

A Comparison of Information Functions and Search Strategies for Sensor Planning in Target Classification

Guoxian Zhang, *Member, IEEE*, Silvia Ferrari, *Senior Member, IEEE*, and Chenghui Cai, *Member, IEEE*

Abstract—This paper investigates the comparative performance of several information-driven search strategies and decision rules using a canonical target classification problem. Five sensor models are considered: one obtained from classical estimation theory and four obtained from Bernoulli, Poisson, binomial, and mixture-of-binomial distributions. A systematic approach is presented for deriving information functions that represent the expected utility of future sensor measurements from mutual information, Rènyi divergence, Kullback–Leibler divergence, information potential, quadratic entropy, and the Cauchy–Schwarz distance. The resulting information-driven strategies are compared to direct-search, alert–confirm, task-driven (TS), and log-likelihood-ratio (LLR) search strategies. Extensive numerical simulations show that quadratic entropy typically leads to the most effective search strategy with respect to correct-classification rates. In the presence of prior information, the quadratic-entropy-driven strategy also displays the lowest rate of false alarms. However, when prior information is absent or very noisy, TS and LLR strategies achieve the lowest false-alarm rates for the Bernoulli, mixture-of-binomial, and classical sensor models.

Index Terms—Classification, detection, information driven, information theory, management, optimal, planning, search, sensor, strategy, target.

I. INTRODUCTION

IN MANY applications, the set of all measurements that can be acquired by a sensor significantly exceeds its available time and processing capabilities [1]. Therefore, sensor planning is used to select the best measurement sequence in order to optimize the sensor performance [2]–[4]. A basic difficulty consists of assessing the sensor performance prior to obtaining the measurement sequence [5]–[7]. In most sensor applications, there exist no closed-form representations for the expected rate of correct target classification or rate of false alarms. Target classification can be reduced to the problem of estimating one or more random variables, referred to as target state, from partial or imperfect measurements [8]. Therefore, the sensor

performance depends on the amount of information, or lack thereof, associated with the target state. In this case, the utility of future measurements may be represented by their expected information value.

Several information theoretic functions have been proposed in the literature to assess the information value of an available set of measurements. Relative entropy was used in [9] to solve a multisensor–multitarget assignment problem and in [10] and [11] to manage agile sensors with Gaussian models for target detection and classification. Shannon entropy and the Mahalanobis distance were applied and compared in [7] for sensor selection in ad hoc sensor networks. Shannon entropy was also used in [12] and [13] for tracking a moving target using a Kalman filter. An approach based on mutual information was presented in [14] for adjusting the parameters of a camera in an object recognition application. However, there presently exists no general approach for computing information theoretic functions prior to obtaining the sensor measurements.

Due to the significant differences between sensor data and applications, little work has been done on comparing information functions on canonical sensor problems and on generalizing the results beyond a particular sensor type and application. The comparative study closest to the one presented in this paper pertains to the optimal choice of the α parameter in the Rènyi divergence [11]. It was determined in [11] that the optimal value of α for a multitarget tracking application involving a simulated moving-target-indicator sensor is 0.5. An empirical study showing that task-driven (TS) approaches may slightly outperform information-driven approaches was recently conducted in [15] using a synthetic-aperture-radar model with two modes of operation and eight possible target locations or cells. However, the study in [15] only considered one sensor model and the Rènyi divergence function.

This paper presents a general approach for estimating the expected information value using objective functions derived from mutual information, Rènyi divergence, Kullback–Leibler (KL) divergence, and quadratic entropy, as well as using two functions known as information potential and Cauchy–Schwarz (CS) distance, which have been recently used for blind-source separation, feature extraction, and machine learning in [16]–[18]. Then, a greedy information-driven strategy selects the measurement sequence that maximizes the chosen information function at every time step. Other effective search strategies that have been proposed in the literature and are implemented in this paper are direct search (DS), alert–confirm (AS) search, TS or Bayes-risk search, and log-likelihood-ratio (LLR) search [15], [19], [20].

Manuscript received December 19, 2010; revised July 9, 2011; accepted July 14, 2011. Date of current version December 7, 2011. This work was supported by the National Science Foundation under Grants CAREER 0448906 and ECS 0823945. This paper was recommended by Associate Editor R. Lynch.

G. Zhang was with the Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708 USA. He is now with MicroStrategy, Inc., Tysons Corner, VA 22182 USA (e-mail: zhangguoxian@gmail.com).

S. Ferrari is with the Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708 USA (e-mail: sferrari@duke.edu).

C. Cai was with the Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708 USA. He is now with Opera Solutions, LLC, New York, NY 10038 USA (e-mail: caichenghui@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCB.2011.2165336

In order to provide a comparative study that is easily validated and generalized, this paper implements the aforementioned search strategies on five simulated sensor models obtained from Bernoulli, Poisson, binomial, and mixture-of-binomial distributions and a noisy power law. The same approach can be applied to sensor data by first obtaining a parametric sensor model, as shown in [21] and [22]. Each strategy is evaluated using the Neyman–Pearson, maximum likelihood, and maximum *a posteriori* decision rules. The effects of prior information are investigated using prior distributions with three levels of information-to-noise ratio. The simulation results and growth curve model (GCM) analysis [23]–[26] show that quadratic entropy typically leads to the most effective search strategy with respect to correct-classification rates. In the presence of prior information, the quadratic-entropy-driven strategy displays the lowest false-alarm rates. However, when prior information is absent or very noisy, TS and LLR strategies achieve the lowest false-alarm rates for Bernoulli, mixture-of-binomial, and classical sensor models.

II. BACKGROUND ON INFORMATION THEORETIC FUNCTIONS

Information theoretic functions have been used to evaluate the information value of sensor measurements in a wide range of applications. As pointed out in [7], a natural choice for measuring information value is entropy. Shannon entropy measures the uncertainty of a discrete and random variable X , with finite range \mathcal{X} , from its probability mass function (PMF) $p_X(x) = \Pr(\{X = x\})$ for $x \in \mathcal{X}$ and is defined as

$$H(X) = - \sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x). \quad (1)$$

Since the computation of (1) requires knowledge of the PMF, it cannot be used to compute the information value because the posterior PMF of X , or *belief state*, is unknown before the measurements are obtained [7]. Furthermore, the optimization of entropy-based functions is usually ill posed because entropy is nonadditive and myopic, i.e., it does not consider the effects of prior measurements on those that are performed subsequently [2]. Another shortcoming of (1) is that it is not a true metric because it is nonsymmetric, and it does not satisfy the triangle inequality [5]. In the remainder of this paper, the PMF $p_X(x)$ is represented by the shorthand notation $p(x)$. Also, uppercase characters are used to denote discrete random variables, and lower case characters are used to denote real numbers, such as the numerical values of random variables.

The Rény information or α -divergence has been proposed in [27] as a means for quantifying the change in the belief state brought about by the sensor measurements. It is based on the Rény's entropy of order α , defined as

$$H_{R_\alpha}(X) = \frac{1}{1-\alpha} \log_2 \sum_{x \in \mathcal{X}} p^\alpha(x) \quad (2)$$

which relates to (1) through the properties $\lim_{\alpha \rightarrow 1} H_{R_\alpha}(X) = H(X)$ and $H_{R_\alpha}(X) \geq H(X) \geq H_{R_\beta}(X)$ if $1 > \alpha > 0$ and $\beta > 1$. Let the current belief state be represented by a PMF $q(x)$, and suppose that a posterior distribution $p(x)$ is expected

as a result of a sensor decision pertaining to the sensor mode and measurement sequence. Then, the α -divergence

$$D_\alpha(p||q) = \frac{1}{\alpha-1} \log_2 \sum_{x \in \mathcal{X}} p^\alpha(x) q^{1-\alpha}(x) \quad (3)$$

can be viewed as a measure of the difference between the two PMFs $q(x)$ and $p(x)$, where the α parameter represents the emphasis placed on the degree of differentiation between the tails of the distributions. In [11], the value $\alpha = 0.5$ was found to be optimal for representing the information value in multitarget tracking applications in which the two PMFs $q(x)$ and $p(x)$ are close. In the limit of $\alpha \rightarrow 1$, (3) can be shown to reduce to the KL divergence or relative entropy, defined as

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)} \quad (4)$$

which was first applied to sensor planning in the seminal work of Kastella [10]. Like entropy, however, the Rény information and KL divergence functions are nonadditive and nonsymmetric and do not satisfy the triangle inequality.

An additive, symmetric, and nonmyopic function based on conditional mutual information was recently proposed by the authors in [28] and successfully applied to multitarget detection and classification in [5]. Mutual information is a measure of the information content of one random variable about another random variable [29]. The conditional mutual information of two random variables X and Z , given Y , is defined as

$$\begin{aligned} I(X; Z|Y) &= H(X|Y) - H(X|Z, Y) \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x, y, z) \log_2 \frac{p(x, z|y)}{p(x|y)p(z|y)} \end{aligned} \quad (5)$$

and represents the reduction in uncertainty in X due to the knowledge of Z , when Y is given. $H(X|Y)$ denotes the conditional entropy of X , given Y , and is defined in [29, p. 16]. While (5) does not obey the triangle inequality, a related metric can be obtained using the function $d(X; Z|Y) = H(X, Z|Y) - I(X; Z|Y)$, which obeys all properties of metrics. Mutual information can also be shown to be a concave function of $p(x|y)$ for fixed $p(z|x, y)$ [29, p. 31].

Although they have not been previously applied to sensor planning, the information potential and CS distance have been recently introduced in [16]–[18] and shown very effective for blind-source separation, feature extraction, and machine learning. Based on the CS inequality, the CS information function defined as

$$C(p, q) = \log_2 \frac{\sum_{x \in \mathcal{X}} p^2(x) \sum_{x \in \mathcal{X}} q^2(x)}{[\sum_{x \in \mathcal{X}} p(x)q(x)]^2} \quad (6)$$

was proposed as a measure of the difference between two PMFs $p(x)$ and $q(x)$ in [16] and [17] and used therein for blind-source separation and feature extraction. The information function in (6) is based on quadratic entropy, which is obtained from (2) by letting $\alpha = 2$. Quadratic forms are particularly well suited to numerical optimization because they are characterized by

high convergence rates and smooth gradient variations near the minimum. Additionally, in the absence of constraints, they typically do not exhibit multiple stationary points. Therefore, in this paper, quadratic entropy and the information potential, defined as

$$V(X) = \sum_{x \in \mathcal{X}} p^2(x) \quad (7)$$

are also applied to sensor planning. In Section IV, the information theoretic functions reviewed earlier are used to derive closed-form representations for the measurements' expected information value in the target classification problem described in the next section.

III. CANONICAL SENSOR PLANNING PROBLEM

In standard estimation theory, a sensor that obtains a vector of measurements, $Z \in \mathbb{R}^r$, in order to estimate an unknown state vector, $X \in \mathbb{R}^n$, at time k is modeled as

$$Z^k = h(X^k, \lambda^k) \quad (8)$$

where $h : \mathbb{R}^{n \times \varphi} \rightarrow \mathbb{R}^r$ is a deterministic vector function that is possibly nonlinear, and time is discretized and indexed by $k = 1, \dots, f$ [8]. The random parameter vector $\lambda \in \mathbb{R}^\varphi$ represents the sensor characteristics, such as sensor mode, environmental conditions, and sensor noise or measurement errors. In many sensor applications, however, the state, the measurements, and the sensor characteristics also are random variables. Therefore, a more general observation or measurement model that has been proposed in the literature is the joint PMF, $p(Z^k, X^k, \lambda^k)$, which typically can be factorized as follows:

$$p(Z^k, X^k, \lambda^k) = p(Z^k | X^k, \lambda^k) p(X^k) p(\lambda^k) \quad (9)$$

assuming X^k and λ^k are independent random variables [5], [21], [30]. Both models (8) and (9) are time invariant. Thus, in later sections, the superscript k is omitted for brevity. The problem dimensions are chosen as $n = r = 1$ and $\varphi = 3$ because they result in efficient and representative numerical simulations.

The probabilistic model in (9) assumes that Z , X , and λ are discrete random variables with finite ranges \mathcal{Z} , \mathcal{X} , and Λ , respectively. Continuous random variables can be considered using an analogous model based on the joint probability density function [11]. The conditional PMF $p(Z|X, \lambda)$ is obtained from the physical principles underlying the measurement process. The priors $p(X)$ and $p(\lambda)$ are computed from prior environmental information when available or are otherwise assumed to be uniformly distributed. Various sensors, including infrared, ground penetrating radars, and synthetic aperture radars, have been modeled from real data using (9), as shown in [10], [11], [15], [21], and [22].

It is assumed that the sensor is deployed for the purpose of classifying multiple targets confined to a discrete set of cells $\mathcal{K} = \{\kappa_1, \dots, \kappa_c\}$. It is also assumed that there is at most one target in each cell and that targets do not move and do not change over time. The state of the i th cell κ_i , denoted by $X_i \in \mathcal{X}$, represents the presence and classification of a

target in κ_i . Thus, the range \mathcal{X} contains mutually exclusive values representing an empty cell and all possible target types, including a "high-risk" target type denoted by x^r . Then, the sensor planning problem considered in this paper consists of selecting one cell from \mathcal{K} at every time $k = 1, \dots, f$ such that the rate of correct target classification is maximized and the rate of false alarms is minimized.

The measurement variable associated with κ_i , denoted by Z_i , is a discrete random variable with finite range \mathcal{Z} , that is unknown *a priori*. If the sensor obtains the measurement value z^k from κ_i at time k , then $Z^k = Z_i = z^k$, $X^k = X_i$, $\lambda^k = \lambda_i$, and z^k is known thereafter. However, since measurements are imperfect and random, the actual value of X_i remains unknown. Thus, X_i must be estimated, or *classified*, from z^k using the decision rules described in Section VI. Over time, the sensor may obtain multiple measurements from the same cell in order to improve its classification performance. At every time k , the search strategy decides which cell to measure in \mathcal{K} , prior to obtaining z^k , based only on the sensor model and the set of measurements $\mathcal{M}^{k-1} = \{z^1, \dots, z^{k-1}\}$ obtained up to time $(k-1)$.

The sensor performance is characterized by the rates of correct classification and false alarms. In most sensor applications, closed-form representations for these rates as functions of future sensor measurements are not available and therefore cannot be directly optimized over time. Once the measurements are obtained and processed by a decision rule, however, they can be computed as follows. At time k , let $n(k)$ denote the total number of cells with a state value that is estimated correctly and $n_r(k)$ denote the number of cells with the high-risk state value x^r that is estimated correctly. The number of false alarms, $n_{fa}(k)$, represents the number of cells that are empty but are incorrectly declared to contain a target by the decision rule. Then, at any time k , the rate of correct classification is

$$F_c(k) = \frac{n(k)}{c} \quad (10)$$

where c is the total number of cells in \mathcal{K} . The rate of correct classification for high-risk targets at time k is

$$F_r(k) = \frac{n_r(k)}{c_r} \quad (11)$$

where c_r is the number of cells in \mathcal{K} with state value x^r . The frequency of false alarms at time k is

$$F_{fa}(k) = \frac{n_{fa}(k)}{c_e} \quad (12)$$

where c_e is the number of empty cells in \mathcal{K} . At the end of the time interval ($k = f$), average rates, denoted by \bar{F}_c , \bar{F}_r , and \bar{F}_{fa} , can be computed by summing $F_c(k)$, $F_r(k)$, and $F_{fa}(k)$, respectively, over k and normalizing the results by f .

Since the classification and false-alarm rates defined earlier are unknown *a priori*, a search strategy can be adopted to select which cell to measure at every time k , based on the information-value functions presented in the next section.

IV. INFORMATION VALUE FUNCTIONS FOR SENSOR PLANNING

Information theoretic functions are a natural choice for representing the value of information because they measure the absolute or relative information content of probability mass (or density) functions. However, as shown in Section II, computing these functions requires knowledge of the PMFs representing the prior and the posterior belief state. Since, in sensor planning, the posterior belief state is unknown, this section presents a general approach for estimating the information value of a cell or target based on prior sensor measurements and the sensor model.

Based on the problem formulation in Section III, at every time k , the sensor must decide from which cell $\kappa_i \in \mathcal{K}$ to obtain an unknown measurement Z_i , given all past measurements \mathcal{M}^{k-1} . The information value of κ_i can be represented by the change in belief state brought about by Z_i and estimated using the sensor model (9). At time k , the change between the prior belief state, $p(X_i|\mathcal{M}^{k-1}, \lambda_i)$, and the posterior belief state, $p(X_i|Z_i, \mathcal{M}^{k-1}, \lambda_i)$, can be estimated by taking the expectation with respect to Z_i , denoted by \mathbb{E}_{Z_i} . Then, from (3), the information value can be represented by the expected α -divergence, defined as

$$\begin{aligned} \hat{\varphi}_{D_\alpha}(X_i; Z_i|\mathcal{M}^{k-1}, \lambda_i) & \equiv \mathbb{E}_{Z_i} \{D_\alpha [p(X_i|Z_i, \mathcal{M}^{k-1}, \lambda_i) \| p(X_i|\mathcal{M}^{k-1}, \lambda_i)]\} \\ & = \sum_{z_j \in \mathcal{Z}} D_\alpha [p(X_i|Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i) \| p(X_i|\mathcal{M}^{k-1}, \lambda_i)] \\ & \quad \times p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i), \text{ for } \forall i, \text{ and } k=2, \dots, f. \end{aligned} \quad (13)$$

By taking the expectation with respect to Z_i , the measurement value z^k is no longer needed, and $\hat{\varphi}_{D_\alpha}$ can be computed from \mathcal{M}^{k-1} and the sensor model, as explained hereinafter.

In many sensor applications, measurements that are obtained at different time instants can be assumed to be conditionally independent given the target state, i.e., $p(z^{k-1}|X_i, z^{k-2}, \dots, z^1, \lambda_i) = p(z^{k-1}|X_i, \lambda_i)$. As a result, an efficient rule can be derived for updating the belief state iteratively over time using the sensor model (9) in combination with Bayes' rule. Suppose that a measurement z^{k-1} is obtained from cell $\kappa_i \in \mathcal{K}$ at time $(k-1)$. Then, the PMF of X_i given \mathcal{M}^{k-1} and λ_i can be updated by the rule

$$\begin{aligned} p(X_i|\mathcal{M}^{k-1}, \lambda_i) & = p(X_i|z^{k-1}, \mathcal{M}^{k-2}, \lambda_i) \\ & = \frac{p(z^{k-1}|X_i, \mathcal{M}^{k-2}, \lambda_i)p(X_i|\mathcal{M}^{k-2}, \lambda_i)}{p(z^{k-1}|\mathcal{M}^{k-2}, \lambda_i)} \\ & = \frac{p(z^{k-1}|X_i, \lambda_i)p(X_i|\mathcal{M}^{k-2}, \lambda_i)}{\sum_{x_\ell \in \mathcal{X}} p(z^{k-1}|X_i = x_\ell, \lambda_i)p(X_i = x_\ell|\mathcal{M}^{k-2}, \lambda_i)}, \\ & \text{for } \forall i, \text{ and } k = 2, \dots, f \end{aligned} \quad (14)$$

where $p(X_i|\mathcal{M}^{k-2}, \lambda_i)$ is known from the previous time step $(k-2)$ and $p(z^{k-1}|X_i, \lambda_i)$ is known from (9). As additional measurements are obtained at subsequent time steps, (14) can be implemented iteratively by updating the time and cell index accordingly. Finally, the posterior belief inside the expectation

in (13) is computed by applying Bayes' rule for every value $z_j \in \mathcal{Z}$, i.e.,

$$\begin{aligned} p(X_i|Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i) & = \frac{p(z_j|X_i, \lambda_i)p(X_i|\mathcal{M}^{k-1}, \lambda_i)}{\sum_{x_\ell \in \mathcal{X}} p(z_j|X_i = x_\ell, \lambda_i)p(X_i = x_\ell|\mathcal{M}^{k-1}, \lambda_i)}. \end{aligned} \quad (15)$$

The expected discrimination gain (EDG), originally proposed in [10], can be derived by taking the expectation of the KL divergence defined in (4), i.e.,

$$\begin{aligned} \hat{\varphi}_D(X_i; Z_i|\mathcal{M}^{k-1}, \lambda_i) & \equiv \mathbb{E}_{Z_i} \{D [p(X_i|Z_i, \mathcal{M}^{k-1}, \lambda_i) \| q(X_i|\mathcal{M}^{k-1}, \lambda_i)]\} \\ & = \sum_{z_j \in \mathcal{Z}} D [p(X_i|Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i) \| q(X_i|\mathcal{M}^{k-1}, \lambda_i)] \\ & \quad \times p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i). \end{aligned} \quad (16)$$

It can be seen that, by this approach, the EDG can be computed from the same PMFs in (14) and (15) used to compute $\hat{\varphi}_{D_\alpha}$.

As shown in [5], an information value function based on conditional mutual information can be used to represent the reduction in uncertainty in X_i , due to the knowledge of Z_i , when \mathcal{M}^{k-1} and λ_i are given. Based on the definition in (5), computing the conditional mutual information $I(X_i; Z_i|\mathcal{M}^{k-1}, \lambda_i)$ requires computing the entropy $H(X_i|Z_i, \mathcal{M}^{k-1}, \lambda_i)$, which, in turn, requires knowledge of Z_i . Therefore, a function better suited to sensor planning is the expected conditional mutual information, defined as

$$\begin{aligned} \hat{\varphi}_I(X_i; Z_i|\mathcal{M}^{k-1}, \lambda_i) & \equiv \mathbb{E}_{Z_i} \{I(X_i; Z_i|\mathcal{M}^{k-1}, \lambda_i)\} \\ & = H(X_i|\mathcal{M}^{k-1}, \lambda_i) - \mathbb{E}_{Z_i} \{H(X_i|Z_i, \mathcal{M}^{k-1}, \lambda_i)\} \\ & = H(X_i|\mathcal{M}^{k-1}, \lambda_i) - \sum_{z_j \in \mathcal{Z}} H(X_i|Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i) \\ & \quad \times p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i) \end{aligned} \quad (17)$$

where the entropy $H(X_i|Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i)$ is computed from (1), using (15).

The expected CS information function is derived from (6) by letting

$$\begin{aligned} \hat{\varphi}_C(X_i; Z_i|\mathcal{M}^{k-1}, \lambda_i) & \equiv \mathbb{E}_{Z_i} \{C [p(X_i|Z_i, \mathcal{M}^{k-1}, \lambda_i), p(X_i|\mathcal{M}^{k-1}, \lambda_i)]\} \\ & = \sum_{z_j \in \mathcal{Z}} C [p(X_i|Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i), p(X_i|\mathcal{M}^{k-1}, \lambda_i)] \\ & \quad \times p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i). \end{aligned} \quad (18)$$

The above CS function provides an alternative measure of the expected distance between the prior and the posterior belief state for a cell κ_i , prior to obtaining Z_i .

By viewing the information potential as an alternative measure of the uncertainty associated with a distribution, the

information value of κ_i can be represented by the expected information potential gain derived from (7) and defined as

$$\begin{aligned} \hat{\varphi}_V(X_i; Z_i | \mathcal{M}^{k-1}, \lambda_i) & \\ & \equiv \mathbb{E}_{Z_i} \{V [p(X_i | Z_i, \mathcal{M}^{k-1}, \lambda_i)] - V [p(X_i | \mathcal{M}^{k-1}, \lambda_i)]\} \\ & = \mathbb{E}_{Z_i} \{V [p(X_i | Z_i, \mathcal{M}^{k-1}, \lambda_i)] - V [p(X_i | \mathcal{M}^{k-1}, \lambda_i)]\} \\ & = \sum_{z_j \in \mathcal{Z}} V [p(X_i | Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i)] p(Z_i = z_j | \mathcal{M}^{k-1}, \lambda_i) \\ & \quad - V [p(X_i | \mathcal{M}^{k-1}, \lambda_i)]. \end{aligned} \quad (19)$$

Similarly, the expected quadratic entropy reduction is defined as

$$\begin{aligned} \hat{\varphi}_{R_2}(X_i; Z_i | \mathcal{M}^{k-1}, \lambda_i) & \\ & \equiv \mathbb{E}_{Z_i} \{H_{R_2} [p(X_i | \mathcal{M}^{k-1}, \lambda_i)] \\ & \quad - H_{R_2} [p(X_i | Z_i, \mathcal{M}^{k-1}, \lambda_i)]\} \\ & = H_{R_2} [p(X_i | \mathcal{M}^{k-1}, \lambda_i)] \\ & \quad - \mathbb{E}_{Z_i} \{H_{R_2} [p(X_i | Z_i, \mathcal{M}^{k-1}, \lambda_i)]\} \\ & = H_{R_2} [p(X_i | \mathcal{M}^{k-1})] \\ & \quad - \sum_{z_j \in \mathcal{Z}} H_{R_2} [p(X_i | Z_i = z_j, \mathcal{M}^{k-1}, \lambda_i)] \\ & \quad \times p(Z_i = z_j | \mathcal{M}^{k-1}, \lambda_i) \end{aligned} \quad (20)$$

where H_{R_2} is obtained from (2) by letting $\alpha = 2$.

All of the information functions derived in this section can be computed prior to acquiring the value of Z_i , using the sensor model (9), and the PMFs in (14) and (15).

V. SEARCH STRATEGIES

Information-driven search (IS) strategies select the measurement sequence based on the maximum expected information value. Assume that the expected information value can be represented by one of the functions derived in Section IV, denoted by $\hat{\varphi}$. Then, at every time k , the cell κ_l with index

$$l = \arg \max_i \{ \hat{\varphi}(X_i; Z_i | \mathcal{M}^{k-1}, \lambda_i) \} \quad (21)$$

is selected from \mathcal{K} . After the measurement value $Z_l = z^k$ is obtained, the measurement set is updated by letting $\mathcal{M}^k = \{\mathcal{M}^{k-1}, z^k\}$, and the PMF of X_l is updated using (14). The IS strategy is applied iteratively over time, until $k = f$. DS, AS, TS, and LLR strategies are also implemented for comparison because they have been shown to outperform other approaches, such as the index rule and the sequential-probability-ratio test [19].

DS consists of selecting cells in the order in which they are provided in \mathcal{K} , obtaining only one measurement from each cell [19]. AS is a sensor management approach implemented in multitarget radar systems which consists of selecting cells in the order in which they are provided in \mathcal{K} until an ‘‘alert’’ is obtained. The alert triggers a ‘‘confirm’’ cycle that obtains additional measurements until a desired confidence level (CL), such as a desired signal-to-noise ratio, is achieved or until the probability of detection exceeds that of false alarm [19], [20]. Only when the confirm cycle is completed does the sensor move on

to the next cell. In this paper, the CL for the AS strategy is represented by a desired belief state or posterior PMF, defined as

$$\text{CL}_i \equiv \max_{x_\ell} \{p(X_i = x_\ell | \mathcal{M}^{k-1}, \lambda_i)\} \quad (22)$$

and is required to be equal to 0.9. Therefore, at time k , the sensor obtains a measurement from κ_i if $\text{CL}_i < 0.9$. Otherwise, the sensor moves to κ_{i+1} and obtains measurements from κ_{i+1} until $\text{CL}_{i+1} \geq 0.9$. Although AS can lead to good sensor performance, its key disadvantage is that it cannot be used to minimize time or energy consumption.

TS, or Bayes-risk, search strategies have been recently proposed for applications in which a sensor is deployed to perform a single task, such as to find mines or minimize tracking errors. TS strategies base the management decisions on a heuristic criterion, such as the associated risk level [15]. These applications are considered here by introducing the high-risk state value x^r , which represents a high-risk target class (such as a mine) or target location. Then, the TS strategy selects the cells with the maximum *a posteriori* probability for the value x^r , optimizing the risk associated with declaring $X_i = x^r$ [15]. Since the *a posteriori* probability of x^r is unknown at the time of the sensor decision, it is estimated as follows

$$\hat{P}_r^k(X_i) \equiv \mathbb{E}_{Z_i} [p(X_i = x^r | \mathcal{M}^{k-1}, \lambda_i, Z_i)] \quad (23)$$

similarly to the approach proposed in [15]. Then, at time k , the TS strategy selects the cell κ_l with index

$$l = \arg \max_i \{ \hat{P}_r^k(X_i) \} \quad (24)$$

from \mathcal{K} . Since $\hat{P}_r^k(X_i)$ is independent of Z_i , (24) can be simplified as follows:

$$\begin{aligned} l & = \arg \max_i \{ \mathbb{E}_{Z_i} [p(X_i = x^r | \mathcal{M}^{k-1}, \lambda_i, Z_i)] \} \\ & = \arg \max_i \left\{ \sum_{z_j \in \mathcal{Z}} p(X_i = x^r | \mathcal{M}^{k-1}, \lambda_i, Z_i = z_j) \right. \\ & \quad \left. \times p(Z_i = z_j | \mathcal{M}^{k-1}, \lambda_i) \right\} \\ & = \arg \max_i \left\{ \sum_{z_j \in \mathcal{Z}} \frac{p(X_i = x^r, \mathcal{M}^{k-1}, \lambda_i, Z_i = z_j)}{p(Z_i = z_j | \mathcal{M}^{k-1}, \lambda_i) p(\mathcal{M}^{k-1}, \lambda_i)} \right. \\ & \quad \left. \times p(Z_i = z_j | \mathcal{M}^{k-1}, \lambda_i) \right\} \\ & = \arg \max_i \left\{ \sum_{z_j \in \mathcal{Z}} p(X_i = x^r, Z_i = z_j | \mathcal{M}^{k-1}, \lambda_i) \right\} \\ & = \arg \max_i \{ p(X_i = x^r | \mathcal{M}^{k-1}, \lambda_i) \}. \end{aligned} \quad (25)$$

Then, the TS strategy selects cells in the order of decreasing probability of the high-risk state x^r .

The LLR criterion has been successfully implemented for the evaluation of track formation hypotheses in multiple target tracking applications [20], [31]. When two possible hypothesis can be used to explain the data, the LLR is proportional to the

ratio of the likelihood of the data given the first hypothesis over the likelihood of the data given the second hypothesis. Typically, the likelihood is multiplied by the prior probability of the hypothesis such that the LLR can be computed recursively using Bayes' rule [20]. In this paper, the LLR definition in [20] is extended to nonbinary state variables with a high-risk value x^r as follows:

$$L^k(X_i) \equiv \ln \left[\frac{p(Z_i|X_i=x^r)p(X_i=x^r)}{p(Z_i|X_i \neq x^r)p(X_i \neq x^r)} \right] \\ = \ln \left[\frac{p(Z_i|X_i=x^r)p(X_i=x^r)}{\sum_{x_\ell \in \mathcal{X}, x_\ell \neq x^r} p(Z_i|X_i=x_\ell)p(X_i=x_\ell)} \right]. \quad (26)$$

Then, since a cell κ_i must be selected before Z_i and its likelihoods are known, the LLR search (LLRS) strategy is developed by taking the expectation of (26) and selecting the cell κ_l with the highest expected LLR, as shown in (27) shown at the bottom of the page. It can be seen from (27) that LLRS selects cells in the order of decreasing log ratio of the expected posterior probability of x^r , divided by the expected posterior probability of all other state values in \mathcal{X} .

VI. SENSOR DECISION RULES

The sensor decision rule refers to the criterion used to decide or estimate the state value of a cell after a sensor measurement is obtained. Although this criterion does not affect the search strategy because it is implemented *a posteriori*, it affects the sensor performance. As reviewed in [31, p. 213], a number of decision rules may be employed to decide what value of X_i to accept based on the posterior belief state. Where, the posterior belief state, $p(X_i|\mathcal{M}^k, \lambda_i)$, is computed from (14) using the latest measurement z^k . The Neyman–Pearson rule accepts a state value only if its likelihood ratio is greater than a desired significance level. The minimax and Bayes cost functions utilize user-defined weights to quantify the risk or cost of choosing one state value over the others based on its likelihood and posterior probability, respectively. When these weights are set equal to one, the minimax and Bayes cost functions reduce respectively to the maximum likelihood and maximum *a posteriori* rules implemented in this paper.

The maximum *a posteriori* rule accepts the state value $x^* \in \mathcal{X}$ if its probability is greater than the posterior probability of

any other value given the data, i.e.,

$$\hat{X}_i = x^*, \quad \text{iff} \quad p(X_i = x^*|\mathcal{M}^k, \lambda_i) \geq p(X_i = x_\ell|\mathcal{M}^k, \lambda_i), \\ \forall x_\ell \in \mathcal{X}, \quad x_\ell \neq x^*. \quad (28)$$

When \mathcal{M}^k contains no measurements about a cell κ_i , the prior is used in place of the posterior probability in (28). The maximum-likelihood estimate (MLE) rule accepts the state value $x^* \in \mathcal{X}$ if its likelihood is greater than that of any other value given the data, i.e.,

$$\hat{X}_i = x^*, \quad \text{iff} \quad p(z^k|X_i = x^*) \geq p(z^k|X_i = x_\ell), \\ \forall x_\ell \in \mathcal{X}, \quad x_\ell \neq x^*. \quad (29)$$

It can be seen from (29) that the MLE rule can only be applied to cells for which at least one measurement has been obtained.

VII. SIMULATED SENSOR MODELS

The comparative performance of the information functions, search strategies, and decision rules presented in the previous sections is investigated by simulating three types of prior distributions and five sensor models. Four probabilistic sensor models are simulated by generating the joint PMF in (9) from Bernoulli, binomial, Poisson, and mixture-of-binomial distributions. One classical sensor model in the form of (8) is simulated using a nonlinear power law and Gaussian noise. The canonical sensor planning problem in Section III is simulated by generating the set of cells \mathcal{K} with corresponding sensor measurements, as explained in the following sections.

A. Probabilistic Sensor Models

The probabilistic sensor model (9) represents the joint probability of the sensor measurements in \mathcal{Z} , the target state values in \mathcal{X} , and the sensor and environmental characteristics in Λ . In practice, this joint probability distribution is determined from the physical processes underlying the sensor measurements and by the nature of the targets and environmental conditions in the region of interest [21]. In this paper, (9) is simulated by means of well-known probability distributions to conduct a comparative study that is independent of the sensing application. A set \mathcal{K} of $c = c_t + c_e$ cells, with c_t targets and c_e empty cells, is generated by sampling uniformly and independently the range of X_i . The cell state X_i has the range $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$,

$$l = \arg \max_i \mathbb{E}_{Z_i} \left\{ \ln \left[\frac{p(Z_i|X_i=x^r, \mathcal{M}^{k-1}, \lambda_i)p(X_i=x^r|\mathcal{M}^{k-1}, \lambda_i)}{p(Z_i|X_i \neq x^r, \mathcal{M}^{k-1}, \lambda_i)p(X_i \neq x^r|\mathcal{M}^{k-1}, \lambda_i)} \right] \right\} \\ = \arg \max_i \sum_{z_j \in \mathcal{Z}} p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i) \ln \left[\frac{p(Z_i = z_j|X_i=x^r, \mathcal{M}^{k-1}, \lambda_i)p(X_i=x^r|\mathcal{M}^{k-1}, \lambda_i)}{p(Z_i = z_j|X_i \neq x^r, \mathcal{M}^{k-1}, \lambda_i)p(X_i \neq x^r|\mathcal{M}^{k-1}, \lambda_i)} \right] \\ = \arg \max_i \sum_{z_j \in \mathcal{Z}} p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i) \ln \left[\frac{p(Z_i = z_j, X_i=x^r, \mathcal{M}^{k-1}, \lambda_i)p(\mathcal{M}^{k-1}, \lambda_i)}{p(Z_i = z_j, X_i \neq x^r, \mathcal{M}^{k-1}, \lambda_i)p(\mathcal{M}^{k-1}, \lambda_i)} \right] \\ = \arg \max_i \sum_{z_j \in \mathcal{Z}} p(Z_i = z_j|\mathcal{M}^{k-1}, \lambda_i) \ln \left[\frac{p(X_i=x^r|\mathcal{M}^{k-1}, \lambda_i, Z_i=z_j)}{\sum_{x_\ell \in \mathcal{X}, x_\ell \neq x^r} p(X_i=x_\ell|\mathcal{M}^{k-1}, \lambda_i, Z_i=z_j)} \right] \quad (27)$$

where x_1 and x_2 denote two types of empty cells (e.g., clear terrain and clutter), x_3 and x_4 denote two types of targets, and the high-risk value is $x^r = x_4$. The sensor characteristics and environmental conditions in κ_i are represented by the random vector $\lambda_i = [\phi_i \beta_i \gamma_i]^T$. Each random element in λ_i is assumed to have three possible values; thus, the range Λ of λ_i has nine possible values.

Let $\theta \in [0, 1]$ denote a parameter of the distribution that can be viewed as the influence of the cell state and environmental conditions on the probability of success of the sensor measurements. Then, θ can be modeled as a function of X_i and λ_i , i.e.,

$$\theta = g(X_i, \lambda_i) \equiv \eta_\ell \phi_i^{a_{\ell 1}} \beta_i^{a_{\ell 2}} \gamma_i^{a_{\ell 3}}, \quad X_i = x_\ell, \quad \forall x_\ell \in \mathcal{X} \quad (30)$$

where $\eta_\ell > 0$ and $A = \{a_{\ell i}\} \in \mathbb{R}^{4 \times 3}$ are parameters that represent the effluence of X_i and λ_i on θ , respectively, and are shown in Appendix A.

1) *Bernoulli Sensor Model*: Bernoulli trials and Poisson models have both been used to model moving target detections by distributed sensor networks in [32]–[34]. A Bernoulli experiment has a random outcome that can take two mutually exclusive values (e.g., success or failure) and, when repeated N independent times, leads to a sequence of N Bernoulli trials [35, p. 134]. Let the measurement Z_i denote the random variable associated with one Bernoulli trial. Since, in the Bernoulli sensor model, Z_i must be binary, all Bernoulli targets are assigned the high-risk value x^r . Then, the probability of a measurement's success ($z_j = 1$) and the probability of failure ($z_j = 0$) can be obtained from the Bernoulli distribution

$$p(Z_i = z_j | \theta) = \theta^{z_j} (1 - \theta)^{1 - z_j}, \quad z_j = 0, 1 \quad (31)$$

where success represents the detection of a high-risk target and failure represents the detection of an empty cell.

2) *Binomial Sensor Model*: In the binomial sensor model, the measurement Z_i denotes the number of observed successes in N Bernoulli trials such that $\mathcal{Z} = \{z_1, \dots, z_N\}$. Then, the posterior PMF in the binomial sensor model can be generated using the binomial distribution

$$p(Z_i = z_j | \theta) = \binom{N}{z_j} \theta^{z_j} (1 - \theta)^{N - z_j}, \quad z_j = 0, \dots, N, \quad (32)$$

where X_i can take all four possible values in \mathcal{X} and $N = 3$.

3) *Poisson Sensor Model*: A Poisson process is a random experiment that generates a number of changes in a fixed interval, such as space or time, and whose probability can be described by an infinite series that converges to an exponential function [35, p. 143]. Let the measurement Z_i denote the number of changes in each interval, and let θ in (30) represent the parameter of the distribution. Then, assuming that the number of changes in nonoverlapping intervals is independent and $\theta > 0$, the Poisson sensor model can be generated from the Poisson distribution

$$p(Z_i = z_j | \theta) = e^{-\theta} \frac{\theta^{z_j}}{z_j!}, \quad z_j = 0, 1, 2, \dots \quad (33)$$

It can be shown that θ is equal to the expected number of changes in the process or $\theta = \mathbb{E}(Z_i)$.

4) *Mixture-of-Binomial Sensor Model*: A more complex sensor model is obtained by means of a mixture of distributions, which compounds multiple PMFs using positive mixing proportions or weights [35, p. 189]. Mixture models are used in a variety of applications, ranging from classification to statistical inference, and are reviewed comprehensively in [36]. Since the random variables considered in this paper are discrete, we consider a mixture of two binomial distributions

$$f(Y) = \sum_{l=1}^2 w_l \binom{N_l}{Y} \theta_l^Y (1 - \theta_l)^{N_l - Y}, \quad 0 \leq w_l \leq 1, \quad \sum_{l=1}^2 w_l = 1 \quad (34)$$

formulated in terms of a discrete random variable Y with range $\mathcal{Y} = \cup_{l=1}^2 \{0, \dots, N_l\}$, where the distributions are indexed by l . The mixing proportions are w_1 and w_2 , and the binomial parameters are θ_1 and θ_2 . Let $\theta_1 = \theta = g(X_i, \lambda_i)$ in (30), with the parameters shown in Appendix A. Also, let $\theta_2 = \sqrt{\theta_1}$, $N_1 = 3$, and $N_2 = 5$. Then, the posterior PMF in the mixture-of-binomial sensor model can be generated according to the distribution

$$p(Z_i = z_j | \theta) = w \binom{N_1}{z_j} \theta^{z_j} (1 - \theta)^{N_1 - z_j} + (1 - w) \binom{N_2}{z_j} \times \theta^{z_j/2} (1 - \sqrt{\theta})^{N_2 - z_j}, \quad 0 \leq w \leq 1, \quad z_j = 0, \dots, 5 \quad (35)$$

where w is a user-defined weight that is set equal to 0.5.

B. Classical Sensor Model

A sensor model that is widely applicable and obeys the classical form (8) is an exponential power law that models the received isotropic energy generated by a constant target source level and attenuated by the environment. This power law is commonly applied to acoustic, magnetic, and optical sensor measurements that are governed by linear wave propagation models [6]. When the received signal exceeds the sensor's threshold, the target is detected, and its distance from the sensor is computed using the sensor model and the known environmental conditions. We consider a multitarget scenario in which each target is located in one cell, for example, $\kappa_i \in \mathcal{K}$, and the state X_i represents the distance between the sensor and the target in κ_i . Then, the distance X_i can be estimated from a measurement Z_i obtained according to the power law

$$Z_i = a_i \|X_i\|^{-\alpha_i} + \nu_i, \quad i = 1, \dots, c_t \quad (36)$$

and subject to additive zero-mean Gaussian noise ν_i , where $\|\cdot\|$ denotes the L_2 -norm, a_i is a known constant that depends on the target characteristics, and α_i is an attenuation coefficient that depends on the environmental conditions. In this paper, it is assumed that X_i is a scalar, $a_i = 10$, and $\alpha_i = 0.3$ for all $\kappa_i \in \mathcal{K}$.

In order to compare the classical model results to those obtained for the probabilistic sensor models, the range of X_i is discretized into four possible values $\mathcal{X} = \{1, 2, 3, 4\}$. The zero-mean Gaussian noise ν_i is fully specified by its standard deviation σ_i . Thus, the sensor and environmental parameters can be represented by $\lambda_i = [a_i \ \alpha_i \ \sigma_i]^T$. It follows that the

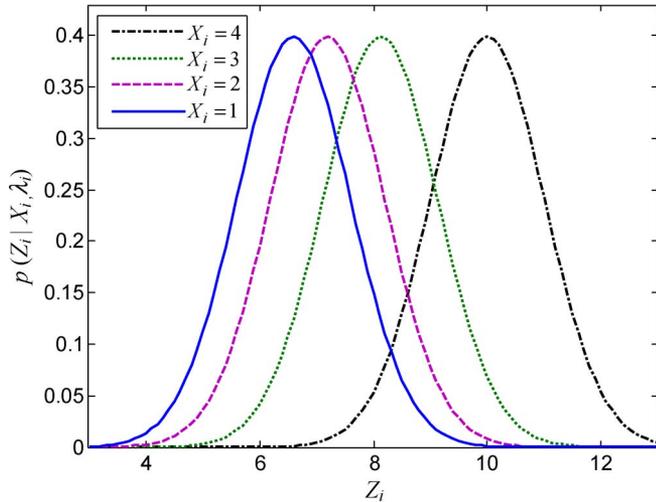


Fig. 1. Posterior probability of classical sensor measurements for $\sigma_i = 1$.

posterior probability of the classical sensor measurements in κ_i , given X_i and λ_i , is

$$p(Z_i = z_j | X_i, \lambda_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\left(\frac{z_j - a_i X_i^{-\alpha_i}}{2\sigma_i}\right)^2}, \quad z_j \in [0, 13] \quad (37)$$

where Z_i and λ_i are continuous random variables and X_i is a discrete random variable. Thus, (37) is referred to as a normal mixed model [37]. In this paper, three noise models are considered with standard deviations $\sigma_i = 1, 3$, and 5 for all $\kappa_i \in \mathcal{K}$. As an example, the posterior probability (37) is plotted in Fig. 1 for $\sigma_i = 1$.

The five sensor models are simulated as follows. When a cell κ_i is selected from \mathcal{K} at time k , one measurement value z^k is sampled from the corresponding joint PMF in (9), given the values of X_i and λ_i in κ_i . The prior $p(X_i)$ represents information available about the cells' state variables at $k = 0$. In this paper, three models of $p(X_i)$ are considered: uninformative prior, informative prior, and informative prior with large noise. The uninformative prior is modeled as a uniform probability distribution over the range \mathcal{X} , indicating that any value of X_i is equally probable for any cell $\kappa_i \in \mathcal{K}$. The informative prior is generated by using the sensor to obtain one measurement per cell. The informative prior with large noise represents expert knowledge or prior measurements obtained by a less accurate sensor. For the probabilistic sensor models in (31)–(35), this prior is generated by adding a random uniformly distributed error to every sensor measurement used to generate the informative prior. For the classical sensor model in (37), the error is sampled from a Gaussian distribution with zero mean and standard deviation equal to $2\sigma_i$.

VIII. NUMERICAL SIMULATIONS AND RESULTS

This section presents the results obtained by implementing all information theoretic functions and search strategies on the canonical sensor problem described in Section III. The canonical sensor problem is simulated by generating a set \mathcal{K} with $c = 1000$ cells as follows. In the Bernoulli sensor simulation, all

$c_t = 500$ cells with targets are assigned the high-risk state value $x_4 = x^r$ (as explained in Section VII), 300 cells are assigned the value x_1 (empty), and 200 cells are assigned the value x_2 (clutter). In the binomial, Poisson, and binomial-mixture sensor simulations, 300 cells are assigned the value x_1 (empty), 200 cells are assigned the value x_2 (clutter), 200 cells are assigned the value x_3 (target), and 300 cells are assigned the value $x_4 = x^r$. In the classical sensor simulation, the state values represent the sensor's distance from the target (see Section VII). Therefore, a target is assigned to every cell in \mathcal{K} such that $c_t = c = 1000$, where 200 cells are assigned the distance x_1 , 300 cells are assigned the distance x_2 , 200 cells are assigned the distance x_3 , and 300 cells are assigned the distance x_4 .

In all sensor simulations, the state values are assigned to the chosen number of cells in random order, by sampling the cell index using a uniform pseudorandom number generator [38]. The value of the parameter vector λ_i is assigned randomly to each cell by sampling the uniform prior distribution $p(\lambda^k)$ over the range Λ . The probability $p(X_i)$ is not uniform and is unknown *a priori*. The three models of uninformative, informative, and large-noise informative priors (see Section VII) are implemented separately to simulate cases in which prior information about X_i is either unavailable, available, or is available but very noisy, respectively. After the set of cells \mathcal{K} is generated, the sensor measurements are obtained using the search strategies presented in Section V. In order to investigate how the search performance varies as a function of time, k is varied from 0 to $f = 3000$. As soon as a measurement is obtained from a cell, it is processed using the belief-state update rule in (14) and the decision rules in Section VI.

The numerical results are organized as follows. For each sensor model, the rates of correct classification (F_c), correct classification for high-risk targets (F_r), and false alarms (F_{fa}) are first evaluated using the IS strategy with each of the information functions presented in Section IV. The IS strategy with the best performance is then compared to DS, AS, TS, and LLRS. For every sensor model and search strategy, a comparative study is conducted using the maximum *a posteriori* and MLE decision rules and the three models of prior information. The results obtained from extensive numerical simulations are organized by sensor model in Sections VIII-A–E. Then, Section VIII-F provides a summary of the search strategies with the best classification and false-alarm rates, on average, over the chosen time interval.

A. Bernoulli Sensor Model

The simulation results from the Bernoulli and other sensor models consistently showed that the maximum *a posteriori* rule significantly outperforms the MLE rule for all information functions. As an example, the correct classification rates for high-risk targets obtained by the IS strategy using the maximum *a posteriori* rule and the MLE rule are plotted in Figs. 2 and 3, respectively. This comparison is representative of all numerical results obtained for F_c and F_{fa} , and for other priors and sensor models, which are omitted here for brevity. Also, while it may be well suited for sensor applications that require a very low rate of false alarms, the Neyman–Pearson rule was found to

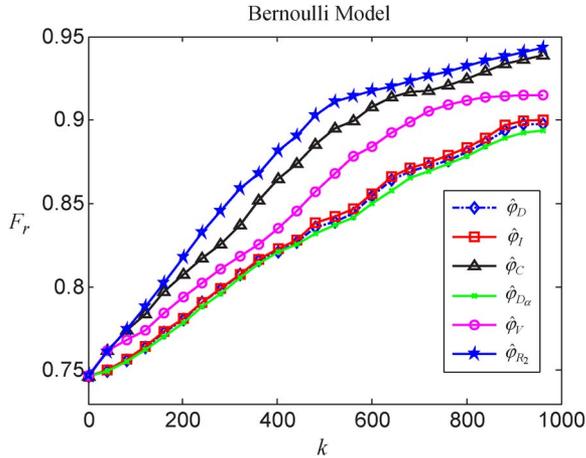


Fig. 2. Correct classification rate for high-risk targets obtained with the maximum *a posteriori* decision rule and an informative prior.

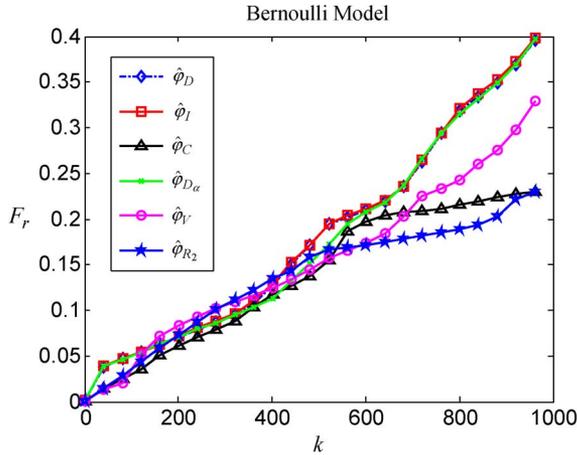


Fig. 3. Correct classification rate for high-risk targets obtained with the MLE decision rule and an informative prior.

perform poorly compared to both the *a posteriori* and MLE rules because of its high rate of unclassified cells that do not meet the required significance level. Therefore, in the remainder of this paper, only the results obtained by the maximum *a posteriori* decision rule are presented.

As shown in Fig. 2, the information function based on quadratic entropy, $\hat{\varphi}_{R_2}$, leads to the best rate of correct classification of high-risk targets, F_r . On the other hand, as shown in Fig. 4, $\hat{\varphi}_{R_2}$ leads to the highest rate of false alarms. In the Bernoulli sensor simulation, all information functions were found to display approximately the same value of correct classification rate, F_c , for all $0 \leq k \leq f$. Therefore, information functions that achieve high values of F_r perform poorly in terms of F_{fa} and vice versa. The information function $\hat{\varphi}_I$ provides a good compromise between F_r and F_{fa} .

The IS strategy was found to achieve the highest rate of correct classification among all strategies (see Fig. 5) and the highest rate of correct classification for high-risk targets (not shown for brevity). The AS, TS, and LLRS strategies display the lowest false-alarm rate because they obtain multiple measurements from few selected cells with low uncertainty until they achieve high CLs for the estimated state values. In the presence of an uninformative prior, the information functions

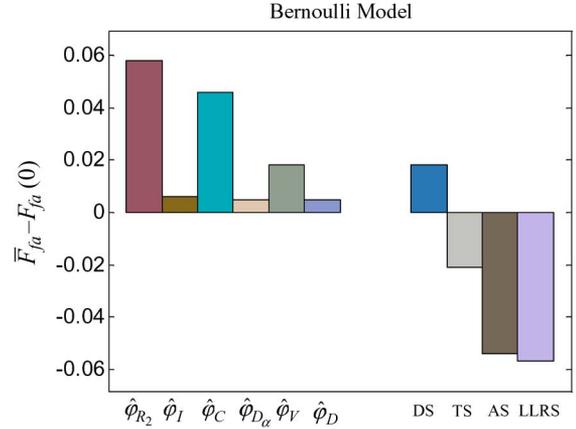


Fig. 4. Average false-alarm rate with informative prior.

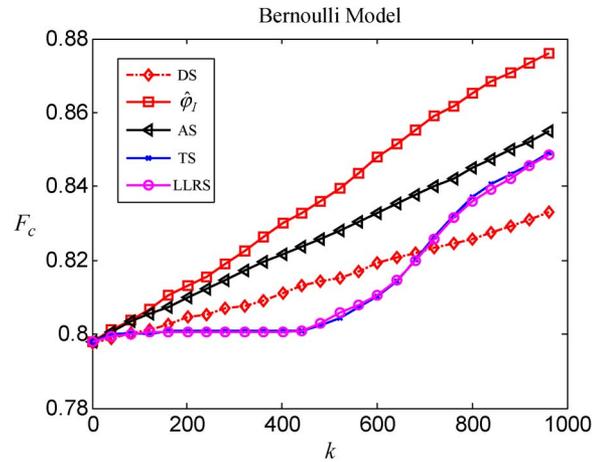


Fig. 5. Correct-classification rate with informative prior.

were found to perform similarly under all three measures of performance, F_r , F_c , and F_{fa} . The IS strategy was found to outperform all other search strategies in classification performance (F_r and F_c) but was significantly outperformed by the AS, TS, and LLRS strategies with respect to the false-alarm rate F_{fa} . Therefore, it can be concluded that, for the Bernoulli sensor model, the IS strategy with the mutual-information function, $\hat{\varphi}_I$, leads to the best classification performance. The AS, TS, and LLRS strategies exhibit similar performance and lead to the lowest rate of false alarms.

B. Poisson Sensor Model

The average rates of correct classification and false alarms for the Poisson sensor model with informative prior are plotted in Figs. 6 and 7, respectively. The simulation results show that, while initially $\hat{\varphi}_I$ and $\hat{\varphi}_{R_2}$ perform similarly, over time, $\hat{\varphi}_{R_2}$ achieves the best performance with respect to both classification and false alarms. When compared to other strategies, the IS strategy based on quadratic entropy achieves the highest rates of correct classification (see Figs. 6 and 8), and the lowest rate of false alarms (see Fig. 7), with an informative prior.

In the presence of an informative prior with large noise (see Fig. 9), all information functions perform similarly with respect to F_c . However, $\hat{\varphi}_{R_2}$ achieves slightly better performance with respect to F_r and significantly better performance with respect

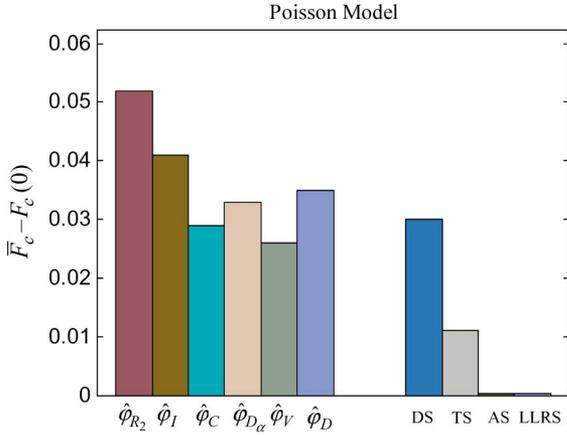


Fig. 6. Average correct-classification rate with informative prior.

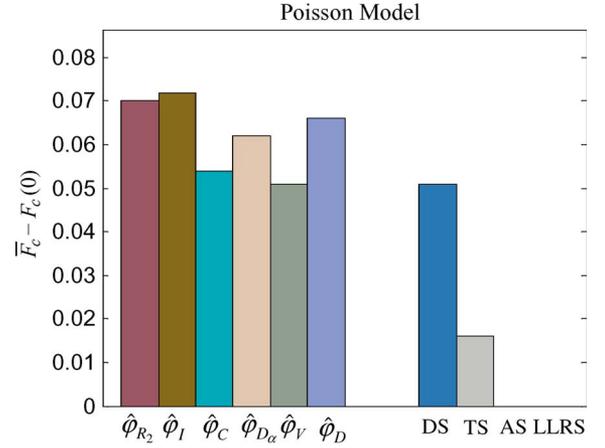


Fig. 9. Average correct-classification rate with large-noise informative prior.

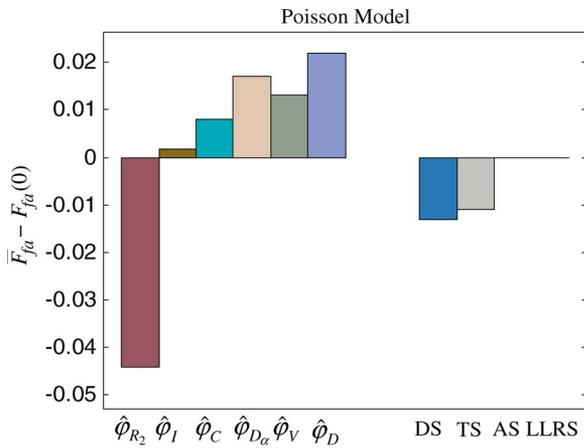


Fig. 7. Average false-alarm rate with informative prior.

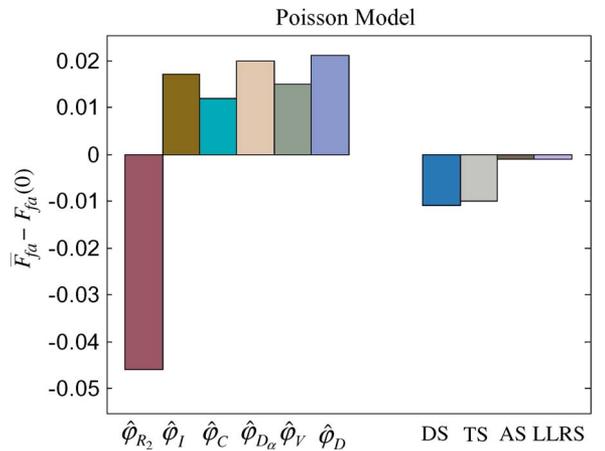


Fig. 10. Average false-alarm rate with large-noise informative prior.

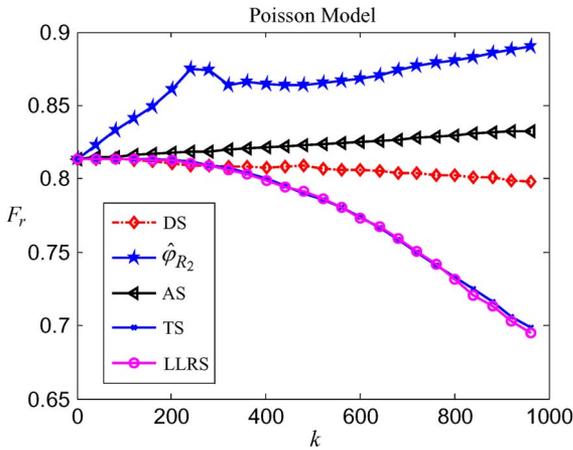


Fig. 8. Correct-classification rate for high-risk state targets with informative prior.

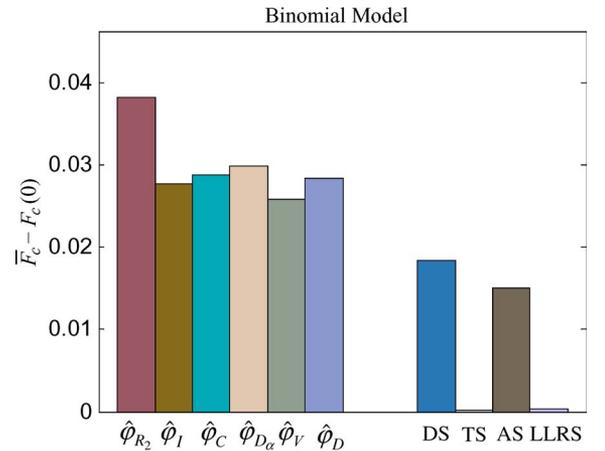


Fig. 11. Average correct-classification rate with informative prior.

to the rate of false alarms, as shown in Fig. 10. When compared to other search strategies, IS leads to significantly higher rates of correct classification, F_c (see Fig. 9) and F_r , and to a significantly lower rate of false alarms, F_{fa} (see Fig. 10). Therefore, for the Poisson sensor model, the IS strategy with quadratic entropy, $\hat{\varphi}_{R_2}$, leads to the best sensor performance overall in the presence of an informative prior and an informative prior with large noise.

C. Binomial Sensor Model

The same set of simulations performed for the Bernoulli and Poisson sensor models was performed using the binomial sensor model. The results (e.g., Fig. 11) showed that, for the binomial model, the IS strategy with the quadratic-entropy function $\hat{\varphi}_{R_2}$ achieves the best performance of all strategies with respect to all three performance measures, F_c , F_r , and F_{fa} . In this section, the IS strategy based on quadratic entropy,

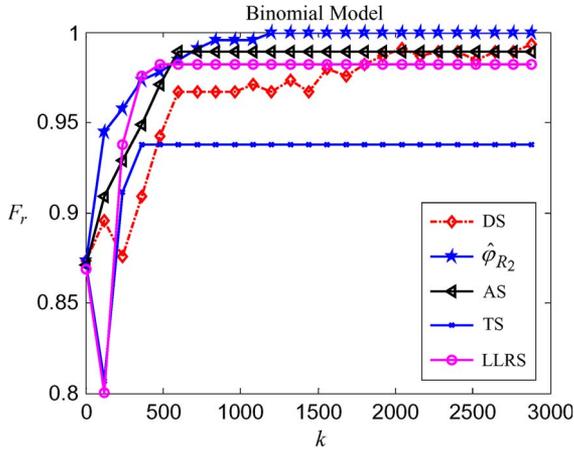


Fig. 12. Correct-classification rate for high-risk state targets, with informative prior, $c = 50$, and $f = 3000$.

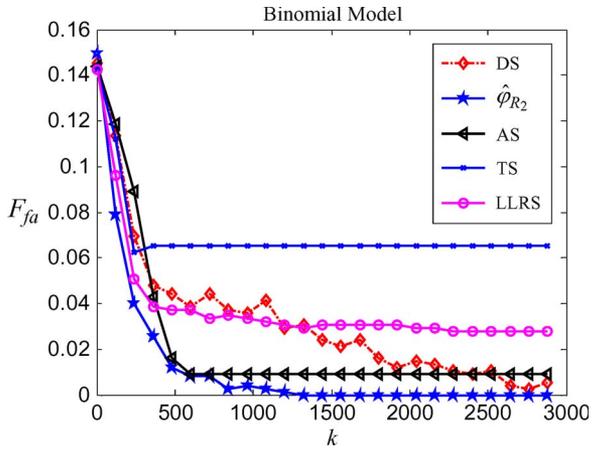


Fig. 13. False-alarm rate with informative prior, $c = 50$, and $f = 3000$.

DS, AS, TS, and LLRS are also implemented using an infinite time horizon, defined as $f \rightarrow \infty$. In practice, an infinite-time horizon is simulated by letting $f \gg c$, such that $F_c \rightarrow 1$ and $F_r \rightarrow 0$, as $k \rightarrow f$. For this study, the number of cells in \mathcal{K} is chosen as $c = 50$, and measurements are obtained up to a final time $f = 3000$.

As shown by the results in Figs. 12 and 13, in the limit of $k \rightarrow f$, the IS strategy classifies all cells correctly and eliminates all false alarms. The AS and DS strategies approach the IS performance with a slower rate of convergence and leave a small percentage of cells improperly classified. Although the TS and LLRS strategies have the advantage of displaying low false-alarm rates (see Section VIII-A), by selecting cells with a high expected CL, TS and LLRS leave a substantial percentage of cells improperly classified even as $k \rightarrow f$ (see Figs. 12 and 13). Therefore, it can be concluded that the IS strategy is better suited to infinite-horizon problems where, given sufficient time and sensor measurements, it is desirable to correctly classify all targets.

D. Mixture-of-Binomial Sensor Model

The average rates of correct classification and false alarms, \bar{F}_r and \bar{F}_{fa} , for the mixture-of-binomial sensor model are plotted respectively in Figs. 14 and 15, using an informative prior.

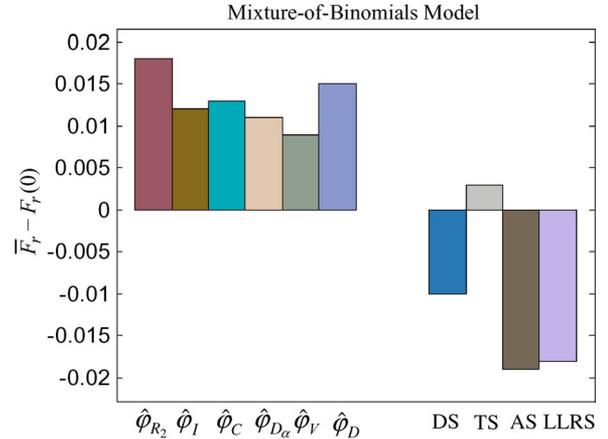


Fig. 14. Average correct-classification rate for high-risk state targets with informative prior.

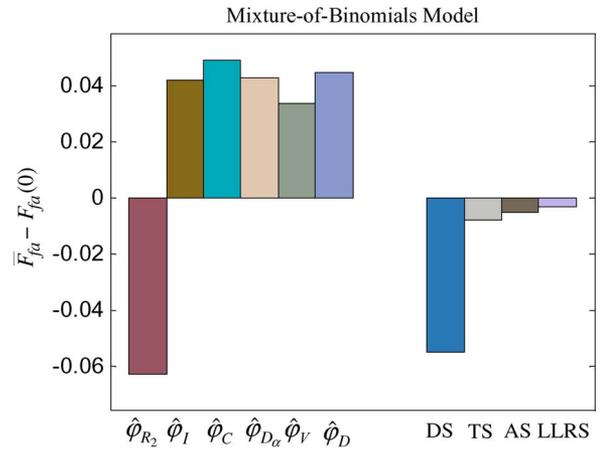


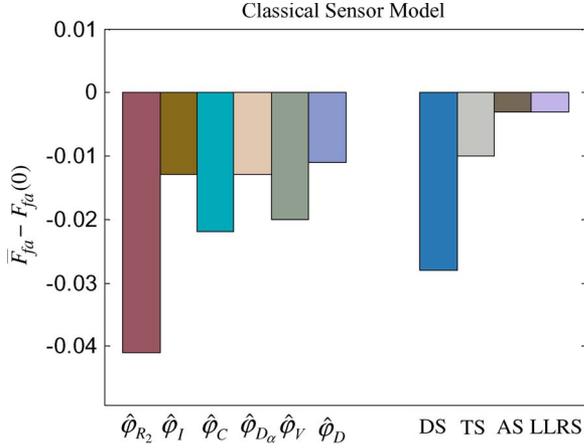
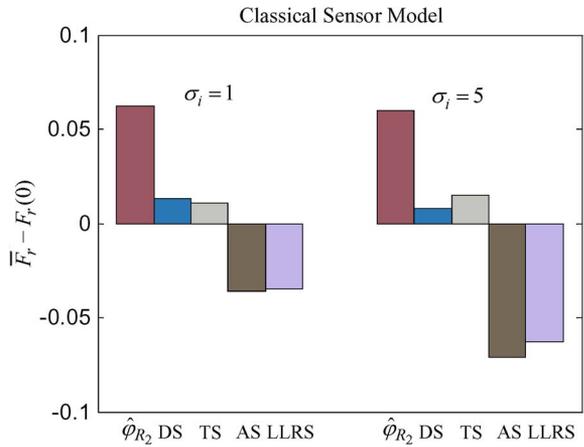
Fig. 15. Average false-alarm rate with informative prior.

It can be seen that $\hat{\varphi}_{R_2}$ clearly outperforms other functions with respect to both \bar{F}_r and \bar{F}_{fa} , as well as with respect to F_c (results not shown for brevity). It can be seen from Figs. 14 and 15 that the IS strategy based on quadratic entropy outperforms all other strategies, both with respect to classification and false alarms.

Extensive mixture-of-binomial sensor simulations were also conducted using the uninformative prior and the informative prior with large noise. It was found that, in these cases, the information functions all perform similarly with respect to F_c , F_r , and F_{fa} , on average and over time. Also, while the IS strategy led to significantly higher values of F_r compared to the other search strategies, IS was outperformed by the DS strategy with respect to F_c and F_{fa} .

E. Classical Sensor Model

For the classical sensor model, the correct classification rate F_c was found to be approximately the same for all six information functions, sensor-noise levels ($\sigma_i = 1, 3$, and 5), and priors $p(X_s)$. As the noise level increases, however, different information functions lead to increasingly different values of F_{fa} and F_r . For example, it was found that $\hat{\varphi}_I$ and $\hat{\varphi}_D$ achieve slightly better values of F_r than other functions for $\sigma_i = 1, 3$, and 5 , respectively, but display the highest rates of false alarms (see Fig. 16). On the other hand, $\hat{\varphi}_{R_2}$ achieves


 Fig. 16. Average false-alarm rate with informative prior and $\sigma_i = 1$.

 Fig. 17. Average correct-classification rate for high-risk targets, with informative prior, $\sigma_i = 1$, and $\sigma_i = 5$.

slightly lower values of F_r but displays the lowest rate of false alarms of all information functions (e.g., see Fig. 16) for all σ_i .

It can be seen from Fig. 17 that the IS strategy based on quadratic entropy achieves the highest value of F_c and F_r (see Fig. 17) and the lowest value of F_{fa} (see Fig. 16). As the noise level in the sensor model is increased, the IS strategy outperforms the other strategies even more significantly with respect to F_c and F_r (see Fig. 17) but is then outperformed by TS and LLRS with respect to F_{fa} . It can be concluded that, for the classical sensor model, the IS strategy leads to the best rates of correct classification, but in the presence of high noise, the TS and LLRS strategies lead to lower rates of false alarms.

F. Summary of Search Strategy Performance Comparison

The numerical results presented in the previous sections illustrate how the correct-classification and false-alarm rates of different search strategies vary as a function of time and how they are influenced by the sensor model, prior information, noise level, and decision rule. In this section, the results of these extensive numerical simulations are summarized by presenting the best search strategy for each sensor and prior information models in Figs. 19 and 20. The search strategies with the best value of \bar{F}_c , \bar{F}_r , and \bar{F}_{fa} are shown in Figs. 18–20, respectively. It can be seen that, in most cases, the IS strategy based on

Prior Distribution \ Sensor Model	Bernoulli	Poisson	Binomial	Binomial Mixture	Classical $\sigma = 1$
	Uninformative	DS -	$\hat{\phi}_{R_2}$ 86.3%	$\hat{\phi}_C$ 2.42%	$\hat{\phi}_{R_2}$ 0.990%
Informative, Large Noise	$\hat{\phi}_V$ 87.2%	$\hat{\phi}_I$ 41.2%	$\hat{\phi}_{R_2}$ 32.4%	$\hat{\phi}_{R_2}$ 75.4%	$\hat{\phi}_{R_2}$ 24.3%
Informative	$\hat{\phi}_V$ 159%	$\hat{\phi}_{R_2}$ 73.3%	$\hat{\phi}_{R_2}$ 108%	$\hat{\phi}_{R_2}$ 87.6%	$\hat{\phi}_V$ 76.5%

 Fig. 18. Strategy with the highest classification rate, \bar{F}_c , and percent improvement over DS.

Prior Distribution \ Sensor Model	Bernoulli	Poisson	Binomial	Binomial Mixture	Classical $\sigma = 1$
	Uninformative	$\hat{\phi}_{R_2}$ 2.12%	$\hat{\phi}_{R_2}$ 632%	$\hat{\phi}_V$ 1.58%	DS -
Informative, Large Noise	$\hat{\phi}_{R_2}$ 155%	$\hat{\phi}_{R_2}$ 185%	$\hat{\phi}_{R_2}$ 337%	$\hat{\phi}_I$ 639%	$\hat{\phi}_{R_2}$ 168%
Informative	$\hat{\phi}_{R_2}$ 151%	$\hat{\phi}_{R_2}$ 843%	$\hat{\phi}_{R_2}$ 1151%	$\hat{\phi}_{R_2}$ 211%	$\hat{\phi}_{R_2}$ 377%

 Fig. 19. Strategy with the highest classification rate for high-risk targets, \bar{F}_r , and percent improvement over DS.

Prior Distribution \ Sensor Model	Bernoulli	Poisson	Binomial	Binomial Mixture	Classical $\sigma = 1$
	Uninformative	AS 86.4%	$\hat{\phi}_{R_2}$ 272%	AS 92.0%	AS 99.3%
Informative, Large Noise	RB 51.4%	$\hat{\phi}_{R_2}$ 318%	$\hat{\phi}_{R_2}$ 31.2%	$\hat{\phi}_{R_2}$ 22.3%	$\hat{\phi}_{R_2}$ 18.5%
Informative	$\hat{\phi}_I$ 66.7%	$\hat{\phi}_{R_2}$ 239%	$\hat{\phi}_{R_2}$ 156%	$\hat{\phi}_{R_2}$ 28.5%	$\hat{\phi}_{R_2}$ 46.4%

 Fig. 20. Strategy with the lowest false-alarm rate, \bar{F}_{fa} , and percent improvement over DS.

quadratic entropy outperforms other strategies. The percent improvement over the DS strategy is also shown because DS is the simplest yet, overall, most efficient of the non-information-driven strategies.

IX. SEARCH STRATEGIES PERFORMANCE ANALYSIS

The GCM has been used for generalized multivariate analysis in a wide variety of applications, ranging from biology to engineering [23]–[25]. Unlike the classical least squares model [39], the GCM is applicable to repeated noisy measurements obtained at a series of known times from multiple experiments that may change over time but potentially are highly correlated. In GCM analysis, each time series is approximated by a polynomial function, and the resulting curves are used to determine whether the difference between experiments is statistically significant. In Section IX-A, GCM analysis is used to determine whether the performance difference observed among IS strategies is statistically significant. Based on Figs. 19 and 20, the Poisson sensor model is chosen as a representative example. Then, the computational complexity of all search strategies is analyzed in Section IX-B.

A. GCM Analysis

The GCM approach proposed in [23] is applied to analyze multiple IS strategies by viewing each strategy as an experiment. It is assumed that the true correct-classification rate of a strategy implementing $\hat{\varphi}$, denoted by $J_{\hat{\varphi}}(k)$ and typically unknown, can be approximated by a second-order polynomial. In order to estimate the coefficients of the polynomial, each strategy (or experiment) is simulated q times. Let $F_c^\ell(k)$ denote the correct-classification rate computed for the ℓ th simulation, with $\ell = 1, \dots, q$. It is also assumed that, at every k , $F_c^\ell(k)$ has a normal distribution with mean $J_{\hat{\varphi}}(k)$ and variance $\sigma_c(k)$, where the uncertainty is caused by the sensor measurement noise. Based on the strong law of large numbers, the probability of $\tilde{F}_c(k) \equiv \lim_{q \rightarrow \infty} \sum_{i=1}^q F_c^\ell(k)/q = J_{\hat{\varphi}}(k)$ equals one [40]. Since, in practice, the sensor simulation cannot be performed an infinite number of times, the GCM is used to fit a curve that represents the change in $\tilde{F}_c(k)$ brought about by different IS strategies, as a function of k . The curve is then used to analyze whether the change is caused by a difference in performance or simply by the sensor measurement noise.

Hereon, the information functions are ordered and indexed as follows: $\{\hat{\varphi}_s\}_{s=1, \dots, 6} \equiv \{\hat{\varphi}_{R_2}, \hat{\varphi}_I, \hat{\varphi}_C, \hat{\varphi}_{D_\alpha}, \hat{\varphi}_V, \hat{\varphi}_D\}$. Based on the previous sections, the GCM simulation parameters are chosen as $f = 1000$ and $q = 30$. N time instants, with $k = 50, 100, \dots, 1000$, are considered in the interval $(0, f]$ and indexed by $K = 1, \dots, N$, with $N = 20$. Let $g_{s,\ell}(k)$ denote the value of $F_c^\ell(k)$ for the s th information function $\hat{\varphi}_s$. Then, over time, this value is approximated by the second-order polynomial in k

$$g_{s,\ell}(k) \approx u_{s,1} + u_{s,2} \left(\frac{k}{50} - \frac{21}{2} \right) + \frac{1}{10} u_{s,3} \left(\frac{k}{50} - \frac{21}{2} \right)^2 + e_{s,\ell}(k) \quad (38)$$

where $s = 1, \dots, 6$, $\ell = 1, \dots, 30$, $e_{i,j}(k)$ is the noise caused by the sensor measurement at time k , and $u_{s,1}$, $u_{s,2}$, and $u_{s,3}$ are the coefficients to be determined.

The correct-classification rates computed from all experiments are organized in an $N \times 180$ matrix, G , defined such that its K th row is a vector of correct-classification values, $[g_{1,1}(K) \dots g_{1,30}(K) \dots g_{6,1}(K) \dots g_{6,30}(K)]$, for the K th time instant. It can be easily shown from (38) that G can be factorized in terms of an $N \times 180$ matrix B , a 3×6 matrix U defined in terms of the polynomial's coefficients, a 6×180 matrix C of ones and zeros, and an $N \times 180$ matrix E of measurement noise values as follows:

$$G = BUC + E. \quad (39)$$

Now, let $J_s(k)$ denote the value of the true correct-classification rate $J_{\hat{\varphi}}(k)$ for the s th strategy. Then, the $N \times 6$ matrix

$$J \equiv \begin{bmatrix} J_1(50) & J_2(50) & \dots & J_6(50) \\ J_1(100) & J_2(100) & \dots & J_6(100) \\ \vdots & \vdots & \ddots & \vdots \\ J_1(1000) & J_2(1000) & \dots & J_6(1000) \end{bmatrix} \quad (40)$$

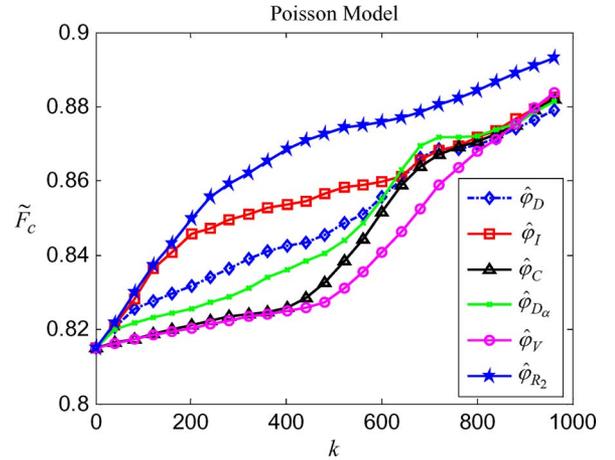


Fig. 21. Correct-classification rate computed experimentally with an informative prior.

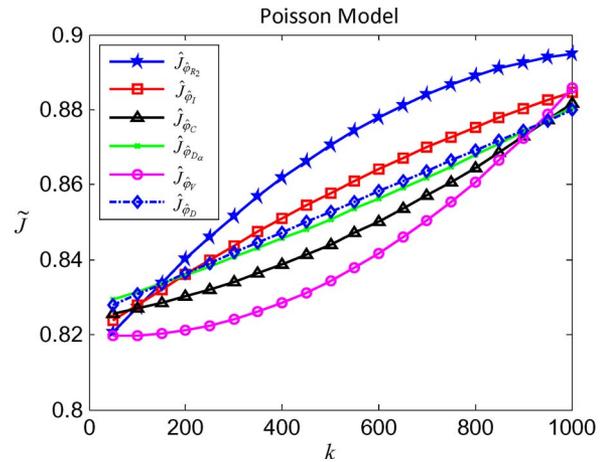


Fig. 22. Correct-classification rate approximated by GCM with an informative prior.

can be written as $J = BU$. The value of U can be approximated by the MLE method [23] and subsequently used to approximate the correct-classification rates by the matrix $\tilde{J} = B\tilde{U}$, where each column corresponds to one IS strategy. The simulations show that the classification rate \tilde{F}_c , computed from experiments and plotted in Fig. 21, follows closely the classification rate \tilde{J} , approximated via curve fitting and plotted in Fig. 22.

Finally, three hypothesis tests are formulated to examine whether the GCM polynomials obtained from different information functions have coefficients that are statistically different. The three hypotheses to test are

$$H_i : u_{1,i} = u_{2,i} = \dots = u_{6,i}, \quad \text{for } i = 1, 2, 3. \quad (41)$$

Using the approach in [24], it was found that all three hypotheses are rejected with a 95% CL. This indicates that correct-classification rates of different IS strategies are statistically different with a probability of error that is less than 0.05.

A similar GCM analysis was used to compare the correct-classification rates obtained by the quadratic-entropy IS strategy and DS, and the rates F_r and F_{fa} obtained by different IS strategies and DS. All of these results, shown in [41] and

omitted here for brevity, showed that the differences in classification performance observed in this paper are statistically significant, and thus, the search strategy and information function should be carefully selected based on the sensor model and the quality of prior information.

B. Complexity Analysis

The computational complexity of the search strategies considered in this paper is derived as a function of the dimensions of the target state, $X \in \mathbb{R}^n$, and the sensor measurement vector, $Z \in \mathbb{R}^r$. For this reason, in this section, $n, r > 1$, and a new notation is adopted by which X_i denotes the i th element of X with range \mathcal{X}_i . $x_{i\ell} \in \mathcal{X}_i$ denotes the ℓ th value of X_i . Z_j denotes the j th element of Z with range \mathcal{Z}_j , and $z_{j\ell} \in \mathcal{Z}_j$ denotes the ℓ th value of Z_j . $|\cdot|$ denotes the cardinality of a set. Computing an information function, such as the EDG in (16), requires computing $p(X|Z^k, M^{k-1}, \lambda)$ from (15) and the PMF

$$p(Z^k|M^{k-1}, \lambda) = \sum_{x_\ell \in \mathcal{X}} p(Z^k|X = x_\ell, \lambda)p(X = x_\ell|M^{k-1}, \lambda) \quad (42)$$

where $x_\ell = [x_{1\ell} \dots x_{n\ell}]^T \in \mathbb{R}^n$. Assume that the computation of each sensor measurement can be implemented in unit time. Then, for every x_ℓ , (42) can be computed directly from the sensor measurement and prior information in time $O(1)$. By the multiplication principle [42], the computational complexity of (42) is obtained by multiplying the cardinality of the ranges \mathcal{X}_i , $i = 1, \dots, n$, thus requiring a time $O(\prod_{i=1}^n |\mathcal{X}_i|)$. The state PMF can be computed from (15), such that

$$p(X|Z^k, M^{k-1}, \lambda) = \frac{p(Z^k|X, M^{k-1}, \lambda)p(X|M^{k-1}, \lambda)}{\sum_{x_\ell \in \mathcal{X}} p(X = x_\ell, Z^k, M^{k-1}, \lambda)} \quad (43)$$

where the numerator can be computed in time $O(1)$ and $p(X|Z^k, M^{k-1}, \lambda)$ can be obtained by normalizing the numerator for all possible outcomes of $x_\ell \in \mathcal{X}$. Since the number of possible outcomes of X is $\prod_{i=1}^n |\mathcal{X}_i|$, the computational complexity of (43) is $O(\prod_{i=1}^n |\mathcal{X}_i|)$.

Computing an information function such as the entropy $H(X|Z^k, M^{k-1}, \lambda)$ requires computing the posterior probability $p(X|Z^k = z_j, M^{k-1}, \lambda)$ for every possible value z_j . Thus, from the multiplication principle, the entropy-based function has a computational complexity $\prod_{j=1}^r |\mathcal{Z}_j| (O(\prod_{i=1}^n |\mathcal{X}_i|)) = O(\prod_{j=1}^r \prod_{i=1}^n |\mathcal{Z}_j| |\mathcal{X}_i|)$. Now, assume that the DS strategy is implemented via a random algorithm that can choose one cell to measure in unit time. Since the DS strategy randomly chooses only one cell at every time k , its computational complexity is $O(1)$. Assume that the search for the highest value in a known prior probability distribution can be achieved in unit time. Since the prior distribution at every time k is used to decide which cell to measure, the computational complexity of the AS and TS strategies is $O(1)$. Implementing the LLR strategy requires the computation of $L^k(X)$ using (26). This requires computing the likelihood of X , and since X has $\prod_{i=1}^n |\mathcal{X}_i|$ possible values, (26) has computational complexity $O(\prod_{i=1}^n |\mathcal{X}_i|)$. Subsequently, (27) requires taking the expectation with respect to Z , which requires a time $\prod_{j=1}^r |\mathcal{Z}_j| (O(\prod_{i=1}^n |\mathcal{X}_i|)) =$

$O(\prod_{j=1}^r \prod_{i=1}^n |\mathcal{Z}_j| |\mathcal{X}_i|)$. Thus, the computational complexity of the LLR strategy is $O(\prod_{j=1}^r \prod_{i=1}^n |\mathcal{Z}_j| |\mathcal{X}_i|)$. It can be concluded that DS, AS, and TS are the strategies with the lowest computational complexity, $O(1)$, while the entropy-based IS and LLRS strategies have the highest computational complexity, $O(\prod_{j=1}^r \prod_{i=1}^n |\mathcal{Z}_j| |\mathcal{X}_i|)$, which also grows with the dimensions of X and Z .

X. SUMMARY AND CONCLUSION

In most applications, sensor performance measures, such as correct-classification and false-alarm rates, are not available in closed form and can only be computed after the sensor measurements have been gathered and processed. Information theoretic functions are a natural choice for representing the information value of a measurement sequence, but they typically require knowledge of the belief state before and after the measurements arrive. This paper presents a general and systematic approach for deriving information functions that represent the expected value of future sensor measurements and thus can be utilized for sensor planning. A corresponding IS strategy can decide the measurement sequence by selecting cells, or targets, with the highest expected information value. In order to provide a comparative study that can be easily validated and generalized, this paper implements the aforementioned search strategies on simulated sensor models, priors, and decision rules that have been previously validated using real data. It is found that quadratic entropy typically leads to the most effective information function and that the corresponding IS strategy outperforms all others in correct-classification performance. In the presence of prior information, IS also displays the lowest rate of false alarms. However, when prior information is absent or very noisy, the TS and LLR strategies achieve the lowest false-alarm rates for the Bernoulli, mixture-of-binomial, and classical sensor models.

APPENDIX A SENSOR MODEL PARAMETERS

The sensor model parameters were chosen to simulate the environmental parameters' influence on the accuracy of the sensor measurements, while guaranteeing $\theta \in [0, 1]$. In the Bernoulli, Poisson, and binomial sensor models, $\eta_1 = 0.1/3$, $\eta_2 = 1/3$, $\eta_3 = 2/3$, and $\eta_4 = 1$. For the Bernoulli sensor model

$$A = \begin{bmatrix} 0.1 & 0.11 & 0.12 \\ 0.1 & 0.11 & 0.12 \\ -0.1 & 0.11 & -0.12 \\ -0.1 & -0.11 & -0.12 \end{bmatrix} \quad (44)$$

and for the Poisson and binomial sensor models

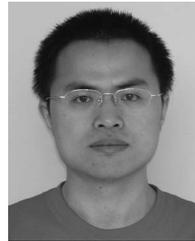
$$A = \begin{bmatrix} 0.1 & 0.11 & -0.12 \\ 0.1 & 0.11 & -0.12 \\ 0.1 & 0.11 & -0.12 \\ 0.1 & 0.11 & -0.12 \end{bmatrix}. \quad (45)$$

ACKNOWLEDGMENT

The authors would like to thank Dr. D. Banks for his advice.

REFERENCES

- [1] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks," *Computer*, vol. 37, no. 8, pp. 41–49, 2004.
- [2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall, 2003.
- [3] G. D. Hager, *Task-Directed Sensor Fusion and Planning: A Computational Approach*. Boston, MA: Kluwer, 1990.
- [4] G. D. Hager and M. Mintz, "Computational methods for task-directed sensor data fusion and sensor planning," *Int. J. Robot. Res.*, vol. 10, no. 4, pp. 285–313, Aug. 1991.
- [5] C. Cai and S. Ferrari, "Information-driven sensor path planning by approximate cell decomposition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 3, pp. 672–689, Jun. 2009.
- [6] M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *Int. J. High Perform. Comput. Appl.*, vol. 16, no. 3, pp. 293–313, Aug. 2002.
- [7] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 61–72, Mar. 2002.
- [8] R. F. Stengel, *Optimal Control and Estimation*. New York: Dover, 1986.
- [9] W. Schmaedeke, "Information based sensor management," in *Proc. SPIE—Signal Process., Sens. Fusion, Target Recognit. II*, Orlando, FL, 1993, vol. 1955, pp. 156–164.
- [10] K. Kastella, "Discrimination gain to optimize detection and classification," *IEEE Trans. Syst., Man Cybern. A, Syst. Humans*, vol. 27, no. 1, pp. 112–116, Jan. 1997.
- [11] C. Kreucher, K. Kastella, and A. Hero, "Sensor management using an active sensing approach," *Signal Process.*, vol. 85, no. 3, pp. 607–624, Mar. 2005.
- [12] K. J. Hintz, "A measure of the information gain attributable to cueing," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 2, pp. 434–442, Mar./Apr. 1991.
- [13] K. J. Hintz, "Multi-process constrained estimation," *IEEE Trans. Syst., Man Cybern.*, vol. 21, no. 1, pp. 237–244, Jan./Feb. 1991.
- [14] J. Denzler and C. M. Brown, "Information theoretic sensor data selection for active object recognition and state estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 145–157, Feb. 2002.
- [15] C. Kreucher, A. Hero, and K. Kastella, "A comparison of task driven and information driven sensor management for target tracking," in *Proc. IEEE Conf. Decision Control*, Seville, Spain, 2005, pp. 4004–4009.
- [16] J. Principe, D. Xu, Q. Zhao, and J. Fisher, "Learning from examples with information theoretic criteria," *J. VLSI Signal Process.*, vol. 26, no. 1/2, pp. 61–77, Aug. 2000.
- [17] E. Gokcay and J. Principe, "Information theoretic clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 158–171, Feb. 2002.
- [18] D. Erdogmus and J. Principe, "Generalized information potential criterion for adaptive system training," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1035–1044, Sep. 2002.
- [19] K. Kastella and S. Musick, "Search for optimal sensor management," in *Proc. SPIE*, 1996, vol. 2759, pp. 318–321.
- [20] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, Jan. 2004.
- [21] S. Ferrari and A. Vaghi, "Demining sensor modeling and feature-level fusion by Bayesian networks," *IEEE J. Sens.*, vol. 6, no. 2, pp. 471–483, Apr. 2006.
- [22] C. Cai, S. Ferrari, and Q. Ming, "Bayesian network modeling of acoustic sensor measurements," in *Proc. IEEE Sens.*, Atlanta, GA, 2007, pp. 345–348.
- [23] C. Khatri, "A note on a manova model applied to problems in growth curve," *Ann. Inst. Statist. Math.*, vol. 18, no. 1, pp. 75–86, Dec. 1966.
- [24] D. Morrison, *Multivariate Statistical Methods*. New York: McGraw-Hill, 1990.
- [25] J. Pan and K. Fang, *Growth Curve Models and Statistical Diagnostics*. New York: Sci. Press, 2002.
- [26] R. Potthoff and S. Roy, "A generalized multivariate analysis of variance model useful especially for growth curve problems," *Biometrika*, vol. 51, no. 3/4, pp. 313–326, 1964.
- [27] A. O. Hero, B. Ma, O. Michael, and J. D. Gorman, "Applications of entropic spanning graphs," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 85–95, Sep. 2002.
- [28] S. Ferrari and C. Cai, "Information-driven search strategies in the board game of CLUE," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 3, pp. 607–625, Jun. 2009.
- [29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [30] G. Zhang, S. Ferrari, and M. Qian, "An information roadmap method for robotic sensor path planning," *J. Intell. Robot. Syst.*, vol. 56, no. 1/2, pp. 69–98, Sep. 2009.
- [31] D. Hall and S. McMullen, *Mathematical Techniques in Multisensor Data Fusion*. Norwood, MA: Artech House, 2004.
- [32] T. A. Wettergren, "Performance of search via track-before-detect for distributed sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 1, pp. 314–325, Jan. 2008.
- [33] A. Poore and N. Rijavec, "A numerical study of some data association problems arising in multitarget tracking," *Comput. Optim. Appl.*, vol. 3, no. 1, pp. 27–57, 1994.
- [34] P. Morse and G. Kimball, *Methods of Operations Research*. New York: Dover, 2003.
- [35] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics*. Upper Saddle River, NJ: Prentice-Hall, 2005.
- [36] G. McLachlan, *Finite Mixture Models*. New York: Wiley-Interscience, 2000.
- [37] C. A. McGilchrist, "Estimation in generalized mixed models," *J. Roy. Statist. Soc.*, vol. 56, no. 1, pp. 61–69, 1994.
- [38] Mathworks, Functions: Rand Matlab, 2004. [Online]. Available: <http://www.mathworks.com>
- [39] R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*. New York: Springer-Verlag, 2005.
- [40] R. Durrett, *Probability: Theory and Examples*. Boston, MA: Duxbury Press, 1995.
- [41] G. Zhang, "The comparison of information functions in sensor planning," M.S. thesis, Statist. Dept., Duke Univ., Durham, NC, 2011.
- [42] J. Lint and R. Wilson, *A Course in Combinatorics*. Cambridge, U.K.: Cambridge Univ. Press, 2001.



Guoxian Zhang (M'07) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2003 and 2006, respectively, and the Ph.D. degree from Duke University, Durham, NC, in 2010.

He is currently a Software Quality Engineer with MicroStrategy, Inc., Tysons Corner, VA. His current research interests include sensor planning, data mining, multiagent systems, Bayesian statistics, and decision making under uncertainty.

Dr. Zhang has been a member of the American Society of Mechanical Engineers since 2008.



Silvia Ferrari (S'01–M'02–SM'08) received the B.S. degree from Embry-Riddle Aeronautical University, Daytona Beach, FL, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ.

She is the Paul Ruffin Scarborough Associate Professor of engineering with Duke University, Durham, NC, where she directs the Laboratory for Intelligent Systems and Controls. Her principal research interests include robust adaptive control of aircraft, learning and approximate dynamic programming, and optimal control of mobile sensor networks.

Dr. Ferrari is a member of the American Society of Mechanical Engineers, the International Society for Optical Engineering, and the American Institute of Aeronautics and Astronautics. She was the recipient of the Office of Naval Research Young Investigator Award (2004), the National Science Foundation CAREER Award (2005), and the Presidential Early Career Award for Scientists and Engineers (2006).



Chenghui Cai (S'02–M'09) received the B.S. and M.S. degrees in mechanical engineering from Tsinghua University, Beijing, China, in 2000 and 2003, respectively, and the Ph.D. degree in mechanical engineering from Duke University, Durham, NC, in 2008.

He is a Junior Analytics Manager with Opera Solutions, LLC, New York, NY. His current research interests include machine learning and data mining technologies and their applications to financial markets.

Dr. Cai is a member of the American Society of Mechanical Engineers, the Society for Industrial and Applied Mathematics, and Sigma Xi.