Issues in Data Fusion

Peter Willett
University of Connecticut
peter.willett@uconn.edu

I. Scan Statistics for Data Fusion

A. Background

In [22] we are interested in a submarine detection with a barrier sensor network [26], which is composed of an acoustic source and a band of passive sensors as shown in Fig. 1. Such a setup is useful for region denial and coastline monitoring. Suppose that each sensor makes a binary decision—'0' or '1'—via a simple comparison of its matched filter output with a given threshold, and then a fusion center (FC) collects the local results to make a system-level inference. Here, we focus on how to fuse these distributed decisions.

Suppose that the submarine is noncooperative and that the sensor network does not have knowledge of its location and acoustic reflection model. As a result, the probabilities of local detection are unavailable, and the optimal fusion rule [11] is irrelevant. A suboptimal alternative is the counting rule test (CRT) [19], which counts the total number of 1's and compares it to a given threshold to infer whether a submarine exists. The CRT is a *global* data utilization approach, and it equally weights all the local decisions without considering their spatial distribution. However, since most reflection energy from a submarine is focused in a narrow conical area [24], only a subset of sensors can produce reliable local detections, while others are generally dominated by background noise, particularly at low signal-noise-ratio (SNR) [26]. If one could find that *significant* area and utilize only it to make a final decision, the system-level performance would be improved. An emerging *local* detection fusion approach is the *scan statistic* [15]. It slides a window around the sensor field, and picks up the most significant compact area to make a system-level detection. Here, the significance is defined by the number of 1's falling into the window. The key feature of a scan statistic is that its threshold can be set accurately to produce any desired system-level false alarm rate. Here we will show that the scan statistic can outperform the CRT if the size of the sliding window is properly selected.

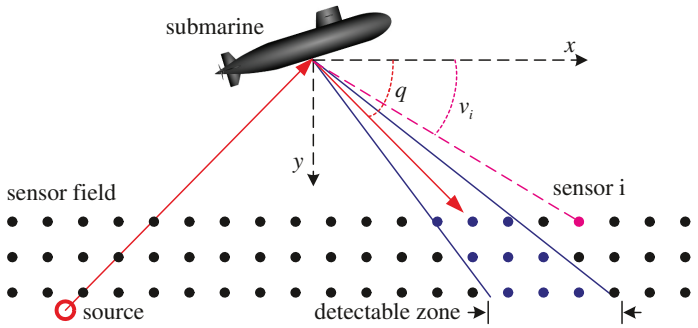


Fig. 1. A barrier sensor network is composed of an acoustic source and a narrow passive sensor band, and it could be used to detect a submarine out of the sensor band. Since the acoustic reflection of a submarine is angle dependent, only those located in a particular zone are able to reliably detect the echoes if the SNR is low.

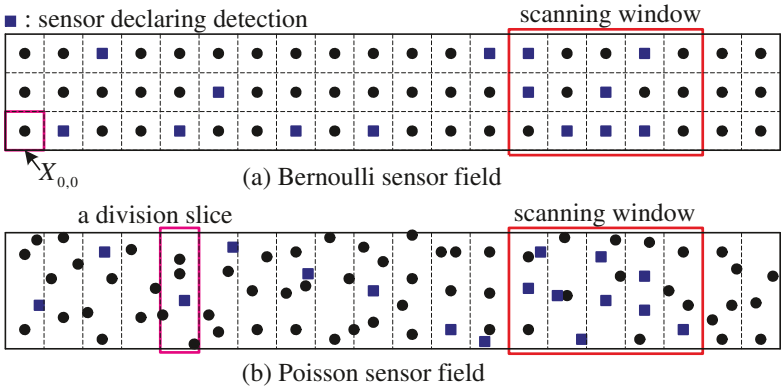


Fig. 2. An illustration of a scan statistics for different sensor fields. If no target appears in the surveillance area, the sensors declaring detection follow a uniform distribution. Otherwise, certain *local detection clusters* may happen. A scan statistic test slides a rectangle window across the sensor filed to check whether a '1' cluster exists.

B. Geometry Dependent Local Detection

Suppose sensor noises be independent and identically distributed (i.i.d.) zero-mean complex Gaussian, and suppose that the submarine has a Rayleigh fading model [24, pp.189]. If each sensor makes a local test based on the magnitude square of its matched filter output, the local false alarm rate and probability of detection for the ith sensor can be written as [26]

$$p_{_f} = \exp(-\tau) \quad \text{and} \quad p_{_{d_i}} = \exp(-\tau/(1+\beta_i)), \tag{1}$$

where τ denotes the universal threshold, and β_i relates to the reflection geometry and transmission power.

We assume that the source and the sensors are at the same depth as the submarine. Let (x_s, y_s), (x_t, y_t), and (x_i, y_i) respectively denote the coordinates of source, submarine, and sensor i. The propagation distance for the path source-submarine-(the ith sensor) is $d_i = \sqrt{(x_s - x_t)^2 + (y_s - y_t)^2} + \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2}$, while $v_i = \arctan(\frac{x_i - x_t}{y_i - y_t})$ is a supplementary angle. Let the principle reflection angle be q; then β_i could be modeled as [26]

$$\beta_i = c_0 f_1(v_i, q)/d_i, \tag{2}$$

where c_0 is a power related constant, d_i is responsible for the propagation attenuation, and the Butterworth beampattern $f_1(v_i, q) = \left[1 + (\frac{v_i - q}{W})^{2K}\right]^{-1}$ models the aspect-dependent reflection of the submarine [26]. In $f_1(\cdot)$, $2W$ denotes the 3dB bandwidth, and K determines the contrast.

Let a sensor take the value of '1' if there is a detection; otherwise it takes the value of '0'. Since the submarine parameter set $\{x_t, y_t, q\}$ is unknown, $p_{_{d_i}}$'s are unavailable; therefore, the optimal Chair-Varshney rule [11] is irrelevant. A suboptimal approach is the CRT, which sums up all the local decisions. However, as a submarine may only affect a subarea of the sensor network due to the aspect-dependent reflection, a smarter local detection fusion approach—the scan statistic—deserves attention.

C. Scan statistic Based Detection Fusion

1) From Submarine Detection To Cluster Detection: A scan statistic was developed to seek clusters in a spatial or (and) time spanned data set [15]. Briefly speaking, it slides a window across the observation domain, and utilizes the most *significant* data subset to infer the hypothesis. The key is that despite the windows being slid (as opposed to "hopping" to adjacent but nonoverlapping partitions) its false alarm

rate is explicit. The shape of the window is usually problem dependent. It can be a square, rectangle, circle, or ellipse, and no uniformly best one exists.

Let the sensors be uniformly distributed in the barrier band. With the i.i.d. noise assumption, the sensors declaring '1' are uniformly dispersed under $\mathcal{H}_0$. If a target invades the surveillance area and if the mainlobe of its reflection pattern falls into the sensor field, a certain number of sensors within this zone will declare '1'. Therefore, a *local detection cluster* may form, and the target detection problem could be converted to a '1' cluster detection within the sensor field:

- $\mathcal{H}_0$: no significant '1' cluster exists;
- $\mathcal{H}_1$: a nonempty zone $\mathcal{Z}$ has a locally significant '1' cluster.

In general, the zone $\mathcal{Z}$ specified in $\mathcal{H}_1^s$ is unknown; some heuristic conjectures such as the size of $\mathcal{Z}$ are necessary.

2) Bernoulli Sensor Field: In this case, N senors are *regularly* deployed into $N = M_1 \times M_2$ uniformly divided cells, and each cell exactly contains one sensor as shown in Fig. 2(a). A shape match between the local detection cluster and scanning window would improve the test performance. Unfortunately, since the cluster varies with the submarine location and the skewness of its pattern mainlobe, a perfect footprint coincidence may not be guaranteed. Here, a suboptimal rectangle window is employed. As the width of the sensor band is thin, it will be included in the window for scanning convenience, see Fig. 2. Let X_{lk} denote the binary local decision of the sensor at a certain cell, and let the discrete length of the window be w, where $1 \leq w \leq M_1$, and then the total number of '1' within the window is

$$Y_m = \sum_{l=m}^{m+w-1} \sum_{k=1}^{M_2} X_{lk} = \sum_{l=m}^{m+w-1} Z_l, \tag{3}$$

where $1 \leq m \leq M_1 - w + 1$, and $Z_l \triangleq \sum_{k=1}^{M_2} X_{lk}$. The two dimensional summation for X_{lk}'s can be equivalently converted into a one-dimensional problem on Z_l; this is an important difference from [16]. The scan statistic $S(w, M_1)$ is defined as [15, p.273]

$$S(w, M_1) = \max \{Y_m; 1 \leq m \leq M_1 - w + 1\}, \tag{4}$$

and the FC makes a system-level decision based on

$$\begin{cases} S(w, M_1) < T_B, & \text{declare } \mathcal{H}_0 \\ S(w, M_1) \geq T_B, & \text{declare } \mathcal{H}_1, \end{cases} \tag{5}$$

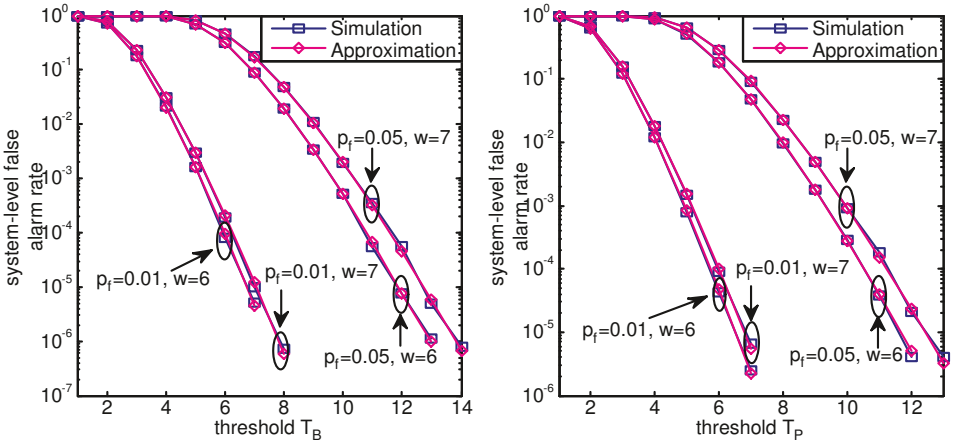
where T_B corresponds to a specified significance level.

Intuitively, a scan statistic divides a sensor field into several partially overlapped subareas, and then selects the most significant one to make a decision. If $w = M_1$, the scan statistic becomes a CRT.

3) Poisson Sensor Field: This subsection concentrates on a random barrier sensor network: *1)* the total number of activated sensors, say N, has a Poisson distribution; *2)* those activated sensors are uniformly distributed within the barrier band. Suppose that the initial distribution of activated sensors is a homogeneous Poisson process with density λ; we have $p(N) = (\lambda A)^N e^{-\lambda A}/N!$, where A denotes the area of the sensor band. Let the sensors be within $x_1 \leq x \leq x_2$ and $y_1 \leq x \leq y_2$, and then we get $A = (x_2 - x_1) \times (y_2 - y_1)$. The locations of sensors are i.i.d., and have a uniform pdf

$$f(x, y) = \begin{cases} \frac{1}{A} & x_1 \leq x \leq x_2 \text{ and } y_1 \leq x \leq y_2 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

With an i.i.d noise assumption, the number of sensors declaring '1' under hypothesis $\mathcal{H}_0$ also has a Poisson distribution with density $\lambda_f = \lambda p_f$, and they are uniformly located within the sensor field. Under $\mathcal{H}_1$, the distribution of sensors declaring '1' is no longer uniform, and a '1' cluster should form.



(a) Bernoulli (b) Poisson

Fig. 3. Comparisons of simulated and approximated system-level false alarm rates of scan statistic in Bernoulli and Poisson sensor fields. Each simulated curve is based on 1000000 runs.

We use a window to scan the sensor band to check whether a '1' cluster exists. Continuous scanning would be the best choice for a given window size; however, we prefer a discrete scanning scenario again for analysis simplicity: uniformly dividing the sensor band into M_1 contiguous narrow slices as shown in Fig. 2(b). Let $\bar{X}_i$ stand for the number of sensors declaring '1' in the ith slice, and let the window cover exactly w slices; hence, the total number of '1' within in the window is $\bar{Y}_m = \sum_{l=m}^{m+w-1} \bar{X}_l$, where $1 \le m \le M_1 - w + 1$. Here the two-dimensional data summation can also be reduced to a scalar case. The scan statistic $S(w, M_1)$ is defined as

$$S(w, M_1) = \max \{\bar{Y}_m; 1 \le m \le M_1 - w + 1\}, \tag{7}$$

and the FC makes the system-level decision based on

$$\begin{cases} S(w, M_1) < T_P, & \text{declare } \mathcal{H}_0 \\ S(w, M_1) \ge T_P, & \text{declare } \mathcal{H}_1, \end{cases} \tag{8}$$

where T_P corresponds to a given significance level again.

D. Performance Analysis for Scan Statistic

1) Bernoulli Sensor Field: The system-level performance including the false alarm rate P_f^s and the probability of detection P_d^s will be discussed for scanning a Bernoulli sensor network. With the i.i.d. sensor noise assumption, the X_{lk}'s are i.i.d. Bernoulli random variables with parameter p_f in the absence of a target. Therefore, the Z_l's are independent and subject to a binomial distribution

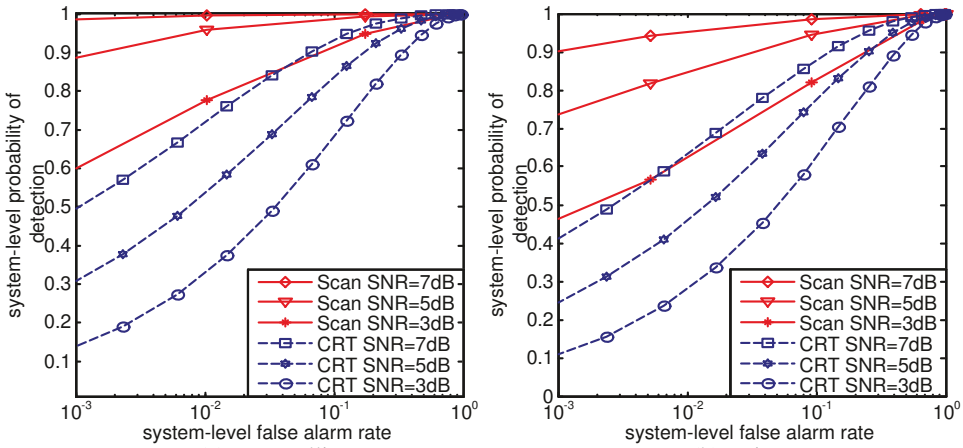
$$\Pr\{Z_l = i\} = \binom{M_2}{i} p_f^i (1 - p_f)^{M_2 - i} \tag{9}$$

under $\mathcal{H}_0$. The system-level false alarm rate is defined as

$$P_f^s = \Pr\{S(w, M_1) \ge T_B | \mathcal{H}_0\}, \tag{10}$$

and it is a function of $\Pr\{Z_l = i\}$ and window size w. The analysis of P_f^s is intricate, and a closed-form expression is unavailable. In [14], an accurate approximation is given as

$$P_f^s \simeq 1 - G_{T_B,w}(2w) \left[\frac{G_{T_B,w}(2w)}{G_{T_B,w}(2w-1)} \right]^{M_1 - 2w}, \tag{11}$$



(a) Bernoulli (b) Poisson

Fig. 4. A comparison of ROCs for scan statistic and CRT in Bernoulli and Poisson sensor fields, where $p_f = 0.05$ and $w = 7$.

where $G_{T_B,w}(2w) \triangleq \Pr\{S(w,2w) < T_B\}$ and $G_{T_B,w}(2w-1) \triangleq \Pr\{S(w,2w-1) < T_B\}$ can be calculated by the procedure in the *Appendix*. Intuitively, $G_{T_B,w}(2w)$ and $G_{T_B,w}(2w-1)$ denote the statistical results for a w-size window scanning two random sequences respectively with length $2w$ and $(2w-1)$. To verify the approximation accuracy of (11), we plot several curves of P_f^s versus the scan statistic threshold T_B in Fig. 3(a), where $M_1 = 101$ and $M_2 = 6$. Each curve is obtained by 1000000 Monte Carlo runs. From this figure, we see that (11) exhibits good approximation performance, and would be effective for false alarm analysis.

The system-level probability of detection is defined as

$$P_d^s = \Pr\big\{S(w, M_1) \geq T_B | \mathcal{H}_1\big\}. \tag{12}$$

However, since the conditional pdf of $p(Z_i|\mathcal{H}_1)$ is unavailable, P_d^s cannot be analytically derived or accurately approximated. In general, it can only be analyzed via simulation; since it is target dependent in any case, this seems acceptable.

2) Poisson Sensor Field: With an i.i.d. sensor noise assumption, the $\bar{X}_l$'s are i.i.d. Poisson random variables

$$\Pr\{\bar{X}_l = k\} = (\lambda_f \bar{A})^k e^{-\lambda_f \bar{A}}/k! \tag{13}$$

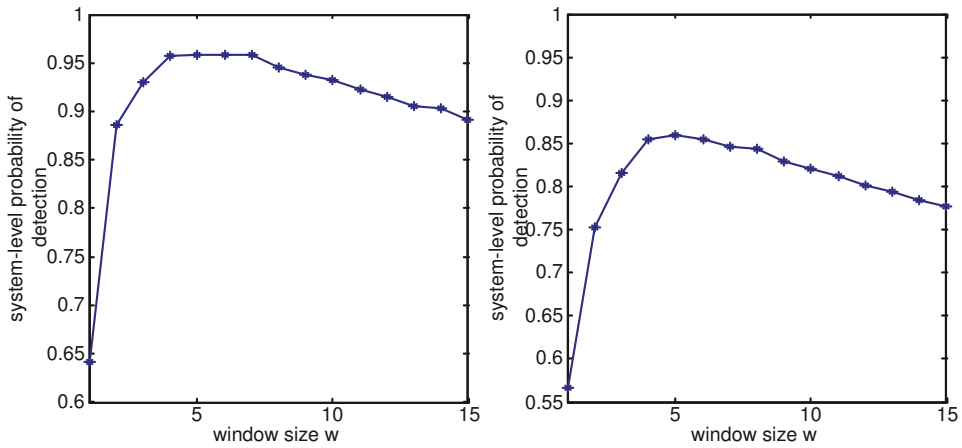
under $\mathcal{H}_0$, where $\bar{A}$ represents the area of each slice. The system-level false alarm rate is

$$P_f^s = \Pr\big\{S(w, M_1) \geq T_P | \mathcal{H}_0\big\}. \tag{14}$$

A closed-form expression for P_f^s is not available, but again it can be approximated with (11). (Actually, approximation (11) is a general result). To approximate (14), we should employ (13) instead of (9) in the calculation of $G_{T_P,w}(2w)$ and $G_{T_P,w}(2w-1)$. The approximation accuracy for the Poisson scenario is demonstrated by Fig. 3(b). In the simulation, we choose $M_1 = 101$ and $\lambda_f \bar{A}/p_f = \lambda \bar{A} = 25$. Each curve is obtained by 1000000 runs.

E. Numerical Results

1) Bernoulli Sensor Field: This subsection compares the performance of the CRT and the scan statistic in submarine detection with a Bernoulli sensor network, which covers a rectangle area $\big\{(x,y)|-2500\text{m} \leq$



(a) Bernoulli (b) Poisson

Fig. 5. The system-level probability of detection P_d^s as a function of window size w for scan statistic in Bernoulli and Poisson sensor fields, where $p_f = 0.05$, $P_f^s = 0.01$, and $\mathrm{SNR} = 5$ dB.

$x \leq 2500\mathrm{m}, 0 \leq y \leq 250\mathrm{m}\}$ in two dimensions. The sensor field is uniformly divided into 6×101 cells, with each cell containing exactly one sensor node. The acoustic source is located at $(x_s = 0, y_s = 0)$. Under $\mathcal{H}_1$, a submarine is fixed at $(x_t = 0, y_t = 1000\mathrm{m})$, while q is uniformly distributed within $[-\pi, \pi]$. c_0 is chosen as $c_0 = 2\varrho\sqrt{(x_t - x_s)^2 + (y_t - y_s)^2}$ to adjust propagation loss, and ϱ is defined as the source-to-receiver SNR.

The detection performance is compared with the help of the receiver operating characteristic (ROC). Let the local false alarm rate be $p_f = 0.05$. Fig. 4(a) gives the ROCs for the CRT and scan statistic with different SNRs, while Fig. 5(a) depicts the relationship between the system-level probability of detection P_d^s and window size w for the latter at a given SNR and system-level false alarm rate P_f^s. From these figures, we see that: ***1)*** for a given SNR, the scan statistic with a properly selected window size has better detection performance than the CRT. The reason for this is that if a submarine can only affect a part of sensor network—the detectable zone, picking up that subarea may be better than interrogating the entire field. ***2)*** The window size w can affect the ROC of a scan statistic. Since the shape of local detection cluster is not available, the best window size is unavailable. From Fig. 5(a), we see that when w varies from 4 to 7, P_d^s's do not exhibit a significant change. In other words, the scan statistic has a certain robustness in window size selection.

2) Poisson Sensor Field: This subsection compares the detection performance of the CRT and the scan statistic for a Poisson sensor network. The sensor band shares the same coverage as that in the previous example. The number of sensors is a Poisson variable with density parameter $\lambda = 1/2500$. The sensor field is sliced into 100 rectangles, and each with size 50×250 m^2. The acoustic source and the submarine setup under $\mathcal{H}_1$ are the same as the previous subsection.

Let $p_f = 0.05$. Fig. 4(b) gives the ROCs of those approaches with different SNRs, while Fig. 5(b) depicts the ROCs for the scan statistic approach with different window sizes at a given SNR and P_f^s. The observations are similar to those of the Bernoulli case.

II. Change Detection by Running Consensus

In [8] the basic change detection problem considered in this work is now formalized according to a very classical setup [1]. In the following, the index $j \in \{1, 2, \ldots, M\}$ identifies a specific sensor, while $n \geq 1$ is the (discrete) time index. The $n-$th observation $x_{n,j}$ collected by the $j-$th node follows the

null-hypothesis distribution $f_0(x)$ until a deterministic but *unknown* time n_0. From n_0 (included) on, the distribution for all j suddenly *changes* to $f_1(x)$.

The goal of the network is to discover the change as soon as possible, with a constraint on the average time between false alarms. Throughout the paper, we make the basic assumption of statistical independence across time and across sensors. We have, for all j,

$$\begin{array}{ll} f_0(x): x_{1,j}, x_{2,j}, \ldots, x_{n_0-1,j} & \searrow \\ f_1(x): & x_{n_0,j}, x_{n_0+1,j}, \ldots \end{array}$$

Note that, at each time slot n, the network globally collects a vector of observations:

$$\mathbf{x}_n = [x_{n,1}, x_{n,2}, \ldots, x_{n,M}].$$

A. *Classical parallel architecture*

If a fusion center is available, the quickest detection problem can be addressed by means of the well-known Page's test [1], which is basically made of the following three elements.

- The CUSUM log-likelihood of the data

$$S_n = \sum_{i=1}^{n} \sum_{j=1}^{M} \log \frac{f_1(x_{i,j})}{f_0(x_{i,j})}. \tag{15}$$

- A recursion rule in the form

$$S_n = \max \left\{ 0, S_{n-1} + \sum_{j=1}^{M} \log \frac{f_1(x_{n,j})}{f_0(x_{n,j})} \right\}, \tag{16}$$

  where we explicitly note that the log-likelihood resets each time it falls below zero, which is thus the point from which Page's test restarts.
- A decision rule prescribing that a change is declared as soon as a threshold γ is crossed, implicitly defining the test stopping time as

$$N = \arg\min_n \{S_n \geq \gamma\}. \tag{17}$$

The usual optimality criterion for assessing the test performance is that of imposing a constraint on the false alarm rate, and accordingly minimizing the detection delay. The former is defined as the reciprocal of the average sample size under the null hypothesis, $1/\mathrm{E}_0[N]$, where $\mathrm{E}_{0,1}[\cdot]$ denotes expectation computed under distribution $f_{0,1}(x)$. The latter is approximated by $\mathrm{E}_1[N]$, which is in fact an upper bound on the real delay, corresponding to the assumption that the CUSUM is exactly zero at time n_0. The precise computation of $\mathrm{E}_1[N]$ would instead require knowledge of the exact value of the CUSUM statistic at n_0, and it is usually intractable [1], [25].

The above key quantities admit closed form approximations mainly relying upon neglecting the excess over the threshold of the test statistic at the stopping time [1], [25]. Specifically, the false alarm rate and detection delay of the centralized system (suffix c consistently appended) are related to the detection threshold via

$$\begin{aligned} R_c(\gamma) &\approx \frac{M\,\Delta_{01}}{e^{\gamma} - \gamma - 1}, \\ D_c(\gamma) &\approx \frac{\gamma + e^{-\gamma} - 1}{M\,\Delta_{10}}, \end{aligned}$$

where Δ_{01} is the Kullback-Leibler divergence [1] from $f_0(x)$ to $f_1(x)$, and Δ_{10} is similarly defined. By combining (18) and (18) the basic operational curve $D_c(R)$ of the detector, that expresses the detection delay as a function of a prescribed false alarm rate R, can be obtained. In the regime of large γ (corresponding to small false alarm rates), the operational curve can be conveniently approximated by the following closed form

$$D_c(R) \approx \frac{\log\left(M\,\Delta_{01}/R\right)}{M\,\Delta_{10}}. \tag{18}$$

Note that the overall divergence pertaining to a single time slot is $M\,\Delta$, accounting for the fact that, at each time slot, M independent observations are collected.

B. Running consensus for quickest detection

As already anticipated, the main strategy proposed in this work for quickest distributed detection in fully flat networks relies upon the running consensus algorithm. Details about this latter can be found in [7], [9] and will not be repeated here for space reasons. In the following we limit ourselves to report the basic elements in order to make the paper self-contained.

The network topology is formalized by an undirected graph $(\mathcal{V}, \mathcal{E}_n)$ where $\mathcal{V} = \{1, 2, \ldots, M\}$ is the vertex set (sensors) and $\mathcal{E}_n$ the edge set that describes sensors' connections. To address the general problem of random and time-varying sensors' connections, we allow $\mathcal{E}_n$ to be random and dependent upon the time slot n. Accordingly, at each n, M data are collected by the network and a realization of $\mathcal{E}_n$ is drawn, meaning that some subset of $\mathcal{V}$ is selected, and the corresponding nodes share their states according to a standard consensus algorithm [6]. The exchanged data are not simply the measurements, but rather the suitable detection statistics computed by the nodes, summarized in the state variables $S_{n,j}$.

Stressing on the flat architecture of the system, we would like to achieve the following goals.

- Each sensor implements its own test by comparing the local statistic $S_{n,j}$ to a detection threshold γ. The $j-$th test accordingly stops at a random time

$$N_j = \arg\min_n \{S_{n,j} \geq \gamma\}. \tag{19}$$

- No post-detection fusion of the local decisions is allowed, the data fusion being instead embodied in the running consensus protocol.
- The decision taken by *any* of the sensors must be representative of the (unavailable) global, centralized decision. Accordingly, it must be possible to retrieve a reliable decision by querying an arbitrary node in the network.

These design goals basically require asymptotic (with n) similarity of $S_{n,j}$ with the centralized detection statistic S_n, for all j. To this aim, we propose the following update rule, that is essentially borrowed from the running consensus data-exchange protocol [7], [9]:

$$\begin{pmatrix} S_{n,1} \\ S_{n,2} \\ \vdots \\ S_{n,M} \end{pmatrix} = \mathbf{W}_n \begin{pmatrix} S_{n-1,1} \\ S_{n-1,2} \\ \vdots \\ S_{n-1,M} \end{pmatrix} + M\,\mathbf{W}_n \begin{pmatrix} \log \frac{f_1(x_{n,1})}{f_0(x_{n,1})} \\ \log \frac{f_1(x_{n,2})}{f_0(x_{n,2})} \\ \vdots \\ \log \frac{f_1(x_{n,M})}{f_0(x_{n,M})} \end{pmatrix}$$

or, in a more compact form

$$S_{n,j} = \mathcal{U}(\{S_{n-1,j}\}_{j=1}^{M}). \tag{20}$$

The M by M consensus matrices $\mathbf{W}_n$, $n = 1, 2, \ldots$, are iid (independent identically distributed) and doubly stochastic. To better highlight the physical meaning of the above matrices, let us consider the

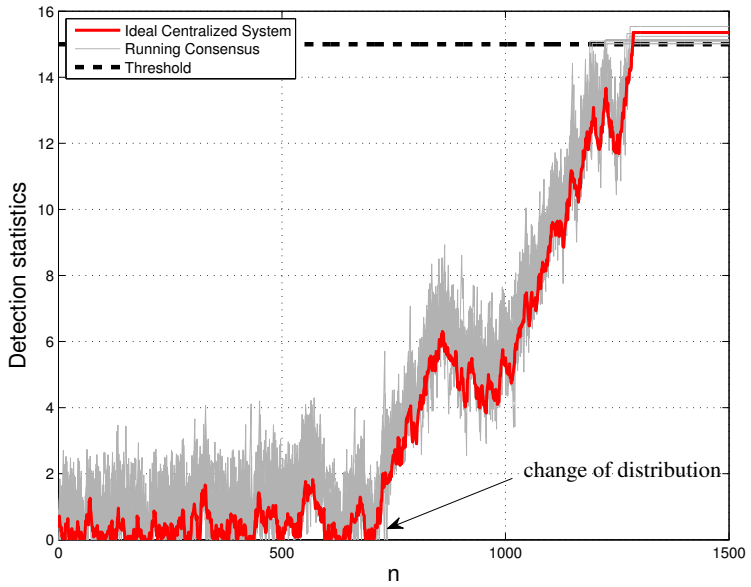


Fig. 6. Empirical realizations of the running consensus statistics, in a network made of 10 sensors (thinner gray lines, almost superimposed to each other). The bold red line refers to the centralized system.

classical example of a pairwise averaging algorithm, according to which, at time n, a pair (h,k) of sensors is uniformly and randomly selected. The corresponding realization of $\mathbf{W}_n$ is

$$\mathbf{W}_n = \mathbf{I} - \frac{(\mathbf{u}_k - \mathbf{u}_h)(\mathbf{u}_k - \mathbf{u}_h)^T}{2}, \tag{21}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{u}_k$ is a vector of all zeros, but for the $k-$th entry which is unity. Using this matrix into the update equation simply amounts to let sensors h and k replace their state by the corresponding arithmetic averages. Formally, in this case $\mathcal{E}_n = \{h,k\}$.

Our solution for quickest detection via running consensus is finally obtained by merging the update rule (20) to the classical Page's recursion (16), the overall recursion (at the $j-$th node) becoming*

$$S_{n,j} = \max\{0, \mathcal{U}(S_{n-1,j})\}. \tag{22}$$

Before going into the details of performance evaluation, it is instructive to start from empirical evidences. Figure 6, obtained by computer experiments, displays the behavior of the ideal centralized statistic S_n (bold red curve), along with the locally computed sensor statistics $S_{n,j}$ (tiny gray curves) of the running consensus Page's detectors. A general trend is observed: in a first portion of the time axis, the statistics often reset to zero; once that the change in distribution takes place, they tend to grow up to eventually cross the detection threshold. As a matter of fact, the different running consensus statistics always behave quite similarly, and, in addition, closely track the statistic of the centralized system. This in turn implies that the instants of detection events, i.e., the times at which the curves cross the positive threshold, are almost the same for the different statistics, leaving hope that the performance of the running consensus quickest detectors may approach the theoretical limit represented by the performance of the centralized system.

*While the update rule $\mathcal{U}$ is linear, the addition of Page's reset rule introduce a nonlinear effect, which is not present in the classical gossip algorithms.

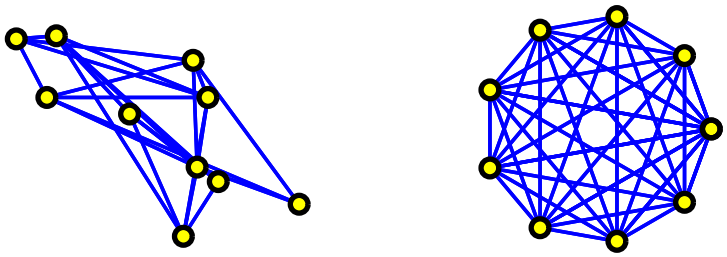


Fig. 7. Network topologies for the two examples considered in Sect. II-D. The circles represent the vertex set $\mathcal{V}$, while the random edge set $\mathcal{E}_n$ is selected among the possible connections shown by lines between the vertices.

This behavior can be explained as follows. Running consensus introduces strong dependencies among nodes by continuously propagating information across the network, and this implies that the change is detected at almost equal times at different sensors. As time elapses, the effect is emphasized and the statistics $S_{n,j}$ at different j become closer and closer each other.

As a consequence, provided that the algorithm evolves for a sufficiently long time, a reliable estimate of the instant at which the distribution-change took place can be obtained by querying any of the M nodes, and the performance of the running consensus scheme can be computed with reference to any of the sensors, according to the genuinely flat nature of the system.

C. Performance evaluation

A complete derivation of the performance formulas is not reported with all the details here; we refer the reader to [10]. The arguments below, however, are sufficient for a complete understanding of the main ideas behind the formal derivations.

It is convenient to regard the local detection statistic as $S_{n,j} = S_n + e_{n,j}$, where the difference between the current state $S_{n,j}$ and its centralized counterpart S_n is measured by an error term, that is assumed for now to be bounded, $|e_{n,j}| \leq \epsilon$, $\forall n$ and $\forall j$.

Sensors initially acquire data following the distribution $f_0(x)$. Until a threshold crossing occurs (either because a real change happened, or because a false alarm is going to be declared), the $j-$th sensor may have experienced a certain number of resets. This number, however, does not depends only upon S_n, but it is also determined by the behavior of the error term $e_{n,j}$. On the other hand, it is reasonable to assume that, for $\gamma \gg \epsilon$, the role of the centralized statistic S_n as to the threshold crossing will be predominant. Formally we have the following: Let us define

$$\begin{aligned} \underline{N} &= \arg\min_n \{S_n > \gamma - \epsilon\}, \\ \overline{N} &= \arg\min_n \{S_n > \gamma + \epsilon\}, \end{aligned}$$

that are nothing but the stopping times pertaining to a centralized Page's test with modified thresholds. Obviously, we have $\mathrm{E}_{0,1}[\underline{N}] \leq \mathrm{E}_{0,1}[N_j] \leq \mathrm{E}_{0,1}[\overline{N}]$. Let us focus on $f_1(x)$. Applying the last inequality, we have, for the detection delay at the $j-th$ sensor:

$$D_c(\gamma - \epsilon) \leq D_j \leq D_c(\gamma + \epsilon).$$

In the regime of large γ (i.e., of small false alarm rate), we can neglect the effect of ϵ (which is $\ll \gamma$), and obtain the approximate operational characteristic of the running scheme:

$$D_r(R) \approx \frac{\log\left(M\,\Delta_{01}/R\right)}{M\,\Delta_{10}}. \tag{23}$$

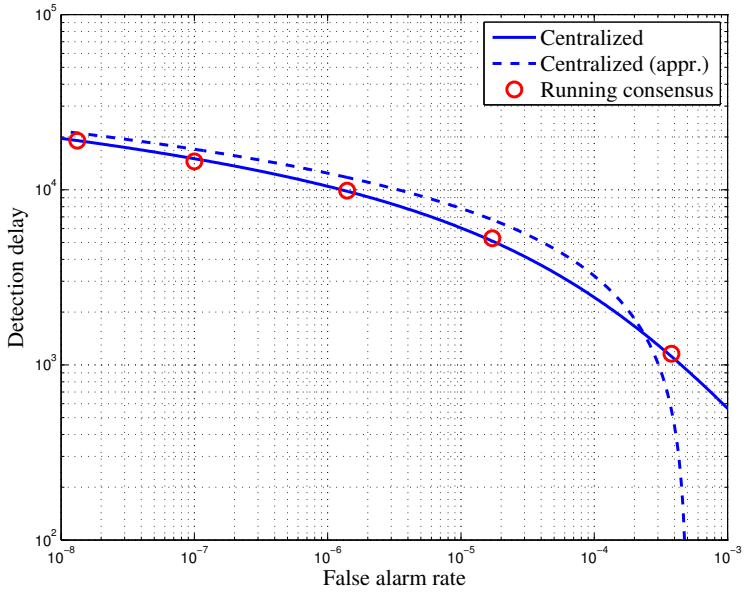


Fig. 8. Operational curve of the running-consensus quickest detector. The detection delay is shown as function of the false alarm rate for the Bernoulli example considered in Sect. II-D. Dots refer to simulations of the detector, while the continuous and dashed curves show the operational characteristic of the ideal centralized system.

We have assumed so far that the error is bounded. Such assumption is usually made in sequential analysis for managing the errors due to the excesses over the thresholds, and provides simple refinements of Wald's approximations, see, e.g. [25]. We would like to mention that an extension of these results to the case that the errors are bounded only on the average can also be pursued, but this would require rather advanced mathematical tools [23].

D. Numerical experiments

In this section we present a summary of the results obtained from Monte Carlo simulations made in two typical detection setups. As a first example, we assume the measurements taken by the sensors as iid binary variables taking value in $\{0, 1\}$, drawn from a Bernoulli distribution. Initially, the outcomes are equiprobable, while after the change the probability of 1 slightly modifies to 0.505. Note that the two hypotheses are "quite close", thus leading to a challenging detection task.

As to the exchanging protocol, we assume that the nodes communicate only with their direct (single hop) neighbors. The network topology is schematically displayed in the left plot of Fig. 7, where neighboring sensors are connected by straight lines. It is also assumed that at each time step v pairs of neighboring sensors are selected to average their own states. In the following examples we use $v = 5$.

The results of 10^4 Monte Carlo simulations, with $M = 10$ sensors, are shown in Fig. 8, where the empirical operational characteristic of the detector is compared to the operational curve $D_c(R)$ obtained by combining (18) and (18); also shown is the closed-form approximation in eq. (18), valid for tight false alarm rates R. As it can be seen, the match with $D_c(R)$ in this case is excellent, and quite accurate is also the match with (18) in the regime of interest.

Consider now a second case study, namely, the classical change detection problem of zero-mean Gaussian observations with different variances. Without loss of generality, we assume that the variance under the null hypothesis is set to 1, and that pertaining to the distribution after the change is σ^2. As

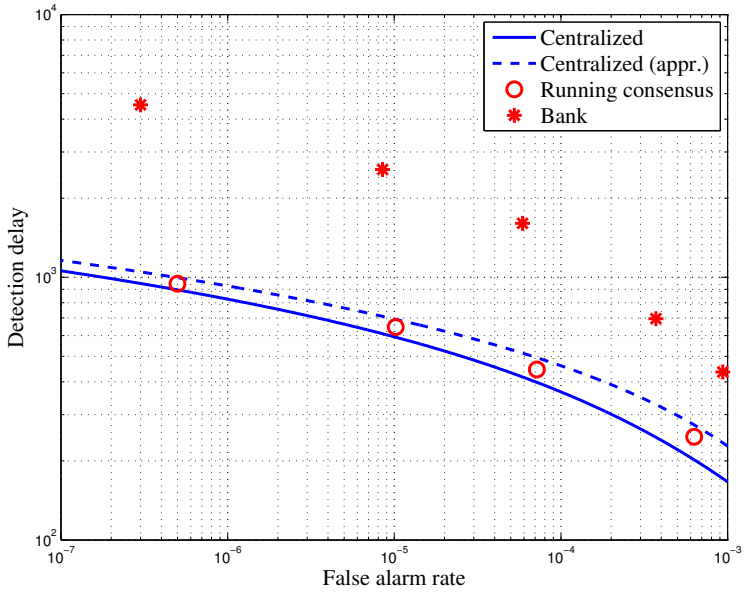


Fig. 9. Operational curve of the running-consensus quickest detector. The detection delay is shown as function of the false alarm rate for the Gaussian example of Sect. II-D. Dots refer to simulations of the running consensus strategy, while the continuous and dashed curves pertain to the ideal centralized system. Asterisks show the operational curve of a bank of Page's detectors.

communication strategy, we adopt here the same repeated pairwise averaging ($v = 5$), but with the topology shown in the right plot of Fig. 7.

In Fig. 9 we report the results from 10^4 Monte Carlo iterations with $M = 10$ sensors, and a value of $\sigma = 1.032$, that is, a value very close to 1 that again leads to a difficult detection task. Comments similar to those of the previous example apply, and the match appears to be satisfying for any practical purposes.

To further highlight the benefits of the (pre-detection) data fusion achieved via running consensus, let us consider a simpler detection scheme working in flat architectures: a bank of Page's detectors that independently process the locally observed data, without any form of on-the-fly cooperation. In this case, as soon as one of these filters declares a change is the distribution of the monitored phenomenon, a broadcast message is sent to the whole system to halt the detection task, and the decision of the quickest sensor is taken as the global decision of the bank. As seen in fig. 9, the running consensus scheme largely outperforms the bank in terms of detection performance, even though this should be expected to be paid in the coin of communication burden.

III. Data Fusion with Intermittent Sensors

A. The Problem

In undersea surveillance of large areas, multistatic sonar networks show promise in the ability to use many sensors to cover a large area with overlapping detection coverage, the achievement of higher data rates from use of multiple sensors (receivers) to process a single active transmission from a source, and in the geometric diversity that can be achieved by selecting receiver locations. However, it has been observed from at-sea testing that sensor detection performance varies significantly over the sensor network and for a single sensor over time. In particular, due to geometric, environmental, and geographic effects a target may be detected by a given sensor with high probability over a number of scans and then suddenly

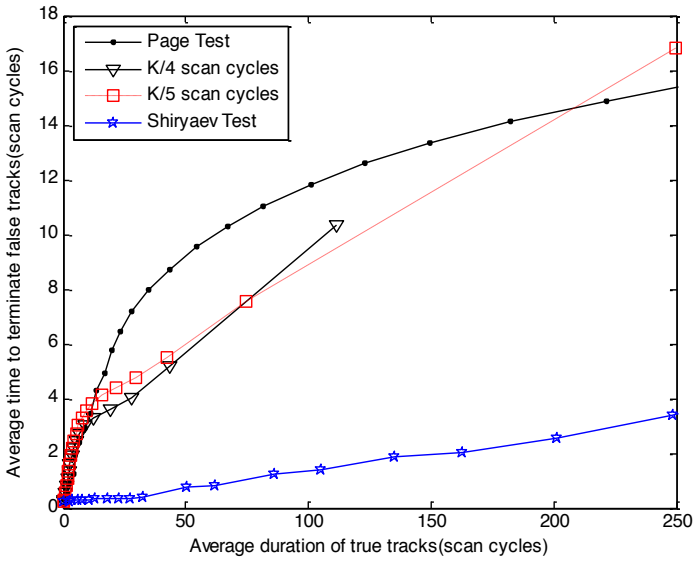


Fig. 10. Comparison of the time to terminate false and true tracks for various track termination tests.

fade from view as the sensor detection probability decreases drastically. A key tracking issue therefore becomes how to adapt the tracking system to account for this fading detection performance in a multistatic active sonar problem. In this paper we address the track termination aspect of this problem. We consider a centralized track management model that processes time ordered measurements from all sensors and include sensor origin information. The sensor detection performance is modeled as a two-state Markov chain with high and low detection states. Target-originated measure- ments (binary detection events) can therefore be described us- ing a Hidden Markov Model (HMM) structure.

There is a well-developed base of literature covering track termination for sensors with a fixed probability of detection (Pd) of the target on a single scan (see e.g., [2]). Examples include K/N tests (a track is terminated if K or fewer detections are received in the last N scans) and track score tests. Track score tests may include the SPRT or Bayesian sequential tests. If the track score (related to the probability that the detection sequence is the result of a true track) falls below a certain value, the track is terminated. However research pertaining to the track termination problem for sensors with Pd based on a Markov model has only recently been considered [12].

In [3] we analyze the performance of K/N-based and sequential track termination tests when target-originated measurements are described by a HMM with high and low Pd Markov states. Using only the binary detection events, it is shown that the K/N test outperforms the Page test over a portion of its operating characteristic region. This result is surprising considering the fact that the Page test is proven to be the optimal sequential test for quickest detection of a change in measurement distribution and we show how when the HMM-based detection statistics are used, a key assumption in the optimality proof for the Page test is no longer satisfied. It is next shown that by using a Bayesian version of a sequential test (the Shiryaev test), significant performance improvement is obtained compared to the K/N test.

B. Results

We compare the performance of the Page test, the Shiryaev test and the K/N rule. Monte Carlo simulations were performed. For each hypothesis, 104 simulations were performed. Simulations under

H_0 yield the average false track life, called average detection delay (ADD). Simulations under H_1 yield the average true track life, called average run length (ARL). Results are plotted for each track termination test over a set of threshold values. Figure 10 presents the track termination performance of each test. Although asymptotically better than the K/N rule, the Page test performed worst over the operating range likely to be used in a track termination module of a track management system (ADD of false tracks 4?15 scan cycles). The Shiryaev test performed the best of the three tests considered. This is likely due to the information provided by the prior knowledge of the distribution of change times used in the Shiryaev test. As a sequential test, the computational cost of the Shiryaev test is small and easily computable as part of an overall track management system. The sub-optimality of the Page test was not expected. Examination of the assumptions used in the optimality proofs of the Page test show that the increments of the cumulative sum must be i.i.d. [1]. That does not hold in this situation.

IV. Distributed Learning

A. Background

Distribution-free or nonparametric inference is a well-established discipline. In the case of the so-called supervised learning, the idea is to build the final inference using a training set $\{(X_i, Y_i)\}_{i=1}^n$ of available i.i.d. (independent, identically distributed) "examples". Specifically, after observing X_0, the goal of the system is to infer what is the "corresponding" Y_0, exploiting the fact that (X_0, Y_0), while it is independent of the training set, has exactly the same joint distribution of each example (X_i, Y_i), although such distribution is completely unknown. To fix terminology, let us say that $Y_0 \in \mathcal{R}$ is called the response variable, while the collected $X_0 \in \mathcal{R}^d$ is termed the observation variable. Also, we here work with continuous random variables.

In the context of distributed inference systems it is usually assumed that the training set is disseminated through the network (assume, for simplicity, that sensor i owns the single example (X_i, Y_i)) and the problem arises of the interplay between the necessity of making available the examples to a common site (the fusion center, FC), and the presence of communication constraints [20], [21]. These latter can be in terms of sensors' energy, channel impairments (e.g., noise, interferences, lack of phase coherence, fading), computational capabilities, and so on.

Complying with the standard assumptions in sensor network applications we assume that the communication links are asymmetric: the communication from the FC to the sensors is essentially unconstrained, such that X_0 is perfectly recovered by the nodes, while on the other hand strict communication constraints act on the sensors' side, due to, as already said, their typical limitations in terms of available energy and hardware/software complexity.

Recently, the authors of [21] showed that a consistent fusion rule can be devised, for the specific case of the classical *naive kernel* estimator, which amounts to compute the empirical mean of the Y_i's corresponding to those samples in the training set that lie in a circumscribed neighborhood of the observation X_0. This means that each sensor knows if its example must be delivered or not (actually only the "labels" Y_i are delivered), regardless of all other examples (X_m, Y_m), $m \neq i$. In the same paper, however, it is noted that implementing the NN regression rule is challenging because, in that case, the estimator exploits the k examples that are closest to X_0, i.e., it uses the k values Y_i corresponding to the k smallest values of the norm $\|X_i - X_0\|$. The information available to each sensor is insufficient to decide whether or not its example should be delivered to the FC. In [18] we show that this challenging problem can be solved: capitalizing on a smart idea originally proposed in [5], and further developed in [4], we are able to implement the NN fusion rule in a fully decentralized way such that universal consistency of the regression rule is ensured.

We have addressed the emerging problem of distributed learning in wireless sensor networks. Our main focus has been to extend the currently available results on the subject basically in two directions. We propose: (i) distributed NN regression rules; and (ii) strategies achieving universal consistency over *noisy channels*. The key enabler for *distributed* NN regression is the "ordered transmission" scheme from [5]. The main results found in this paper can be so summarized:

- We design an access policy such that the FC recovers exactly those labels it needs for the inference task. This is done without inter-sensor coordination.
- A *single-tranmission* NN regression rule is proposed, achieving an asymptotic performance of twice the MMSE.
- A general scheme for k_n-NN consistent regression rules with *quantized labels* is devised, which can be implemented in a distributed fashion. This strategy is studied in both the noiseless case, and in the case that the links connecting the remote sensors to the FC are modeled as independent BSCs. Notably, the proposed algorithms do not require knowledge of the channel noise parameters, and do work in the absence of channel state informations. They require only knowledge of the error probabilities of the BSCs.
- We introduce the concept of *uncoded* k_n-NN consistent regression rules, and propose two possible implementations suited for coherent and noncoherent channel models, respectively; numerical experiments show that, in both implementations, the uncoded schemes can outperform the previously considered coded ones. Again, remarkably, the proposed algorithms do not require knowledge of the channel noise parameters, and work well without needing channel state information (needless to say, in the coherent case, the phase information is assumed to be retrievable). Only the first (resp., second) moment of the fading distribution must be known at the receiver for the coherent (resp., noncoherent) scenario.

While the attention has been focused on NN rules, some results found in this paper appear to be generalizable to other regression rules, such as, e.g., the distributed naive kernel proposed in [21]. These results include the general procedure for preserving consistency over BSCs, and the idea of uncoded communication for distributed regression.

B. Classical Nearest-Neighbor Regression

In the distribution-free approach to statistical inference, a classical way to build estimators relies upon the so-called *local averaging* regression functions [17]:

$$\sum_{i=1}^{n} W_{ni}(x_0)\, Y_i, \tag{24}$$

where the weights are functions of x_0 and of the observation variables in the training set, namely, $X_1, X_2, \ldots, X_n$. Within this class, a major role is played by the NN rules, which are now briefly summarized.

Let us introduce some notation. All throughout the paper, the probability density function (pdf) and the cumulative distribution function (cdf) of a random variable X will be denoted, respectively, by $f_X(x)$ and $F_X(x)$. The pdf of the observation-label pair (X, Y) is accordingly $f_{XY}(x, y)$.

For $x_0 \in \mathcal{R}^d$, let:

$$\left(X_{(1,n)}(x_0), Y_{(1,n)}(x_0)\right), \ldots, \left(X_{(n,n)}(x_0), Y_{(n,n)}(x_0)\right),$$

be the sequence of pairs ordered according to

$$\|X_{(1,n)}(x_0) - x_0\| \leq \cdots \leq \|X_{(n,n)}(x_0) - x_0\|,$$

where $\|\cdot\|$ denotes the standard Euclidean norm in $\mathcal{R}^d$, and, again, n is the training set cardinality. In our model, ties will be excluded by the assumption of continuous random variables. The NN regression rule can be now formalized as [17]:

$$\frac{1}{k_n}\sum_{i=1}^{k_n} Y_{(i,n)}(x_0),$$

so that the weights in (24) are

$$W_{ni}(x_0) = \begin{cases} \dfrac{1}{k_n}, & \text{if } X_i \text{ is one of the } k_n\text{-NN of } x_0, \\ \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

In the above, the notation k_n is used because the number of neighbors might depend on the number of observations n. Of particular interest is the case that $n, k_n \to \infty$ while $k_n/n \to 0$, because, with this choice, the NN regression rule achieves consistency [17]. In the following, we shall also consider the case that $k_n = const.$, with emphasis on the extreme case $k_n = 1$, i.e., only one (but the most informative) label is used.

C. Nearest-Neighbor Access

Let us now focus on the *decentralized* implementation of a local averaging regression function. To this aim, we note that in eq. (24) the effect of the response variables Y_i is decoupled from that of the observation variables X_i, which are contained only in the weights $W_{ni}(x_0)$. In addition, in the NN rule (25), the weights exhibit an on-off structure, which immediately suggests a convenient way to manage the access to the channel that is necessary in the decentralized setup: only those sensors with nonzero weights should convey information toward the FC. This yields as a fundamental design guideline the following decoupled approach: the *access policy* is designed to reproduce the weights at the FC, while the data *coding* strategy acts only on the relevant labels Y_i to be transmitted.

In the simplest case of the naive kernel, this idea directly translates to the scheme of distributed regression with abstention employed in [21]. There, the event that the i-th label must be considered for estimation depends only on the i-th observation (and on X_0), such that abstention is determined by a *locally available* knowledge. In the case of NN regression functions of our interest, this is no longer true, leading to the difficulty quoted from [21]. To overcome this issue, with the general "decoupling" philosophy in mind, a different access policy specifically tailored to the NN rules must be conceived.

The transmission scheme for implementing k_n-NN-based rules works as follows, see Fig. 11. The i-th sensor computes the *local* distance between the observation X_i available in its own example, and the measurement X_0, namely $\|X_i - X_0\|$. After that, it sends its own label Y_i over the channel, at a transmission instant which is directly related to the distance $\|X_i - X_0\|$. The time-origin of the network is set when the sensors receive a broadcast message from the FC (for example, when X_0 is received), and we assume that the time scale of the system is such that differences in the transmission delays between different sensors and the FC can be safely neglected. Also, once that the desired number of labels have been received, further transmissions by the sensors are inhibited by means of a broadcast stop message sent by the FC.

In this way, sensors having collected observations with smaller distances from X_0 transmit first, so that the FC receives the labels $Y_{(i,n)}(x_0)$, $i = 1, \ldots, k_n$, i.e., the labels are just ordered in terms of the desired NN criterion, and only the first k_n labels are sent. Otherwise stated, the FC is able to compute the weights relevant for the evaluation of the regression function (24), and this is obtained without transmitting the value of the observation variables X_i. The messages sent over the channel are the labels Y_i's. The specific

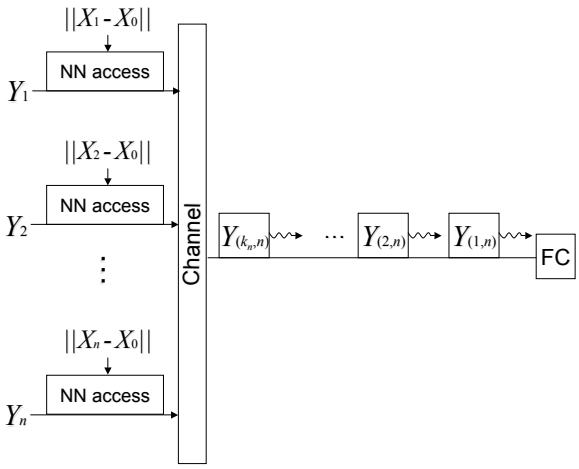


Fig. 11. Distributed learning with NN access. The i-th sensor decides to transmit its label Y_i, accessing the channel medium at a time instant which is based upon the observed distance $||X_i - X_0||$. This allows the FC to recover the labels ordered according to the desired k_n-NN criterion.

way to convey this information toward the FC, coping with the communication constraints, is discussed in [18].

Note that the procedure described is based on a *cross-layer* approach, mixing the layers of communications and inference: sensors attempt to access the channel in a random fashion, but at specific time instants related to the quality of their data for the final inference.

Before going on, a key point to be emphasized concerns the access policy: As n grows, the distances $\|X_{(i,n)} - X_0\|$ shrink to zero, and therefore it can be difficult for the FC to identify correctly the first k_n deliveries and to stop further transmissions. To avoid these difficulties, the distances $\|X_{(i,n)} - X_0\|$ must be properly scaled by a proportionality factor depending upon n, in order to make successive transmissions sufficiently spaced apart. In our distribution-free setup, this is particularly challenging, in that a *universal* scaling law must be found, namely, one valid for any observation model.

D. Numerical examples

The results of computer experiments are now described. We consider the following example:

$$\begin{aligned} &X \text{ uniform in } [-2,2], \\ &Y = r^*(X) + \frac{E}{\sqrt{\text{VAR}\{E\}}}\sqrt{\text{MMSE}}, \end{aligned}$$

where the optimal regression function $r^*(x_0)$ is chosen as shown in Fig. 12 (see inset), that is, the concatenation of a sinusoidal and a triangular wave. In the above, the error term E is a mixture of a zero-mean Gaussian and a symmetrized Weibull distribution, with pdf:

$$f_E(x) = \lambda \mathcal{N}(x; a_1) + (1-\lambda)\mathcal{W}(x; a_2, a_3),$$

where $\mathcal{N}(x, a_1)$ is the pdf of a zero-mean Gaussian with variance a_1^2, and $\mathcal{W}(x; a_2, a_3)$ denotes the pdf of a symmetrized Weibull distribution with scale parameter a_2 and shape parameter a_3, namely:

$$\mathcal{W}(x; a_2, a_3) = \frac{1}{2}\frac{a_3}{a_2}\left(\frac{|x|}{a_2}\right)^{a_3-1} \exp\left\{-\left(\frac{|x|}{a_2}\right)^{a_3}\right\}.$$

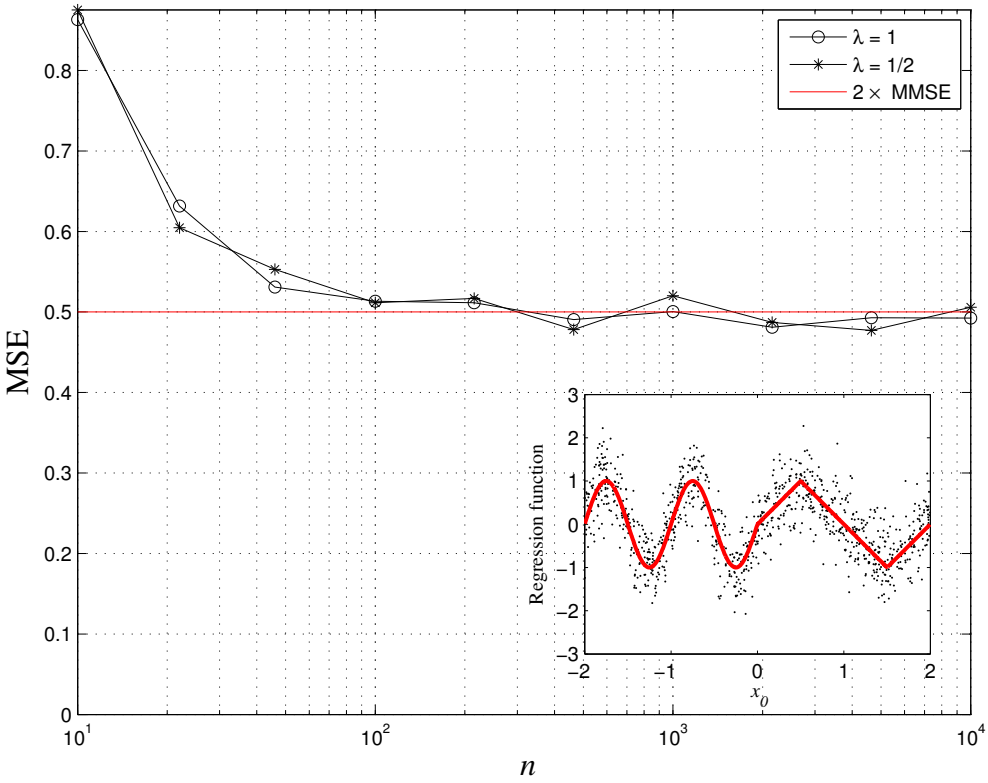


Fig. 12. Single-transmission NN strategy, applied to the estimation problem (26), with parameters $a_1 = 0.1, a_2 = 1, a_3 = 0.9$, MMSE= $1/4$, for two values of the mixing probability, namely, $\lambda = 1$ and $\lambda = 1/2$. In the inset box, the optimal regression function $r^*(x_0)$ (bold line) is displayed, along with samples (circles), corresponding to different Monte Carlo runs, of the regression function $Y_{1,n}(x_0)$ estimated by $n = 10^4$ training set examples. In the main plot, the MSE performance of the distributed strategy, estimated over 10^4 Monte Carlo runs, is compared to the asymptotic limit of $2\times$MMSE (shown as a straight line).

Without any pretence of exhaustiveness, we adopt the above mixture model in order to appreciate how the system performs in the presence of deviations from a typical scenario, such as the Gaussian one. In the following, we report the results arising from two cases, corresponding to $\lambda = 1$ (purely Gaussian) and $\lambda = 1/2$ (Gaussian/Weibull half-mixed). The employed parameters are $a_1 = 0.1, a_2 = 1, a_3 = 0.9$, and MMSE=$1/4$.

In the forthcoming analysis, we first investigate the estimation performance of the proposed schemes, and then focus on the transmission issues. Let us start by exploring the unquantized, single-transmission strategy. The box panel in Fig. 12 displays the optimal regression function (bold line), along with the surrounding cloud of points representing the estimated regression function observed in a number of Monte Carlo trials. As it can be seen, the system is able to learn the shape of the optimal regression function, and the spread about the true regression function yields the MSE performance shown in the figure, reaching the asymptotic limit of $2\times$MMSE (horizontal line), even for moderately large values of n, in the order of a few tens. Here, as already pointed out in the theoretical analysis, we basically work under the assumption that the communication protocol affords the FC perfect recovery of the NN label, so that the above findings are expected from the standard theory of NN estimation [13].

REFERENCES

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, N.J: Prentice-Hall, 1993.

[2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[3] W. Blanding, P. Willett, Y. Bar-Shalom, and S. Coraluppi, "Multisensor track management for targets with fluctuating snr," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 4, pp. 1275–1292, Oct. 2009.

[4] R. S. Blum, "Ordering for estimation and optimization in energy efficient sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2847–2856, 2011.

[5] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3329–3235, 2008.

[6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[7] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, 2008.

[8] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 814–825, Feb. 2010.

[9] ——, "Asymptotic optimality of running consensus in testing statistical hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, 2010.

[10] ——, "Consensus-based page's test in sensor networks," *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.

[11] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 98–101, Jan. 1986.

[12] S. Coraluppi and C. Carthel, "Distributed tracking in multistatic sonar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 1138–1147, Jul. 2005.

[13] T. M. Cover, "Estimation by the nearest neighbor rule," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 50–55, Jan. 1968.

[14] J. Glaz and J. Naus, "Tight bounds and approximations for scan statistic probabilities for discrete data," *Ann. Appl. Probab.*, vol. 1, no. 2, pp. 306–318, May 1991.

[15] J. Glaz, J. Naus, and S. Wallenstein, *Scan Statistics*. Springer, 2001.

[16] M. Guerriero, P. Willett, and J. Glaz, "Distributed target detection in sensor networks using scan statistics," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2629–2639, Jul. 2009.

[17] L. Gyorfi, M. Kohler, A. Krzyzak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002.

[18] S. Marano, V. Matta, and P. Willett, "Nearest-neighbor distributed learning by ordered transmissions," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 5217–5230, Sep. 2013.

[19] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP J. Wireless Commun. Netw.*, no. 4, pp. 339–349, 2005.

[20] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, 2006.

[21] ——, "A collaborative training algorithm for distributed learning," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1856–1871, 2009.

[22] X. Song, P. Willett, J. Glaz, and S. Zhou, "Active detection with a barrier sensor network using a scan statisti," *IEEE Journal on Oceanic Engineering*, vol. 37, no. 1, pp. 66–74, Jan. 2012.

[23] A. G. Tartakovsky and I. A. Ivanova, "Approximations in sequential rules for testing composite hypotheses and their accuracy in the problem of signal detection from post-detector data," *Prob. Inform. Transm.*, vol. 28, no. 1, pp. 55–64, Jan.-Mar. 1992.

[24] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. New York: McGraw-Hill, 1983.

[25] A. Wald, *Sequential Analysis*. New York: Dover, 1947.

[26] S. Zhou and P. Willett, "Submarine location estimation via a network of detection-only sensors," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3104–3115, Jun. 2007.