit channel fading (see Figure 10).