

RESEARCH

Open Access



Scan statistics with local vote for target detection in distributed system

Junhai Luo^{1,2*}  and Qi Wu¹

Abstract

Target detection has occupied a pivotal position in distributed system. Scan statistics, as one of the most efficient detection methods, has been applied to a variety of anomaly detection problems and significantly improves the probability of detection. However, scan statistics cannot achieve the expected performance when the noise intensity is strong, or the signal emitted by the target is weak. The local vote algorithm can also achieve higher target detection rate. After the local vote, the counting rule is always adopted for decision fusion. The counting rule does not use the information about the contiguity of sensors but takes all sensors' data into consideration, which makes the result undesirable. In this paper, we propose a scan statistics with local vote (SSLV) method. This method combines scan statistics with local vote decision. Before scan statistics, each sensor executes local vote decision according to the data of its neighbors and its own. By combining the advantages of both, our method can obtain higher detection rate in low signal-to-noise ratio environment than the scan statistics. After the local vote decision, the distribution of sensors which have detected the target becomes more intensive. To make full use of local vote decision, we introduce a variable-step-parameter for the SSLV. It significantly shortens the scan period especially when the target is absent. Analysis and simulations are presented to demonstrate the performance of our method.

Keywords: Scan statistics, Target detection, Local vote decision, Data fusion, Variable-step-parameter

1 Introduction

Target detection has an important research significance on military and civil applications in distributed system, such as intrusion detection and fire detection. The reliability of target detection result suffers from the problem of local false alarm, while data fusion can improve the precision of the target detection. For multiple sensor systems, sensors send their sense data to a fusion center, and the fusion center makes the final decision to improve the global probability of detection. Distributed detection using multiple sensors and optional fusion rules has been extensively investigated.

Chair and Varshney [1] present an optimum fusion structure to classical Bayesian detection problem in distributed sensor networks. To obtain the global decision, the fusion center weighs the reliability of every sensor and

compares them with a threshold. The reliability of sensor is supported by the probability of detection and false alarm rate. Although this method can get optimum performance, it has to know the probability of detection and false alarm rate previously. Since we cannot get the target location before detection, this method cannot be applied to practical applications.

Niu and Varshney [2] put forward the counting rule, where the fusion center employs the total number of detections reported by local sensors for hypothesis testing and analyzes the performance of the counting rule with a significant number of sensors. In [3], the authors give performance analysis when sensors are deployed in a random sensor field. The counting rule does not need the probabilities of local detection and a false alarm in advance. It makes the counting rule more suitable for the practical environment. For the counting rule, it takes the results of sensors equally. However, sensors close to the target get higher accuracy than sensors far away. Sensors far away from the target degrade global probability of detection. In [4, 5], the authors propose a method which gives different weight to each sensor depending on the evaluated

*Correspondence: junhai_luo@uestc.edu.cn

¹School of Electronic Engineering, University of Electronic Science and Technology of China, No.2006, Xiyuan Ave, West Hi-Tech Zone, 61000 Chengdu, China

²Department of Electrical Engineering and Computer Science, The University of Tennessee Knoxville, 37919 Knoxville, USA

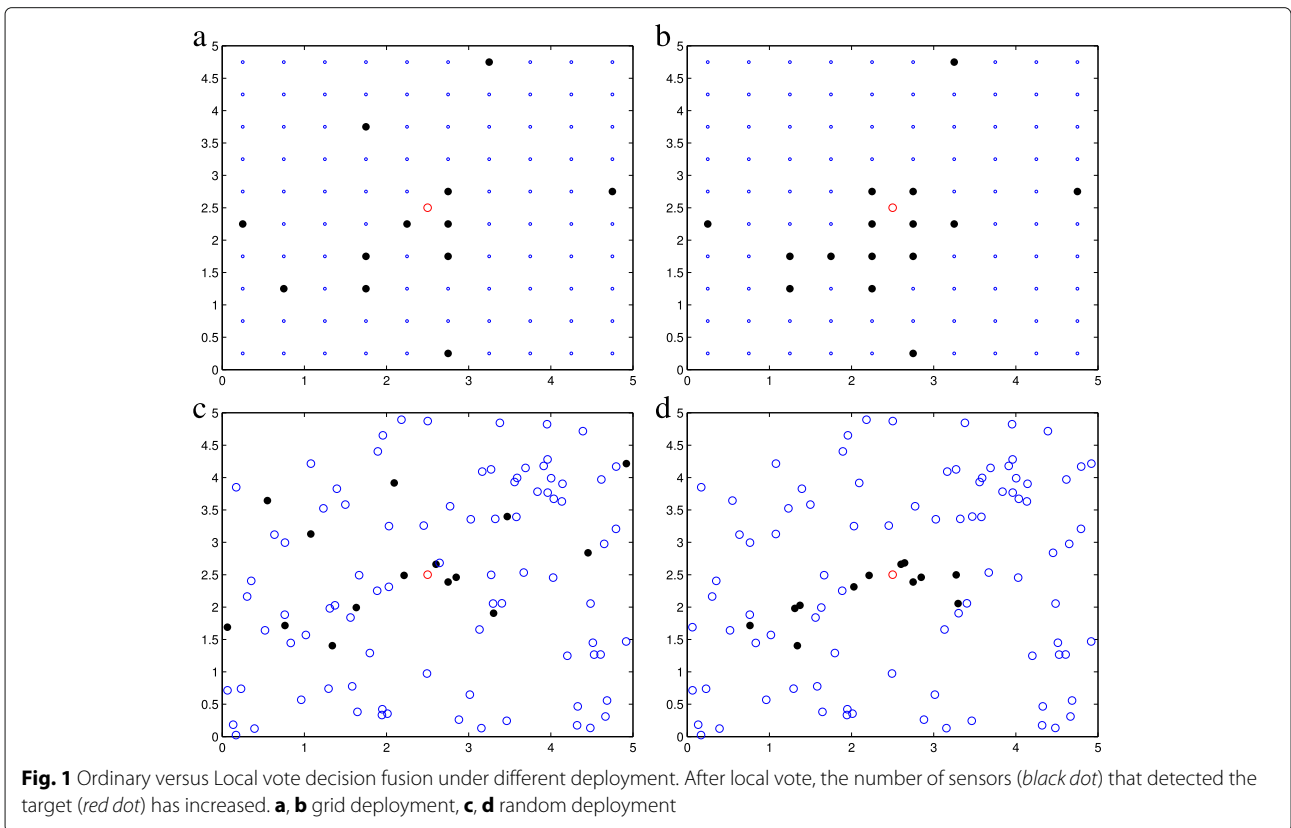
distance to the target or the signal-to-noise ratio (SNR). It makes the final decision more accurate at the cost of sending more data to the fusion center.

The authors in [6] propose the local vote algorithm using decisions of neighboring sensors and making a collective decision as a network. The authors examine both distance-based and nearest neighbor-based versions of local vote algorithm for grid and random sensor deployments and show that in many situations, for a fixed system false alarm, the local vote correction achieves significantly higher target detection rate than the decision fusion based on uncorrected decisions (see Fig. 1). The authors in [7] propose an improved threshold approximation for the local vote decision fusion and demonstrate that this method can achieve a more accurate result.

Scan statistics has been used to an epidemic or computer intrusion in [8–11]. Moreover, Guerriero [12] puts the scan statistics to the signal processing community firstly. The detection is carried out in a mobile fusion center as a mobile agent (MA) which successively counts the number of binary decisions reported by local sensors lying inside its moving field of view. The MA, playing the role of the fusion center, makes the final decision about the presence of a target. The authors also demonstrate the existence of optimal size for the field of view and the disjoint-window test. In disjoint-window scan

test, the MA travels across the sensor network and scans the network using no overlapping windows. In [13–15], the authors introduce the variable window scan statistics and investigate the performance of those variable window scan statistics methods. The disjoint-window scan statistics can shorten the scan period. However, it has poor performance compared with the scan statistics (SS).

How to improve the probability of detection while reducing the false alarm rate is an eternal topic. To handle complex network environments, improving the global performance in low SNR is our primary goal. The research mentioned above can improve the detection probability and decrease the false alarm rate. However, those algorithms cannot meet the expected performance in low SNR. The local vote algorithm can significantly improve the global performance, especially in low SNR. After the local vote, the counting rule is adopted for decision fusion. The counting rule does not consider the spatial correlation that sensors near the target have a higher probability of reporting detections. It weakens the advantage which is brought by local vote. In this paper, we combine two previous ideas: local vote decision and scan statistics. Sensors make a local vote, and the MA performs scan statistics. According to Fig. 1, we can see that local vote makes the distribution of sensors which have detected the target more concentrated. It inspires us to introduce a



variable-step-parameter for the scan statistics with a local vote (SSLV). Our contributions in this paper are described as follows.

- A model of the SSLV is proposed. We analyze the difference between the SSLV and the traditional SS. The deduction of global false alarm ratio for the SSLV is developed.
- We apply the SSLV to a grid sensor network and compare its performance with the SS. According to the simulation, we can prove that the SSLV overwhelms the SS in low SNR. We also verify that an optimal M_x at a given situation does exist.
- We introduce a variable-step-parameter for the SSLV and analyze its influence on our method. From the simulation, we know that the variable-step-parameter has little negative effect on detection performance of the SSLV. However, it significantly shortens the scan period especially when a target is absent.

The remainder of the paper is organized as follows. Section 2 demonstrates two-dimensional scan statistics as a foundation for the SSLV. Section 3 describes the system model of scan statistics with the local vote and introduces a variable-step-parameter into the SSLV. Section 4 applies the SSLV to a grid sensor network, and various simulations and analysis are provided. Finally, Section 5 concludes our research.

2 Scan statistics

In this part, we will introduce the classical two-dimensional scan statistics algorithm. The scan statistics is a kind of distributed detection method. Each sensor makes its hypotheses according to its sense data and sends the result to the fusion center. The traditional counting rule algorithm collects data from all sensors in the field of interest and makes the global judgment through these data. Unlike the counting rule algorithm, the SS makes an MA sequentially collect the data from the agent area, and the MA makes the final decision for the global network. When a target is present, sensors near the target are more likely to make the right judgments. The SS considers this spatial correlation that makes it more accurate than the counting rule.

We assume that all sensors follow the same hypotheses: either H_0 (target absent) is valid or H_1 (target present) is uniformly accurate. R presents the region of interest (ROI). We deploy sensors in region R , and the region is defined by $[0, T_1] \times [0, T_2]$. More specifically, let $h_i = T_i/N'_i > 0$, where N'_i are positive integers, and $i = 1, 2$. For $1 \leq i \leq N'_1$ and $1 \leq j \leq N'_2$, let X'_{ij} be the count of sensors that have been observed in the rectangular basic regions $[(i-1)h_1, ih_1] \times [(j-1)h_2, jh_2]$. In the process of scan, the MA records the results of

sum in its agent region, and the size of agent region is m_1 by m_2 .

$$v'_{i_1, i_2} = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} X'_{xj} \quad (1)$$

where m_1 and m_2 are the width and height of the MA respectively. We also call the m_1 and m_2 the window sizes of the MA. The MA collects data and finds the maximum value to compare with a pre-set threshold value k .

$$S_{m_1 \times m_2 N_1 \times N_2} = \max\{v'_{i_1, i_2}; 1 \leq i_1 \leq N_1 - m_1 + 1, 1 \leq i_2 \leq N_2 - m_2 + 1\} \quad (2)$$

If there is a maximum value greater than k , we can say that k events are clustered within the inspected region. Therefore, the global probability of detection PD and global probability of false alarm PF can be respectively expressed as

$$\begin{aligned} PD &= P(S_{m_1 \times m_2 N_1 \times N_2} \geq k | H_1) \\ PF &= P(S_{m_1 \times m_2 N_1 \times N_2} \geq k | H_0) \end{aligned} \quad (3)$$

It is important for us to give the expression of $P(S_{m_1 \times m_2 N_1 \times N_2} \geq k)$. Although there is no exact expression for $P(S_{m_1 \times m_2 N_1 \times N_2} \geq k)$, we can evaluate the approximation for it. When the X'_{ij} is the Bernoulli random variable with parameter $P = \alpha$, where $0 < \alpha < 1$, the accurate approximation for $P(S_{m_1 \times m_2 N_1 \times N_2} \geq k)$ can be expressed as

$$\begin{aligned} P(S_{m \times m} \geq k) &\approx 1 - \frac{[P\{S_{m \times m}(m, m) \leq k-1\}]^{(N-m-1)^2}}{[P\{S_{m \times m}(m, m+1) \leq k-1\}]^{2(N-m-1)(N-m)}} \\ &\quad \times [P\{S_{m \times m}(m+1, m+1) \leq k-1\}]^{(N-m)^2} \end{aligned} \quad (4)$$

The full content can be found in [12, 16]. In [12], the authors also demonstrate the expression when the X'_{ij} conforms to Poisson distribution.

3 Two-dimensional scan statistics with local vote decision

In this section, we give a detailed description of the SSLV that we proposed and the problems which are brought by local vote. Section 3.1 presents the scan statistics with local vote decision algorithm. The correlation of local sensors is presented in Section 3.2. According to our analysis, it turns out that the results provided by sensors are not independent and identically distributed (i.i.d) anymore after the local vote. It makes an expression (4) cannot be used in the SSLV, but a new expression for the SSLV is deduced in Section 3.3. Section 3.4 introduces a variable-step-parameter into the SSLV in order to take full advantage of the local vote.

3.1 Scan statistics with local vote decision

Precisely, in this part, we will give an introduction about the SSLV in two-dimensional region. The underlying assumptions in the last part are still suitable here. The difference is that we let sensors make a local vote decision before scan statistics. According to [6], we can take various neighborhood algorithms, such as fixed distance r or fixed size. Any one of those algorithms can be selected, then all corresponding parameters can be confirmed. For better description, we will redefine some variables. Let $X_{i,j}$ be the event that has been observed in the rectangular sub-regions $[ih_1, (i + 1)h_1] \times [jh_2, (j + 1)h_2]$ after local vote where $1 \leq i \leq N_1$ and $1 \leq j \leq N_2, N_i = N'_i - 2$, and for simplicity, we exclude the rectangular sub-regions on the edge of the field in this part. For $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$, let m_1 and m_2 be the positive integers, $1 \leq m_1 \leq N_1$ and $1 \leq m_2 \leq N_2$.

$$v_{i_1, i_2} = \sum_{j=i_2}^{i_2+m_2-1} \sum_{i=i_1}^{i_1+m_1-1} X_{i,j} \quad (5)$$

Similarly, if v_{i_1, i_2} exceeds a pre-set value of k , the MA makes the final decision that a target is present. The largest number of events in an agent region can be expressed as

$$S_{m_1 \times m_2; N_1 \times N_2} = \max \{v_{i_1, i_2}; 1 \leq i_1 \leq N_1 - m_1 + 1, 1 \leq i_2 \leq N_2 - m_2 + 1\} \quad (6)$$

For simplicity, we abbreviate $S_{m_1 \times m_2; N_1 \times N_2}$ to S . The next step is to obtain the expression of $P(S_{m_1 \times m_2; N_1 \times N_2} \geq k)$ to make the SSLV useful.

3.2 Correlation of sensors

Our algorithm introduces local vote decision into the traditional scan statistics. Therefore, we should figure out what has changed after the combination of two algorithms. The dependence among sensors should be examined first. For any sensor detection event Z_i , we start by calculating the expected value μ_i and variance σ_i^2 of the updated decision.

$$\mu_i = P(Z_i = 1) = \sum_{n=M_x}^{M_i} \binom{M_i}{n} \alpha^n (1 - \alpha)^{M_i - n} \quad (7)$$

where M_i is the number of neighbors which depends on local vote decision algorithm. M_x is a variable that has a significant influence on the performance. $\sigma_i^2 = \mu_i (1 - \mu_i)$.

The dependence between Z_i and Z_j has relations with the intersection of their respective neighborhoods $U(i)$ and $U(j)$. The number of sensors in the intersection $U(i) \cap U(j)$ can be denoted by $n_{i,j}$. According to the expression of covariance, we first compute $E(Z_i Z_j) = P(Z_i = Z_j = 1)$

and then calculate the covariance between Z_i and Z_j . We divide the neighborhoods into three parts. Suppose that A is the number of positive decisions in $U(i) \cap U(j)$ and B is the number of positive decisions in $U(i)$, but not in $U(j)$, while C is the number of positive decisions in $U(j)$, but not in $U(i)$. Noting that A, B, and C are independent, we can have

$$E(Z_i Z_j) = \sum_{k=0}^{n_{i,j}} P(A = k) P(B > M_{x_i} - k) \times P(C > M_{x_j} - k) \quad (8)$$

$$P(A = k) = \binom{n_{i,j}}{k} \alpha^k (1 - \alpha)^{n_{i,j} - k} \quad (9)$$

$$P(B > M_{x_i} - k) = \sum_{q=M_{x_i}-k+1}^{M_i - n_{i,j}} \binom{M_i - n_{i,j}}{q} \alpha^q (1 - \alpha)^{M_i - n_{i,j} - q} \quad (9)$$

$$P(C > M_{x_j} - k) = \sum_{q=M_{x_j}-k+1}^{M_j - n_{i,j}} \binom{M_j - n_{i,j}}{q} \alpha^q (1 - \alpha)^{M_j - n_{i,j} - q} \quad (10)$$

The covariance is then given by

$$Cov(Z_i, Z_j) = [E(Z_i, Z_j) - \mu_i \mu_j] I(n_{i,j} > 0) \quad (11)$$

According to the deductions above, we can find out that decision $X_{i,j}$ is not i.i.d anymore after the local vote.

3.3 Approximation for $P(S \geq k)$

In [16], the authors give the proof of approximation when the $X_{i,j}$ is i.i.d with the Markov Chain imbeddable systems [17]. Obviously, it is not applicable here. Luckily, there are different ways to give the accurate approximation for $P(S \geq k)$ and one of them is using the Haiman theorem [18–20].

Theorem 1 *Let $\{X_i\}$ be a stationary 1-dependent sequence of r.v's and for $x < w, w = \sup\{u; P(X_1 \leq u) < 1\}$, let $q_n = q_n(x) = P(\max(X_1, \dots, X_n) \leq x)$. For any x such that $P(X_1 > x) = 1 - q_1 \leq 0.025$ and any integer $n > 3$ such that $3.3n(1 - q_1)^2 \leq 1$, we have*

$$\left| \frac{q_n - \frac{(2q_1 - q_2)}{(1 + q_1 - q_2 + 2(q_1 - q_2)^2)^n}}{q_n} \right| \leq 3.3n(1 - q_1)^2 \quad (12)$$

According to the Haiman theorem, we need to construct a stationary 1-dependent sequence. Supposing that

$N_1 = Km_1$ and $N_2 = Lm_2$, where K and L are positive integers, we have

$$Z_k = \max_{\substack{(k-1)m_1 < t \leq km_1 \\ 0 < s \leq (L-1)m_2}} v_{ts}, k = 1, 2, \dots, K-1 \tag{13}$$

$\{Z_k\}_{k=1, \dots, K-1}$ is a 1-dependent stationary sequence and $P(S \leq n) = P(\max_{k=1, \dots, K-1} \{Z_k\} \leq n)$. Let $Q_2 = P(Z_1 \leq n)$ and $Q_3 = P(Z_1 \leq n, Z_2 \leq n)$. Then, if $1-Q_2 \leq 0.025$, we can get approximation from Haiman theorem

$$P(S \leq n) \approx (2Q_2 - Q_3)[1 + Q_2 - Q_3 + 2(Q_2 - Q_3)^2]^{-(K-1)} \tag{14}$$

with an error of about $3.3(K-1)(1-Q_2)^2$. To evaluate (15), one needs approximations for Q_2 and Q_3 . Hence, the question is transformed into evaluating Q_2 and Q_3 . We may apply Theorem 1 again considering the two sequences of random variables defined by

$$Y_l = \max_{\substack{0 < t \leq m_1 \\ (l-1)m_2 < s \leq lm_2}} v_{ts}$$

and

$$Z_l = \max_{\substack{0 < t \leq 2m_1 \\ (l-1)m_2 < s \leq lm_2}} v_{ts}, l = 1, 2, \dots, L-1$$

which are also stationary and 1-dependent. Put $Q_{22} = P(Y_1 \leq n)$, $Q_{23} = P(Y_1 \leq n, Y_2 \leq n)$, $Q_{32} = P(Z_1 \leq n)$ and $Q_{33} = P(Z_1 \leq n, Z_2 \leq n)$. We have

$$\begin{aligned} Q_{22} &= P(S(m_1, m_2, 2m_1, 2m_2) \leq n), \\ Q_{23} &= P(S(m_1, m_2, 2m_2, 3m_2) \leq n), \\ Q_{32} &= P(S(m_1, m_2, 3m_1, 2m_2) \leq n), \\ Q_{33} &= P(S(m_1, m_2, 3m_1, 3m_2) \leq n). \end{aligned}$$

Then, if $1-Q_{22} \leq 0.025$ and $1-Q_{32} \leq 0.025$, we can still get the approximations from Theorem 1.

$$Q_2 \approx (2Q_{22} - Q_{23})[1 + Q_{22} - Q_{23} + 2(Q_{22} - Q_{23})^2]^{-(L-1)} \tag{15}$$

with an error of about $3.3(L-1)(1-Q_{22})^2$ and

$$Q_3 \approx (2Q_{32} - Q_{33})[1 + Q_{32} - Q_{33} + 2(Q_{32} - Q_{33})^2]^{-(L-1)} \tag{16}$$

with an error of about $3.3(L-1)(1-Q_{32})^2$. Assuming that $L \leq K$ and substituting (17) and (16) into (15), we can get the final expression we need.

The total error on the resulting approximation of $P(S \leq n)$ is bounded by about

$$E_{app} = 3.3(L-1)(K-1)((1-Q_{22})^2 + (1-Q_{32})^2 + (L-1)(Q_{22} - Q_{23})^2). \tag{17}$$

The exact formulas for $Q_{uv}, u, v \in \{2, 3\}$ is hard to be obtained. Thus, we can use Monte Carlo simulation to evaluate these quantities. The final expression can be given by

$$P(S \geq k) = 1 - P(S < k) = 1 - P(S \leq k-1) \tag{18}$$

where $P(S \leq k-1)$ can be approximated by (15).

3.4 The SSLV with variable-step-parameter

The traditional scan statistics is a kind of continuous scan. Disjoint-window scan statistics means the MA travels across the ROI and scans the area using no overlapping windows. In [12], the authors investigate the disjoint-window test and compare its performance with the scan statistics. Obviously, the scan statistics overwhelms the disjoint-window, and its performance is more stable. However, the disjoint-window can shorten scan period. In this section, we will introduce a variable-step-parameter for the SSLV. In the process of scan, the MA makes a choice for the next start position according to the result of the current scan. Since the detection probability is based on the distance between the target and sensors, sensors near the target have a higher probability of detecting the target. If the result of detection is small, we can magnify the value of step to avoid the redundant scan especially when the target is absent. The variable step is given by

$$step = \max \left\{ \left\lfloor \left(1 - \frac{v_{ts}}{k}\right) f_{ov} \right\rfloor, 1 \right\} \tag{19}$$

The scan region can be a rectangular region given by $R(i_1, i_2) = [i_1 h_1, (i_1 + m) h_1] \times [i_2 h_2, (i_2 + m) h_2]$. Assuming $R(i_1, i_2)$ is the scan region at the current time, then the next scan region is $R[(i_1 + step) h_1, (i_1 + step + m) h_1] \times [i_2 h_2, (i_2 + m) h_2]$, $i_1 + step \leq N - m + 1$. We only introduce the step at one-dimensional field for better performance. Whereas, the global false alarm rate can still be evaluated by (15).

4 Application of the SSLV in distributed system

In this section, we apply the SSLV into a particular situation and provide a detailed description concerning observation model, local vote decision model, false alarm probability at the MA, and the optimal M_x .

4.1 Observations model

In this section, we will present the observation model depicted in Fig. 1. All sensors are assigned in a grid pattern

in Fig. 1a and in a random pattern in Fig. 1c. We consider the two-dimensional field is a square of areas b^2 . The number of total sensors can be expressed as M . (x_s, y_s) for $s = 1, \dots, M$ present the coordinates of sensor s . The coordinate of each sensor is known. Noises at local sensors are i.i.d and follow the Gaussian distribution with zero mean and variance σ_w^2 .

$$w_s \sim N(0, \sigma_w^2) \quad s = 1, \dots, M \quad (20)$$

We design each sensor s to decide between the following hypotheses

$$\begin{aligned} H_0 : r_s &= w_s \\ H_1 : r_s &= y_s + w_s \end{aligned} \quad (21)$$

where r_s is the received signal at sensor s . Sensors make their decisions according to the value of r_s . $y_s = a_s/d_s$, and a_s is i.i.d which follows the Gaussian distribution with zero mean and variance σ^2 (σ^2 represents the power of the signal that is emitted by the target at distance $d_s = 1\text{m}$), and d_s is the Euclidean distance between the local sensor s and the target

$$d_s = \sqrt{(x_s - x_t)^2 + (y_s - y_t)^2} \quad (22)$$

and (x_t, y_t) are the unknown coordinates of the target. Sensors near the target receive more signals than those far away. Receiving more signals means higher probability of detection. In our simulations, we assume that the location of the target follows a uniform distribution, and all local sensors make their judgments by using the same threshold τ . According to the Neyman-Pearson lemma [21], the local sensor-level false alarm rate and probability of detection can be respectively obtained by

$$p_{fa} = 2Q\left(\sqrt{\frac{\tau}{\sigma_w^2}}\right) \quad (23)$$

$$p_{ds} = 2Q\left(\sqrt{\frac{\tau}{\sigma_w^2 + \frac{\sigma^2}{d_s^2}}}\right) \quad (24)$$

where $Q(x) = \int_x^\infty 1/\sqrt{2\pi} e^{-\xi^2/2} d\xi$ is the unit Gaussian exceedance function.

4.2 Local vote decision model

We divide the ROI into M (the total number of sensors) little sub-squares. The location of the sensor inside each small sub-square is known. Let $h = b/N'$, where N' satisfies $N'^2 = M$, and we divide the square of area b^2 into M cells so that each cell of area h^2 contains only one sensor. Let us denote the cell $[ih, (i + 1)h] \times [jh, (j + 1)h]$ by $c(i, j)$. We define X'_{ij} as the binary data from the local sensor s inside $c(i, j)$ with $0 \leq i \leq N'-1$ and $0 \leq j \leq N'-1$.

If sensors are deployed along a regular grid, sensors at the vertex of a square can be selected as the neighbors for

local vote algorithm. The number of the neighbors is fixed. Each sensor contains nine neighbors (including itself) in our simulation. When sensors are randomly deployed in the field, sensors within a fixed distance can be selected as neighbors for local vote algorithm. The number of the neighbors is not fixed in this version. Every sensor receives the decision from its neighbors ignoring the sensors on the edge of the field. If the sum of these decisions exceeds the given threshold M_x , the sensor makes the decision of target presence. After the local vote decision, the local sensor-level false alarm rate and probability of detection are given by

$$p'_{fa} = \sum_{n=M_x}^{M_i} \binom{M_i}{n} p_{fa}^n (1 - p_{fa})^{M_i - n} \quad (25)$$

$$p'_{ds} = \sum_{n=M_x}^{M_i} \binom{M_i}{n} p_{ds}^n (1 - p_{ds})^{M_i - n} \quad (26)$$

where M_i is the number of neighbors and M_x is the pre-set threshold. We define $X_{i,j}$ as the binary data from the local sensor s inside $c(i, j)$ with $1 \leq i \leq N$ and $1 \leq j \leq N$ after local vote where $N = N' - 2$. We observe that for each $1 \leq i \leq N$, the sequence $(X_{i,j})_{1 \leq j \leq N}$ is c -dependent, and for each $1 \leq j \leq N$, the sequence $(X_{i,j})_{1 \leq i \leq N}$ is also c -dependent, where $c = 2$ in our simulation. c has relations with the choice of local vote algorithm. In (15), we construct a 1-dependent sequence to evaluate global false alarm probability. Only when $c \leq m$ can we guarantee the sequence is 1-dependent.

4.3 False alarm probability at the MA

The binary data from local sensors can be expressed as $I_s = \{0, 1\}$ ($s = 1, \dots, M'$). M' represents the number of all sensors except those on the edge of the field. When there is a detected target, I_s takes the value 1; otherwise, it takes the value 0 after the local vote. It is easy to verify that

$$\sum_{s=1}^{M'} I_s = \sum_{j=1}^N \sum_{i=1}^N X_{ij} \quad (27)$$

The MA travels across the area and sequentially collects the local binary decisions from sensors located inside its agent region, which we consider to be squares of size $f_{ov}h$. f_{ov} is the size of the window. The sequential fusion rule at the MA for $1 \leq i_1 \leq N - f_{ov} + 1$ and $1 \leq i_2 \leq N - f_{ov} + 1$ is given by

$$\begin{cases} v_{i_1, i_2} \geq k \Rightarrow \text{decide } H_1 \\ \text{otherwise} \Rightarrow \text{MA continues to scan} \end{cases} \quad (28)$$

where $v_{i_1, i_2} = \sum_{j=i_2}^{i_2 + f_{ov} + 1} \sum_{i=i_1}^{i_1 + f_{ov} + 1} X_{ij}$. At the MA, the probability of global false alarm PF is

$$PF = P(S_{f_{ov} \times f_{ov}} \geq k | H_0) \quad (29)$$

We note that (30) can be evaluated using the approximation as in (15) after substituting f_{ov} with m and p_{fa} with α .

In Fig. 2, we plot the global probability of false alarm *PF* for the MA versus the local probability of false alarm p_{fa} for the sensor. The curves obtained by using the SSLV approximation in (15) and simulations (based on 5000 Monte Carlo runs) are plotted. Fig. 2 shows that the approximation in (15) is very accurate.

4.4 Optimal M_x

Obviously, the global probabilities of false alarm and detection have relations with the value of M_x . In this section, we are looking for the optimal M_x and trying to show the existence of optimal value. The choice of M_x must maximize the probability of detection at the given global false alarm rate so the expression can be written as

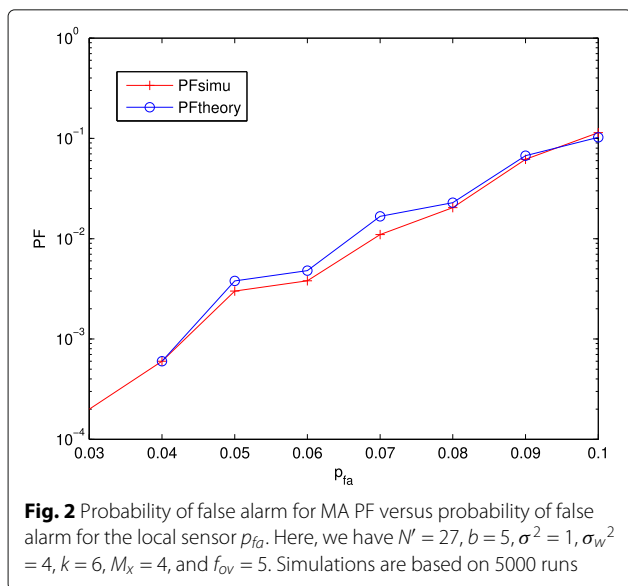
$$\max_{M_x: PF = \alpha} PD \tag{30}$$

where *PD* is the probability of detection at the MA. For the given exact value of α , it is hard to confirm the related parameter k according to α because it involves discrete distributions. To solve this constrained optimization problem in (31), we should use a randomized test [22]. By defining α_1 and α_2 as follows

$$\begin{aligned} P(S_{fov \times fov} \geq k_1) &\approx \alpha_1 < \alpha \\ P(S_{fov \times fov} \geq k_1 - 1) &\approx \alpha_2 > \alpha, \end{aligned} \tag{31}$$

we use a ‘‘coin-flip’’ decision with probability

$$\begin{aligned} k &= k_1 && \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1} > 1/2 \\ k &= k_1 - 1 && \text{otherwise.} \end{aligned} \tag{32}$$



We can confirm the parameter k according to the shortest distance from α to α_1 and α_2 .

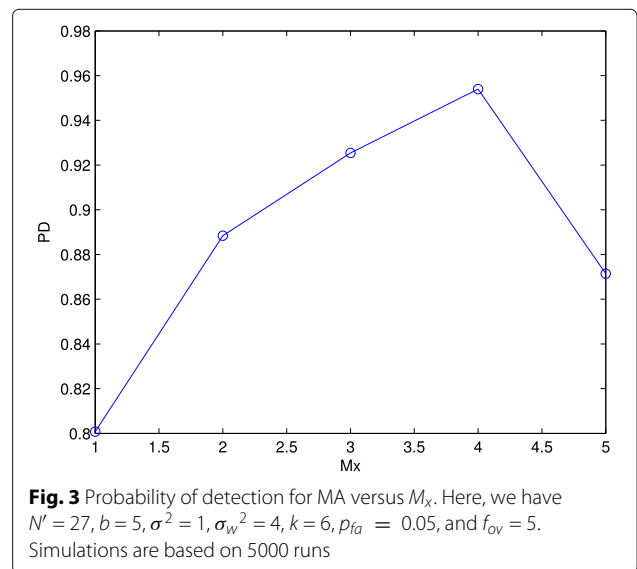
When an M_x and the exact global probability of false alarm is given, we can get the corresponding k . The global probability of detection can be confirmed by k . In Fig. 3, combining with (3) and (30), we plot the global probability of detection versus M_x . α is set to be 0.1. As shown in Fig. 3, there does exist an optimal M_x that maximizes the probability of detection *PD* for the MA at the given condition. By employing this optimum M_x , a significant improvement in *PD* can be achieved. In our simulation, with the increase of M_x , *PD* increases as well. When $M_x = 4$, *PD* reaches the maximum. After that, *PD* decreases with the increase of M_x . The optimal M_x has relations with other parameters and is different in different environments. However, under the given condition, there is indeed an optimal value that maximizes *PD*.

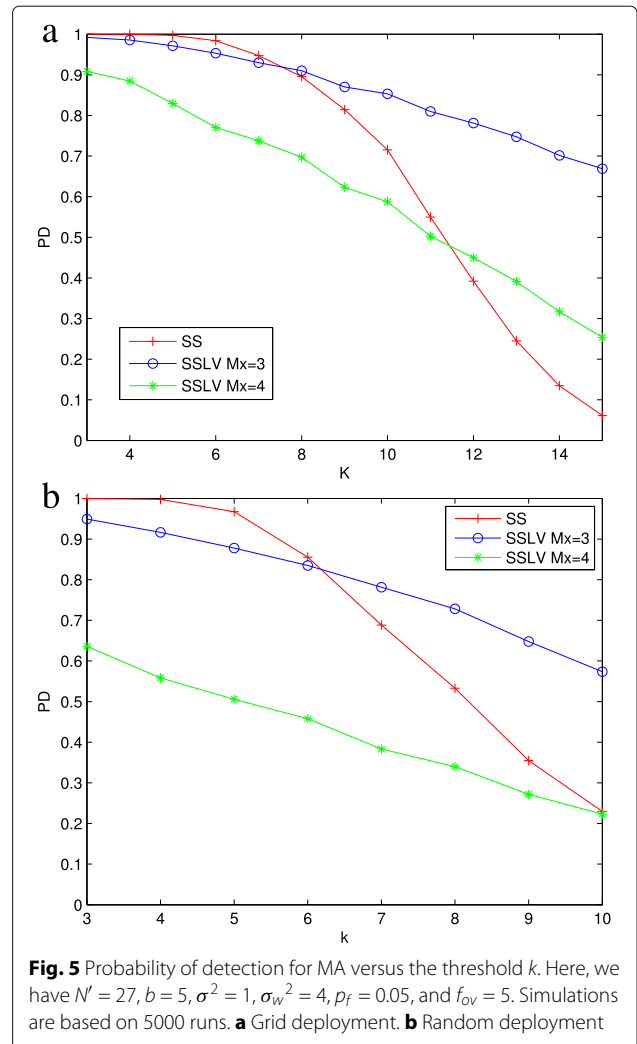
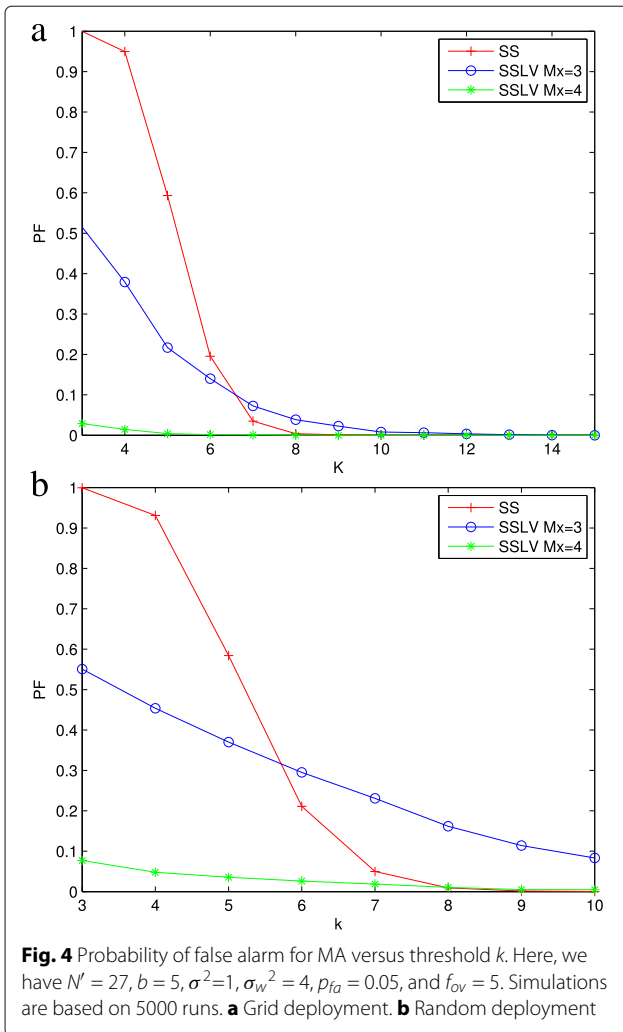
From the perspective of the theory, the increase of M_x decreases the value of k for fixed α . The decrease of k can increase *PD*, meanwhile, the increase of M_x can decrease p'_{ds} from (27). The decrease of p'_{ds} constrains the increase of *PD*, and *PD* mainly relies on p'_{ds} . The decrease of k can compensate the influence which is brought by the decrease of p'_{ds} at the beginning. Hence, *PD* shows the unimodal characteristic.

5 Performance analysis

After all above analysis, we should compare the SSLV with the scan statistics and find out the difference in performance between them. Numerous simulations and analysis are given in this section.

In Fig. 4, we plot the global probability of false alarm for the MA versus the threshold k under different deployments. In Fig. 4a, sensors are deployed along the regular





grid, and $M_i = 9$. In Fig. 4b, sensors are randomly deployed in the field, and the neighborhood distance is set to be 0.1. We can see from Fig. 4 that with the increase of M_x , the global probability of false alarm decreases significantly. What we need is to get lower false alarm rate. There is a crossing point in Fig. 4, which means the scan statistics has the same global probability of false alarm with the SSLV at that point. Before that critical point, the SSLV gets a lower PF than the scan statistics. After that, the scan statistics overwhelms the SSLV. It is because the local vote increases the count of sensors that has detected event compared with the scan statistics. This critical point is not an integer and does not exist in fact. However, the nearest two positive integers of crossing point are the practical key points.

In Fig. 5, we plot the global probability of detection for the MA versus the threshold k . From Fig. 5, we know that PD of the SSLV does not always overwhelm the scan statistics. However, it shows that for the large value of the threshold k , the SSLV performs better than the scan

statistics. The larger threshold k means the smaller global probability of false alarm is demanded. It is related to the value of M_x . With the increase of M_x , we get lower false alarm rate; however, the probability of detection decreases as well. Hence, it is important to select the value of M_x according to the various detection environments. At the given condition, we can use the method in Section 3.4 to evaluate the optimal M_x . Overall, the SSLV can substantially decrease the probability of false alarm and improve the global probability of detection compared with the scan statistics.

In Fig. 6, we plot the global probability of detection for the MA versus σ^2 (power of the signal that is emitted by the target at a distance $d_s = 1$ m) at different local false alarm rates. The increase of σ^2 enables more sensors to sense the signal, and the local probability of detection increases as well. The local probability of detection affects the global probability of detection. Simulations are based on 5000 runs. From Fig. 6, we know the probability of detection increases with the increase of signal strength.

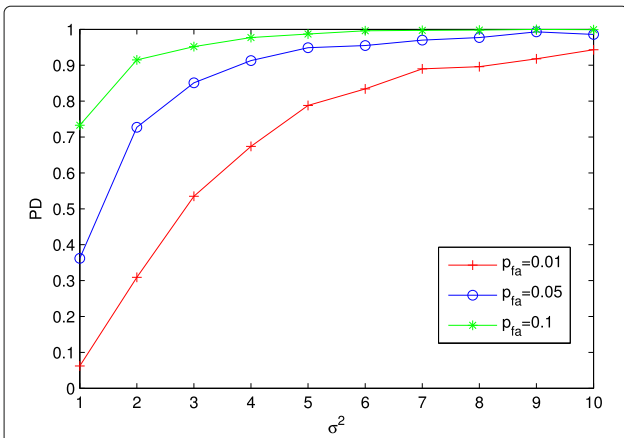


Fig. 6 Probability of detection for MA versus σ^2 . Here, we have $N' = 27$, $b = 5$, $M_x = 3$, $\sigma_w^2 = 6$, and $f_{ov} = 5$. Simulations are based on 5000 runs

In Fig. 7, we plot the global probability of detection versus σ^2 for different methods. Figure 7 illustrates the SSLV has higher PD than scan statistics when the SNR is low. It means our SSLV is more suitable for the tough environment. When the SNR is high, the advantage of the SSLV is not evident.

After introducing a variable-step-parameter into the SSLV, we should figure out its influence on the SSLV. Figure 8 presents the receiver operating characteristic curves (ROC) of the SSLV and the SSLV with the variable-step-parameter. From Figure 8, we can see that the variable-step-parameter has little negative effect on the performance of detection. The local vote makes the distribution of sensors report event so concentrated that we can use this parameter without worry. Figure 9 shows the scan times of different methods to show the advantage

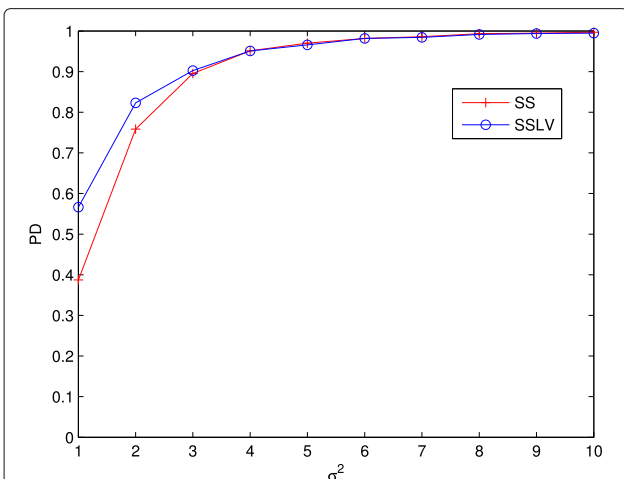


Fig. 7 Probability of detection for MA versus σ^2 . Here, we have $N' = 27$, $b = 5$, $\sigma_w^2 = 4$, $k = 8$, $p_{fa} = 0.05$, and $f_{ov} = 5$. Simulations are based on 5000 runs

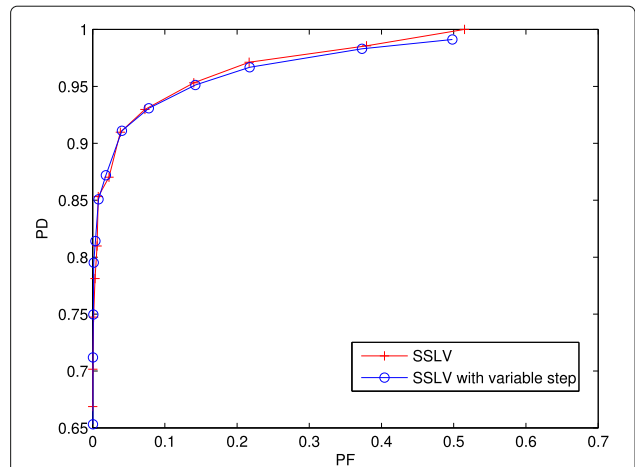


Fig. 8 ROC of SSLV and SSLV with variable step. Here, we have $N' = 27$, $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $p_{fa} = 0.05$, $k = 6$, $M_x = 3$, and $f_{ov} = 5$. Simulations are based on 5000 runs

of variable-step-parameter. When the target is absent, the SSLV with variable step significantly decreases the times of scan, meanwhile, when the target is present, the SSLV with variable step reduces the times of scan to some extent. In our simulation, the target is placed at the center of the field.

6 Conclusions

This paper has introduced the SSLV algorithm specially designed to work with target detection in low SNR condition. The correlation between sensors and the expression for global false alarm ratio after the local vote have been described in detail. Moreover, based on the SSLV, the SSLV with variable-step-parameter has been proposed. The two algorithms have been examined in simulation

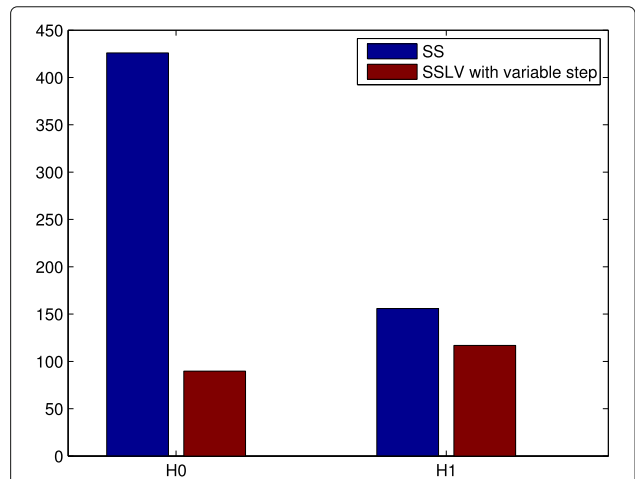


Fig. 9 Scan times versus different hypotheses. Here, we have $N' = 27$, $b = 5$, $\sigma^2 = 1$, $\sigma_w^2 = 4$, $k = 6$, $M_x = 3$, $p_{fa} = 0.05$, and $f_{ov} = 5$. The target is placed at the center of field under the H1 hypothesis

studies which revealed that they produce similar detection accuracies, but the SSLV with variable step method is substantially faster during once scan cycle. Nevertheless, there are some potential research topics which will be further discussed. Firstly, it is evident that getting the optimal M_x from the simulation is not the optimal method and a new expression for the optimal M_x should be deduced. Furthermore, the variable-step-parameter for the SSLV can be extended to two-dimensional without losing any detection performance.

Acknowledgements

This work was supported in part by the Overseas Academic Training Funds, University of Electronic Science and Technology of China (OATF, UESTC) (Grant No.201506075013), and the Program for Science and Technology Support in Sichuan Province (Grant Nos. 2014GZ0100 and 2016GZ0088).

Authors' contributions

JHL and QW proposed the algorithm and carried out the simulations. JHL and QW analyzed the experimental results. JHL gave the critical revision and final approval. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 December 2016 Accepted: 21 April 2017

Published online: 03 May 2017

References

1. Z Chair, PK Varshney, Optimal data fusion in multiple sensor detection systems. *IEEE Trans. Aerosp Electron Syst.* **22**(1), 98–101 (1986)
2. R Niu, PK Varshney, MH Moore, D Klamer, *Decision fusion in a wireless sensor network with a large number of sensors. presented at the 7th IEEE Int. Conf. Information Fusion (ICIF)*. (Stockholm, Sweden, 2004)
3. R Niu, P Varshney, Performance analysis of distributed detection in a random sensor field. *IEEE Trans. Signal Process.* **56**(1), 339–349 (2008)
4. D Marco, Y-H Hu, Distance-based decision fusion in a distributed wireless sensor network. *Telecommun Syst.* **26**(2–4), 339–350 (2004)
5. SH Javadi, A Peiravi, Fusion of weighted decisions in wireless sensor networks. *IET Wirel Sens. Syst.* **5**(2), 97–105 (2015)
6. N Katenka, E Levina, G Michailidis, Local vote decision fusion for target detection in wireless sensor networks. *IEEE Trans. Signal Process.* **56**(1), 329–338 (2008)
7. MS Ridout, An improved threshold approximation for local vote decision fusion. *IEEE Trans. Signal Process.* **61**(5), 1104–1106 (2013)
8. T Tango, K Takahashi, A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. *Stat. Med.* **31**(30), 4207 (2012)
9. J Fu, W Lou, *Distribution theory of runs and patterns and its applications*. (World Scientific, Singapore, 2003)
10. DU Pfeiffer, KB Stevens, Spatial and temporal epidemiological analysis in the Big Data era. *Prev. Vet. Med.* **122**(1–2), 213–220 (2015)
11. C Teljeur, A Kelly, M Loane, et al, *Using scan statistics for congenital anomalies surveillance: the EUROCAT methodology*, vol. 30, (2015), p. 1165
12. M Guerriero, P Willett, J Glaz, Distributed target detection in sensor networks using scan statistics. *IEEE Trans. Signal Process.* **57**(7), 2629–2639 (2009)
13. TL Wu, J Glaz, A new adaptive procedure for multiple window scan statistics. *Comput. Stat. Data Anal.* **82**(82), 164–172 (2015)
14. BJ Reich, Multiple window discrete scan statistic for higher-order Markovian sequences. *J. Appl. Stat.* **42**(8), 1–16 (2015)
15. X Wang, J Glaz, Variable window scan statistics for normal data. *Commun. Statistics-Theory Methods.* **43**, 2489–2504 (2014)
16. MV Boutsikas, M Koutras, Reliability approximations for Markov chain imbeddable systems. *Method. Comput. Appl. Probab.* **2**, 393–412 (2000)
17. WC Lee, Power of discrete scan statistics: a finite Markov chain imbedding approach. *Methodol. Comput. Appl. Probab.* **17**(3), 833–841 (2015)
18. A Amarioarei, C Preda, Approximations for two-dimensional discrete scan statistics in some block-factor type dependent models. *J. Stat. Plann. Infer.* **151**(3), 107–120 (2014)
19. G Haiman, C Preda, Estimation for the distribution of two-dimensional discrete scan statistics. *Methodol. Compu. Appl. Probab.* **8**(373), 382 (2006)
20. G Haiman, 1-dependent stationary sequences for some given joint distributions of two consecutive random variables. *Methodol. Comput. Appl. Probab.* **14**, 445–458 (2012)
21. HV Poor, An introduction to signal detection and estimation. *Springer Texts Electr. Eng.* **333**(1), 127–139(13) (1998)
22. S Kotz, DL Banks, CB Read, *Read, Encyclopedia of statistical sciences*. (Wiley-Interscience, New York, 1997)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com